Front. Comput. Sci., 2018, 12(2): 297–315
https://doi.org/10.1007/s11704-016-5353-5

RESEARCH ARTICLE

On the selection of solutions for mutation in differential evolution

Yong WANG (✉)[1,2], Zhi-Zhong LIU[1], Jianbin LI[3], Han-Xiong LI[4,5], Jiahai WANG[6]

1 School of Information Science and Engineering, Central South University, Changsha 410083, China
2 School of Computer Science and Informatics, De Montfort University, Leicester LE1 9BH, UK
3 Institute of Information Security and Big Data, Central South University, Changsha 410083, China
4 Department of Systems Engineering and Engineering Management, City University of Hong Kong, Hong Kong, China
5 State Key Laboratory of High Performance Complex Manufacturing, Central South University, Changsha 410083, China
6 School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510006, China

Abstract Differential evolution (DE) is a kind of evolutionary algorithms, which is suitable for solving complex optimization problems. Mutation is a crucial step in DE that generates new solutions from old ones. It was argued and has been commonly adopted in DE that the solutions selected for mutation should have mutually different indices. This restrained condition, however, has not been verified either theoretically or empirically yet. In this paper, we empirically investigate the selection of solutions for mutation in DE. From the observation of the extensive experiments, we suggest that the restrained condition could be relaxed for some classical DE versions as well as some advanced DE variants. Moreover, relaxing the restrained condition may also be useful in designing better future DE algorithms.

Keywords differential evolution, mutation, the selection of solutions for mutation, evolutionary algorithms

1 Introduction

Differential evolution (DE), proposed by Storn and Price in 1995 [1,2], is one of the most popular evolutionary algorithm (EA) paradigms in the community of evolutionary computation. Like other EA paradigms, DE is a population based optimization method, which contains a lot of solutions. In DE, each solution in the population is called a target vector. DE includes three main operators, i.e., mutation, crossover, and selection. In the classical DE, for each target vector, a mutant vector is generated by making use of the mutation operator. Afterward, the crossover operator is implemented on the target vector and the mutant vector, and thus, a trial vector is obtained. Finally, the target vector is compared with the trial vector, and the better one will be selected for the next population. The mutation operator and the crossover operator together are called the trial vector generation strategy, since they are utilized to generate the trial vector. DE also contains three important control parameters, i.e., the population size, the scaling factor in the mutation operator, and the crossover control parameter in the crossover operator.

Received September 2, 2015; accepted June 17, 2016

E-mail: ywang@csu.edu.cn

Recent years have witnessed the significant progress in the area of DE. Some representatives are briefly introduced as follows:

- How to improve the trial vector generation strategy of DE has attracted considerable interest. For example, Fan and Lampinen [3] proposed a trigonometric mutation as a local search operator. Zhang and Sanderson [4] presented a new mutation operator called DE/current-to-*p*best/1. Das et al. [5] proposed a neighborhood-based mutation operator. Wang et al. [6] used an orthogonal crossover to enhance the search ability of DE and suggested a generic DE framework. Very recently, Guo and Yang [7] and Wang et al. [8] utilized the population distribution information to establish an Eigen coordinate system, and implemented the crossover operator in the Eigen coordinate system with a predefined probability.

- Improving DE's performance by adapting the control parameter setting has also been an active research direction. For example, Liu and Lampinen [9] used fuzzy logic controllers to adapt the scaling factor and the crossover control parameter. Brest et al. [10] designed an efficient technique to self-adapt the scaling factor and the crossover control parameter. In Ref. [4], the scaling factor is generated according to a Cauchy distribution and the crossover control parameter is generated according to a normal distribution.
- Some researchers investigated hybridizing DE with other search techniques. For example, Noman and Iba [11] combined an adaptive local search with DE. Rahnamayan et al. [12] adopted opposition-based learning to improve the convergence rate of DE. Sun et al. [13] proposed a combination of DE and estimation of distribution algorithm (EDA).
- Recently, much attention has been paid to integrate multiple trial vector generation strategies with multiple control parameter settings in DE. For example, Qin et al. [14] proposed a self-adaptive DE, in which both the trial vector generation strategies and the control parameter settings are gradually self-adapted according to the previous experiences. Mallipeddi et al. [15] employed an ensemble of control parameter settings and trial vector generation strategies with DE. Wang et al. [16] exploited DE researchers' experiences to construct the strategy candidate pool and the parameter candidate pool, and randomly combined the trial vector generation strategies with the control parameter settings to create multiple trial vectors for each target vector.

In the first DE paper [1], Storn and Price argued that the solutions chosen for mutation should have mutually different indices. Later, this restrained condition has been broadly recognized by DE researchers during the past twenty years [17]. However, the rationality of this restrained condition has not been verified either theoretically or experimentally. Motivated by the above consideration, in this paper we investigate the selection of solutions for mutation in DE empirically. From the results of extensive experiments, some interesting phenomena have been observed:

- If this restrained condition is relaxed, the mutation operators of DE might degenerate due to the fact that the differential vector will be equal to zero with a small probability. This phenomenon slightly decreases the diversity of the population and has an advantage of enhancing the convergence speed.
- This restrained condition could be relaxed for some classical DE versions with high randomness and for some advanced DE variants with very competitive performance. It is because the performance of such DE versions and variants can be further improved by accelerating the convergence.
- However, this restrained condition cannot be removed from the relatively greedy DE, in which the information of the best solution in the population is exploited frequently. It is not difficult to understand since the greedy DE has a very fast convergence speed and poor diversity, and relaxing this restrained condition further reduces the diversity and results in performance degradation. Moreover, this restrained condition cannot be removed from DE with a small population size, and DE with a small scaling factor and meanwhile a large crossover control parameter.

By investigating the effect of the solution selection for mutation on the performance of different versions and variants of DE, this paper is helpful for revealing and understanding the search mechanism of DE, since mutation is the main operator and characteristic of DE.

The remainder of this paper is organized as follows. In Section 2, DE is briefly introduced, including its mutation, crossover and selection operators. Section 3 discusses the selection of solutions for mutation in DE. The experimental results are provided in Section 4. Section 5 concludes this paper.

2 Differential evolution (DE)

DE is a population-based optimizer. The population of DE at generation G can be formulated as follows:

$$P_G = \{\vec{x}_{i,G} = (x_{i,1,G}, x_{i,2,G}, \ldots, x_{i,D,G}), i = 1, 2, \ldots, NP\}, \quad (1)$$

where NP is the population size, $\vec{x}_{i,G}$ is the ith solution (also called the ith target vector) in the population, and D is the number of decision variables contained by each target vector.

