

Received December 13, 2019, accepted December 26, 2019, date of publication January 6, 2020, date of current version January 15, 2020. *Digital Object Identifier 10.1109/ACCESS.2020.2964222*

Advanced Cauchy Mutation for Differential Evolution in Numerical Optimization

TAE JONG CHO[I](https://orcid.org/0000-0001-8398-1673)^{®1,2}, JULIAN TOGELIUS², (Member, IEEE), AND YUN-GYUNG CHEONG³, (Member, IEEE)

¹Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon-si 16419, South Korea 2Department of Computer Science and Engineering, Tandon School of Engineering, New York University, Brooklyn, NY 11201, USA ³College of Software, Sungkyunkwan University, Suwon-si 16419, South Korea

Corresponding author: Yun-Gyung Cheong (aimecca@skku.edu)

This work was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) under Grant NRF-2017R1C1B2012752, and in part by the Ministry of Education under Grant NRF-2019R1A2C1006316.

ABSTRACT Among many evolutionary algorithms, differential evolution (DE) has received much attention over the last two decades. DE is a simple yet powerful evolutionary algorithm that has been used successfully to optimize various real-world problems. Since it was introduced, many researchers have developed new methods for DE, and one of them makes use of a mutation based on the Cauchy distribution to increase the convergence speed of DE. The method monitors the results of each individual in the selection operator and performs the Cauchy mutation on consecutively failed individuals, which generates mutant vectors by perturbing the best individual with the Cauchy distribution. Therefore, the method can locate the consecutively failed individuals to new positions close to the best individual. Although this approach is interesting, it fails to take into account establishing a balance between exploration and exploitation. In this paper, we propose a sigmoid based parameter control that alters the failure threshold for performing the Cauchy mutation in a time-varying schedule, which can establish a good ratio between exploration and exploitation. Experiments and comparisons have been done with six conventional and six advanced DE variants on a set of 30 benchmark problems, which indicate that the DE variants assisted by the proposed algorithm are highly competitive, especially for multimodal functions.

INDEX TERMS Artificial intelligence, evolutionary algorithm, differential evolution, numerical optimization.

I. INTRODUCTION

An evolutionary algorithm (EA) is a population-based metaheuristic inspired by biological evolution and genetic variations observed in nature. An EA takes a population of individuals and searches for a global optimum by artificially designed evolutionary operators. An EA does not make any assumption about a given problem, such as continuity and differentiability. Thus, it can be applied to any problem.

Among many EAs, differential evolution (DE) [1], [2] is widely considered to be one of the most effective EAs to optimize multidimensional real-valued functions. DE takes a population of individuals and iterates the mutation, crossover, and selection operators to search for a global optimum. Since it was introduced, DE has received much attention because it is simple and straightforward to implement.

The associate editor coordinating the review of this manuscript and approv[i](https://orcid.org/0000-0001-9186-6472)ng it for publication was Hisao Ishibuchi¹⁰.

The effectiveness of DE has been demonstrated successfully in various real-world problems [3], [4].

To increase the convergence speed of DE, Ali and Pant [5] proposed a variant of DE called modified DE (MDE), which makes use of a mutation based on the Cauchy distribution as an additional operator. MDE monitors the results of each individual in the selection operator. When an individual consecutively failed to find a better position than its current position for a predefined number of generations, MDE performs the Cauchy mutation on the individual, which generates a mutant vector by perturbing the best individual with the Cauchy distribution. Therefore, MDE can locate the consecutively failed individual to a new position close to the best individual. However, MDE uses the same failure threshold for performing the Cauchy mutation throughout the whole search process, which causes a serious limitation of establishing a balance between exploration and exploitation. Moreover, MDE uses the best individual based Cauchy mutation, which may reduce

the diversity of individuals drastically. Here, exploration is the property of searching for entirely new areas of a search space, while exploitation is the property of searching for the areas close to previously searched [6], [7].

EAs need to establish a balance between exploration and exploitation to be successful [6], [7]. In this paper, we propose a variant of MDE called advanced Cauchy mutation DE (ACM-DE), which alters the failure threshold for performing the Cauchy mutation in a time-varying schedule. The earlier work of ACM-DE can be found in [8]. ACM-DE uses a sigmoid based parameter control, which assigns a high failure threshold at the beginning of the search process and gradually reduces it over generations. Therefore, ACM-DE performs the Cauchy mutation with a low probability at the early stage of the search process to preserve the diversity of individuals and a high probability at the late stage of the search process to increase the convergence speed. ACM-DE also uses the *p*-best individual [9] based Cauchy mutation to prevent premature convergence. Therefore, ACM-DE can establish a good ratio between exploration and exploitation compared to MDE.

Experiments and comparisons were carried out with six conventional and six advanced DE variants on the Congress on Evolutionary Computation (CEC) 2017 benchmark problems [10], which contains a set of 30 benchmark problems. The experimental results indicate that the DE variants assisted by the proposed Cauchy mutation significantly outperform the DE variants assisted by the previous Cauchy mutation as well as the original DE variants. As a result, the main contributions of this paper are as follows.

- 1) ACM is simple and easy to embed into any DE variant as an additional operator.
- 2) ACM can establish a good ratio between exploration and exploitation.
- 3) ACM is a clear advance on the previous Cauchy mutation, and the effectiveness of ACM is demonstrated successfully by experiments and comparisons.

The rest of this paper is organized as follows: we introduce DE and the Cauchy distribution in Section [II.](#page-1-0) In Section [III,](#page-2-0) we review six advanced DE variants used for experiments and comparisons in this paper. Section [IV](#page-4-0) presents the technical details of the proposed algorithm. Section [V](#page-5-0) presents the experimental setups. In Sections [VI](#page-6-0) and [VII,](#page-11-0) we discuss the experimental results of the proposed algorithm. Finally, we provide the conclusion and the future work of this paper in Section [VIII.](#page-13-0)

II. BACKGROUND

A. DIFFERENTIAL EVOLUTION

DE is a population-based metaheuristic that takes a population of *NP* target vectors. Each target vector is a *D*-dimensional vector, denoted by $\vec{x_i g} = (x_{i,g}^1, x_{i,g}^2, \dots, x_{i,g}^D)$, where $g = 1, 2, \cdots, G_{max}$. Here, G_{max} denotes the maximum number of generations. DE consists of four operators: initialization, mutation, crossover, and selection. At the

beginning of the search process, the initialization operator uniformly distributes the population over a search space. Then, the mutation and crossover operators generate a population of *NP* trial vectors, and the selection operator makes up a new population for the next generation by comparing each target vector with its corresponding trial vector.

1) INITIALIZATION

First, let us define the lower and upper bounds as follows: $\vec{x_{min}} = (x_{min}^1, x_{min}^2, \cdots, x_{min}^D)$ and $\vec{x_{max}} =$ $(x_{max}^1, x_{max}^2, \dots, x_{max}^D)$, where $\overrightarrow{x_{min}}$ and $\overrightarrow{x_{max}}$ denote the lower and upper bounds, respectively. Each component (*j*) of each target vector (*i*) is initialized as follows:

$$
x_{i,0}^j = x_{min}^j + rand_i^j \cdot (x_{max}^j - x_{min}^j)
$$
 (1)

where *rand^j i* denotes a uniformly distributed random value in the interval [0, 1].

2) MUTATION

The role of the mutation operator is to generate a set of *NP* mutant vectors. Each mutant vector is denoted by $v_{i,g}$. The following list shows the six most commonly used conventional mutation strategies.

DE/rand/1:

$$
\vec{v_{i,g}} = \vec{x_{r_1,g}} + F \cdot (\vec{x_{r_2,g}} - \vec{x_{r_3,g}}) \tag{2}
$$

DE/best/1:

$$
\vec{v_{i,g}} = x_{best,g} + F \cdot (\vec{x_{r_1,g}} - \vec{x_{r_2,g}}) \tag{3}
$$

DE/current-to-best/1:

$$
\vec{v_{i,g}} = \vec{x_{i,g}} + F \cdot (x_{best,g} - \vec{x_{i,g}}) + F \cdot (\vec{x_{r_1,g}} - \vec{x_{r_2,g}}) \quad (4)
$$

DE/current-to-rand/1:

$$
\vec{v_{i,g}} = \vec{x_{i,g}} + K \cdot (\vec{x_{r_1,g}} - \vec{x_{i,g}}) + F \cdot (\vec{x_{r_2,g}} - \vec{x_{r_3,g}}) \tag{5}
$$

DE/rand/2:

$$
\vec{v_{i,g}} = \vec{x_{r_1,g}} + F \cdot (\vec{x_{r_2,g}} - \vec{x_{r_3,g}}) + F \cdot (\vec{x_{r_4,g}} - \vec{x_{r_5,g}}) \quad (6)
$$

DE/current-to-best/2:

$$
\vec{v_{i,g}} = \vec{x_{i,g}} + F \cdot (x_{best,g} - \vec{x_{i,g}}) + F \cdot (\vec{x_{r_1,g}} - \vec{x_{r_2,g}}) + F \cdot (\vec{x_{r_3,g}} - \vec{x_{r_4,g}}) \tag{7}
$$

where $\vec{x}_{r_1,g}$, $\vec{x}_{r_2,g}$, $\vec{x}_{r_3,g}$, $\vec{x}_{r_4,g}$, and $\vec{x}_{r_5,g}$ denote randomly selected donor vectors, which are mutually different and not equal to their corresponding target vector $x_{i,g}$. Additionally, $\vec{x}_{best,g}$ denotes the best individual. Finally, *F* and *K* denote the scaling factor and a uniformly distributed random value in the interval [0, 1], respectively.

