An Improved LSHADE-RSP Algorithm with the Cauchy Perturbation: iLSHADE-RSP

Tae Jong Choi, Chang Wook Ahn

\[Department of Artificial Intelligence and Software, Kyungil University\]
50, Gamasil-gil, Hayang-eup, Gyeongsan-si, Gyeongsangbuk-do, Republic of Korea

\[Artificial Intelligence Graduate School, Gwangju Institute of Science and Technology (GIST)\]
123, Cheomdangwagi-ro, Buk-gu, Gwangju, Republic of Korea

Abstract

A new method for improving the optimization performance of a state-of-the-art differential evolution (DE) variant is proposed in this paper. The technique can increase the exploration by adopting the long-tailed property of the Cauchy distribution, which helps the algorithm to generate a trial vector with great diversity. Compared to the previous approaches, the proposed approach perturbs a target vector instead of a mutant vector based on a jumping rate. We applied the proposed approach to LSHADE-RSP ranked second place in the CEC 2018 competition on single objective real-valued optimization. A set of 30 different and difficult optimization problems is used to evaluate the optimization performance of the improved LSHADE-RSP. Our experimental results verify that the improved LSHADE-RSP significantly outperformed not only its predecessor LSHADE-RSP but also several cutting-edge DE variants in terms of convergence speed and solution accuracy.

Keywords: Artificial Intelligence, Evolutionary Algorithm, Differential Evolution, Mathematical Optimization

1. Introduction

A population-based metaheuristic optimization method called evolutionary algorithms (EAs) is designed based on Darwin’s theory of natural selection. EAs generate a set of initial candidate solutions and update them iteratively with artificially designed evolutionary operators. As compared to traditional search algorithms, EAs are global, robust, and can be applied to any problem.

It is important to establish a balance between exploration and exploitation to improve the optimization performance of EAs. Matej et al. \[1\] stated that “Exploration is the process of visiting entirely new regions of a search space, whilst exploitation is the process of visiting those regions of a search space within the neighborhood of previously visited points.” If an EA has too strong exploration, it might not be beneficial from existing candidate solutions [2]. On the other hand, if an EA has too strong exploitation, the probability of finding an optimal solution might be decreased [2]. Many researchers have investigated a number of approaches for balancing the two cornerstones [1].

Differential evolution (DE) proposed by Storn and Price \[5\] is one of the most successful EAs to deal with mathematical optimization. DE distributes its candidate solutions over the search boundaries of an optimization problem and updates them iteratively with vector difference based evolutionary operators. DE has two main advantages over other EAs: 1) it has a simple structure and a few control parameters, and 2) the effectiveness of DE has been demonstrated on various real-world problems [5, 6]. Among numerous DE variants, L-SHADE variants \[7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18\] frequently perform very well on various optimization problems. LSHADE-RSP \[17\] was recently proposed and ranked second place in the CEC 2018 competition on single objective real-valued optimization.

Although LSHADE-RSP has shown excellent performance, it has the following problem: LSHADE-RSP uses a rank-based selective pressure scheme, which increases the greediness and boosts the convergence speed. However, it may cause premature convergence in which all the candidate solutions fall into the local optimum of an optimization problem and cannot escape from there [19, 20, 21, 22]. Although LSHADE-RSP uses a setting for increasing the number of pbest individuals in an effort to mitigate the problem, it may not be sufficient. In other words, LSHADE-RSP may fail to achieve exploration and exploitation.

In this paper, we proposed a new method for improving the optimization performance of LSHADE-RSP. The technique perturbs a target vector with the Cauchy distribution based on a jumping rate, which helps the algorithm to generate a trial vector with great diversity. Therefore, the technique can increase the probability of finding an optimal solution by adopting the long-tailed property of the Cauchy distribution. In the literature of DE, some researchers have demonstrated the effectiveness of using the Cauchy distribution in the phase of the recombination \[23, 24, 25, 26, 27\]. The novelty of the proposed approach lies in the perturbation of a target vector instead of a mutant vector. We named the combination of LSHADE-RSP and the proposed approach as iLSHADE-RSP.

We carried out experiments to evaluate the optimization performance of the proposed algorithm on the CEC 2017 test suite \[28\]. Our experimental results verify that the proposed algorithm significantly outperformed not only its predecessor LSHADE-RSP but also several state-of-the-art DE variants in
terms of convergence speed and solution accuracy.

The rest of this paper is organized as follows: We introduce the background of this paper in Section 2. In Section 3 we review the relevant literature to know, especially for L-SHADE variants. In Section 4, the details of the proposed algorithm is explained. We describe the experimental setup in Section 5. We present the experimental results and discussion in Section 6. Finally, we conclude this paper in Section 7.

2. Background

2.1. Differential Evolution

Since it was introduced, DE has received much attention because of simplicity and applicability. At the beginning of an optimization process, DE generates a set of NP initial candidate solutions as follows.

\[ \mathbf{P}_g = (\mathbf{x}_{1,g}, \mathbf{x}_{2,g}, \cdots, \mathbf{x}_{NP,g}) \]  

where \( \mathbf{P}_g \) denotes a population at generation \( g \). Each candidate solution denoted by \( \mathbf{x}_{i,g} = (x_{i,1}^1, x_{i,2}^2, \cdots, x_{i,D}^D) \) is a \( D \)-dimensional vector. DE updates the candidate solutions iteratively with vector difference based evolutionary operators, such as mutation, crossover, and selection, to search for the global optimum of an optimization problem. The mutation and crossover operators create a set of NP offspring, and the selection operator creates a population for the next generation by comparing the fitness value of a candidate solution and that of its corresponding offspring. DE returns the current best solution when it reaches the maximum number of generations \( G_{\text{max}} \) or function evaluations \( NFE_{\text{max}} \).

2.1.1. Initialization

At the beginning of an optimization process, DE distributes its candidate solutions over the search boundaries of an optimization problem with the initialization operator. Each candidate solution is initialized as follows.

\[ x_{i,0}^j = x_{\text{min}}^j + \text{rand}_i^j \cdot (x_{\text{max}}^j - x_{\text{min}}^j) \]  

where \( x_{\text{min}} = (x_{\text{min}}^1, x_{\text{min}}^2, \cdots, x_{\text{min}}^D) \) and \( x_{\text{max}} = (x_{\text{max}}^1, x_{\text{max}}^2, \cdots, x_{\text{max}}^D) \) denote the lower and upper search boundaries of an optimization problem, respectively. Also, \( \text{rand}_i^j \) denotes a uniformly distributed random number between [0, 1].

2.1.2. Mutation

A mutant vector \( \mathbf{v}_{i,g} \) is created in the mutation operator. The six frequently used classical mutation strategies are listed as follows.

- DE/best/1:
  \[ \mathbf{v}_{i,g} = \mathbf{x}_{\text{best},g} + \mathbf{F} \cdot (\mathbf{x}_{i,g} - \mathbf{x}_{r_1,g}) \]
- DE/best/2:
  \[ \mathbf{v}_{i,g} = \mathbf{x}_{\text{best},g} + \mathbf{F} \cdot (\mathbf{x}_{i,g} - \mathbf{x}_{r_2,g}) + \mathbf{F} \cdot (\mathbf{x}_{r_1,g} - \mathbf{x}_{r_2,g}) \]
- DE/current-to-best/1:
  \[ \mathbf{v}_{i,g} = \mathbf{x}_{i,g} + \mathbf{F} \cdot (\mathbf{x}_{\text{best},g} - \mathbf{x}_{i,g}) + \mathbf{F} \cdot (\mathbf{x}_{r_1,g} - \mathbf{x}_{r_2,g}) \]
- DE/current-to-best/2:
  \[ \mathbf{v}_{i,g} = \mathbf{x}_{i,g} + \mathbf{F} \cdot (\mathbf{x}_{r_1,g} - \mathbf{x}_{r_2,g}) + \mathbf{F} \cdot (\mathbf{x}_{r_1,g} - \mathbf{x}_{r_2,g}) \]
- DE/rand/1:
  \[ \mathbf{v}_{i,g} = \mathbf{x}_{i,g} + \mathbf{F} \cdot (\mathbf{x}_{r_1,g} - \mathbf{x}_{r_2,g}) \]
- DE/rand/2:
  \[ \mathbf{v}_{i,g} = \mathbf{x}_{i,g} + \mathbf{F} \cdot (\mathbf{x}_{r_1,g} - \mathbf{x}_{r_2,g}) + \mathbf{F} \cdot (\mathbf{x}_{r_1,g} - \mathbf{x}_{r_2,g}) \]

where \( r_1, r_2, r_3, r_4, r_5 \) denote mutually different random indices within \([1, 2, \cdots, NP]\), which are also different from \( i \). Moreover, \( \mathbf{x}_{\text{best},g} \) denotes the current best candidate solution. Furthermore, \( \mathbf{F} \) denotes a scaling factor, and \( K \) denotes a uniformly distributed random number between [0, 1].