For each target vector $\vec{x}_{i,G}$, the following six classical mutation operators have been broadly applied to create a mutant vector $\vec{v}_{i,G} = (v_{i,1,G}, v_{i,2,G}, \ldots, v_{i,D,G})$:

- DE/rand/1:

$$\vec{v}_{i,G} = \vec{x}_{r1,G} + F \times (\vec{x}_{r2,G} - \vec{x}_{r3,G}), \quad (2)$$

- DE/rand/2:

$$\vec{v}_{i,G} = \vec{x}_{r1,G} + F \times (\vec{x}_{r2,G} - \vec{x}_{r3,G}) + F \times (\vec{x}_{r4,G} - \vec{x}_{r5,G}), \quad (3)$$

- DE/best/1:

$$\vec{v}_{i,G} = \vec{x}_{best,G} + F \times (\vec{x}_{r1,G} - \vec{x}_{r2,G}), \quad (4)$$

- DE/best/2:

$$\vec{v}_{i,G} = \vec{x}_{best,G} + F \times (\vec{x}_{r1,G} - \vec{x}_{r2,G}) + F \times (\vec{x}_{r3,G} - \vec{x}_{r4,G}), \quad (5)$$

- DE/current-to-best/1:

$$\vec{v}_{i,G} = \vec{x}_{i,G} + F \times (\vec{x}_{best,G} - \vec{x}_{i,G}) + F \times (\vec{x}_{r1,G} - \vec{x}_{r2,G}), \quad (6)$$

- DE/current-to-rand/1:

$$\vec{v}_{i,G} = \vec{x}_{i,G} + rand \times (\vec{x}_{r1,G} - \vec{x}_{i,G}) + F \times (\vec{x}_{r2,G} - \vec{x}_{r3,G}), \quad (7)$$

where F is the scaling factor, $\vec{x}_{best,G}$ is the best target vector in the population, and $rand$ denotes a uniformly distributed random number between 0 and 1. Moreover, in the first, second, and sixth mutation operators, indices $r1$, $r2$, $r3$, $r4$, and $r5$ represent the mutually different integers randomly chosen from $\{1, 2, \ldots, NP\} \backslash i$, and in the remaining three mutation operators, indices $r1$, $r2$, $r3$, and $r4$ represent the mutually different integers randomly chosen from $\{1, 2, \ldots, NP\} \backslash \{i, best\}$. It is necessary to note that in the classcial DE, usually one mutant vector is produced for each target vector.

After the mutation, the binomial crossover is usually implemented on the target vector $\vec{x}_{i,G}$ and the mutant vector $\vec{v}_{i,G}$ to generate a trial vector $\vec{u}_{i,G} = (u_{i,1,G}, u_{i,2,G}, \ldots, u_{i,D,G})$ as follows:

$$u_{i,j,G} = \begin{cases} v_{i,j,G}, \text{ if } rand_j \leqslant CR \text{ or } j = j_{rand}; \\ x_{i,j,G}, \text{ otherwise.} \end{cases} \quad (8)$$

where CR is the crossover control parameter, j_{rand} is a randomly chosen integer from $[1, D]$, and $rand_j$ denotes a uniformly distributed random number from $[0, 1]$ and regenerated for each j. The purpose of j_{rand} is to make $\vec{u}_{i,G}$ different from $\vec{x}_{i,G}$ by at least one dimension.

The aim of the selection operator is to choose the better one from the target vector $\vec{x}_{i,G}$ and the trial vector $\vec{u}_{i,G}$ as follows (in the minimization sense):

$$\vec{x}_{i,G+1} = \begin{cases} \vec{u}_{i,G}, \text{ if } f(\vec{u}_{i,G}) \leqslant f(\vec{x}_{i,G}); \\ \vec{x}_{i,G}, \text{ otherwise.} \end{cases} \quad (9)$$

The general framework of DE has been given in Algorithm 1 [18, 19].

Algorithm 1 The general framework of DE

1 $G = 1$; /* G denotes the generation number */
2 Randomly generate an initial population $P_G = \{\vec{x}_{1,G}, \vec{x}_{2,G}, \ldots, \vec{x}_{NP,G}\}$ throughout the decision space;
3 Evaluate each solution in P_G according to the fitness function;
4 $FEs = NP$; /* FEs denotes the number of fitness evaluations */
5 $P_{G+1} = \emptyset$;
6 **for** each solution $\vec{x}_{i,G}$ (also called a target vector) in P_G **do**
7   Implement the mutation operator to generate the mutant vector $\vec{v}_{i,G}$;
8   Implement the crossover operator on $\vec{x}_{i,G}$ and $\vec{v}_{i,G}$ to produce the trial vector $\vec{u}_{i,G}$;
9   Evaluate $\vec{u}_{i,G}$ according to the fitness function;
10   Set $FEs = FEs + 1$;
11   Implement the selection operator to select a better one from $\vec{x}_{i,G}$ and $\vec{u}_{i,G}$, and store it into P_{G+1};
12 $G = G + 1$;
13 **Stopping Criterion**: If the maximum number of fitness evaluations is reached, then stop and output the best solution in P_G; otherwise go to Step 5.

3 On the selection of solutions for mutation in DE

According to the introduction in Section 2, it is clear that the solutions chosen for mutation should have mutually different indices. A question which arises naturally is why the above restrained condition should be satisfied for DE.

During the design of composite differential evolution (CoDE) [16], we have noticed that DE/current-to-rand/1 without the restrained condition is able to greatly enhance the performance (Tvrdík [20] also pointed out the above phenomenon after carefully implementing CoDE). However, due to space limitations, we have not investigated this issue in [16] in depth. Note that Price et al. [21] provided a preliminary analysis on the difference between DE with and without the restrained condition. However, they only tested the performance of DE/rand/1/bin with and without the restrained condition on the sphere function, and therefore, their conclusion is limited. Recently, Liu et al. [22] also noticed the impact of the above restrained condition on the performance of DE. They proposed an unrestrained method to generate the mutant vector, which allows the solutions in the population to appear repeatedly in the mutation operator. Unfortunately, they did not provide the related experimental results to verify the effectiveness of the unrestrained method.

Recognizing the current situation, in this paper we attempt to empirically study the selection of solutions for mutation in DE, which has been pending during the past twenty years. In our experiments, we compare the performance between DE with and without the restrained condition by extensive experiments. DE without the restrained condition means that the indices $r1$, $r2$, $r3$, $r4$, and $r5$ can be randomly chosen from $[1, NP]$.

Prior to the experiments, the properties of DE without the restrained condition are given as follows.

- **Property 1** At each generation, the probability that one differential vector (such as $(\vec{x}_{r2,G} - \vec{x}_{r3,G})$) degenerates to zero is $1/NP$ and the probability that two differential vectors degenerate to zero is $1/NP^2$.
- **Property 2** For DE/rand/1 and DE/rand/2, at each generation the probability that the mutant vector is equal to the base vector is $1/NP$ and $1/NP^2$, respectively.
- **Property 3** At each generation, the probability that DE/rand/2 degenerates to DE/rand/1 and DE/best/2 degenerates to DE/best/1 is $2(NP-1)/NP^2$.
- **Property 4** For DE/best/1 and DE/best/2, at each generation the probability that the mutant vector is equal to the best solution in the population is $1/NP$ and $1/NP^2$, respectively.
- **Property 5** At each generation, for DE/current-to-best/1, the probability that the mutant vector is a linear combination of the ith target vector and the best solution in the population is $1/NP$, and for DE/current-to-rand/1, the probability that the mutant vector is a linear combination of the ith target vector and a randomly selected solution in the population is also $1/NP$.
- **Property 6** For DE/rand/1 and DE/rand/2, the centers of all possible mutant vectors in DE with and without the restrained condition are $\frac{1}{(NP-1)}\sum_{j\in\{1,2,\ldots,NP\}\backslash i}\vec{x}_{j,G}$ and $\frac{1}{NP}\sum_{j\in\{1,2,\ldots,NP\}}\vec{x}_{j,G}$, respectively.
- **Property 7** For DE/best/1, DE/best/2, and DE/current-to-best/1, the centers of all possible mutant vectors of DE with the restrained condition are equal to that of DE without the restrained condition.