3) CROSSOVER

The role of the crossover operator is to generate a set of *NP* trial vectors. Each trial vector is denoted by $\vec{u}_{i,g}$. There are two classical crossover operators, binomial and exponential. In the binomial crossover, an integer value *jrand* is randomly selected within $\{1, 2, \cdots, D\}$. The binomial crossover generates each component (*j*) of each trial vector (*i*) as follows.

$$
u_{i,g}^j = \begin{cases} v_{i,g}^j & \text{if } rand_i^j < CR \text{ or } j = j_{rand} \\ x_{i,g}^j & \text{otherwise} \end{cases} \tag{8}
$$

where *CR* denotes the crossover rate.

In the exponential crossover, two integer values *n* and *L* are initialized first. The starting component *n* is randomly selected within $\{1, 2, \cdots, D\}$. The number of components *L* is obtained as follows.

$$
L = 0
$$

DO { $L = L + 1$ } WHILE $((\text{rand}_i^j < CR)$ AND $(L < D)$)

With the two integer values, the exponential crossover generates each component (*j*) of each trial vector (*i*) as follows.

$$
u_{i,g}^j = \begin{cases} v_{i,g}^j & \text{if } j = \langle n \rangle_D, \langle n+1 \rangle_D, \cdots, \langle n+L-1 \rangle_D \\ v_{i,g}^j & \text{otherwise} \end{cases}
$$
(9)

where $\langle \cdot \rangle_D$ denotes the modulo operator with the divisor *D*.

4) SELECTION

The selection operator compares each target vector with its corresponding trial vector and picks the better one. In other words, if the fitness value of a trial vector is better than or equal to that of its corresponding target vector, the trial vector is selected as a member of the population for the next generation. Otherwise, the trial vector is discarded, and the target vector is selected. The selection operator makes up a new population for the next generation as follows.

$$
\vec{x}_{i,g+1} = \begin{cases}\n\vec{u}_{i,g} & \text{if } f(\vec{u}_{i,g}) \le f(\vec{x}_{i,g}) \\
\vec{x}_{i,g} & \text{otherwise.}\n\end{cases} \tag{10}
$$

where $f(\vec{x})$ denotes an objective function.

DE iterates the mutation, crossover, and selection operators until one of the termination criteria is satisfied. The most commonly used termination criterion is to reach the maximum number of generations *Gmax* or the maximum number of function evaluations *NFEmax* .

B. ANALYSIS OF CAUCHY DISTRIBUTION

The Cauchy distribution is a continuous probability distribution that has two parameters, x_0 and γ . x_0 is the location parameter, and γ is the scale parameter that determines the shape of the Cauchy distribution. For example, if a higher value is assigned to γ , the height of the peak of the probability density function (PDF) will be shorter, and its width will be wider. On the other hand, if a lower value is assigned to γ , the height of the peak of the PDF will be taller, and its width

FIGURE 1. Various probability density functions of Cauchy distribution.

will be narrower. The PDF of the Cauchy distribution can be defined as follows.

$$
f(x; x_0, \gamma) = \frac{1}{\pi \gamma [1 + (\frac{x - x_0}{\gamma})^2]} = \frac{1}{\pi} \left[\frac{\gamma}{(x - x_0)^2 + \gamma^2} \right]
$$
(11)

Additionally, the cumulative distribution function of the Cauchy distribution can be defined as follows.

$$
F(x; x_0, \gamma) = \frac{1}{\pi} \arctan\left(\frac{x - x_0}{\gamma}\right) + \frac{1}{2} \tag{12}
$$

Fig. [1](#page-2-1) presents the various PDFs of the Cauchy distribution.

III. RELATED WORK

Since it was introduced, many researchers have developed new methods for DE [9], [11]–[39]. For more detailed information, please refer to the following papers [3], [4], [40], [41]. In this section, we review six advanced DE variants used for experiments and comparisons in this paper.

A. SaDE

Qin et al. proposed a self-adaptive DE variant called SaDE [11], which automatically adjusts mutation strategies and control parameters during the search process. SaDE uses four mutation strategies, ''DE/rand/1'', ''DE/current-tobest/1'', ''DE/rand/2'', and ''DE/current-to-rand/1.'' At each generation, SaDE assigns each individual one of the mutation strategies according to strategy probabilities p_k , $k =$ $\{1, 2, 3, 4\}$. The strategy probabilities are initialized to $\frac{1}{K}$, $K = 4$ at the beginning of the search process. To update the strategy probabilities, SaDE monitors the success and failure results of each individual in the selection operator. At each generation, after a predefined number of generations *LP*, SaDE updates the strategy probabilities as follows.

$$
p_{k,g} = \frac{S_{k,g}}{\sum_{k=1}^{K} S_{k,g}}
$$
(13)

where

$$
S_{k,g} = \frac{\sum_{t=g-LP}^{g-1} n s_{k,t}}{\sum_{t=g-LP}^{g-1} n s_{k,t} + \sum_{t=g-LP}^{g-1} n f_{k,t}} + \epsilon \qquad (14)
$$

where $ns_{k,g}$ and $nf_{k,g}$ denote the success and failure memories with the *k*th mutation strategy, respectively. Additionally, $\epsilon =$ 0.001 is used to avoid division by zero.

At each generation, SaDE assigns each individual a scaling factor $F_{i,g}$ with the Gaussian distribution as follows.

$$
F_{i,g} = rndn_i(0.5, 0.3)
$$
 (15)

Similarly, at each generation, SaDE assigns each individual a crossover rate *CRi*,*^g* with the Gaussian distribution as follows.

$$
CR_{i,G} = rndn_i(CRm_k, 0.1)
$$
 (16)

where CRm_k is the median of the crossover rates used by successfully evolved individuals with the *k*th mutation strategy. After that, the crossover rate is truncated to [0, 1].

B. EPSDE

Mallipeddi et al. proposed a self-adaptive DE variant called EPSDE [12], which uses a pool of mutation strategies and two pools of control parameters. The pool of mutation strategies contains ''DE/best/2'', ''DE/rand/1'', and ''DE/current-torand/1.'' Additionally, the pool of scaling factors contains the values in the range of 0.4−0.9 in steps of 0.1, and the pool of crossover rates contains the values in the range of $0.1 - 0.9$ in the steps of 0.1. At each generation, EPSDE assigns each individual a combination of mutation strategy and control parameters taken from the respective pools. If an individual successfully evolved with the combination, the individual uses the same combination for the next generation. Otherwise, EPSDE assigns the individual a new combination or one of the combinations used by successfully evolved individuals with equal probability for the next generation.

C. CoDE

Wang et al. proposed an advanced DE variant called CoDE [13], which generates three candidate trial vectors for each individual. CoDE uses three mutation strategies, which are "DE/rand/1", "DE/rand/2", and "DE/currentto-rand/1.'' Additionally, CoDE uses three pairs of control parameters, which are $F = 1.0$, $CR = 0.1$ ", $F =$ 1.0, $CR = 0.9$ ", and " $F = 0.8$, $CR = 0.2$." At each generation, CoDE generates three candidate trial vectors for each individual by using different mutation strategies and three randomly selected pairs of control parameters. Therefore, the three candidate trial vectors have different characteristics. Among the three candidate trial vectors, the best one is selected for the selection operator.

D. SHADE

Tanabe and Fukunaga proposed an advanced DE variant called SHADE [18], which is an extension of an advanced DE variant called JADE [9]. We first introduce JADE and then SHADE. JADE uses an advanced mutation strategy called DE/current-to-*p*best, which can be defined as follows.

DE/current-to-*p*best:

$$
\vec{v_{i,g}} = \vec{x_{i,g}} + F \cdot (\vec{x_{\text{pbest},g}} - \vec{x_{i,g}}) + F \cdot (\vec{x_{r_1,g}} - \vec{r_{r_2,g}}) \tag{17}
$$

where $\vec{x_{pbest}, g}$ and $\vec{r_{r2}, g}$ denote one of the top 100*p*% individuals with $p \in (0, 1]$ and a randomly selected donor vector from a population or an archive of recently discarded target vectors, respectively.