2.1.3. Crossover

A trial vector \( \mathbf{u}_{i,g} \) is created in the crossover operator. The binomial crossover frequently used creates a trial vector as follows.

\[ \mathbf{u}_{i,g}^j = \begin{cases} v_{i,g}^j & \text{if } \text{rand}_i^j < \text{CR} \text{ or } j = j_{\text{rand}} \\ x_{i,g}^j & \text{otherwise} \end{cases} \]  

where \( j_{\text{rand}} \) denotes a random index within \([1, 2, \cdots, D]\). Also, \( \text{CR} \) denotes a crossover rate. The exponential crossover creates a trial vector as follows.

\[ \mathbf{u}_{i,g}^j = \begin{cases} v_{i,g}^j & \text{if } j = \langle n \rangle_D, \langle n + 1 \rangle_D, \cdots, \langle n + L - 1 \rangle_D \\ x_{i,g}^j & \text{otherwise} \end{cases} \]  

where \( n \) denotes a random index within \([1, 2, \cdots, D]\), and \( L \) denotes the number of elements, which can be calculated as follows.

\[ L = 0 \]

DO \{ \text{IF } L = L + 1 \}

WHILE ((\text{rand}_i^j < \text{CR}) \text{ AND } (L < D))

Also, \( \langle \cdot \rangle_D \) denotes the modulo of \( D \).

2.1.4. Selection

DE creates a population for the next generation with the selection operator. The selection operator compares the fitness value of a candidate solution \( (\mathbf{x}_{i,g}) \) and that of its corresponding offspring \( (\mathbf{u}_{i,g}) \) and picks the better one in terms of solution accuracy as follows.

\[ \mathbf{x}_{i,g+1} = \begin{cases} \mathbf{u}_{i,g} & \text{if } f(\mathbf{u}_{i,g}) \leq f(\mathbf{x}_{i,g}) \\ \mathbf{x}_{i,g} & \text{otherwise} \end{cases} \]  

where \( f(\mathbf{x}) \) denotes an optimization problem to be minimized.

2.2. Analysis of Cauchy Distribution

The Cauchy distribution is a family of continuous probability distributions, which is stable and has a probability density function (PDF), which can be expressed analytically. As compared to the Gaussian distribution, the Cauchy distribution has
γ and parameter is low. The PDF of the Cauchy distribution with x scale parameter is high, while a tall and narrow PDF if the scale parameter is low. The PDF of the Cauchy distribution with x0 and γ can be defined as follows.

\[ f(x; x_0, γ) = \frac{1}{πγ[1 + \left(\frac{x - x_0}{γ}\right)^2]} = \frac{1}{π\left(1 + \left(\frac{x - x_0}{γ}\right)^2 + γ^2\right)} \]  

(6)

Additionally, the cumulative distribution function of the Cauchy distribution with x0 and γ can be defined as follows.

\[ F(x; x_0, γ) = \frac{1}{π}\arctan\left(\frac{x - x_0}{γ}\right) + \frac{1}{2} \]  

(7)

Fig. 1 shows the four different PDFs of the Cauchy distribution.

3. Literature Review

Since it was introduced, many researchers have developed new methods for DE, such as

- Adaptive parameters with single mutation strategy DE algorithms [29, 30, 31, 32, 33, 34, 35, 36, 37].
- Adaptive parameters with multiple mutation strategy DE algorithms [38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48].
- Hybrid DE algorithms [49, 50, 51, 52, 53, 54, 55, 56].
- DE algorithms with sampling explicit probabilistic models [53, 55, 56].

For more detailed information, please refer to the following papers [5, 6, 37, 58].

Among numerous DE variants, L-SHADE variants [7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18] frequently perform very well on various optimization problems. JADE [7] proposed by Zhang and Sanderson is considered to be the origin of L-SHADE variants. JADE uses a new mutation strategy DE/current-to-pbest/1 with an external archive. The new mutation strategy is the generalization of a classical mutation strategy DE/current-to-best/1, and the external archive supplies the progress of an evolutionary process. These modifications improve the exploration of the algorithm. JADE also uses a learning process-based adaptive parameter control to adjust the control parameters F and CR automatically. In the experiments, JADE outperformed several algorithms, including jDE [29], SaDE [38], and PSO [59].

Tanabe and Fukunaga proposed SHADE [8], an enhancement to JADE. The main difference between SHADE and JADE is that SHADE utilizes historical memories, which store the average of successfully evolved individuals’ control parameters F and CR to adjust the control parameters F and CR automatically. In the experiments, SHADE outperformed several algorithms, including CoDE [40], EPSDE [59], and dynNPD-DE [50]. Tanabe and Fukunaga later proposed L-SHADE [9], an enhancement to SHADE. The main difference between L-SHADE and SHADE is that L-SHADE utilizes linear population size reduction (LPSR), which gradually reduces the population size as a linear function to establish a balance between exploration and exploitation. In the experiments, L-SHADE outperformed several algorithms, including NBIPOP-2xCMAES [60] and iCMAES-ILS [61].

Brest et al. proposed an improved version of L-SHADE called iL-SHADE [10]. iL-SHADE updates historical memories µF and µCR by calculating the average of old and new values. iL-SHADE also gradually reduces the p value of DE/current-to-pbest/1 as a linear function. iL-SHADE was ranked third place in the CEC 2016 competition on single objective real-valued optimization. Brest et al. later proposed an improved version of iL-SHADE called jSO [11]. jSO uses a new mutation strategy DE/current-to-pbest/w/1, which assigns a lower scaling factor Fw at the early stage of an evolutionary process and a higher scaling factor Fw at the late stage of an evolutionary process. jSO was ranked second place in the CEC 2017 competitions on single objective real-valued optimization.

Awad et al. proposed LSHADE-EpSin [12] based on L-SHADE, which utilizes a new ensemble sinusoidal approach for tuning the scaling factor F in an adaptive manner. The new ensemble sinusoidal approach is the combination of two sinusoidal waves whose objective is to establish a balance between exploitation and exploration. LSHADE-EpSin also utilizes a random walk at the late stage of an evolutionary process. Awad et al. proposed L-convSHADE [13] based on L-SHADE, which utilizes a new crossover operator based on covariance matrix adaptation with Euclidean neighborhood for rotation invariance. Awad et al. later proposed LSHADE-cnEpSin [14], which is the combination of LSHADE-EpSin and L-convSHADE with two major modifications: 1) a new ensemble sinusoidal approach with a learning process-based adaptive parameter control and 2) a new crossover operator based on covariance matrix adaptation with Euclidean neighborhood for rotation invariance. Finally, Awad et al. proposed EsDE-NR
Based on LSHADE-EpSin, which gradually reduces the population size with a niching-based approach.

Mohamed et al. proposed LSHADE-SPACMA [16], an enhancement to L-SHADE, which is a hybrid algorithm between LSHADE-SPA [16] and a modified version of CMA-ES [62]. LSHADE-SPA uses a new semi-parameter control for tuning the scaling factor $F$ in an adaptive manner. The modified version of CMA-ES undergoes the phase of the crossover, which improves the exploration of the algorithm.

Yet et al. propose mL-SHADE [18], an enhancement to L-SHADE, in which three major modifications are made: 1) a terminal value for the control parameter $CR$ is excluded, 2) a polynomial mutation strategy is included, and 3) a perturbation for historical memories is included. These modifications improve the exploration of the algorithm.

4. Proposed Algorithm

This section describes an improved LSHADE-RSP called iLSHADE-RSP, which employs a modified recombination operator, which calculates a perturbation of a target vector with the Cauchy distribution.

4.1. Mutation Strategy

The proposed algorithm uses a new mutation strategy DE/current-to-pbest/r [17]. The strategy is designed based on a rank-based selective pressure scheme [63], which proportionally selects two donor vectors $x_{i1}$ and $x_{i2}$, with respect to the fitness value. The higher the ranking of a candidate solution has, the more opportunity it will be selected. The strategy can be defined as follows.

- DE/current-to-pbest/r:
  
  $$v_{ig} = x_{ig} + F_u \cdot (x_{pbest,g} - x_{ig}) + F \cdot (x_{pr1,g} - x_{pr2,g})$$

  where

  $$F_u = \begin{cases} 
  0.7 \cdot F & \text{if } 0 \leq NFE < 0.2 \cdot NFE_{\text{max}} \\
  0.8 \cdot F & \text{if } 0.2 \cdot NFE_{\text{max}} \leq NFE < 0.4 \cdot NFE_{\text{max}} \\
  1.2 \cdot F & \text{otherwise}
  \end{cases}$$

  where $x_{pbest,g}$ denotes one of the top 100$p\%$ individuals with $p \in (0, 1]$. Also, $x_{pr1,g}$ denotes a random donor vector from a population based on rank-based probabilities, and $x_{pr2,g}$ denotes a random donor vector from a population based on rank-based probabilities or from an external archive. The probability of the $i$th individual being selected can be calculated as follows.

  $$pr_i = \frac{\text{Rank}_i}{\sum_{j=1}^{NP} \text{Rank}_j}$$

  where

  $$\text{Rank}_i = k \cdot (NP_g - i) + 1$$

  where $k$ denotes a rank greediness factor. Additionally, LSHADE-RSP uses a setting for increasing the number of $p$best individuals, which can be calculated as follows.

  $$p = 0.085 \cdot \left(1 + \frac{NFE}{NFE_{\text{max}}}\right)$$

4.2. Linear Population Size Reduction

The proposed algorithm uses linear population size reduction (LPSR) [9] to establish a balance between exploration and exploitation. The idea behind LPSR is to use a higher population size at the beginning of an optimization process and gradually reduce its value as a linear function. At the end of each generation, LPSR calculates the population size for the next generation $NP_{g+1}$ as follows.

  $$NP_{g+1} = \text{round} \left[ NP_{\text{init}} - \frac{N_{\text{ES}}}{N_{\text{ES}}_{\text{max}}} \cdot (NP_{\text{init}} - NP_{f\text{in}}) \right]$$

  where $NP_{\text{init}}$ and $NP_{f\text{in}}$ denote the initial and final population sizes, respectively. If the next population size $NP_{g+1}$ is smaller than the current one $NP_g$, the worst $NP_g - NP_{g+1}$ candidate solutions with respect to the fitness value are discarded. For the initial and final population sizes, LSHADE-RSP uses $NP_{\text{init}} = \text{round} \left( \sqrt{D} \cdot \log(D) \cdot 25 \right)$ and $NP_{f\text{in}} = 4$.