The differences between the current work and the previous work in Refs. [16, 21, 22] are the following:

- In Refs. [16, 21, 22], the researchers have studied on the restrained condition in DE. However, the experiments in those papers are insufficient. In this paper, systematic experiments have been conducted to compare DE with and without the restrained condition on two sets of benchmark test functions, namely, 14 test functions with 30 dimensions at the 2005 IEEE Congress on Evolutionary Computation (IEEE CEC2005) [23] and 28 test functions with 30 and 50 dimensions at the 2013 IEEE Congress on Evolutionary Computation (IEEE CEC2013) [24]. Moreover, six classical DE versions and seven advanced DE variants have been chosen to produce the experimental results.
- We have investigated the effect of the control parameter settings on the performance of DE with and without restrained condition.
- We have also given some guidelines on the types of DE where the restrained condition is favorable or could be relaxed.

4 Experimental study

In the experiments of this section, the function error value $(f(\vec{x}_{best}) - f(\vec{x}^*))$ has been recorded in each run for each test function, where $f(\vec{x}^*)$ is the objective function value of the global optimal solution and $f(\vec{x}_{best})$ is the objective function value of the best solution found when the evolution halts. We used the average and standard deviation of the function error values over all the runs to compare the experimental results. In order to make the comparison statistically sound, Wilcoxon's rank sum test at a 0.05 significance level was performed between pairwise methods. For each test function, the maximum number of function evaluations (FEs) was set to $10,000 \times D$.

4.1 On classical DE versions

Firstly, the first 14 test functions with 30 dimensions (30D) developed for IEEE CEC2005 [23] have been employed to investigate the influence of the solution selection for mutation on the performance of six classical DE versions, i.e., DE/rand/1/bin, DE/rand/2/bin/, DE/best/1/bin, DE/best/2/bin, DE/current-to-best/1/bin, and DE/current-to-rand/1/bin. These 14 test functions are denoted as $F_1 - F_{14}$, and can be divided into three categories: five unimodal functions ($F_1 - F_5$), seven basic multimodal functions ($F_6 - F_{12}$), and two expanded multimodal functions ($F_{13} - F_{14}$). We only considered these test functions with 30D in this paper. Note that in some papers (such as [16]), the binomial crossover is not applied to DE/current-to-rand/1 in order to keep its rotation invariance. However, in this paper the binomial crossover is also applied to DE/current-to-rand/1, and the corresponding DE version is called DE/current-to-rand/1/bin. For DE without the restrained condition, two letters "U-" are added to the original DE. For example, U-DE/rand/1/bin denotes DE/rand/1/bin without the restrained condition. For all the classical DE versions in this subsection, the following parameter settings were used: $NP = D$, $F = 0.9$, and $CR = 0.9$. The above parameter settings were the same as in Ref. [11]. For each test function, according to the suggestion in Ref. [23], 25 independent runs were performed.

In Section 3, we have introduced the probabilities of some situations in Property 1–Property 5. One may be interested in the real probability that each of situations in Property 1–Property 5 really happens in the experiments. To this end, Table 5 compares the theoretic probability and the real probability. From Table 5, it is clear that the real probability is nearly consistent with the theoretic probability, which demonstrates the contributions of these five properties directly.

The experimental results of the 14 test functions with 30D have been summarized in the supplemental file (Appendix:

Table 1 The theoretic probability and the real probability of each situation in Property 1–Property 5

Situation	Theoretic probability (NP=30)	Real probability (NP=30)
Property 1: one differential vector degenerates to zero	$1/NP \approx 0.03333$	0.03332
Property 1: two differential vectors degenerate to zero	$1/NP^2 \approx 0.00111$	0.00114
Property 2: in DE/rand/1, the mutant vector is equal to the base vector	$1/NP \approx 0.03333$	0.03334
Property 2: in DE/rand/2, the mutant vector is equal to the base vector	$1/NP^2 \approx 0.00111$	0.00112
Property 3: DE/rand/2 degenerates to DE/rand/1	$2(NP-1)/NP^2 \approx 0.06444$	0.06436
Property 3: DE/best/2 degenerates to DE/best/1	$2(NP-1)/NP^2 \approx 0.06444$	0.06433
Property 4: in DE/best/1, the mutant vector is equal to the best solution in the population	$1/NP \approx 0.03333$	0.03331
Property 4: in DE/best/2, the mutant vector is equal to the best solution in the population	$1/NP^2 \approx 0.00111$	0.00113
Property 5: in DE/current-to-best/1, the mutant vector is a linear combination of the ith target vector and the best solution in the population	$1/NP \approx 0.03333$	0.03335
Property 5: in DE/current-to-rand/1, the mutant vector is a linear combination of the ith target vector and a randomly selected solution in the population	$1/NP \approx 0.03333$	0.03332

Table 2 Statistical test results of six classical DE versions with and without the restrained condition over 25 independent runs on the 14 test functions with 30D from IEEE CEC2005 using 300,000 FEs

Test functions (30D)		U-DE/rand/1/bin vs DE/rand/1/bin	U-DE/rand/2/bin vs DE/rand/2/bin	U-DE/best/1/bin vs DE/best/1/bin	U-DE/best/2/bin vs DE/best/2/bin	U-DE/current-to-best/1/bin vs DE/current-to-best/1/bin	U-DE/current-to-rand/1/bin vs DE/current-to-rand/1/bin
Unimodal functions	F_1	+	+	−	+	−	+
	F_2	+	+	−	+	+	+
	F_3	+	+	−	+	+	+
	F_4	+	+	−	+	−	+
	F_5	−	+	−	+	−	+
Basic multimodal functions	F_6	+	+	−	+	−	+
	F_7	−	+	−	+	=	−
	F_8	=	+	−	+	=	=
	F_9	−	+	−	+	−	+
	F_{10}	+	+	−	+	−	+
	F_{11}	+	+	−	+	−	+
	F_{12}	=	+	−	+	=	=
Expanded multimodal functions	F_{13}	−	+	−	+	−	+
	F_{14}	+	+	−	+	+	+
+		8	14	0	14	3	11
−		4	0	14	0	8	1
=		2	0	0	0	3	2

Note: Wilcoxon's rank sum test at a 0.05 significance level is performed between a classical DE version with and without the restrained condition. "+", "−", and "=" denote that the performance of DE without the restrained condition is better than, worse than, and similar to that of DE with the restrained condition, respectively

Tables S1–S3) and the statistical test results have been summarized in Table 2. From Table 2, we can give the following comments:

1) U-DE/rand/1/bin outperforms DE/rand/1/bin on four out of five unimodal functions, which means that U-DE/rand/1/bin exhibits a faster convergence speed. The above phenomenon can be attributed to two aspects. In one aspect, at each generation the mutant vector will be equal to the base vector once on average according to Property 2. Under this condition, if the base vector (i.e., $\vec{x}_{r1,G}$) is better than the target vector (i.e., $\vec{x}_{i,G}$), the crossover operator might have the capability to improve some dimensions of the target vector directly. On the other hand, at each generation the base vector will be equal to its target vector once on average, since the probability that $\vec{x}_{r1,G} = \vec{x}_{i,G}$ is $1/NP$. As a result, the mutation is similar to a local search operator. In addition, U-DE/rand/1/bin performs better than DE/rand/1/bin on basic multimodal functions and has similar performance with DE/rand/1/bin on expanded multimodal functions. As far as the overall performance is considered, U-DE/rand/1/bin is better than DE/rand/1/bin.
2) For the mutation operators with two differential vectors (i.e., DE/rand/2 and DE/best/2), DE without the restrained condition has an edge over DE with the restrained condition on all the test functions, see, for example, U-DE/rand/2/bin versus DE/rand/2/bin, and U-DE/best/2/bin versus DE/best/2/bin. It is not difficult to understand since without the restraint, DE/rand/2/bin and DE/best/2/bin will degenerate to DE/rand/1/bin and DE/best/1/bin with the probability 0.06444 (i.e., 30×0.06444×10,000≈19,332 times on average during the evolution due to the fact that the total generation number is equal to 10,000) according to Property 3 as shown in Table 1, and in general, DE with one differential vector converges faster than DE with two differential vectors.
3) Like DE/rand/2/bin and DE/best/2/bin, the performance of DE/current-to-rand/1/bin is significantly outperformed by that of U-DE/current-to-rand/1/bin.
4) DE/best/1/bin beats U-DE/best/1/bin on all the test functions. In addition, DE/current-to-best/1/bin performs better than U-DE/current-to-best/1/bin in terms of the overall performance. It is because at each generation the mutant vector will be equal to the best solution in the population once on average for U-DE/best/1/bin according to Property 4, and the mutant vector will be a linear combination of the ith target vector and the best solution in the population once on average for U-DE/current-to-best/1/bin according to Property 5. Thus, the population might be easily trapped into a local optimum due to the information of the best solution being exploited frequently.

The above discussion suggests that the restrained condition could be relaxed for four classical DE versions, i.e., DE/rand/1/bin, DE/rand/2/bin, DE/best/2/bin, and DE/current-to-rand/1/bin, and is unavoidable for two classical DE versions, i.e., DE/best/1/bin and DE/current-to-best/1/bin. Figures 1 and 2 present the evolution of the average function error values derived from six classical DE versions with and without the restrained condition on F_2 with 30D and F_6 with 30D.

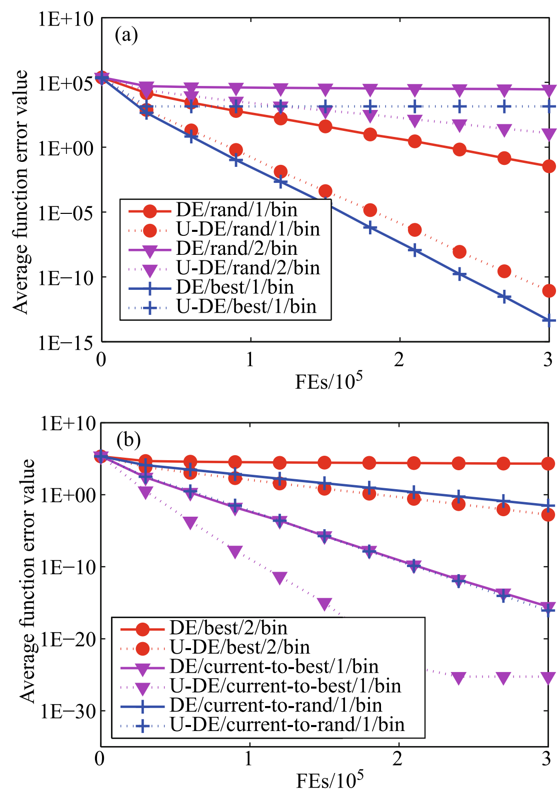


Fig. 1 Convergence graphs of the average function error values derived from six classcial DE versions with and without the restrained condition on F_2 with 30D. (a) Convergence curves of DE/rand/1/bin, U-DE/rand/1/bin, DE/rand/2/bin/, U-DE/rand/2/bin/, DE/best/1/bin, and U-DE/best/1/bin; (b) convergence curves of DE/best/2/bin, U-DE/best/2/bin, DE/current-to-best/1/bin, U-DE/current-to-best/1/bin, DE/current-to-rand/1/bin, and U-DE/current-to-rand/1/bin

4.2 On advanced DE variants

In this subsection, the 28 test functions with 30 dimensions

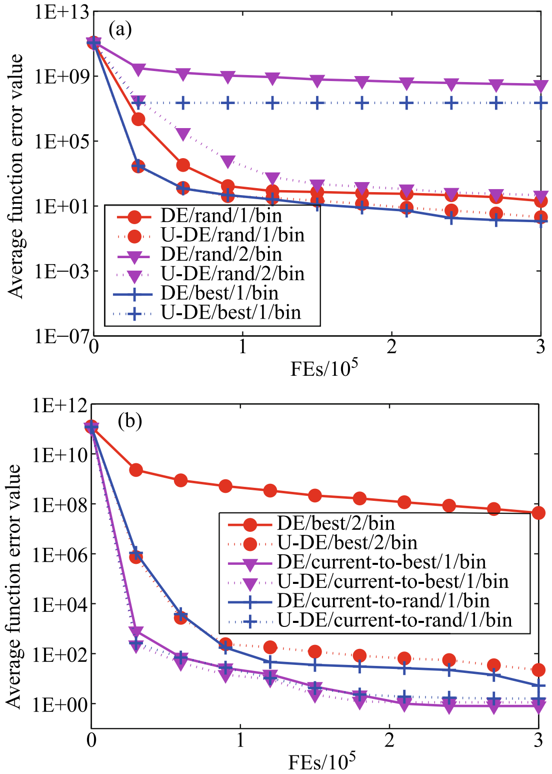


Fig. 2 Convergence graphs of the average function error values derived from six classical DE versions with and without the restrained condition on F_6 with 30D. (a) Convergence curves of DE/rand/1/bin, U-DE/rand/1/bin, DE/rand/2/bin/, U-DE/rand/2/bin/, DE/best/1/bin, and U-DE/best/1/bin; (b) convergence curves of DE/best/2/bin, U-DE/best/2/bin, DE/current-to-best/1/bin, U-DE/current-to-best/1/bin, DE/current-to-rand/1/bin, and U-DE/current-to-rand/1/bin

(30D) and 50 dimensions (50D) designed for IEEE CEC2013 [24] have been further used to investigate the impact of the solution selection for mutation on the performance of seven advanced DE variants, i.e., JADE [4], jDE [10], SaDE [14], EPSDE [15], CoDE [16], LSHADE [25], and JADE/eig [7]. These 28 test functions are denoted as $cf_1 - cf_{28}$ and can be divided into three categories: five unimodal functions ($cf_1 - cf_5$), 15 basic multimodal functions ($cf_6 - cf_{20}$), and eight composition functions ($cf_{21} - cf_{28}$). The mutation operator adopted by JADE, LSHADE, and JADE/eig is a generalized DE/current-to-best/1, called DE/current-to-$p$best/1, jDE uses the classical DE/rand/1 mutation operator, and SaDE, EPSDE, and CoDE establish a candidate pool which consists of several mutation operators. The above seven DE variants without the restrained condition are denoted as U-JADE, U-jDE, U-SaDE, U-EPSDE, U-CoDE, U-LSHADE, and U-JADE/eig, respectively. Actually, in the original CoDE [16], DE/current-to-rand/1 without the restrained condition has been utilized. Therefore, CoDE in [16] is called U-CoDE in this paper, and in this paper CoDE means that the indices of the solutions for mutation are mutually different in all the mutation operators.