At each generation, JADE assigns each individual a pair of control parameters as follows.

$$
F_{i,g} = \text{rndc}_i(\mu_F, 0.1) \tag{18}
$$

$$
CR_{i,g} = rndn_i(\mu_{CR}, 0.1) \tag{19}
$$

where *rndcⁱ* and *rndnⁱ* denote the Cauchy distribution and the Gaussian distribution, respectively. After that, the scaling factor is regenerated if $F_{i,g} \leq 0$ or truncated to 1 if $F_{i,g} > 1$, and the crossover rate is truncated to [0, 1]. The value of μ_F and the value of μ_{CR} are updated as follows.

$$
\mu_F = (1 - c) \cdot \mu_F + c \cdot mean_L(S_F) \tag{20}
$$

$$
\mu_{CR} = (1 - c) \cdot \mu_{CR} + c \cdot mean_A(S_{CR}) \tag{21}
$$

where c is a constant in the interval $[0, 1]$. Additionally, *mean^L* and *mean^A* denote the Lehmer mean and the arithmetic mean, respectively. Finally, *S^F* and *SCR* denote the scaling factors and the crossover rates used by successfully evolved individuals, respectively.

SHADE uses an improved parameter control of JADE, which contains historical memories with *H* entries for the values of μ_F and the values of μ_{CR} , denoted by $M_{F,k}$ and $M_{CR,k}$. The historical memories are initialized to 0.5 at the beginning of the search process. At each generation, SHADE assigns each individual a pair of control parameters as follows.

$$
F_{i,g} = \text{rndc}_i(M_{F,r_i}, 0.1) \tag{22}
$$

$$
CR_{i,g} = rndn_i(M_{CR,r_i}, 0.1) \tag{23}
$$

where r_i is randomly selected within $\{1, 2, \cdots, H\}$. After the selection operator, SHADE updates the historical memories as follows.

$$
M_{F,\langle g \rangle_H} = \begin{cases} \text{mean}w_L(S_F) & \text{if } S_F \neq \emptyset \\ M_{F,\langle g \rangle_H} & \text{otherwise.} \end{cases} \tag{24}
$$

$$
M_{CR,\langle g \rangle_H} = \begin{cases} \text{mean}_{A}(S_{CR}) & \text{if } S_{CR} \neq \emptyset \\ M_{CR,\langle g \rangle_H} & \text{otherwise.} \end{cases} \tag{25}
$$

where *meanw^L* and *meanw^A* denote the weighted Lehmer mean and the weighted arithmetic mean, respectively.

E. MPEDE

Wu et al. proposed a multi-population-based DE variant called MPEDE [20], which uses three mutation strategies, i.e., ''DE/current-to-pbest/1'', ''DE/current-to-rand/1'', and ''DE/rand/1.'' MPEDE splits a population into four subpopulations, which are three equally sized smaller subpopulations and one larger reward subpopulation. Each subpopulation uses a distinct mutation strategy to generate mutant vectors. After every predefined number of generations *LP*, the best mutation strategy is determined by comparing the average fitness improvement. The reward subpopulation then uses the best mutation strategy to generate mutant vectors for the next *LP* generations. Additionally, MPEDE uses the parameter control of JADE [9] to automatically adjust control parameters during the search process. As a result, MPEDE can find the best mutation strategy and assign it to the reward subpopulation during the search process.

F. EDEV

Wu et al. proposed a multi-population-based DE variant called EDEV [21], which uses three DE variants, i.e., JADE [9], CoDE [13], and EPSDE [12]. Similar to MPEDE, EDEV splits a population into four subpopulations, which are three equally sized smaller subpopulations and one larger reward subpopulation. Each subpopulation uses a distinct DE variant to generate mutant vectors. After every predefined number of generations *LP*, the best DE variant is determined by comparing the average fitness improvement. The reward subpopulation then uses the best DE variant to generate mutant vectors for the next *LP* generations. As a result, EDEV can find the best DE variant and assign it to the reward subpopulation during the search process.

IV. PROPOSED ALGORITHM

This section outlines the proposed Cauchy mutation and discusses the details of each algorithmic component.

A. REVIEW OF MDE

MDE [5] is the basis of the proposed algorithm, which keeps track of the results of each individual in the selection operator. When an individual consecutively failed to find a better position than its current position for a predefined number of generations, MDE assumes that the individual gets trapped in a local optimum. To help the individual escape from the local optimum, MDE performs the Cauchy mutation on the individual, which generates a mutant vector by perturbing the best individual with the Cauchy distribution. The Cauchy mutation of MDE generates a trial vector as follows.

$$
u_{i,g}^j = \begin{cases} \n\text{rndc}_i^j(x_{best,g}^j, 0.1) & \text{if } \text{rand}_i^j < 0.5 \text{ or } j = j_{\text{rand}} \\
x_{i,g}^j & \text{otherwise}\n\end{cases} \tag{26}
$$

where rnd_i^j and rand_i^j denote the Cauchy distribution and a uniformly distributed random value in the interval [0, 1], respectively. Therefore, MDE can locate the consecutively failed individual to a new position close to the best individual.

Compared to other DE variants that make use of a mutation based on the Cauchy distribution, MDE has two main advantages: First, MDE carries out the Cauchy mutation on an individual basis, while other DE variants [42], [43] do so on a population basis. In other words, the DE variants [42], [43] calculate the running metric of convergence [44] in every predefined number of generations and carry out the Cauchy mutation if the calculated convergence is lower than a predefined threshold. This approach is effective when all individuals get stuck in the same local optimum but ineffective when different individuals get stuck in different local optima. Secondly, there is no increment in computational cost.

The Cauchy mutation of MDE requires a computational cost similar to that of the mutation and crossover operators. On the other hand, the Cauchy mutation of the DE variants [42], [43] requires a considerable computational cost because they need to calculate the running metric of convergence and the diversity of each component in every predefined number of generations.

B. ADVANCED CAUCHY MUTATION

Although the effectiveness of MDE has been demonstrated successfully on the classical benchmark problems, MDE is not well suited to optimize complex problems. One of the major drawbacks of MDE is that it uses the same failure threshold for performing the Cauchy mutation throughout the whole search process, which causes a serious limitation of establishing a balance between exploration and exploitation. Moreover, MDE uses the best individual based Cauchy mutation, which may reduce the diversity of individuals drastically. The aim of this paper is thus to propose an improved approach, which removes all of the difficulties that MDE faces.

ACM-DE consists of two algorithmic components: a sigmoid based parameter control and the *p*-best individual [9] based Cauchy mutation. The formal can alters the failure threshold for performing the Cauchy mutation in a time-varying schedule. The latter can prevent premature convergence by preventing consecutively failed individuals from converging into one region. The following sections provide a detailed description of these algorithmic components.

1) SIGMOID BASED PARAMETER CONTROL

FOR FAILURE THRESHOLD

MDE uses a fixed failure threshold for performing the Cauchy mutation, which causes one of the following problems.

- Premature convergence: MDE with a low failure threshold may execute the Cauchy mutation too frequently at the early stage of the search process, which may hinder from discovering promising regions.
- Ineffectiveness: MDE with a high failure threshold may execute the Cauchy mutation too rarely at the end of the search process, which may not take advantage of the Cauchy mutation to increase the convergence speed.

To address these problems, we propose a sigmoid based parameter control, which gradually decreases the failure threshold as a sigmoid function of the number of generations. First, let us define two failure threshold parameters, one for the initial failure threshold *FTinit* and the other for the final failure threshold *FTfin*. At each generation, a new failure threshold *FT^g* is obtained as follows.

$$
FT_g = FT_{init} + S\left(\frac{g}{G_{max}}\right) \cdot (FT_{fin} - FT_{init}) \tag{27}
$$

where

$$
S(x) = \frac{1}{1 + e^{-(lb + x \cdot (ub - lb))}}
$$
(28)

where *lb* and *ub* are two constants, -6 and 6, respectively.

FIGURE 2. Cauchy Mutation of ACM-DE.

Therefore, ACM-DE assigns a high failure threshold at the beginning of the search process and gradually reduces it over generations. In doing so, ACM-DE performs the Cauchy mutation on consecutively failed individuals with a low probability at the early stage of the search process to preserve the diversity of individuals and a high probability at the late stage of the search process to increase the convergence speed.

2) *P*-BEST INDIVIDUAL BASED CAUCHY MUTATION

MDE uses the components of the best individual in the phase of the Cauchy mutation. Although consecutively failed individuals can quickly move toward the region in which the best individual is located, the population diversity of MDE may be reduced drastically. If the population diversity of MDE becomes too low, premature convergence may occur.