4.2.1. Adaptive Parameter Control

The proposed algorithm uses a learning process-based adaptive parameter control [11] to adjust the control parameters $F$ and $CR$ automatically. Each candidate solution has its control parameters $F_i,g$ and $CR_i,g$. At each generation, the control parameters $F_i,g$ and $CR_i,g$ are calculated as follows.

  $$F_i,g = \text{rndc}(M_{F,g}, 0.1)$$

  $$CR_i,g = \text{rrnd}(M_{CR,g}, 0.1)$$

  where $\text{rndc}$ and $\text{r rnd}$ denote the Cauchy and Gaussian distributions, respectively. Also, $M_{F,g}$ and $M_{CR,g}$ denote randomly selected values from historical memories $\mu_F$ and $\mu_{CR}$, respectively. The scaling factor is recalculated if $F_{i,g} \leq 0$ or truncated to 1 if $F_{i,g} > 1$. The crossover rate is first truncated to $[0, 1]$. After that, the crossover rate is modified as follows.

  $$CR_i,g = \begin{cases} 
  0.7 & \text{if } CR_i < 0.7 \text{ and } NFE < 0.25 \cdot NFE_{\text{max}} \\
  0.6 & \text{if } CR_i < 0.6 \text{ and } NFE < 0.5 \cdot NFE_{\text{max}} \\
  CR_i & \text{otherwise}
  \end{cases}$$

The historical memories $\mu_F$ and $\mu_{CR}$ store the successfully evolved candidate solutions’ control parameters. The capacity of the historical memories is $H$. At the beginning of an optimization process, all the entries, except the last one, of the memory $\mu_F$ are initialized to 0.3. Similarly, all the entries, except the last one, of the memory $\mu_{CR}$ are initialized to 0.8. The last entry of the memories $\mu_F$ and $\mu_{CR}$ always keep 0.9 during the optimization process. After the selection operator, the successfully evolved candidate solutions’ control parameters are stored in $S_F$ and $S_{CR}$. Then, one of the entries of the memories is updated as follows.
component of a random donor vector from a population based on rank-

where \( \text{meanWL} \) denotes the weighted Lehmer mean, which takes into consideration the improvement in fitness values between candidate solutions and their corresponding offspring.

4.3. Cauchy Perturbation

LSHADE-RSP \[17\] uses a rank-based selective pressure scheme, which tends to select higher ranking candidate solutions as donor vectors. Therefore, the scheme can increase the greediness, which can boost the convergence speed. However, the scheme may decrease the solution accuracy because of premature convergence in which all the candidate solutions fall into the local optimum of an optimization problem and cannot escape from there \[19, 20, 21, 22\]. Although LSHADE-RSP uses a setting for increasing the number of \( p \text{best} \) individuals, it may not be sufficient to compensate for the increased greediness.

To improve the exploration property of EAs, many researchers have developed new methods. Among them, using a long-tailed stable distribution, such as the Cauchy or Lévy distribution, in the phase of the recombination is one of the popular ones. By using a long-tailed stable distribution, EAs can generate candidate solutions over large distances, which can improve the exploration property. In the literature of DE, some researchers have demonstrated the effectiveness of using a long-tailed stable distribution in the phase of the recombination \[26, 27, 25\]. This motivated us to devise a modified recombination operator for LSHADE-RSP.

The idea behind the modified recombination operator is simple. When creating a trial vector, the operator first perturbs a target vector with the Cauchy distribution. After that, the operator creates a trial vector by recombining the perturbed target vector and its corresponding mutant vector. Therefore, a much different trial vector can be created by adopting the long-tail property of the Cauchy distribution. The novelty of the proposed approach lies in the perturbation of a target vector instead of a mutant vector. Fig. \[2\] illustrates the behavior of the modified recombination operator with DE/rand/1/bin.

As we mentioned earlier, the modified recombination operator is an extension of the recombination operator of LSHADE-RSP. The original operator can be defined as follows.

\[
M_{F,k} = \begin{cases} 
\text{meanWL}(S_F) & \text{if } S_F \neq \emptyset \\
M_{F,k} & \text{otherwise}
\end{cases}
\]

\[
M_{CR,k} = \begin{cases} 
\text{meanWL}(S_{CR}) & \text{if } S_{CR} \neq \emptyset \\
M_{CR,k} & \text{otherwise}
\end{cases}
\]

where \( \text{meanWL} \) denotes the weighted Lehmer mean, which takes into consideration the improvement in fitness values between candidate solutions and their corresponding offspring.

\[ \frac{M_{F,k}}{M_{F,k}} \]

\[ \frac{M_{CR,k}}{M_{CR,k}} \]

\[ \begin{cases} 
\text{meanWL}(S_F) & \text{if } S_F \neq \emptyset \\
M_{F,k} & \text{otherwise}
\end{cases} \]

\[ \begin{cases} 
\text{meanWL}(S_{CR}) & \text{if } S_{CR} \neq \emptyset \\
M_{CR,k} & \text{otherwise}
\end{cases} \]

on rank-based probabilities, and \( \tilde{x}_{pr,g}^j \) denotes the \( j \)-th component of a random donor vector from a population based on rank-based probabilities or from an external archive. The modified operator can be defined as follows.

\[
u_{i,g}^j = \begin{cases} 
x_{i,g}^j + F \cdot (x_{\text{best},g}^j - x_{i,g}^j) & \text{if } \text{rand}_i^j < CR \text{ or } j = j_{\text{rand}} \\
+F \cdot (x_{pr,1,g}^j - \tilde{x}_{pr,2,g}^j) & \text{otherwise}
\end{cases}
\]

where \( \text{rand}_i^j \) denotes the Cauchy distribution.

The proposed algorithm alternately applies one of the two recombination operators according to the jumping rate \( p_j \). When creating a trial vector, the proposed algorithm applies the original operator if a random number is higher than or equal to the rate. Otherwise, the proposed algorithm applies the modified operator. Algorithm \[\text{Algorithm} 1\] shows the pseudo-code of the proposed algorithm.
5.2. Test Algorithms

We used the following test algorithms for the comparative analysis.

- iLSAHDE-RSP: the proposed algorithm.

Algorithm 1: iLSAHDE-RSP

Input: Objective function \( f(x) \), lower bound \( x_{\min} \), upper bound \( x_{\max} \), maximum number of function evaluations \( NFE_{\max} \), and jumping rate \( p_j \)

Output: Final best objective value \( f(x_{\text{best, } G_{\max}}) \)

/* Initialization */
1 Set function evaluation \( NFE \leftarrow 0 \);
2 Set generation \( g \leftarrow 1 \);
3 Initialize population \( P_g = (x_{1,g}, \ldots, x_{N_{P,g}}) \) randomly;
4 Set archive \( A \leftarrow 0 \);
5 Set all elements in \( \mu_F \) to 0.3;
6 Set all elements in \( \mu_{CR} \) to 0.8;

/* Iteration */
7 while None of termination criteria is satisfied do
   /* Recombination operator */
   8 Set \( S_F \leftarrow 0, S_{CR} \leftarrow 0 \);
   9 for \( i = 0; i < NP; i = i + 1 \) do
      10 if \( f(u_{ig}) \leq f(x_{ig}) \) then
         11 Assign \( F_{i,g}, CR_{i,g} \) using Algorithm 2;
         12 \( u_{ig} \leftarrow \) the modified operator using Eq. (21);
         13 else \( u_{ig} \leftarrow \) the original operator using Eq. (16);
      end
      /* Selection operator */
      17 if \( f(u_{ig}) \leq f(x_{ig}) \) then
         18 \( x_{ig+1} \leftarrow u_{ig} \);
         19 \( x_{ig} \rightarrow A \);
         20 \( F_{i,g} \rightarrow S_F, CR_{i,g} \rightarrow S_{CR} \);
         21 else
            22 \( x_{ig+1} \leftarrow x_{ig} \);
      end
   end
26 Shrink \( P_{g+1} \) by discarding worst solutions;
27 Shrink \( A \) by discarding random solutions;
28 Update \( \mu_F, \mu_{CR} \);
29 \( p \leftarrow 0.085 \cdot \left( 1 + \frac{NFE}{NFE_{\max}} \right) \);
30 \( g \leftarrow g + 1 \);
end