For each DE variant, according to the suggestion in [24], 51 independent runs were performed on each test function with 30D and 50D. The function error value ($f(\vec{x}_{best}) - f(\vec{x}^*)$) smaller than 10^{-8} was taken as zero. The experimental results have been summarized in the supplemental file (Appendix: Tables S4–S9) and the statistical test results have been summarized in Tables 3 and 4. It is noteworthy that the parameter settings of JADE, jDE, SaDE, EPSDE, CoDE, LSHADE, and JADE/eig were the same as in the original papers. In order to ensure the comparison fair, the parameter settings of U-JADE, U-jDE, U-SaDE, U-EPSDE, U-CoDE, U-LSHADE, and U-JADE/eig were kept the same as those of JADE, jDE, SaDE, EPSDE, CoDE, LSHADE, and JADE/eig, respectively.

From Table 3, it can be seen that for the 28 test functions with 30D, the seven advanced DE variants are statistically outperformed by their unrestrained competitors respectively. In particular, JADE, SaDE, EPSDE, and LSHADE cannot perform better than their unrestrained competitors even on one test function. jDE is better than U-jDE on only one basic multimodal function, CoDE is better than U-CoDE on two test functions including one basic multimodal function and one composition function, and JADE/eig beats U-JADE/eig on only one basic multimodal function. However, U-jDE, U-CoDE, and U-JADE/eig surpass their restrained versions on 9, 22, and 12 test functions, respectively. It is interesting to note that relaxing the restrained condition fails to improve the performance of DE/current-to-best/1, while it is effective for JADE, LSHADE, and JADE/eig, which use a generalized DE/current-to-best/1.

As shown in Table 4, the overall performance of JADE, jDE, SaDE, EPSDE, CoDE, LSHADE, and JADE/eig can be greatly improved by relaxing the restrained condition for the 28 test functions with 50D. More specifically, U-JADE, U-jDE, U-SaDE, U-EPSDE, U-CoDE, U-LSHADE, and U-JADE/eig exhibit either similar or better performance on all the test functions compared with their restrained versions except that CoDE beats U-CoDE on two composition test functions and JADE/eig performs better than U-JADE/eig on one basic multimodal function.

The above extensive empirical evidences confirm that relaxing the restrained condition could be an effective way to further refine the performance of some advanced DE variants. Figure 3 presents the evolution of the average function error values derived from seven advanced DE variants with and without the restrained condition on cf_3 with 30D.

Table 3 Statistical test results of seven advanced DE variants with and without the restrained condition over 51 independent runs on the 28 test functions with 30D from IEEE CEC2013 using 300,000 FEs

Test functions (30D)		U-JADE vs JADE	U-jDE vs jDE	U-SaDE vs SaDE	U-EPSDE Vs EPSDE	U-CoDE vs CoDE	U-LSHADE vs LSHADE	U-JADE/eig vs JADE/eig
Unimodal functions	cf_1	=	=	=	=	=	=	=
	cf_2	+	=	=	=	+	=	+
	cf_3	+	=	+	+	+	+	+
	cf_4	+	=	=	+	+	=	=
	cf_5	=	=	=	=	=	=	=
Basic multimodal functions	cf_6	+	+	+	+	+	=	+
	cf_7	+	+	+	+	+	+	–
	cf_8	=	=	=	=	+	=	=
	cf_9	=	+	=	=	+	+	=
	cf_{10}	=	=	=	=	–	=	+
	cf_{11}	=	=	=	=	=	=	=
	cf_{12}	=	+	+	=	+	=	+
	cf_{13}	=	+	=	=	+	+	+
	cf_{14}	=	–	=	+	+	+	=
	cf_{15}	=	+	=	=	+	=	+
	cf_{16}	=	=	=	=	+	+	+
	cf_{17}	=	=	=	=	+	=	=
	cf_{18}	=	+	+	=	+	=	+
	cf_{19}	=	=	+	=	+	=	=
	cf_{20}	=	=	=	+	+	=	=
Composition functions	cf_{21}	+	=	=	=	+	=	+
	cf_{22}	=	+	+	+	+	=	=
	cf_{23}	=	=	+	+	+	=	=
	cf_{24}	+	=	=	=	+	=	=
	cf_{25}	=	=	=	=	+	=	=
	cf_{26}	+	=	+	=	–	=	=
	cf_{27}	=	+	+	=	+	=	+
	cf_{28}	=	=	+	+	=	=	+
+		8	9	11	9	22	6	12
–		0	1	0	0	2	0	1
=		20	18	17	19	4	22	15

Note: Wilcoxon's rank sum test at a 0.05 significance level is performed between an advanced DE variant with and without the restrained condition. "+", "–", and "=" denote that the performance of DE without the restrained condition is better than, worse than, and similar to that of DE with the restrained condition, respectively

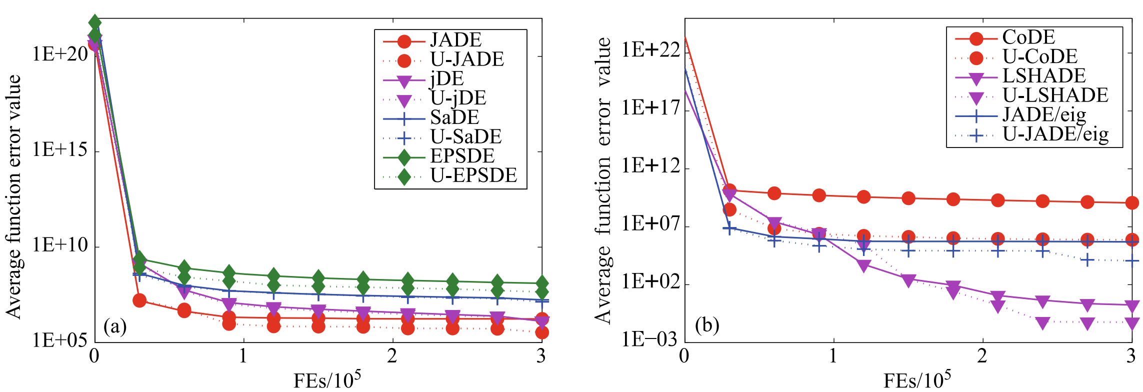


Fig. 3 Convergence graphs of the average function error values derived from seven advanced DE variants with and without the restrained condition on cf_3 with 30D. (a) Convergence curves of JADE, U-JADE, jDE, U-jDE, SaDE, U-SaDE, EPSDE, and U-EPSDE; (b) convergence curves of CoDE, U-CoDE, LSHADE, U-LSHADE, JADE/eig, and U-JADE/eig

Table 4 Statistical test results of seven advanced DE variants with and without the restrained condition over 51 independent runs on the 28 test functions with 50D from IEEE CEC2013 using 500,000 FEs