To address this problem, we replaced the best individual with the *p*-best individual [9], which uses the components of any of the top *p*% individuals in the phase of the Cauchy mutation. The Cauchy mutation of ACM-DE generates a trial vector as follows.

$$
u_{i,g}^j = \begin{cases} \n\text{rndc}_i^j(x_{pbest,g}^j, 0.1) & \text{if } \text{rand}_i^j < \text{CR}_{i,g}^c \text{ or } j = j_{\text{rand}} \\
x_{i,g}^j & \text{otherwise}\n\end{cases} \tag{29}
$$

where $CR_{i,g}^c$ denotes the crossover rate for the Cauchy mutation, which is either 0.1 or 0.9 with equal probability.

Therefore, the Cauchy mutation of ACM-DE uses the components of any of the top *p*% individuals to move consecutively failed individuals to the regions in which the top *p*% individuals are located. In doing so, the Cauchy mutation of ACM-DE can prevent the consecutively failed individuals from converging into one region, which can prevent premature convergence.

C. COMBINATION

ACM-DE is the combination of DE with the proposed Cauchy mutation. During the iteration phase, the standard mutation or the proposed Cauchy mutation is used to search for a global optimum. The switch between the two mutations occurs according to the failure counter of each individual *FCⁱ* . If the

FIGURE 3. Sigmoid Based Parameter Control of ACM-DE.

failure counter of an individual is equal to the failure threshold *FTg*, ACM-DE performs the proposed Cauchy mutation on the individual. Otherwise, ACM-DE performs the standard mutation on the individual. The offspring generated by either the standard mutation or the proposed Cauchy mutation inherits the information from its corresponding target vector through the binomial crossover. The pseudo-code of ACM-DE is presented in Algorithm [1.](#page-6-1) Additionally, Figs. [2](#page-5-1) and [3](#page-5-2) show an illustration of the Cauchy mutation of ACM-DE compared to that of MDE and an illustration of the failure threshold decrease and increase with the sigmoid based parameter control, respectively.

V. EXPERIMENTAL SETUP

A. TEST FUNCTIONS

To evaluate the performance of test algorithms, we employed 30 test functions from CEC 2017 benchmark problems [10]. The CEC 2017 benchmark problems consist of three unimodal functions (F_1-F_3) , seven simple multimodal functions (F_4-F_{10}) , ten expanded multimodal functions $(F_{11}-F_{20})$, and ten hybrid composition functions $(F_{21}-F_{30})$. For more detailed information, please refer to the following paper [10].

All experimental settings such as the number of function evaluations, the lower and upper bounds, and the termination criteria for the test functions are initialized as same as in [10].

Algorithm 1 ACM-DE (DE/rand/1/bin)

Algorithm 2 DE/rand/1/bin

1 Select three random donor vectors $\vec{x}_{r_1, g}, \vec{x}_{r_2, g}, \vec{x}_{r_3, g}$ where $r_1 \neq r_2 \neq r_3 \neq i$; **²** Select random integer *jrand* within [1, *D*]; **3 for** $j = 0$; $j < D$; $j = j + 1$ **do 4 | if** $rand_i^j \leq CR$ or $j = j_{rand}$ **then 5** $u^j_{i,g} = x^j_{r_1,g} + F \cdot (x^j_{r_2,g} - x^j_{r_3,g});$ **⁶ else 7** $u_{i,g}^j = x_i^j$ *i*,*g* ; **⁸ end ⁹ end**

Algorithm 3 ACM/bin

- **1** Select any of top $p\%$ individuals $\vec{x_{p}}_{best,g}$;
- **²** Select random integer *jrand* within [1, *D*];
- **³** Select crossover rate *CR^c i*,*g* from either 0.1 or 0.9;

4 for
$$
j = 0
$$
; $j < D$; $j = j + 1$ do

5
\n**if**
$$
rand_i^j \le CR_{i,g}^c
$$
 or $j = j_{rand}$ then
\n6
\n**if** $val_i^j = rndc_i^j(x_{pbest,g}^j, 0.1);$
\n7
\n**else**
\n8
\n9
\n**end**
\n10
\n**end**

B. PERFORMANCE METRICS

To evaluate the performance of test algorithms, we employed the following performance criteria.

1) FUNCTION ERROR VALUE

We employed the function error value (FEV) to evaluate the accuracy of a test algorithm. The FEV metric corresponds to the absolute difference between the global optimum of an objective function and the final best objective value of a test algorithm. The FEV metric can be defined as follows.

$$
FEV = |f(\vec{x_*}) - f(\vec{x_{best}, G_{max}})|
$$
\n(30)

where $f(\vec{x})$ denotes an objective function. Additionally, \vec{x}_* and $\vec{x}_{best}, \vec{G}_{max}$ denote the global optimum of the objective function and the final best objective value of a test algorithm, respectively. As the value of the FEV metric decreases, the accuracy of a test algorithm increases.

2) STATISTICAL TEST

We employed the Wilcoxon signed-rank test at a significance level of 0.05 to determine whether the performance differences between two algorithms are statistically significant [45]. Hereafter, the symbols used in the tables indicate the following.

- $1)$ +: The proposed algorithm is significantly superior than the comparison algorithm according to the Wilcoxon signed-rank test.
- $2)$ =: The difference between the proposed algorithm and the comparison algorithm is not significant according to the Wilcoxon signed-rank test.
- 3) −: The proposed algorithm is significantly inferior than the comparison algorithm according to the Wilcoxon signed-rank test.

VI. EXPERIMENTAL RESULTS

Experiments and comparisons were carried out with six conventional and six advanced DE variants to demonstrate the effectiveness of the proposed Cauchy mutation.

IEEE Access

TABLE 1. Averages and standard deviations of function error values with conventional DE variants ($D = 30$).

signed-r represents that AC ith o significantly perior than the corresponding variant. "=" recantly inferior than the corresponding varian repre cally significant. And. represents that ACM variant is sig between ACM variant and the

 $14/8/8$

28/2/0

 \mathbf{I}

FIGURE 4. Median and interquartile ranges (25th and 75th) of function error values with DE/rand/1/bin variants (D = 30).

A. CONVENTIONAL DE VARIANTS

21/6/3

22/8/0

 \mathbf{I}

We applied the proposed Cauchy mutation to six conventional DE variants, namely DE/rand/1/bin, DE/best/1/bin, DE/current-to-best/1/bin, DE/current-to-rand/1, DE/rand/2/ bin, and DE/current-to-best/2/bin. For all of the test algorithms, the scaling factor *F*, the crossover rate *CR*, and the population size *NP* are initialized to 0.5, 0.5, and 100, respectively. Additionally, FT_{init} = 100 and $FT_{fin} = 5$ are used for the proposed Cauchy mutation (ACM), and $FT = 5$ is used for the previous Cauchy mutation (CM) as same as in [5].

Table [1](#page-7-0) presents the averages and the standard deviations of the FEVs obtained by independently running each algorithm 51 times at 30 dimensions. The summary of the experimental results is as follows.

• DE/rand/1/bin and DE/rand/2/bin: These mutation strategies are designed for higher exploration. ACM-DE/ rand/1/bin finds 17/4 statistically better/worse solutions

 $25/3/2$

20/6/4

TABLE 2. Averages and standard deviations of function error values with conventional DE variants ($D = 50$).

 \parallel references
that ACM variant is significantly superior than the corresponding variant. "=" re-
And, "-" represents that ACM variant is significantly inferior than the corresponding variant the statistical results of the Wilcoxon signed-rank test with α difference between ACM variant and the corresponding vari "=" represents that the performan $= 0.05$ significance level. lically significant. And, nding variant is not statis

FIGURE 5. Median and interquartile ranges (25th and 75th) of function error values with DE/rand/1/bin variants (D = 50).

compared to CM-DE/rand/1/bin and 24/2 compared to DE/rand/1/bin. Also, ACM-DE/rand/2/bin finds 14/8 statistically better/worse solutions compared to CM-DE/rand/2/bin and 28/0 compared to DE/rand/2/bin.

• DE/best/1/bin, DE/current-to-best/1/bin, and DE/ current-to-best/2/bin: These mutation strategies are designed for higher exploitation. ACM-DE/best/ 1/bin finds 15/1 statistically better/worse solutions compared to CM-DE/best/1/bin and 6/1 compared to DE/best/1/bin. Also, ACM-DE/current-to-best/1/bin finds 18/4 statistically better/worse solutions compared to CM-DE/current-to-best/1/bin and 19/0 compared to DE/current-to-best/1/bin, and ACM-DE/current-to-

IEEE Access®

TABLE 3. Averages and standard deviations of function error values with advanced DE variants (D = 30).

ificance level. "+" represents that AC
ically significant. And, "-" represents ts that ACM variant is significantly superior than the corresponding variant. "=" re
represents that ACM variant is significantly inferior than the corresponding variant tatistical result signed-ra vith o repre between ACM variant and the

FIGURE 6. Median and interquartile ranges (25th and 75th) of function error values with EDEV variants (D = 30).

best/2/bin finds 20/4 statistically better/worse solutions compared to CM-DE/current-to-best/2/bin and 25/2 compared to DE/current-to-best/2/bin.