Algorithm 2: Parameter Assignment

1 Select \( r_i \) from \([1, H]\) randomly;
2 if \( r_i = H \) then
   3 \( M_{FR_i} \leftarrow 0.9 \);
   4 \( M_{CR_i} \leftarrow 0.9 \);
5 end
6 \( F_{i,g} \leftarrow \text{rndc}(M_{FR_i}, 0.1) \);
7 if \( g < 0.6 \cdot NFE_{\max} \) and \( F_{i,g} > 0.7 \) then
   8 \( F_{i,g} < 0.7 \);
9 end
10 if \( M_{CR_i} < 0 \) then
11 \( CR_{i,g} \leftarrow 0 \);
12 else
13 \( CR_{i,g} \leftarrow \text{rndn}(M_{CR_i}, 0.1) \);
14 end
15 if \( g < 0.25 \cdot NFE_{\max} \) then
16 \( CR_{i,g} \leftarrow \text{max}(CR_{i,g}, 0.7) \);
17 else
18 if \( g < 0.5 \cdot NFE_{\max} \) then
19 \( CR_{i,g} \leftarrow \text{max}(CR_{i,g}, 0.6) \);
20 end
21 end

5. Experimental Setup

5.1. System Configuration

All the following experiments were performed on Windows 10 Pro 64 bit of a PC with AMD Ryzen Threadripper 2990WX @ 3.0GHz. The proposed and comparison algorithms were developed in the C++ programming language with Visual Studio 2019 64 bit.

5.2. Test Algorithms

We used the following test algorithms for the comparative analysis.

- LSHADE-RSP [17]: ranked the second place in the CEC 2018 competition on single objective optimization.
- jSO [11]: ranked the second place in the CEC 2017 competition on single objective optimization.
- L-SHADE [9]: ranked the first place in the CEC 2014 competition on single objective optimization.
- SHADE [8]: ranked the fourth place in the CEC 2013 competition on single objective optimization.
- JADE [21]: the origin of L-SHADE variants.
- EDEV [46]: a multi-population-based DE variant.
- MPEDE [45]: a multi-population-based DE variant.
- CoDE [40]: a composite DE variant.
- EPSDE [39]: an ensemble DE variant.
- SaDE [38]: a self-adaptive DE variant.
- dynNP-DE [30]: a self-adaptive DE variant.

The algorithms are six L-SHADE variants, two multi-population-based DE variants, and four well-known classical DE variants. The proposed algorithm introduces the jumping rate \( p_j \). As shown in Section 6, the proposed algorithm works best with \( p_j \in [0.15, 0.35] \). The proposed algorithm uses \( p_j = 0.2 \) in all the following experiments. Except for the jumping rate, the proposed algorithm uses the same values for the control parameters as its predecessor LSHADE-RSP.
5.3. Test Functions

To compare the proposed and comparison algorithms experimentally, we carried out experiments on the CEC 2017 test suite in 10, 30, 50, and 100 dimensions. The CEC 2017 test suite has 30 different and difficult optimization problems, such as three unimodal test functions \((F_1, F_3, F_5)\), seven simple multimodal test functions \((F_2, F_10, F_15, F_20, F_23, F_30, F_33)\), ten expanded multimodal test functions \((F_21, F_31, F_32, F_35, F_36, F_38, F_39, F_40, F_41, F_42)\), and ten hybrid composition test functions \((F_26, F_28, F_29, F_34, F_43, F_44, F_45, F_46, F_47, F_48)\). A function is said to be unimodal if it has no local optima, while a function is said to be multimodal if it has multiple local optima.

According to the experimental setups of the test suite, the maximum number of function evaluations \(NFE_{\text{max}}\) was set to 10,000 \(\cdot\) \(D\). Moreover, the search boundaries of the test suite were set to \([-100,100]^D\). For more detailed information, please refer to the following papers \([28]\).

5.4. Performance Metrics

5.4.1. Function Error Value

The function error value (FEV) is utilized to assess the test algorithm’s accuracy. The FEV is the difference between the final best objective value of a test algorithm and the global optimum of an optimization problem, which can be defined as follows.

\[
\text{FEV} = f(x_{\text{best,global}}) - f(x) \tag{18}
\]

where \(f(x)\) denotes an objective function. Also, \(x_{\text{best,global}}\) and \(x\) denote the final best objective value and the global optimum, respectively.

5.4.2. Statistical Test

We utilized the Wilcoxon rank-sum test and the Friedman test with Hochberg’s post hoc for the comparative analysis. The former is used to test the statistical significance of two test algorithms, while the latter is used to test the statistical significance of multiple test algorithms \([64]\).

6. Experimental Results and Discussion

In this section, we present the experimental results and discussion on the CEC 2017 test suite in 10, 30, 50, and 100 dimensions.

6.1. Comparative Analysis

We present the comparative analysis of the test algorithms in this subsection. Tables \(1, 3, 5\) and \(7\) present the means and standard deviations of the FEVs of the test algorithms in 10, 30, 50, and 100 dimension, respectively. In the tables, the symbols “+”, “=”, and “-” denote that the corresponding algorithm has statistically better, similar or worse performance compared to the proposed algorithm, respectively. Moreover, Tables \(2, 4, 6\) and \(8\) present the results of the Friedman test with Hochberg’s post hoc in 10, 30, 50, and 100 dimension, respectively. Furthermore, Figs. \(3, 5, 6, 7\) provide the convergence graphs of the test algorithms in 100 dimension.

Table \(1\) presents the means and standard deviations of the FEVs of the test algorithms in 10 dimension, obtained by 51 independent runs. As can be seen from the table, the proposed algorithm performs better performance than all of the other test algorithms. Specifically, iLSHADE-RSP found significantly better solutions with lower FEVs than SHADE, JADE, EDEV, MPEDE, EPSDE, and SaDE on more than 50 percent of the test functions. In particular, MPEDE and CoDE were significantly outperformed by iLSHADE-RSP on approximately 80 percent of the test functions. As compared to its predecessor LSHADE-RSP, the proposed algorithm considerably outperformed on 2 test functions and underperformed it on 0 test functions. In addition, Table \(2\) presents the Friedman test with Hochberg’s post hoc, which supports the comparative analysis in Table \(1\) where iLSHADE-RSP ranked the first among the test algorithms, and the outperformance over SHADE, JADE, EDEV, MPEDE, CoDE, EPSDE, and SaDE was statistically significant.

The means and standard deviations of the FEVs of the proposed and comparison algorithms in 30 dimension are shown in Table \(3\) collected by 51 independent runs. As can be seen from the table, the proposed algorithm performs better performance than all of the other test algorithms. Specifically, iLSHADE-RSP found significantly better solutions with lower FEVs than SHADE, JADE, EDEV, MPEDE, CoDE, EPSDE, SaDE, and dynNP-DE on more than 80 percent of the test functions. In particular, JADE, EDEV, MPEDE, CoDE, EPSDE, and SaDE were not able to outperform iLSHADE-RSP on any of the test functions. As compared to its predecessor LSHADE-RSP, the proposed algorithm considerably outperformed on 6 test functions and underperformed it on 1 test functions. Additionally, Table \(4\) presents the Friedman test with Hochberg’s post hoc, which supports the comparative analysis in Table \(3\) where iLSHADE-RSP ranked the first among the test algorithms, and the outperformance over SHADE, JADE, EDEV, MPEDE, CoDE, EPSDE, SaDE, and dynNP-DE was statistically significant.

Table \(5\) presents the means and standard deviations of the FEVs of the test algorithms in 50 dimension, obtained by 51 independent runs. As can be seen from the table, the proposed algorithm performs better performance than all of the other test algorithms. Specifically, iLSHADE-RSP found significantly better solutions with lower FEVs than all of the other test algorithms except LSHADE-RSP on more than 50 percent of the test functions. In particular, SHADE, JADE, EDEV, MPEDE, CoDE, EPSDE, SaDE, and dynNP-DE were significantly outperformed by iLSHADE-RSP on approximately 90 percent of the test functions. As compared to its predecessor LSHADE-RSP, the proposed algorithm considerably outperformed on 8 test functions and underperformed it on 4 test functions. In addition, Table \(6\) presents the Friedman test with Hochberg’s post hoc, which supports the comparative analysis in Table \(5\) where iLSHADE-RSP ranked the first among the test algorithms, and the outperformance over SHADE, JADE, EDEV, MPEDE, CoDE, EPSDE, SaDE, and dynNP-DE was statistically significant.