Test functions (50D)		U-JADE vs JADE	U-jDE vs jDE	U-SaDE vs SaDE	U-EPSDE Vs EPSDE	U-CoDE vs CoDE	U-LSHADE vs LSHADE	U-JADE/eig vs JADE/eig
Unimodal functions	cf_1	=	=	=	=	+	=	=
	cf_2	=	=	=	+	+	=	+
	cf_3	+	+	+	+	+	+	+
	cf_4	+	=	=	+	+	=	=
	cf_5	=	=	=	=	+	=	=
Basic multimodal functions	cf_6	+	=	=	+	+	=	=
	cf_7	=	=	=	+	+	+	−
	cf_8	=	=	=	=	=	=	=
	cf_9	=	+	+	+	+	=	+
	cf_{10}	=	=	=	+	+	=	+
	cf_{11}	=	=	=	+	+	=	=
	cf_{12}	=	+	+	=	+	=	+
	cf_{13}	=	+	+	=	+	=	=
	cf_{14}	=	=	+	+	+	=	+
	cf_{15}	=	+	=	+	+	=	+
	cf_{16}	=	+	=	+	+	+	+
	cf_{17}	=	=	=	=	+	=	=
	cf_{18}	=	+	=	=	+	=	=
	cf_{19}	+	+	=	=	+	=	=
	cf_{20}	=	+	=	=	+	=	=
Composition functions	cf_{21}	+	=	=	=	−	+	=
	cf_{22}	+	+	+	=	+	=	+
	cf_{23}	+	=	=	=	+	=	=
	cf_{24}	=	=	=	=	+	=	=
	cf_{25}	=	+	=	=	+	=	=
	cf_{26}	=	=	+	=	−	=	=
	cf_{27}	+	=	=	+	+	+	=
	cf_{28}	=	+	+	=	+	=	+
+		8	12	8	12	25	5	10
−		0	0	0	0	2	0	1
=		20	16	20	16	1	23	17

Note: Wilcoxon's rank sum test at a 0.05 significance level is performed between an advanced DE variant with and without the restrained condition. "+", "−", and "=" denote that the performance of DE without the restrained condition is better than, worse than, and similar to that of DE with the restrained condition, respectively

4.3 Effect of the control parameter settings

The aim of this subsection is to investigate the effect of the control parameter settings on the performance of DE with and without the restrained condition. We chose DE/rand/1/bin as the instance algorithm and employed the first 14 test functions with 30D from IEEE CEC2005 [23] to produce the experimental results.

Firstly, we tested the effect of the population size on the performance of DE with and without restrained condition. The population size was set to a relatively small value, i.e., 10. Other parameter settings were kept the same as in Section 4.1. Summarized in Table 5 are the experimental results of DE/rand/1/bin and U-DE/rand/1/bin. It is clear from Table 5 that DE/rand/1/bin performs better than U-DE/rand/1/bin on all the test functions, which means that a small population size leads to drastic performance degradation for U-DE/rand/1/bin. However, one should note that a small population size also severely deteriorates the overall performance of DE/rand/1/bin. The poor performance of U-DE/rand/1/bin could be attributed to two aspects: 1) DE/rand/1/bin with a small population size is not capable of maintaining the diversity of the population, and 2) relaxing the restrained condition further reduces the diversity of DE/rand/1/bin. The above comparison implies that the restrained condition cannon be removed from DE with a small population size.

Subsequently, the effects of the parameters F and CR on

Table 5 Experimental results of DE/rand/1/bin and U-DE/rand/1/bin over 25 independent runs on the 14 test functions with 30D from IEEE CEC2005 using 300,000 FEs

Test functions (30D)		DE/rand/1/bin Mean Error±Std Dev	U-DE/rand/1/bin Mean Error±Std Dev	Statistical test
Unimodal functions	F_1	5.65E+00±2.82E+01	1.22E+03±3.13E+03	–
	F_2	1.95E+01±9.76E+01	2.17E+03±6.34E+03	–
	F_3	2.12E+05±1.05E+05	1.37E+07±5.53E+07	–
	F_4	1.29E+03±2.92E+03	1.18E+04±1.01E+04	–
	F_5	2.60E+03±6.98E+02	7.32E+03±3.05E+03	–
Basic multimodal functions	F_6	1.33E+02±2.20E+02	8.74E+08±1.33E+09	–
	F_7	2.64E–02±3.12E–02	8.44E+01±1.67E+02	–
	F_8	2.09E+01±6.05E–02	2.11E+01±7.69E–02	–
	F_9	7.53E+01±3.10E+01	1.39E+02±5.82E+01	–
	F_{10}	8.33E+01±2.36E+01	2.34E+02±8.25E+01	–
	F_{11}	3.05E+01±6.31E+00	3.44E+01±3.75E+00	–
	F_{12}	1.29E+04±1.19E+04	3.16E+04±2.19E+04	–
Expanded multimodal functions	F_{13}	1.04E+01±5.24E+00	3.69E+01±1.86E+01	–
	F_{14}	1.32E+01±3.96E–01	1.37E+01±3.02E–01	–

Note: The population size was set to 10. "Mean Error" and "Std Dev" indicate the average and standard deviation of the function error values obtained in 25 runs, respectively. Wilcoxon's rank sum test at a 0.05 significance level is performed between DE/rand/1/bin and U-DE/rand/1/bin. "–" denotes that the performance of U-DE/rand/1bin is worse than that of DE/rand/1/bin

the performance of DE with and without restrained condition were investigated. In our experiments, we tested three different F: 0.5, 0.7, and 0.9, and five different CR: 0.1, 0.3, 0.5, 0.7, and 0.9. Thus, we obtained 15 different combinations of F and CR. The first 14 test functions with 30D from IEEE CEC2005 [23] were also used to test the performance of DE with and without restrained condition for these 15 combinations. In all experiments, the population size was set to 30. The experimental results are presented in Fig. 4, in which "better", "similar" and "worse" mean that U-DE/rand/1/bin performs better than, similar to, and worse than DE/rand/1/bin, respectively. From Fig. 4, we can give the following remarks:

1) When F is set to a small value (i.e., 0.5), U-DE/rand/1/bin outperforms DE/rand/1/bin in the cases of CR=0.1, 0.3, and 0.5. They both have similar performance when CR=0.7, and U-DE/rand/1/bin is worse than DE/rand/1/bin when CR=0.9.
2) When F is set to a middle value (i.e., 0.7), U-DE/rand/1/bin achieves better overall performance than DE/rand/1/bin in the cases of CR=0.1, 0.3, 0.5, and 0.7, and DE/rand/1/bin beats U-DE/rand/1/bin when CR=0.9.
3) It is interesting to note that when F is set to a big value (i.e., 0.9), U-DE/rand/1/bin performs better than DE/rand/1/bin, regardless of the setting of CR.