• DE/current-to-rand/1: This mutation strategy is designed for rotation invariant. ACM-DE/current-to-rand/1 finds 21/3 statistically better/worse solutions compared to CM-DE/current-to-rand/1 and 22/0 compared to DE/current-to-rand/1.

Although CM-DE variants are more greedy than ACM-DE variants, the performance differences between them on the unimodal functions (F_1-F_3) are insignificant. On the contrary, ACM-DE variants yield significantly better performance than CM-DE variants on the multimodal functions (F_4-F_{30}) . These observations are supported by the convergence graphs of DE/rand/1/bin variants from Fig. [4.](#page-7-1) We selected DE/rand/1/bin because it is the most widely

TABLE 4. Averages and standard deviations of function error values with advanced DE variants (D = 50).

signed-ra ficance level. "+" represents that AC
cally significant. And, "-" represents significantly ith o corresponding va repre is that ACM variant is significantly
represents that ACM variant is sig between ACM variant and the cantly inferior than the corresponding varian

FIGURE 7. Median and interquartile ranges (25th and 75th) of function error values with EDEV variants (D = 50).

used one among the six conventional DE variants. As shown in the figures, the convergence speed of CM-DE/rand/1/bin is faster than ACM-DE/rand/1/bin, but it gets trapped in a local optimum during the search process. On the other hand, the convergence speed of ACM-DE/rand/1/bin is more robust than CM-DE/rand/1/bin. We can see a similar tendency at 50 dimensions from Table [2](#page-8-0) and Fig. [5.](#page-8-1)

As a result, the proposed Cauchy mutation can enhance the performance of conventional DE variants that have different characteristics such as higher exploration, higher exploitation, and rotation invariant, especially for multimodal functions.

B. ADVANCED DE VARIANTS

Although conventional DE variants serve as the basis for many advanced DE variants, further investigation with advanced DE variants may be intriguing. We applied the proposed Cauchy mutation to six advanced DE variants,

TABLE 5. DE/rand/1/bin with failure threshold decrease and increase approaches $(D = 30)$.

F1 F2 F ₃ F ₄	SFTD MEAN (STD DEV) 2.71E+02 (1.44E+02) 1.73E+07 (9.97E+07) 4.76E+03 (1.44E+03) 8.52E+01 (2.36E-01)	SFTI MEAN (STD DEV) $5.87E+00$ (4.15E+00) - $9.66E+05(4.27E+06)$ -	LFTD MEAN (STD DEV) $2.06E+02(8.35E+01)$ -	LFTI MEAN (STD DEV) $1.93E+01(1.08E+01)$ -
			$4.05E+09(1.17E+10) +$	$7.84E+08(1.99E+09) +$
		$4.51E+03(2.52E+03) =$	$1.16E+04(3.36E+03) +$	$1.18E+04(3.13E+03) +$
		$8.07E+01(2.28E+01) =$	$8.52E+01(2.28E-01) =$	$8.34E+01(2.37E+01) =$
F5	4.27E+01 (1.14E+01)	$6.33E+01$ (1.76E+01) +	$3.99E+01(1.20E+01) =$	$5.97E+01$ (1.61E+01) +
F ₆	3.95E-12 (3.52E-12)	$1.65E-13(6.12E-14)$ -	$1.68E-12(1.16E-12)$ -	$1.94E-13(6.11E-14) -$
F7	$6.68E+01(1.05E+01)$	$9.19E+01(1.67E+01) +$	$6.79E+01(1.08E+01) =$	8.72E+01 (1.49E+01) +
F8	3.79E+01 (1.21E+01)	$6.59E+01(1.86E+01) +$	$3.82E+01(1.16E+01) =$	$6.18E+01(1.71E+01) +$
F9	4.47E-15 (2.23E-14)	$2.61E+00$ (1.26E+01) =	$2.24E-15(1.60E-14) =$	$2.24E-15(1.60E-14) =$
F10	$2.32E+03(6.53E+02)$	$2.90E+03(5.48E+02) +$	$2.33E+03(5.95E+02) =$	$2.85E+03(5.26E+02) +$
F11	2.66E+01 (2.47E+01)	$6.65E+01(3.89E+01) +$	$2.51E+01(2.24E+01) =$	$5.10E+01(3.40E+01) +$
F12	1.22E+05 (7.00E+04)	7.79E+04 (3.48E+04) -	$1.65E+05(8.15E+04) +$	$9.33E+04$ (4.54E+04) -
F13	$1.29E+04(5.15E+03)$	$1.16E+04(1.25E+04) =$	$1.38E+04(5.26E+03) =$	$1.18E+04(1.29E+04) =$
F14	$9.87E+01(9.72E+00)$	$2.86E+02(5.35E+02) +$	$1.01E+02(8.52E+00) =$	$1.25E+02(6.54E+01) +$
F15	1.93E+02 (3.55E+01)	$7.08E+02(8.10E+02) +$	$2.52E+02(4.18E+01) +$	$1.12E+03(2.38E+03) +$
F16	$5.34E+02(2.11E+02)$	$8.73E+02(3.06E+02) +$	$5.59E+02(2.36E+02)$ =	$9.22E+02(2.98E+02) +$
F17	7.07E+01 (6.99E+01)	$3.25E+02$ (1.60E+02) +	$6.90E+01(8.02E+01) =$	$2.87E+02(1.65E+02) +$
F18	$1.12E+05(1.06E+05)$	$2.14E+05(1.79E+05) +$	$9.98E+04(9.78E+04) =$	$1.65E+05$ (1.43E+05) +
F19	$8.64E+01$ (1.70E+01)	$1.23E+03(2.89E+03) +$	$9.36E+01(1.87E+01) =$	$7.12E+02$ $(1.46E+03)$ +
F20	8.32E+01 (9.78E+01)	$3.74E+02$ (1.52E+02) +	$8.67E+01(9.01E+01) =$	$3.59E+02$ (1.87E+02) +
F21	2.40E+02 (1.04E+01)	$2.64E+02(1.54E+01) +$	$2.43E+02(1.15E+01) =$	$2.67E+02(1.66E+01) +$
F ₂₂	$1.00E+02(0.00E+00)$	$2.36E+02(6.85E+02) =$	$1.00E+02(0.00E+00) =$	$2.31E+02(6.61E+02)$
F ₂ 3	3.85E+02 (1.33E+01)	$4.15E+02$ (1.82E+01) +	$3.88E+02$ (1.16E+01) =	$4.08E+02(1.89E+01) +$
F ₂₄	$4.62E+02$ (1.11E+01)	$4.86E+02(1.93E+01) +$	$4.63E+02$ (1.06E+01) =	$4.78E+02$ (1.46E+01) +
F ₂₅	3.87E+02 (0.00E+00)	$3.87E+02(0.00E+00) =$	$3.87E+02(0.00E+00) =$	$3.87E+02$ (7.84E-01) =
F ₂₆	1.36E+03 (1.31E+02)	$1.78E+03(2.00E+02) +$	$1.37E+03$ $(1.48E+02)$ =	$1.68E+03(2.35E+02) +$
F ₂₇	$5.00E+02(6.24E+00)$	$5.18E+02(9.17E+00)+$	$5.01E+02$ (5.14E+00) =	$5.11E+02(7.86E+00) +$
F ₂₈	$3.63E+02$ (4.46E+01)	$3.73E+02(6.41E+01) =$	$3.80E+02(4.35E+01) +$	$3.63E+02(5.52E+01) =$
F ₂₉	5.03E+02 (9.24E+01)	$7.78E+02(1.66E+02) +$	$4.93E+02$ (1.06E+02) =	$6.85E+02$ (1.84E+02) +
F30	$1.05E+04(2.42E+03)$	$8.60E+03(3.46E+03)$ -	$1.43E+04(2.79E+03) +$	$9.26E+03(3.12E+03)$ -
$+/-/$		18/7/5	6/22/2 The symbols "+/=/-" show the statistical results of the Wilcoxon signed-rank test with $\alpha = 0.05$ significance level.	20/6/4

+" represents that SFTD variant is significantly superior than the corresponding variant. "=" represents that the The presents und 31 TD variant is significantly soperon una une concesponding variant. = represents unat the performance difference between SFTD variant and the corresponding variant is not statistically significant. And,

TABLE 6. EDEV with failure threshold decrease and increase approaches $(D = 30)$.