The means and standard deviations of the FEVs of the pro-
Table 1: Means and standard deviations of FEVs of test algorithms on CEC 2017 test suite in 10 dimension

| Algorithm     | Mean (STD DEV) | Mean (STD DEV) | Mean (STD DEV) | Mean (STD DEV) | Mean (STD DEV) |
|---------------|----------------|----------------|----------------|----------------|----------------|
| iLSHADE-RSP   | 1.06E-01 (4.94E-01) | 3.18E-01 (9.48E-02) | 0.00E (0.00E)  | 9.18E-02 (2.22E-02) | 0.00E (0.00E)  |
| L-SHADE-RSP   | 1.06E-01 (3.48E-01) | 3.75E-01 (3.54E-01) | 0.00E (0.00E)  | 1.25E-02 (3.43E-03) | 0.00E (0.00E)  |
| JOO           | 4.41E-02 (2.81E-02) | 3.42E-01 (3.14E-01) | 0.00E (0.00E)  | 1.04E-02 (5.31E-03) | 0.00E (0.00E)  |
| L-SHADE       | 1.06E-01 (2.81E-02) | 4.27E-02 (7.59E-02) | 0.00E (0.00E)  | 8.14E-02 (4.02E-02) | 0.00E (0.00E)  |
| SHADE         | 4.17E-02 (1.81E-02) | 3.56E-01 (3.66E-01) | 0.00E (0.00E)  | 7.77E-02 (3.28E-02) | 0.00E (0.00E)  |
| JADE          | 1.06E-01 (1.98E-01) | 8.06E-01 (3.00E-01) | 0.00E (0.00E)  | 1.24E-01 (3.28E-02) | 0.00E (0.00E)  |
| EDEV          | 1.06E-01 (1.98E-01) | 8.06E-01 (3.00E-01) | 0.00E (0.00E)  | 1.24E-01 (3.28E-02) | 0.00E (0.00E)  |

The symbols "+" and "−" indicate that the corresponding algorithm performed significantly better (+), not significantly better or worse (=), or significantly worse (−) compared to iLSHADE-RSP using the Wilcoxon rank-sum test with α = 0.05 significance level.

Table 2: Friedman test with Hochberg's post hoc for test algorithms on CEC 2017 test suite in 10 dimension

| Algorithm     | Average ranking | p-value | p-value | Adjust. p-value (Hochberg) | Sig. | Test statistics |
|---------------|-----------------|---------|---------|---------------------------|------|----------------|
| 1             | iLSHADE-RSP     | 4.15    |         |                            |      |                |
| 2             | L-SHADE-RSP     | 4.42    | 2.69E-01| 0.075E-01                   | 90   |                |
| 3             | JOO             | 5.30    | 1.24E-01| 0.075E-01                   | 90   |                |
| 4             | L-SHADE         | 5.21    | 1.16E-01| 0.245E-01                   | 75   |                |
| 5             | SHADE           | 6.12    | 2.11E-01| 1.75E-01                    | 75   |                |
| 6             | JADE            | 7.05    | 3.12E-01| 1.84E-02                    | 75   |                |
| 7             | EDEV            | 7.35    | 3.44E-01| 5.87E-04                    | 75   |                |
| 8             | MPEDE           | 9.97    | 6.29E-02| 4.55E-02                    | 75   |                |
| 9             | CoDe            | 7.97    | 3.99E-01| 5.84E-04                    | 75   |                |
| 10            | EPSDE           | 7.62    | 3.72E-01| 1.96E-04                    | 75   |                |
| 11            | SaDE            | 7.62    | 6.03E-01| 5.62E-04                    | 75   |                |
| 12            | dynNP-DE        | 5.03    | 4.49E-01| 6.85E-04                    | 75   |                |

The results indicate that iLSHADE-RSP performs significantly better than the other test algorithms except for JADE, which is not significantly different from iLSHADE-RSP.
Table 3: Means and standard deviations of FEVs of test algorithms on CEC 2017 test suite in 30 dimension

| Algorithm     | Average ranking | p-value | p-value (Hochberg) | Sig. | Test statistics |
|---------------|----------------|---------|--------------------|------|----------------|
| LSHADE-RSP    | 3.95           | 0.0000  | 0.0000             | No   | N              |
| LSHADE-RSP    | 3.90           | 0.0000  | 0.0000             | No   | N              |
| LSHADE-RSP    | 3.62           | 0.0000  | 0.0000             | No   | N              |
| LSHADE-RSP    | 3.80           | 0.0000  | 0.0000             | No   | N              |
| LSHADE-RSP    | 7.80           | 0.0000  | 0.0000             | Yes  | 1.8E-07         |
| LSHADE-RSP    | 4.63           | 0.0001  | 0.0001             | Yes  | 1.8E-07         |
| Jade          | 7.93           | 0.0000  | 0.0000             | Yes  | 1.8E-07         |
| Jade          | 7.96           | 0.0000  | 0.0000             | Yes  | 1.8E-07         |
| Jade          | 9.12           | 0.0000  | 0.0000             | Yes  | 1.8E-07         |
| Jade          | 6.28           | 0.0000  | 0.0000             | Yes  | 1.8E-07         |

The symbols "+"/- indicate that the corresponding algorithm performed significantly better (+), not significantly better or worse (=), or significantly worse (-) compared to iLSHADE-RSP using the Wilcoxon rank-sum test with α = 0.05 significance level.

Table 4: Friedman test with Hochberg’s post hoc for test algorithms on CEC 2017 test suite in 30 dimension

| Algorithm     | Average ranking | p-value | p-value (Hochberg) | Sig. | Test statistics |
|---------------|----------------|---------|--------------------|------|----------------|
| LSHADE-RSP    | 3.95           | 0.0000  | 0.0000             | No   | N              |
| LSHADE-RSP    | 3.90           | 0.0000  | 0.0000             | No   | N              |
| LSHADE-RSP    | 3.62           | 0.0000  | 0.0000             | No   | N              |
| LSHADE-RSP    | 3.80           | 0.0000  | 0.0000             | No   | N              |
| LSHADE-RSP    | 7.80           | 0.0000  | 0.0000             | Yes  | 1.8E-07         |
| LSHADE-RSP    | 4.63           | 0.0001  | 0.0001             | Yes  | 1.8E-07         |
| Jade          | 7.93           | 0.0000  | 0.0000             | Yes  | 1.8E-07         |
| Jade          | 7.96           | 0.0000  | 0.0000             | Yes  | 1.8E-07         |
| Jade          | 9.12           | 0.0000  | 0.0000             | Yes  | 1.8E-07         |
| Jade          | 6.28           | 0.0000  | 0.0000             | Yes  | 1.8E-07         |

The symbols "+"/- indicate that the corresponding algorithm performed significantly better (+), not significantly better or worse (=), or significantly worse (-) compared to iLSHADE-RSP using the Wilcoxon rank-sum test with α = 0.05 significance level.

The convergence graphs of the proposed and comparison algorithms in 100 dimension are provided in Figs 3, 4, 5 and 7.
Each convergence graph provides the median and interquartile ranges (one-fourth and three-fourth) of the FEVs of the proposed and comparison algorithms. The first part of the convergence graphs (F₁ - F₁₀) is given in Fig. 3. The second part of the convergence graphs (F₁₁ - F₁₂) is given in Fig. 4. The third part of the convergence graphs (F₁₃ - F₁₈) is given in Fig. 5.

The symbols "*" and "+" indicate that the corresponding algorithm performed significantly better (+), not significantly better or worse (=), or significantly worse (−) compared to iLSHADE-RSP using the Wilcoxon rank-sum test with ρ = 0.05 significance level.

| Algorithm | Mean rank | p-value | p-value (Hochberg) | Sig. | Test statistics |
|-----------|-----------|---------|-------------------|-----|----------------|
| 1 | iLSHADE-RSP | 2.47 | | | |
| 2 | LSHADE-RSP | 2.63 | 0.10 | 0.01 | No |
| 3 | SMADE | 3.96 | 0.00 | 0.00 | Yes |
| 4 | LSHADE | 3.71 | 0.01 | 0.00 | No |
| 5 | MIDE | 4.08 | 0.00 | 0.00 | Yes |
| 6 | JADE | 4.93 | 0.00 | 0.00 | Yes |
| 7 | DE | 6.14 | 0.19 | 0.05 | Yes |
| 8 | MIDE | 6.77 | 0.01 | 0.01 | Yes |
| 9 | CAGE | 6.07 | 0.00 | 0.00 | Yes |
| 10 | EPDE | 9.94 | 0.00 | 0.00 | Yes |
| 11 | NDE | 9.83 | 0.00 | 0.00 | Yes |
| 12 | dyNE-DDE | 7.08 | 0.00 | 0.00 | Yes |

The symbols "*" and "+" indicate that the corresponding algorithm performed significantly better (+), not significantly better or worse (=), or significantly worse (−) compared to iLSHADE-RSP using the Wilcoxon rank-sum test with ρ = 0.05 significance level.