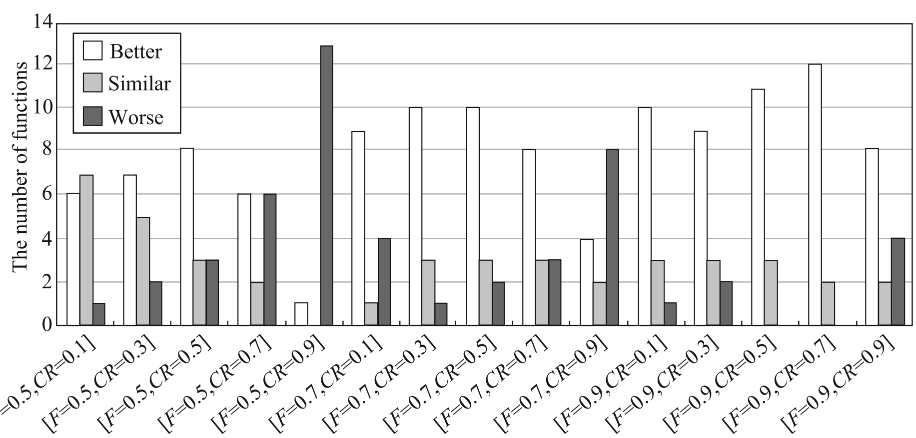


Fig. 4 Experimental results of DE/rand/1/bin and U-DE/rand/1/bin over 25 independent runs on the 14 test functions with 30D from IEEE CEC2005 using 300,000 FEs (In the experiments, 15 combination of F and CR were tested. "Better", "Similar" and "Worse" mean that U-DE/rand/1/bin performs better than, similar to, and worse than DE/rand/1/bin, respectively)

According to the above observation, we can conclude that the restrained condition could be removed from DE in most cases, but cannot be relaxed for DE with a small F and a big CR. It is because in most cases, relaxing the restrained condition is an effective way to improve DE's convergence as discussed in Section 4.1. The reason why the restrained condition cannot be relaxed for DE with a small F and a big CR is explained as follows. DE with a small F can only add a small perturbation to the base vector to generate the mutant vector. On the other hand, if a big CR is utilized in the crossover operator, the trial vector will inherit more information from the mutant vector. Therefore, the exploration ability of DE with a small F and a big CR is limited. As pointed out previously, removing the restrained condition has a side effect on the diversity of DE. Thus, the exploration ability of DE with a small F and a big CR will further degrade if the restrained condition is removed.

4.4 Discussion

Based on the above experimental results, we give some guidelines on the types of DE where the restrained condition should be applied or could be relaxed:

- As shown in Table 2, the overall performance of U-DE/best/1/bin and U-DE/current-to-best/1/bin is worse than that of DE/best/1/bin and DE/current-to-best/1/bin, respectively. In DE/best/1/bin and DE/current-to-best/1/bin, the information of the best solution in the population is exploited explicitly. Therefore, they exhibit greedy characteristics and are suitable for unimodal problems. In this case, removing the restrained condition has a negative influence on the diversity of the population, which inevitably induces the poor performance. On the other hand, as pointed out in Section 4.3, the unrestrained DE does not benefit from a small population size, as well as a small F and meanwhile a big CR. Consequently, we can conclude that the restrained condition should be applied to the greedy DE, DE with a small population size, and DE with a small F and a big CR.
- As shown in Tables 2–4, the restrained condition could be relaxed for four classical DE versions (i.e., DE/rand/1/bin, DE/rand/2/bin, DE/best/2/bin, and DE/current-to-rand/1/bin) and seven advanced DE variants (i.e., JADE, jDE, SaDE, EPSDE, CoDE, LSHADE, and JADE/eig). In terms of DE/rand/1/bin, DE/rand/2/bin, and DE/current-to-rand/1/bin, the target vector learns the information from other randomly chosen solutions. Therefore, they have the capability to maintain the diversity of the population and are suitable for multimodal problems. Under this condition, the removal of the restrained condition accelerates the convergence to a certain degree. Note that DE/best/2/bin also utilizes the information provided by the best solution in the population. However, compared with DE/best/1/bin, two differential vectors are incorporated into DE/best/2/bin, which is able to alleviate the greediness and compensate for the convergence pressure. On the other hand, by adding some extra mechanisms, the seven advanced DE variants have very competitive performance on optimization problems with complex characteristics, such as multimodal problems, non-separable problems, rotated problems, and ill-conditioned problems. For example, in JADE, LSHADE, and JADE/eig, the archiving technique and the self-adaptive parameter adaptation are used to improve the search ability. With respect to SaDE, EPSDE, and CoDE, trail vector generation strategies and control parameter settings with different properties are combined to improve the performance. In addition, jDE tunes the control parameter settings with a self-adaptive manner. For these seven advanced DE variants, the restrained condition could be relaxed to further enhance the performance by accelerating the convergence. According to the above analysis, relaxing the restrained condition does play an important role in two types of DE, i.e., DE with high randomness and DE with very competitive performance.

5 Conclusion

This paper investigates an interesting issue in DE, i.e., the selection of solutions for mutation. In the existing DE, the indices of the chosen solutions for mutation should be mutually different. The above restrained condition has been extensively used during the past twenty years. However, in this paper we verify that this restrained condition could be eliminated for some classical DE versions and some state-of-the-art DE variants by a large number of experiments. Moreover, for some of them, the performance could be remarkably improved by removing this restrained condition. We also identify the types of DE in which the restrained condition should be kept untouched or could be removed. With this paper, we suggest that DE researchers make an attempt to ascertain whether relaxing the restrained condition can improve the performance when designing a DE variant. We guess that relaxing the restrained condition is useful for large-scale optimization [26], since under this condition a very large population size is usually adopted due to the high-dimensional search space. In the future, we will carry out an in-depth theoretical analysis on the insights derived from experimentation.

Acknowledgements The authors would like to thank the anonymous reviewers for their very constructive and helpful suggestions. This work was supported in part by the National Basic Research Program (973 Program) of China (2011CB013104), in part by the Innovation-driven Plan in Central South University (2015CXS012 and 2015CX007), in part by the National Natural Science Foundation of China (Grant Nos. 61273314 and 61673397), in part by the EU Horizon 2020 Marie Skłodowska-Curie Individual Fellowships (Project ID: 661327), in part by the Hunan Provincial Natural Science Fund for Distinguished Young Scholars (2016JJ1018), in part by the Program for New Century Excellent Talents in University (NCET-13-0596), and in part by State Key Laboratory of Intelligent Control and Decision of Complex Systems, Beijing Institute of Technology.

Appendix

Table S1 Experimental results of DE/rand/1/bin, U-DE/rand/1/bin, DE/rand/2/bin, and U-DE/rand/2/bin over 25 independent runs on the 14 test functions with 30D from IEEE CEC2005 using 300,000 FEs

Test functions (30D)		DE/rand/1/bin	U-DE/rand/1/bin	DE/rand/2/bin	U-DE/rand/2/bin
		Mean Error±Std Dev	Mean Error±Std Dev	Mean Error±Std Dev	Mean Error±Std Dev
Unimodal functions	F_1	6.54E–18±1.19E–17	1.84E–27±1.37E–27+	5.99E+03±9.25E+02	5.46E–10±7.06E–10+
	F_2	4.97E–02±4.99E–02	7.33E–12±9.99E–12+	2.91E+04±3.69E+03	1.17E+01±8.80E+00+
	F_3	6.41E+05±3.78E+05	1.30E+05±5.78E+04+	1.85E+08±4.87E+07	1.55E+06±7.54E+00+
	F_4	1.37E+01±1.56E+01	7.10E–02±1.50E–01+	3.75E+04±5.48E+03	2.68E+02±1.51E+02+
	F_5	1.30E+02±1.55E+02	5.07E+02±3.22E+02–	1.20E+04±9.40E+02	6.84E+02±4.77E+02+
Basic multimodal functions	F_6	2.51E+01±2.62E+01	2.12E+00±2.20E+00+	3.04E+08±6.29E+07	6.25E+01±5.66E+01+
	F_7	5.51E–03±8.35E–03	1.57E–02±1.25E–02–	4.43E+03±7.45E+02	9.34E–03±9.24E–03+
	F_8	2.09E+01±6.91E–02	2.09E+01±6.25E–02=	2.09E+01±3.93E–02	2.06E+01±3.15E–01+
	F_9	2.18E+01±7.73E+00	4.28E+01±1.19E+01–	2.40E+02±1.41E+01	2.62E+01±8.11E+00+
	F_{10}	1.32E+02±8.66E+01	4.94E+01±1.35E+01+	2.88E+02±1.60E+01	3.83E+01±1.44E+01+
	F_{11}	3.76E+01±5.37E+00	1.79E+01±6.80E+00+	3.97E+01±1.23E+00	1.54E+01±7.18E+00+
	F_{12}	4.59E+03±5.07E+03	5.11E+03±4.38E+03=	7.39E+05±8.16E+04	4.51E+03±5.48E+03+
Expanded multimodal functions	F_{13}	3.04E+00±8.86E–01	4.35E+00±2.10E+00–	5.20E+01±7.19E+00	3.08E+00±8.67E–01+
	F_{14}	1.34E+01±1.43E–01	1.32E+01±3.58E–01+	1.34E+01±1.88E–01	1.30E+01±3.80E–01+
		+	8	+	14
		–	4	–	0
		=	2	=	0