	EDEV					
	SFTD	SFTI	LFTD	LFTI		
	MEAN (STD DEV)	MEAN (STD DEV)	MEAN (STD DEV)	MEAN (STD DEV)		
F1	5.57E-16 (2.78E-15)	$8.35E-16(3.37E-15) =$	$0.00E+00(0.00E+00) =$	$8.35E-16(3.37E-15) =$		
F2	7.69E+07 (5.16E+08)	$1.26E-06$ $(3.59E-06) =$	$1.43E+10(8.13E+10) =$	$2.70E+10(1.15E+11) =$		
F ₃	3.81E+02 (1.57E+03)	$2.94E+02$ (1.91E+03) =	$1.60E+03(5.30E+03) =$	$1.27E+03(5.65E+03) =$		
F ₄	5.14E+01 (2.06E+01)	$4.83E+01(2.40E+01) =$	$5.49E+01(1.57E+01) =$	$5.19E+01(2.08E+01) =$		
F5	2.53E+01 (5.26E+00)	$5.86E+01(1.72E+01) +$	$2.40E+01$ (5.55E+00) =	$4.65E+01$ $(1.45E+01) +$		
F ₆	1.11E-11 (5.05E-11)	4.34E-13 (5.60E-13) -	$4.30E-12(1.27E-11) =$	2.02E-12 (6.21E-12) -		
F7	$5.70E+01(5.11E+00)$	$7.55E+01(1.27E+01) +$	$5.95E+01(4.62E+00) +$	$6.01E+01(8.63E+00) =$		
F8	2.47E+01 (6.09E+00)	$5.74E+01(1.64E+01) +$	$2.66E+01(6.32E+00) =$	$4.91E+01(1.70E+01) +$		
F9	1.07E-02 (6.45E-02)	$1.87E-01(8.92E-01) =$	$1.75E-03(1.25E-02) =$	$3.51E-03$ $(1.75E-02)$ =		
F10	$1.62E+03(3.82E+02)$	$2.80E+03(5.90E+02) +$	$1.52E+03(4.61E+02) =$	$2.49E+03(4.60E+02) +$		
F11	1.94E+01 (2.19E+01)	$4.38E+01(2.84E+01) +$	$1.54E+01$ $(1.54E+01) =$	$2.58E+01(2.36E+01) =$		
F12	$1.32E+03(6.03E+02)$	$1.96E+03(2.49E+03) =$	$2.21E+03(3.79E+03) =$	$1.72E+03(1.63E+03) =$		
F13	6.17E+01 (5.47E+01)	$6.98E+01(5.76E+01) =$	$6.04E+01(8.58E+01) =$	$6.40E+01(5.33E+01) =$		
F14	3.64E+01 (1.56E+01)	$3.76E+01(9.77E+00) =$	$3.65E+01(1.66E+01) =$	$3.52E+01(1.12E+01) =$		
F ₁₅	2.88E+01 (2.26E+01)	$2.51E+01$ $(1.83E+01) =$	$2.72E+01(1.74E+01) =$	$2.79E+01(1.88E+01) =$		
F16	3.08E+02 (1.25E+02)	$8.64E+02$ (2.53E+02) +	$3.33E+02$ (1.53E+02) =	$8.54E+02$ (3.13E+02) +		
F17	4.13E+01 (2.93E+01)	$2.89E+02$ (1.77E+02) +	$4.57E+01(2.09E+01) +$	$2.05E+02$ (1.27E+02) +		
F18	7.44E+02 (3.81E+03)	$1.56E+02(4.26E+02) =$	$6.78E+02(3.17E+03) =$	$8.66E+02(4.04E+03) =$		
F19	1.88E+01 (1.64E+01)	$1.48E+01$ (4.10E+00) =	$1.61E+01(8.53E+00) =$	$2.12E+01(1.91E+01) =$		
F ₂₀	3.66E+01 (4.67E+01)	$2.75E+02$ (1.45E+02) +	$6.17E+01(5.35E+01) +$	$2.10E+02(1.48E+02) +$		
F21	2.28E+02 (6.25E+00)	$2.61E+02(2.82E+01) +$	$2.27E+02(5.76E+00) =$	$2.54E+02(1.75E+01) +$		
F ₂₂	$1.00E+02(0.00E+00)$	$1.00E+02(2.80E-01) =$	$1.00E+02(0.00E+00) =$	$1.00E+02(0.00E+00) =$		
F ₂₃	3.71E+02 (7.99E+00)	$4.07E+02$ (1.82E+01) +	$3.72E+02(7.29E+00) =$	$3.98E+02(1.41E+01) +$		
F24	4.42E+02 (5.47E+00)	$4.78E+02$ (1.92E+01) +	$4.41E+02(5.97E+00) =$	$4.74E+02$ (1.57E+01) +		
F ₂₅	$3.87E+02(0.00E+00)$	$3.87E+02$ (5.60E-01) =	$3.87E+02(0.00E+00) =$	$3.87E+02(0.00E+00)=$		
F ₂₆	$1.18E+03(1.49E+02)$	$1.67E+03(2.06E+02) +$	$1.18E+03(8.62E+01) =$	$1.50E+03(1.71E+02) +$		
F ₂₇	5.00E+02 (8.39E+00)	$5.11E+02(9.11E+00) +$	$5.00E+02(7.55E+00) =$	$5.07E+02(8.12E+00) +$		
F ₂₈	3.29E+02 (5.14E+01)	$3.28E+02(5.33E+01) =$	$3.34E+02(5.09E+01) =$	$3.30E+02(5.03E+01) =$		
F ₂₉	$4.31E+02$ (3.01E+01)	$6.53E+02$ (1.50E+02) +	$4.32E+02(3.56E+01) =$	$5.63E+02(1.22E+02) +$		
F30	2.12E+03 (1.90E+02)	$2.13E+03$ (1.48E+02) =	$2.46E+03$ (1.60E+03) =	$2.16E+03$ (1.50E+02) +		
$+/-/$		14/15/1	3/27/0	13/16/1		
The symbols "+/=/-" show the statistical results of the Wilcoxon signed-rank test with $\alpha = 0.05$ significance level. "+" cancacante that SETD variant is significantly superior than the corresponding variant. "=" represents that the						

performance difference between SFTD variant and the corresponding variant is not statistically significant. And, "
""
represents that SFTD variant and the corresponding variant is not statistically significant. And, "
""
r

namely SaDE [11], EPSDE [12], CoDE [13], SHADE [18], MPEDE [20], and EDEV [21]. For all of the test algorithms, the control parameters are initialized to the recommended values by their authors. Additionally, $FT_{init} = 100$ and $FT_{fin} = 5$ are used for the proposed Cauchy mutation, and $FT = 5$ is used for the previous Cauchy mutation as same as in [5].

Table [3](#page-9-0) presents the averages and the standard deviations of the FEVs obtained by independently running each algorithm 51 times at 30 dimensions. The summary of the experimental results is as follows.

- SaDE [11] and EPSDE [12]: These are the self-adaptive DE variants. ACM-SaDE finds 17/4 statistically better/worse solutions compared to CM-SaDE and 14/3 compared to SaDE. Also, ACM-EPSDE finds 14/3 statistically better/worse solutions compared to
- CM-EPSDE and 11/0 compared to EPSDE. • CoDE [13]: This DE variant uses three composite mutation strategies. ACM-CoDE finds 16/8 statistically better/worse solutions compared to CM-CoDE and 23/0 compared to CoDE.
- SHADE [18]: This DE variant uses an advanced mutation strategy. ACM-SHADE finds 15/0 statistically better/worse solutions compared to CM-SHADE and 8/1 compared to SHADE.
- MPEDE [20] and EDEV [21]: These are the multipopulation-based DE variants. ACM-MPEDE finds 15/0 statistically better/worse solutions compared to CM-MPEDE and 18/0 compared to MPEDE. Also, ACM-EDEV finds 15/1 statistically better/worse solutions compared to CM-EDEV and 12/0 compared to EDEV.

Although CM-DE variants are more greedy than ACM-DE variants, the performance differences between them on the unimodal functions (F_1-F_3) are insignificant. On the contrary, ACM-DE variants yield significantly better performance than CM-DE variants on the multimodal functions (F_4-F_{30}) . These observations are supported by the convergence graphs of EDEV variants from Fig. [6.](#page-9-1) We selected EDEV because it is the most recently proposed one among the six advanced DE variants. Additionally, EDEV is a multi-population-based DE variant that uses three DE variants that have different characteristics. As shown in the figures, the convergence speed of CM-EDEV is faster than ACM-EDEV, but it gets trapped in a local optimum during the search process. On the other hand, the convergence speed of ACM-EDEV is more robust than CM-EDEV. We can see a similar tendency at 50 dimensions from Table [4](#page-10-0) and Fig. [7.](#page-10-1)

As a result, the proposed Cauchy mutation can enhance the performance of advanced DE variants, especially for multimodal functions.

VII. ANALYSIS OF ACM-DE

We have insisted that the performance of ACM-DE is due to the sigmoid based parameter control and the *p*-best individual based Cauchy mutation. The advantages of using the *p*-best individual information over using the best individual information have been verified successfully [9], [18], [19]. Therefore, we performed a series of experiments to verify the contribution of the sigmoid based parameter control only.