Table 6: Friedman test with Hochberg's post hoc tests for algorithms on CEC 2017 test suite in 50 dimension
| Algorithm | Average ranking | p-value | p-value | Adj. p-value (Bonferroni) |
|-----------|----------------|---------|---------|-------------------------|
| iLSHADE-RSP | 2.55 |   |   |               |
| LSHADE-RSP | 2.87 | 0.01 | 0.96 | 0.23 |
| 4 | 3.77 | 0.01 | 0.96 | 0.23 |
| 7 | 3.77 | 0.01 | 0.96 | 0.23 |
| 10 | 6.07 | 0.01 | 0.96 | 0.23 |
| 13 | 5.60 | 0.01 | 0.96 | 0.23 |
| 16 | 6.07 | 0.01 | 0.96 | 0.23 |
| 19 | 6.07 | 0.01 | 0.96 | 0.23 |
| 21 | 6.07 | 0.01 | 0.96 | 0.23 |

The symbols "*" and "**" indicate that the corresponding algorithm performed significantly better (+), not significantly better or worse (=), or significantly worse (−) compared to iLSHADE-RSP using the Wilcoxon rank-sum test with α = 0.05 significance level.

Table 8: Friedman test with Hochberg’s post hoc test for algorithms on CEC 2017 test suite in 100 dimension

The results show a discussion on the comparative analysis between the proposed algorithm and each of the two state-of-the-art L-SHADE variants, LSHADE-RSP and ISO. After that, we discuss the experimental results with the proposed algorithm and the other

accuracy, especially for the test functions F5, F7 - F9, F11 - F21, and F23 - F30. In particular, iLSHADE-RSP can escape from the local optimum of the test functions F20 and F30, while the other test algorithms cannot.

6.2. Discussion of Comparative Analysis

We first discuss the comparative analysis between the proposed algorithm and each of the two state-of-the-art L-SHADE variants, LSHADE-RSP and ISO. After that, we discuss the experimental results with the proposed algorithm and the other
test algorithms.

- The performance difference between the proposed algorithm and LSHADE-RSP is negligible in 10 dimensions. The proposed algorithm has only two improvements against zero deteriorations. However, the performance difference is much different in 30, 50, and 100 dimensions. The proposed algorithm has six improvements against one deterioration in 30 dimension, eight improvements against...
four deteriorations in 50 dimension, and eight improvements against five deteriorations in 100 dimension. Additionally, we investigated the performance difference with respect to the characteristics of the test functions. We found out that the proposed algorithm has worse performance on the unimodal ($F_1$-$F_3$) and some of the simple multimodal test functions ($F_4$-$F_{10}$) but better performance on the expanded multimodal ($F_{11}$-$F_{20}$) and the hy-
Figure 7: Convergence graphs of test algorithms on CEC 2017 test suites in 100 dimension ($F_{25}$ - $F_{30}$).

- The performance difference between the proposed algorithm and jSO is negligible in 10 dimension. The proposed algorithm has only five improvements against three deteriorations. However, the performance difference is much different in 30, 50, and 100 dimensions. The proposed algorithm has 11 improvements against six deteriorations in 30 dimension, 17 improvements against two deteriorations in 50 dimension, and 18 improvements against six deteriorations in 100 dimension. Additionally, we investigated the performance difference with respect to the characteristics of the test functions. We found out that the proposed algorithm has worse performance on the unimodal ($F_1$-$F_3$) and some of the simple multimodal test functions ($F_4$-$F_{10}$) but better performance on the expanded multimodal ($F_{11}$-$F_{20}$) and the hybrid composition test functions ($F_{21}$-$F_{30}$). In a word, the proposed algorithm performs better than jSO on more complicated optimization problems.

- The proposed algorithm found more significantly accurate solutions compared with the test algorithms, including L-SHADE, SHADE, JADE, EDEV, MPEDE, CoDE, EPSDE, SaDE, and dynNP-DE, in all the dimensions.

Note that, the experimental results with the proposed algorithm and LSHADE-RSP lend weight to the effectiveness of the modified recombination operator of the proposed algorithm, which can increase the probability of finding an optimal solution by adopting the long-tailed property of the Cauchy distribution, and thus, it can improve the optimization performance of LSHADE-RSP significantly.

6.3. Analysis of iLSHADE-RSP

6.3.1. Parameter Values for Jumping Rate

As we mentioned earlier, the proposed algorithm alternately applies one of the two recombination operators according to the jumping rate $p_j$. The jumping rate determines the additional exploration of the proposed algorithm. If the jumping rate is too high, the modified recombination operator is applied too often, and thus, the proposed algorithm might not be beneficial from existing candidate solutions. On the other hand, if the jumping rate is too low, the modified recombination operator is applied too rarely, and thus, the additional exploration of the proposed algorithm might be negligible. We carried out experiments to find out appropriate parameter values for the jumping rate. Table 9 shows the means and standard deviations of the FEVs of the proposed algorithm with different parameter values for the jumping rate. As can be seen from the table, the parameter values $p_j \in [0.15, 0.35]$ can lead to satisfactory results.

6.3.2. Lévy Perturbation

As we mentioned earlier, the proposed algorithm employs a modified recombination operator, which calculates a perturbation of a target vector with the Cauchy distribution. The Cauchy distribution is a special case of the Lévy $\alpha$-stable distribution. Therefore, it is interesting to test the other cases of the Lévy $\alpha$-stable distribution for the modified recombination operator.
The symbols "+/−" indicate that the corresponding algorithm performed significantly better (+), not significantly better or worse (=), or significantly worse (−) compared to iLSHADE-RSP with $p_{ij} = 0.2$ using the Wilcoxon rank-sum test with $\alpha = 0.05$ significance level.

The symbols "+/−/−" indicate that the corresponding algorithm performed significantly better (+), not significantly better or worse (=), or significantly worse (−) compared to iLSHADE-RSP with $\alpha = 1.0$ using the Wilcoxon rank-sum test with $\alpha = 0.05$ significance level.

Lévy $\alpha$-stable distribution can be defined by the characteristic function, derived by Lévy [65] and Hall [66] as follows.

$$
\log(\phi) = \begin{cases} 
-\epsilon^\alpha|\!|1 - i\beta\sin(\alpha\theta)\tan\frac{\alpha\theta}{2}\!| + \mu\| & \text{if } \alpha \neq 1 \\
-\epsilon^\alpha|\!|1 + i\beta\sin(\alpha\theta)\!|\log|\!|\!| + \mu\| & \text{if } \alpha = 1
\end{cases} 
$$

The Lévy $\alpha$-stable distribution denoted by $S_\alpha(\beta, c, \mu)$ has four parameters: the stability parameter $\alpha \in [0, 2]$, the skewness parameter $\beta \in [-1, 1]$, the scale parameter $c \in (0, \infty)$, and the location parameter $\mu \in (-\infty, \infty)$. The Lévy $\alpha$-stable distribution has three special cases as follows.

$$
\log(\phi) = \begin{cases} 
-\epsilon^\alpha|\!|1 - i\beta\sin(\alpha\theta)\tan\frac{\alpha\theta}{2}\!| + \mu\| & \text{if } \alpha \neq 1 \\
-\epsilon^\alpha|\!|1 + i\beta\sin(\alpha\theta)\!|\log|\!|\!| + \mu\| & \text{if } \alpha = 1
\end{cases} 
$$

where

$$
\log(\phi) = \begin{cases} 
-\epsilon^\alpha|\!|1 - i\beta\sin(\alpha\theta)\tan\frac{\alpha\theta}{2}\!| + \mu\| & \text{if } \alpha \neq 1 \\
-\epsilon^\alpha|\!|1 + i\beta\sin(\alpha\theta)\!|\log|\!|\!| + \mu\| & \text{if } \alpha = 1
\end{cases} 
$$

The Lévy $\alpha$-stable distribution denoted by $S_\alpha(\beta, c, \mu)$ has four parameters: the stability parameter $\alpha \in [0, 2]$, the skewness parameter $\beta \in [-1, 1]$, the scale parameter $c \in (0, \infty)$, and the location parameter $\mu \in (-\infty, \infty)$. The Lévy $\alpha$-stable distribution has three special cases as follows.
• The Gaussian distribution: \( S \) \( \beta, 2c^2, \mu \).
• The Cauchy distribution: \( S \) \( 1, c, \mu \).
• The Lévy distribution: \( S_{0.5} \) \( 1, c, \mu \).

The modified recombination operator, which calculates a perturbation of a target vector with the Lévy \( \alpha \)-stable distribution can be defined as follows.

\[
W^j = \begin{cases} 
    x_{j,g}^i + F_w \cdot (x_{\text{pbest},g}^j - x_{j,g}^i) & \text{if } \text{rand}_j < CR \text{ or } j = j_{\text{rand}} \\
    S_{\alpha}(0, 0.1, x_{j,g}^i) & \text{otherwise}
\end{cases}
\]

We used Chambers-Mallows-Stuck method \([67, 68, 69, 70]\) to generate Lévy \( \alpha \)-stable random numbers as follows.

**Step 1.** Generate a random number from the uniform distribution \( V \in [-\frac{\pi}{2}, \frac{\pi}{2}] \) and a random number from the exponential distribution \( W \) with mean 1.

**Step 2.** If \( \alpha \neq 1 \) then:

\[
X = S_{\alpha,\beta} \sin[\alpha(V + B_{\alpha,\beta})] \left[ \frac{\cos(V - \alpha(V + B_{\alpha,\beta}))}{W} \right]^{1-\alpha}
\]

where

\[
B_{\alpha,\beta} = \frac{\arctan(\beta \tan \frac{\pi \alpha}{2})}{\alpha}
\]

\[
S_{\alpha,\beta} = \left[ 1 + \beta^2 \tan^2 \left( \frac{\pi \alpha}{2} \right) \right]^{\frac{1}{2\alpha}}
\]

**Step 3.** If \( \alpha = 1 \) then:

\[
X = \frac{2}{\alpha} \cdot \left( \frac{\pi}{2} + \beta V \right) \tan V - \frac{2}{\alpha} \cdot \beta \ln \left( \frac{\alpha W \cos V}{2} + \beta V \right)
\]

**Step 4.** If \( X \sim S_{\alpha}(\beta, 1, 0) \) then:

\[
Y = \begin{cases} 
    cX + \mu & \text{if } \alpha \neq 1 \\
    cX + \frac{1}{2} \beta \text{clogc} + \mu & \text{if } \alpha = 1
\end{cases}
\]

is \( S_{\alpha}(\beta, c, \mu) \).

Table \( \text{T1} \) shows the means and standard deviations of the FEVs of the proposed algorithm with different stability parameters for the Lévy \( \alpha \)-stable distribution with the jumping rate \( p_j = 0.2 \). For simplicity’s sake, we only considered symmetric distributions. The Lévy \( \alpha \)-stable distribution has a short and wide PDF if the stability parameter is high, while a tall and narrow PDF if the stability parameter is low. As can be seen from the table, the parameter values \( \alpha = [1, 1.5] \) can lead to satisfactory results. In other words, the modified recombination operator with from the Cauchy distribution (\( \alpha = 1 \)) to the Lévy \( \alpha \)-stable distribution (\( \alpha = 1.5 \)) works best for the proposed algorithm. We chose the Cauchy distribution for the modified recombination operator because generating Cauchy random numbers is much easier than generating Lévy \( \alpha \)-stable random numbers.

7. Conclusion

Differential evolution is a popular evolutionary algorithm for multidimensional real-valued functions. Like other evolutionary algorithms, it is important for differential evolution to establish a balance between exploration and exploitation to be successful. Recently, a state-of-the-art DE variant called LSHADE-RSP was proposed. Although it has shown excellent performance, the greediness of LSHADE-RSP may cause premature convergence in which all the candidate solutions fall into the local optimum of an optimization problem and cannot escape from there.

To mitigate the problem, we have devised a modified recombination operator for LSHADE-RSP, which perturbs a target vector with the Cauchy distribution. Therefore, the modified recombination operator can increase the probability of finding an optimal solution by adopting the long-tailed property of the Cauchy distribution. We called the resulting algorithm iLSHADE-RSP, which alternately applies the original and modified recombination operators according to a jumping rate.

The proposed algorithm has been tested on the CEC 2017 test suite in 10, 30, 50, and 100 dimensions. Our experimental results verify that the improved LSHADE-RSP significantly outperformed not only its predecessor LSHADE-RSP but also several cutting-edge DE variants in terms of convergence speed and solution accuracy. In particular, the proposed algorithm performs better than the comparison algorithms on more complicated optimization problems, such as expanded multimodal test functions and hybrid composition test functions, in all the dimensions.

Possible directions for future work include 1) testing the proposed algorithm for constrained optimization problems; 2) testing the proposed algorithm for large-scale optimization problems; 3) applying the proposed algorithm to various real-world problems.

Acknowledgement

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. NRF-2017R1C1B2012752). The correspondence should be addressed to Dr. Chang Wook Ahn.

References

[1] M. Črepinšek, S.-H. Liu, M. Mernik, Exploration and exploitation in evolutionary algorithms: A survey, ACM computing surveys (CSUR) 45 (3) (2013) 1–33.
[2] A. E. Eiben, C. A. Schippers, On evolutionary exploration and exploitation, Fundamenta Informaticae 35 (1-4) (1998) 35–50.
[3] R. Storn, K. Price, Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces, Journal of global optimization 11 (4) (1997) 341–359.
[4] K. Price, R. M. Storn, J. A. Lampinen, Differential evolution: a practical approach to global optimization, Springer Science & Business Media, 2006.
[5] S. Das, P. N. Suganthan, Differential evolution: A survey of the state-of-the-art, IEEE transactions on evolutionary computation 15 (1) (2010) 4–31.
[6] S. Das, S. S. Mullick, P. N. Suganthan, Recent advances in differential evolution—an updated survey, Swarm and Evolutionary Computation 27 (2018) 1–30.
[7] J. Zhang, A. C. Sanderson, JADE: adaptive differential evolution with optional external archive, IEEE Transactions on evolutionary computation 13 (5) (2009) 945–958.
[8] R. Tanabe, A. Fukunaga, Success-history based parameter adaptation for differential evolution, in: 2013 IEEE congress on evolutionary computation, IEEE, 2013, pp. 71–78.
[9] R. Tanabe, A. S. Fukunaga, Improving the search performance of shade using linear population size reduction, in: 2014 IEEE congress on evolutionary computation (CEC), IEEE, 2014, pp. 1658–1665.
[10] J. Brest, M. S. Maucic, B. Boskovic, l-SHADE: Improved L-SHADE algorithm for single objective real-parameter optimization, in: 2016 IEEE Congress on Evolutionary Computation (CEC), IEEE, 2016, pp. 1188–1195.
[11] J. Brest, M. S. Maucic, B. Boskovic, Single objective real-parameter optimization: algorithm jSO, in: 2017 IEEE congress on evolutionary computation (CEC), IEEE, 2017, pp. 1311–1318.
[12] N. H. Awad, M. Z. Ali, P. N. Suganthan, R. G. Reynolds, An ensemble sinusoidal parameter adaptation incorporated with L-SHADE for solving CEC2014 benchmark problems, in: 2016 IEEE congress on evolutionary computation (CEC), IEEE, 2016, pp. 2958–2965.
[13] N. H. Awad, M. Z. Ali, P. N. Suganthan, R. G. Reynolds, A. M. Shatnawi, A novel differential crossover strategy based on covariance matrix learning with Euclidean neighborhood for solving real-world problems, in: 2017 IEEE Congress on Evolutionary Computation (CEC), IEEE, 2017, pp. 380–386.
[14] N. H. Awad, M. Z. Ali, P. N. Suganthan, Ensemble sinusoidal differential covariance matrix adaptation with Euclidean neighborhood for solving CEC2017 benchmark problems, in: 2017 IEEE Congress on Evolutionary Computation (CEC), IEEE, 2017, pp. 372–379.
[15] N. H. Awad, M. Z. Ali, P. N. Suganthan, Ensemble of parameters in a sinusoidal differential evolution with niching-based population reduction, Swarm and evolutionary computation 39 (2018) 141–156.
[16] A. W. Mohamed, A. A. Hadi, A. M. Fattouh, K. M. Jambi, LSHADE with semi-parameter adaptation hybrid with CMA-ES for solving CEC 2017 benchmark problems, in: 2017 IEEE Congress on evolutionary computation (CEC), IEEE, 2017, pp. 145–152.
[17] V. Stanovov, S. Akhmedova, E. Semenkin, LSHADE algorithm with rank-based selective pressure strategy for solving CEC 2017 benchmark problems, in: 2018 IEEE Congress on Evolutionary Computation (CEC), IEEE, 2018, pp. 1–8.
[18] J. F. Yeh, T.-Y. Chen, T.-C. Chiang, Modified L-SHADE for single objective real-parameter optimization, in: 2019 IEEE Congress on Evolutionary Computation (CEC), IEEE, 2019, pp. 381–386.
[19] J. E. Baker, Adaptive selection methods for genetic algorithms, in: Proceedings of an International Conference on Genetic Algorithms and their applications, Hillsdale, New Jersey, 1985, pp. 101–111.
[20] K. A. De Jong, Analysis of the behavior of a class of genetic adaptive systems, Tech. rep. (1975).
[21] S. J. Louis, G. J. Rawlins, Syntactic analysis of convergence in genetic algorithms, in: Foundations of genetic algorithms, Vol. 2, Elsevier, 1993, pp. 141–151.
[22] D. E. Goldberg, P. Segrest, Finite markov chain analysis of genetic algorithms, in: Proceedings of the second international conference on genetic algorithms, Vol. 1, 1987, p. 1.
[23] T. J. Choi, J. Togelius, Y.-G. Cheong, Advanced Cauchy mutation for differential evolution in numerical optimization, IEEE Access 8 (2020) 8720–8734.
[24] T. J. Choi, J. Togelius, Y.-G. Cheong, ACM-DE: Adaptive p-best Cauchy mutation with linear failure threshold reduction for differential evolution in numerical optimization, arXiv preprint arXiv:1907.01095.
[25] H. Zhang, X. Lei, C. Wang, D. Yue, X. Xie, Adaptive grid based multi-objective Cauchy differential evolution for stochastic dynamic economic dispatch with wind power uncertainty, PloS one 12 (9).
[26] M. Ali, M. Pant, Improving the performance of differential evolution algorithm using Cauchy mutation, Soft Computing 15 (5) (2011) 991–1007.
[27] H. Qin, J. Zhou, Y. Lu, Y. Wang, Y. Zhang, Multi-objective differential evolution with adaptive Cauchy mutation for short-term multi-objective optimal hydro-thermal scheduling, Energy Conversion and Management 51 (4) (2010) 788–794.
[28] N. Awad, M. Ali, J. Liang, B. Qu, P. Suganthan, Problem definitions and evaluation criteria for the CEC 2017 special session and competition on single objective real-parameter numerical optimization, Tech. Rep.
[29] J. Brest, S. Greiner, B. Boskovic, M. Mernik, V. Zumer, Self-adapting control parameters in differential evolution: A comparative study on numerical benchmark problems, IEEE transactions on evolutionary computation 10 (6) (2006) 646–657.
[30] J. Brest, M. S. Maucic, Population size reduction for the differential evolution algorithm, Applied Intelligence 29 (3) (2008) 228–247.
[31] T. J. Choi, C. W. Ahn, J. An, An adaptive Cauchy differential evolution algorithm for global numerical optimization, The Scientific World Journal 2013.
[32] M. Leon, N. Xiong, Adapting differential evolution algorithms for continuous optimization via gradual adjustment of control parameters, Journal of Artificial Intelligence and Soft Computing Research 6 (2) (2016) 103–118.
[33] T. J. Choi, C. W. Ahn, An improved differential evolution algorithm and its application to large-scale artificial neural networks, in: Journal of Physics: Conference Series, Vol. 806, IOP Publishing, 2017, p. 012010.
[34] T. J. Choi, C. W. Ahn, Adaptive α-stable differential evolution in numerical optimization, Natural Computing 16 (4) (2017) 637–657.
[35] P. Bujok, J. Tvedik, Enhanced individual-dependent differential evolution with population size adaptation, in: 2017 IEEE congress on evolutionary computation (CEC), IEEE, 2017, pp. 1358–1365.
[36] T. J. Choi, Y. Lee, Asynchronous differential evolution with self-adaptive parameter control for global numerical optimization, in: MATEC Web of Conferences, Vol. 189, EDP Sciences, 2018, p. 03020.
[37] T. J. Choi, Y.-G. Cheong, C. W. Ahn, A performance comparison of crossover variations in differential evolution for training multi-layer perceptron neural networks, in: International Conference on Bio-Inspired Computing: Theories and Applications, Springer, 2018, pp. 477–488.
[38] A. K. Qin, V. L. Huang, P. N. Suganthan, Differential evolution algorithm with strategy adaptation for global numerical optimization, IEEE transactions on Evolutionary Computation 13 (2) (2008) 398–417.
[39] R. Mallipeddi, P. N. Suganthan, Q.-K. Pan, M. F. Tasgetiren, Differential evolution algorithm with ensemble of parameters and mutation strategies, Applied soft computing 11 (2) (2011) 1679–1696.
[40] Y. Wang, Z. Cai, Q. Zhang, Differential evolution with composite trial vector generation strategies and control parameters, IEEE transactions on evolutionary computation 15 (1) (2011) 55–66.
[41] T. J. Choi, C. W. Ahn, An adaptive Cauchy differential evolution algorithm with bias strategy adaptation mechanism for global numerical optimization., JCP 9 (9) (2014) 2139–2145.
[42] T. J. Choi, C. W. Ahn, An adaptive differential evolution algorithm with automatic population resizing for global numerical optimization, in: Bio-Inspired Computing-Theories and Applications, Springer, 2014, pp. 68–72.
[43] T. J. Choi, C. W. Ahn, An adaptive population resizing scheme for differential evolution in numerical optimization, Journal of Computational and Theoretical Nanoscience 12 (7) (2015) 1336–1350.
[44] T. J. Choi, C. W. Ahn, An adaptive Cauchy differential evolution algorithm with population size reduction and modified multiple mutation strategies, in: Proceedings of the 18th Asia Pacific Symposium on Intelligent and Evolutionary Systems-Volume 2, Springer, 2015, pp. 13–26.
[45] G. Wu, R. Mallipeddi, P. N. Suganthan, R. Wang, H. Chen, Differential evolution with multiple-population based ensemble of mutation strategies, Information Sciences 329 (2016) 329–345.
[46] G. Wu, X. Shen, H. Li, H. Chen, A. Lin, P. N. Suganthan, Ensemble of differential evolution variants, Information Sciences 423 (2018) 172–186.
[47] T. J. Choi, C. W. Ahn, Adaptive Cauchy differential evolution with strategy adaptation and its application to training large-scale artificial neural network.
networks, in: International Conference on Bio-Inspired Computing: Theories and Applications, Springer, 2017, pp. 502–510.