A. SIGMOID BASED PARAMETER CONTROL

We carried out experiments with four parameter controls for the failure threshold as follows.

- 1) SFTD: decreases *FT* as a sigmoid function of *g*.
- 2) SFTI: increases *FT* as a sigmoid function of *g*.

TABLE 7. ACM-DE/rand/1/bin with different failure threshold settings ($D = 30$).

ents that the The symb mbois "++-" snow the statsical results of the Wicoxon support-rank test with $\alpha = 0.10$ significant clevel. "+" represents that $F^T T_{\tilde{i}n} t = 100$ variant is significantly singerior than the corresponding variant."
perf

TABLE 8. ACM-EDEV with different failure threshold settings ($D = 30$).

The symbols "+/=/" show the statistical results of the Wilcoxon signed-rank test with $\alpha = 0.05$ significance level. "+" represents that $FT_{init} = 100$ variant is significantly superior than the corresponding variant. "=" re

3) LFTD: decreases *FT* as a linear function of *g*.

4) LFTI: increases *FT* as a linear function of *g*.

There are two failure threshold decrease approaches (SFTD and LFTD) and two failure threshold increase approaches (SFTI and LFTI). SFTD and LFTD assign a high failure threshold at the beginning of the search process and gradually reduces it over generations, while SFTI and LFTI do the opposite. We used DE/rand/1/bin as a conventional DE variant and EDEV as an advanced DE variant. For ACM-DE/rand/1/bin, the scaling factor *F*, the crossover rate *CR*, and the population size *NP* are initialized to 0.5, 0.5, and 100, respectively, and for ACM-EDEV, the control parameters are initialized to the recommended values by their authors. Tables [5](#page-11-1) and [6](#page-11-2) present the averages and the standard deviations of the FEVs obtained by independently running DE/rand/1/bin variants and EDEV variants 51 times at 30 dimensions. As we can see from the tables, both DE/rand/1/bin and EDEV work best with SFTD. SFTD and LFTD yield significantly better performance than SFTI and LFTI because SFTD and LFTD can establish a balance between exploration and exploitation, while SFTI and LFTI cannot. Also, SFTD yields slightly better performance than LFTD because SFTD has higher exploration in the first half of the search process and higher exploitation in the second half of the search process.

B. FAILURE THRESHOLD

ACM-DE introduces two control parameters *FTinit* and *FTfin* that determine a new failure threshold at each generation. Therefore, further investigation to find appropriate control parameters may be intriguing. We carried out experiments with seven initial failure thresholds $FT_{init} \in$ {30, 50, 80, 100, 130, 150, 180} and one final failure threshold $FT_{fin} = 5$. We used DE/rand/1/bin as a conventional DE variant and EDEV as an advanced DE variant. For ACM-DE/rand/1/bin, the scaling factor *F*, the crossover rate *CR*, and the population size *NP* are initialized to 0.5, 0.5, and 100, respectively, and for ACM-EDEV, the control parameters are initialized to the recommended values by their authors. Tables [7](#page-12-0) and [8](#page-12-1) present the averages and the standard deviations of the FEVs obtained by independently running DE/rand/1/bin variants and EDEV variants 51 times at 30 dimensions. As we can see from the tables, both ACM-DE/ rand/1/bin and ACM-EDEV work best with FT_{init} ∈ [80, 130]. As a result, the setting of FT_{init} = 100 and $FT_{fin} = 5$ is considered as a standard setting of ACM-DE.

VIII. CONCLUSION

EAs need to establish a balance between exploration and exploitation to be successful [6], [7]. The previous Cauchy mutation of MDE was proposed to increase the convergence speed of DE by using the best individual information and the Cauchy distribution, which can locate the consecutively failed individuals to new positions close to the best individual. Although the effectiveness of MDE has been demonstrated successfully on the classical benchmark problems, MDE lacks robustness and faces difficulty in optimizing complex problems because of its strong exploitation.

We have proposed a variant of MDE called advanced Cauchy mutation DE (ACM-DE). We employed a sigmoid based parameter control, which alters the failure threshold for performing the Cauchy mutation in a time-varying schedule. That is, ACM-DE assigns a high failure threshold at the beginning of the search process and gradually reduces it over generations. We also employed the *p*-best individual based Cauchy mutation to prevent premature convergence.

ACM-DE has been tested on a set of 30 different and difficult CEC 2017 benchmark problems. Experiments and comparisons were carried out with six conventional and six advanced DE variants to demonstrate the effectiveness of the proposed Cauchy mutation. The experiment results indicate that the DE variants assisted by the proposed Cauchy mutation have better or at least competitive performance in terms of accuracy and robustness, particularly for multimodal functions compared to the DE variants assisted by the previous Cauchy mutation as well as the original DE variants.

Possible directions for future work include 1) designing a mutation based on the multivariate Cauchy distribution; 2) extending the proposed Cauchy mutation for multi-objective optimization; 3) applying the proposed Cauchy mutation to other population-based metaheuristics.