[48] T. J. Choi, Y. Lee, Asynchronous differential evolution with strategy adaptation for global numerical optimization, in: Proceedings of the 2018 2nd High Performance Computing and Cluster Technologies Conference, 2018, pp. 15–18.

[49] S. Rahnamayan, H. R. Tizhoosh, M. M. Salama, Opposition-based differential evolution, IEEE Transactions on Evolutionary computation 12 (1) (2008) 64–79.

[50] H. Wang, Z. Wu, S. Rahnamayan, Enhanced opposition-based differential evolution for solving high-dimensional continuous optimization problems, Soft Computing 15 (11) (2011) 2127–2140.

[51] S.-Y. Park, J.-J. Lee, Stochastic opposition-based learning using a beta distribution in differential evolution, IEEE transactions on cybernetics 46 (10) (2015) 2184–2194.

[52] T. J. Choi, J. Togelius, Y.-G. Cheong, A fast and efficient stochastic opposition-based learning for differential evolution in numerical optimization, arXiv preprint arXiv:1908.08011.

[53] T. J. Choi, J.-H. Lee, H. Y. Youn, C. W. Ahn, Adaptive differential evolution with elite opposition-based learning and its application to training artificial neural networks, Fundamenta Informaticae 164 (2-3) (2019) 227–242.

[54] E. Mininno, F. Neri, F. Cupertino, D. Naso, Compact differential evolution, IEEE Transactions on Evolutionary Computation 15 (1) (2010) 32–54.

[55] J.-h. Zhong, J. Zhang, SDE: a stochastic coding differential evolution for global optimization, in: Proceedings of the 14th annual conference on Genetic and evolutionary computation, 2012, pp. 975–982.

[56] H. Wang, S. Rahnamayan, H. Sun, M. G. Omran, Gaussian bare-bones differential evolution, IEEE Transactions on Cybernetics 43 (2) (2013) 634–647.

[57] R. D. Al-Dabbagh, F. Neri, N. Idris, M. S. Baba, Algorithmic design issues in adaptive differential evolution schemes: Review and taxonomy, Swarm and Evolutionary Computation 43 (2018) 284–311.

[58] K. R. Opara, J. Arabas, Differential evolution: A survey of theoretical analyses, Swarm and evolutionary computation 44 (2019) 546–558.

[59] I. C. Trelea, The particle swarm optimization algorithm: convergence analysis and parameter selection, Information processing letters 85 (6) (2003) 317–325.

[60] T. Liao, T. Stuetzle, Benchmark results for a simple hybrid algorithm on the CEC 2013 benchmark set for real-parameter optimization, in: 2013 IEEE Congress on Evolutionary Computation, IEEE, 2013, pp. 1938–1944.

[61] B. Lacroix, D. Molina, F. Herrera, Dynamically updated region based memetic algorithm for the 2013 CEC special session and competition on real parameter single objective optimization, in: 2013 IEEE Congress on Evolutionary Computation, IEEE, 2013, pp. 1945–1951.

[62] N. Hansen, The CMA evolution strategy: a comparing review, in: Towards a new evolutionary computation, Springer, 2006, pp. 75–102.

[63] W. Gong, Z. Cai, Differential evolution with ranking-based mutation operators, IEEE Transactions on Cybernetics 43 (6) (2013) 2066–2081.

[64] J. Demsar, Statistical comparisons of classifiers over multiple data sets, Journal of Machine learning research 7 (Jan) (2006) 1–30.

[65] P. Lévy, Sur les intégrales dont les éléments sont des variables aléatoires indépendantes, Annali della Scuola Normale Superiore di Pisa Classe di Scienze 3 (3-4) (1934) 337–366.

[66] P. Hall, A comedy of errors: the canonical form for a stable characteristic function, Bulletin of the London Mathematical Society 13 (1) (1981) 23–27.

[67] J. M. Chambers, C. L. Mallows, B. Stuck, A method for simulating stable random variables, Journal of the American statistical association 71 (354) (1976) 340–344.

[68] R. Weron, On the Chambers-Mallows-Stuck method for simulating skewed stable random variables, Statistics & Probability Letters 28 (2) (1996) 165–171.

[69] R. Weron, et al., Correction to “on the Chambers–Mallows–Stuck method for simulating skewed stable random variables”, Tech. rep., University Library of Munich, Germany (2010).

[70] R. Weron, Computationally intensive value at risk calculations, Tech. rep., Papers/Humboldt-Universitat Berlin, Center for Applied Statistics and (2004).