REFERENCES

- [1] R. Storn and K. Price, "Differential evolution-A simple and efficient heuristic for global optimization over continuous spaces,'' *J. Global Optim.*, vol. 11, no. 4, pp. 341–359, 1997.
- [2] K. Price, R. M. Storn, and J. A. Lampinen, *Differential Evolution: A Practical Approach to Global Optimization*. Berlin, Germany: Springer, 2006.
- [3] S. Das and P. N. Suganthan, ''Differential evolution: A survey of the stateof-the-art,'' *IEEE Trans. Evol. Comput.*, vol. 15, no. 1, pp. 4–31, Feb. 2011.
- [4] S. Das, S. S. Mullick, and P. Suganthan, "Recent advances in differential evolution—An updated survey,'' *Swarm Evol. Comput.*, vol. 27, pp. 1–30, Apr. 2016.
- [5] M. Ali and M. Pant, ''Improving the performance of differential evolution algorithm using Cauchy mutation,'' *Soft Comput.*, vol. 15, no. 5, pp. 991–1007, May 2011.
- [6] A. Eiben and C. Schippers, ''On evolutionary exploration and exploitation,'' *Fundamenta Informaticae*, vol. 35, nos. 1–4, pp. 35–50, 1998.
- [7] M. Črepinšek, S.-H. Liu, and M. Mernik, "Exploration and exploitation in evolutionary algorithms: A survey,'' *CSURACM Comput. Surv.*, vol. 45, no. 3, pp. 1–33, Jun. 2013.
- [8] T. J. Choi and C. W. Ahn, ''Accelerating differential evolution using multiple exponential cauchy mutation,'' in *Proc. Genetic Evol. Comput. Conf. Companion*, 2018, pp. 207–208.
- [9] J. Zhang and A. Sanderson, ''JADE: Adaptive differential evolution with optional external archive,'' *IEEE Trans. Evol. Comput.*, vol. 13, no. 5, pp. 945–958, Oct. 2009.
- [10] N. H. Awad, M. Z. Ali, J. J. Liang, B. Y. Qu, and P. N. Suganthan, ''Problem definitions and evaluation criteria for the CEC 2017 special session and competition on single objective bound constrained real-parameter numerical optimization,'' Nanyang Technol. Univ., Singapore, Tech. Rep., Nov. 2016. [Online]. Available: http://www.ntu.edu.sg/home/epnsugan/
- [11] A. Qin, V. Huang, and P. Suganthan, ''Differential evolution algorithm with strategy adaptation for global numerical optimization,'' *IEEE Trans. Evol. Comput.*, vol. 13, no. 2, pp. 398–417, Apr. 2009.
- [12] R. Mallipeddi, P. N. Suganthan, Q.-K. Pan, and M. F. Tasgetiren, "Differential evolution algorithm with ensemble of parameters and mutation strategies,'' *Appl. Soft Comput.*, vol. 11, no. 2, pp. 1679–1696, 2011.
- [13] Y. Wang, Z. Cai, and Q. Zhang, ''Differential evolution with composite trial vector generation strategies and control parameters,'' *IEEE Trans. Evol. Comput.*, vol. 15, no. 1, pp. 55–66, Feb. 2011.
- [14] T. J. Choi, J.-H. Lee, H. Y. Youn, and C. W. Ahn, "Adaptive differential evolution with elite opposition-based learning and its application to training artificial neural networks,'' *Fundamenta Informaticae*, vol. 164, nos. 2–3, pp. 227–242, 2019.
- [15] T. J. Choi, Y.-G. Cheong, and C. W. Ahn, "A performance comparison of crossover variations in differential evolution for training multi-layer perceptron neural networks,'' in *Proc. Int. Conf. Bio-Inspired Comput., Theories Appl.* Singapore: Springer, 2018, pp. 477–488.
- [16] T. J. Choi and C. W. Ahn, ''Adaptive cauchy differential evolution with strategy adaptation and its application to training large-scale artificial neural networks,'' in *Proc. Int. Conf. Bio-Inspired Comput., Theories Appl.* Singapore: Springer, 2017, pp. 502–510.
- [17] T. J. Choi and C. W. Ahn, ''An improved differential evolution algorithm and its application to large-scale artificial neural networks,'' *J. Phys., Conf. Ser.*, vol. 806, Feb. 2017, Art. no. 012010.
- [18] R. Tanabe and A. Fukunaga, "Success-history based parameter adaptation for differential evolution,'' in *Proc. IEEE Congr. Evol. Comput.*, Jun. 2013, pp. 71–78.
- [19] R. Tanabe and A. S. Fukunaga, ''Improving the search performance of shade using linear population size reduction,'' in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Jul. 2014, pp. 1658–1665.
- [20] G. Wu, R. Mallipeddi, P. Suganthan, R. Wang, and H. Chen, "Differential evolution with multi-population based ensemble of mutation strategies,'' *Inf. Sci.*, vol. 329, pp. 329–345, Feb. 2016.
- [21] G. Wu, X. Shen, H. Li, H. Chen, A. Lin, and P. Suganthan, "Ensemble of differential evolution variants,'' *Inf. Sci.*, vol. 423, pp. 172–186, Jan. 2018.
- [22] T. J. Choi and C. W. Ahn, "An adaptive population resizing scheme for differential evolution in numerical optimization,'' *J. Comput. Theor. Nanosci.*, vol. 12, no. 7, pp. 1336–1350, 2015.
- [23] T. J. Choi and C. W. Ahn, "An adaptive differential evolution algorithm with automatic population resizing for global numerical optimization,'' in *Bio-Inspired Computing—Theories and Applications*. Berlin, Germany: Springer, 2014, pp. 68–72.
- [24] V. Stanovov, S. Akhmedova, and E. Semenkin, "LSHADE algorithm with rank-based selective pressure strategy for solving CEC 2017 benchmark problems,'' in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Jul. 2018, pp. 1–8.
- [25] N. H. Awad, M. Z. Ali, and P. N. Suganthan, "Ensemble sinusoidal differential covariance matrix adaptation with Euclidean neighborhood for solving CEC2017 benchmark problems,'' in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Jun. 2017, pp. 372–379.
- [26] J. Brest, M. S. Maučec, and B. Bošković, ''Single objective real-parameter optimization: Algorithm JSO,'' in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Jun. 2017, pp. 1311–1318.
- [27] P. Bujok and J. Tvrdík, ''Enhanced individual-dependent differential evolution with population size adaptation,'' in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Jun. 2017, pp. 1358–1365.
- [28] A. W. Mohamed, A. A. Hadi, A. M. Fattouh, and K. M. Jambi, ''Lshade with semi-parameter adaptation hybrid with CMA-ES for solving CEC 2017 benchmark problems,'' in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Jun. 2017, pp. 145–152.
- [29] D. Jagodziński and J. Arabas, ''A differential evolution strategy,'' in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Jun. 2017, pp. 1872–1876.
- [30] T. J. Choi and C. W. Ahn, ''Adaptive α-stable differential evolution in numerical optimization,'' *Natural Comput.*, vol. 16, no. 4, pp. 637–657, 2017.
- [31] T. J. Choi and C. W. Ahn, ''An adaptive cauchy differential evolution algorithm with population size reduction and modified multiple mutation strategies,'' in *Proc. 18th Asia–Pacific Symp. Intell. Evol. Syst.*, vol. 2. Cham, Switzerland: Springer, 2015, pp. 13–26.
- [32] T. J. Choi and C. W. Ahn, "An adaptive cauchy differential evolution algorithm with bias strategy adaptation mechanism for global numerical optimization,'' *J. Comput.*, vol. 9, no. 9, pp. 2139–2145, 2014.
- [33] T. J. Choi, C. W. Ahn, and J. An, ''An adaptive cauchy differential evolution algorithm for global numerical optimization,'' *Sci. World J.*, vol. 2013, pp. 1–12, Jun. 2013.
- [34] T. J. Choi and Y. Lee, ''Asynchronous differential evolution with strategy adaptation for global numerical optimization,'' in *Proc. 2nd High Perform. Comput. Cluster Technol. Conf.*, 2018, pp. 15–18.
- [35] T. J. Choi and Y. Lee, "Asynchronous differential evolution with selfadaptive parameter control for global numerical optimization,'' in *Proc. MATEC Web Conf.*, vol. 189, 2018, p. 03020.
- [36] A. Moraglio and J. Togelius, ''Geometric differential evolution,'' in *Proc. 11th Annu. Conf. Genetic Evol. Comput.*, 2009, pp. 1705–1712.
- [37] A. Moraglio, J. Togelius, and S. Silva, ''Geometric differential evolution for combinatorial and programs spaces,'' *Evol. Comput.*, vol. 21, no. 4, pp. 591–624, 2013.
- [38] M. Yang, C. Li, Z. Cai, and J. Guan, "Differential evolution with autoenhanced population diversity,'' *IEEE Trans. Cybern.*, vol. 45, no. 2, pp. 302–315, Feb. 2015.
- [39] M. Yang, J. Guan, and C. Li, "Differential evolution with auto-enhanced population diversity: The experiments on the cec'2016 competition,'' in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Jul. 2016, pp. 4785–4789.
- [40] R. D. Al-Dabbagh, F. Neri, N. Idris, and M. S. Baba, ''Algorithmic design issues in adaptive differential evolution schemes: Review and taxonomy,'' *Swarm Evol. Comput.*, vol. 43, pp. 284–311, Dec. 2018.
- [41] K. R. Opara and J. Arabas, "Differential Evolution: A survey of theoretical analyses,'' *Swarm Evol. Comput.*, vol. 44, pp. 546–558, Feb. 2019.
- [42] H. Qin, J. Zhou, Y. Lu, Y. Wang, and Y. Zhang, ''Multi-objective differential evolution with adaptive cauchy mutation for short-term multi-objective optimal hydro-thermal scheduling,'' *Energy Convers. Manage.*, vol. 51, no. 4, pp. 788–794, 2010.
- [43] H. Zhang, X. Lei, C. Wang, D. Yue, and X. Xie, "Adaptive grid based multi-objective Cauchy differential evolution for stochastic dynamic economic emission dispatch with wind power uncertainty,'' *PLoS ONE*, vol. 12, no. 9, Sep. 2017, Art. no. e0185454.
- [44] K. Deb and S. Jain, "Running performance metrics for evolutionary multiobjective optimization,'' Kanpur Genetic Algorithms Lab., Indian Inst. Technol. Kanpur, Kanpur, India, KanGAL Rep. 2002004, 2002.
- [45] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, Jan. 2006.

TAE JONG CHOI received the Ph.D. degree in electrical and computer engineering from Sungkyunkwan University (SKKU), in 2017. He is currently a Postdoctoral Fellow with the Department of Computer Science and Engineering, New York University Tandon School of Engineering. He is currently working on designing a fast and robust large-scale evolutionary algorithm, designing an effective illumination algorithm empowered by evolutionary algorithms, and

neuro-evolutions. His research interest includes artificial intelligence, especially for evolutionary algorithms and their applications.

JULIAN TOGELIUS received the B.A. degree from Lund University, the M.Sc. degree from the University of Sussex, and the Ph.D. degree from the University of Essex. He was with IDSIA, Lugano, and also with the IT University of Copenhagen. He is currently an Associate Professor with the Department of Computer Science and Engineering, New York University, USA. He works on artificial intelligence for games and on games for artificial intelligence. His current main

research directions involve procedural content generation in games, general video game playing, player modeling, and fair and relevant benchmarking of AI through game-based competitions. Additionally, he works on topics in evolutionary computation, quality-diversity algorithms, and reinforcement learning. He is the Editor-in-Chief of the IEEE TRANSACTIONS ON GAMES.

YUN-GYUNG CHEONG received the B.S. degree and the M.S. degree in information engineering from Sungkyunkwan University (SKKU), in 1996 and 1998, respectively, and the Ph.D. degree in computer science from North Carolina State University, Raleigh, NC, USA, in 2007. She was a Postdoctoral Fellow with the Center for Computer Games Research, IT University of Copenhagen, from 2010 to 2014 and a Researcher with the Samsung Advanced Institute of Tech-

nology, from 2007 to 2010. She is currently an Associate Professor with Sungkyunkwan University, South Korea. She is working on natural language techniques and interactive storytelling techniques on various platforms, including 3D animation, VR, and games. She is interested in computational models of stories and text modeling. Her research interest includes artificial intelligence with an emphasis on its use in discourse planning for narrative, text, and games.

 $\frac{1}{2}$