Note: "Mean Error" and "Std Dev" indicate the average and standard deviation of the function error values obtained in 25 runs, respectively. Wilcoxon's rank sum test at a 0.05 significance level is performed between DE/rand/1/bin and U-DE/rand/1/bin, and between DE/rand/2/bin and U-DE/rand/2/bin. "+", "–", and "=" denote that the performance of DE without the restrained condition is better than, worse than, and similar to that of DE with the restrained condition, respectively

Table S2 Experimental results of DE/best/1/bin, U-DE/best/1/bin, DE/best/2/bin, and U-DE/best/2/bin over 25 independent runs on the 14 test functions with 30D from IEEE CEC2005 using 300,000 FEs

Test functions (30D)		DE/best/1/bin	U-DE/best/1/bin	DE/best/2/bin	U-DE/best/2/bin
		Mean Error±Std Dev	Mean Error±Std Dev	Mean Error±Std Dev	Mean Error±Std Dev
Unimodal functions	F_1	2.11E–27±9.05E–28	2.89E+01±6.59E+01–	1.62E+03±7.05E+02	6.48E–02±3.24E–01+
	F_2	4.94E–14±1.15E–13	9.24E+02±2.51E+03–	2.02E+04±5.54E+03	2.16E–03±2.85E–03+
	F_3	1.30E+05±7.26E+04	6.10E+06±1.09E+07–	1.29E+08±3.53E+07	5.54E+05±3.80E+05+
	F_4	3.25E+00±1.27E+01	3.08E+03±6.19E+03–	3.00E+04±5.56E+03	3.93E+01±5.23E+01+
	F_5	2.38E+02±2.00E+02	4.27E+03±1.36E+03–	9.19E+03±9.80E+02	4.97E+02±3.60E+02+
Basic multimodal functions	F_6	2.03E+00±2.24E+00	4.02E+07±1.32E+08–	5.64E+07±3.62E+07	1.22E+01±1.21E+01+
	F_7	2.02E–02±2.20E–02	9.14E+01±1.61E+02–	1.58E+03±5.84E+02	2.45E–02±2.85E–02+
	F_8	2.09E+01±5.99E–02	2.10E+01±9.62E–02–	2.09E+01±4.12E–02	2.01E+01±2.12E–01+
	F_9	6.35E+01±2.60E+01	1.54E+02±4.04E+01–	2.37E+02±1.53E+01	6.69E+01±2.06E+01+
	F_{10}	8.99E+01±3.23E+01	1.86E+02±4.76E+01–	2.80E+02±1.98E+01	7.39E+01±1.85E+01+
	F_{11}	2.05E+01±5.01E+00	3.37E+01±3.40E+00–	3.96E+01±.9.58E–01	2.36E+01±7.88E+01+
	F_{12}	1.15E+04±1.17E+04	6.46E+04±6.34E+04–	6.82E+05±7.45E+04	4.80E+03±8.92E+03+

(*Continued*)

Test functions (30D)		DE/best/1/bin Mean Error±Std Dev	U-DE/best/1/bin Mean Error±Std Dev	DE/best/2/bin Mean Error±Std Dev	U-DE/best/2/bin Mean Error±Std Dev
Expanded multimodal	F_{13}	5.99E+00±2.60E+00	2.21E+01±7.20E+00 –	3.32E+01±7.79E+00	7.33E+00±2.50E+00+
functions	F_{14}	1.29E+01±4.69E–01	1.36E+01±1.85E–01–	1.34E+01±1.62E–01	1.30E+01±2.98E–01+
		+	0	+	14
		–	14	–	0
		=	0	=	0

Note: "Mean Error" and "Std Dev" indicate the average and standard deviation of the function error values obtained in 25 runs, respectively. Wilcoxon's rank sum test at a 0.05 significance level is performed between DE/best/1/bin and U-DE/best/1/bin, and between DE/best/2/bin and U-DE/best/2/bin. "+", "–", and "=" denote that the performance of DE without the restrained condition is better than, worse than, and similar to that of DE with the restrained condition, respectively

Table S3 Experimental results of DE/current-to-best/1/bin, U-DE/current-to-best/1/bin, DE/current-to-rand/1/bin, and U-DE/current-to-rand/1/bin over 25 independent runs on the 14 test functions with 30D from IEEE CEC2005 using 300,000 FEs

Test functions (30D)		DE/current-to-best/1/bin Mean Error±Std Dev	U-DE/current-to-best/1/bin Mean Error±Std Dev	DE/current-to-rand/1/bin Mean Error±Std Dev	U-DE/current-to-rand/1/bin Mean Error±Std Dev
	F_1	1.53E–27±1.09E–27	2.53E–26±4.51E–26–	1.79E–17±3.22E–17	2.10E–29±6.35E–29+
	F_2	2.63E–16±7.21E–16	5.48E–25±2.26E–24+	2.56E–02±2.16E–02	9.61E–17±4.00E–16+
Unimodal functions	F_3	1.24E+05±5.81E+04	3.10E+04±1.62E+04+	7.04E+05±3.76E+05	1.00E+05±6.46E+04+
	F_4	1.11E–01±2.53E–01	2.55E+00±1.22E+01–	1.95E+01±1.79E+01	2.58E–03±3.72E–03+
	F_5	2.30E+02±2.07E+02	1.43E+03±4.89E+02–	1.82E+02±9.87E+01	9.03E+01±1.52E+02+
	F_6	1.27E+00±1.89E+00	1.59E+00±1.99E+00–	4.74E+00±3.11E+00	9.56E–01±1.73E+00+
	F_7	1.69E–02±1.24E–02	2.28E–02±2.63E–02=	1.65E–03±3.74E–03	1.35E–02±7.21E–03–
	F_8	2.09E+01±5.22E–02	2.09E+01±4.83E–02=	2.09E+01±3.28E–02	2.09E+01±5.36E–02=
Basic multimodal	F_9	3.77E+01±9.71E+00	9.60E+01±2.74E+01–	1.90E+02±1.10E+01	2.88E+01±8.17E+00+
functions	F_{10}	5.49E+01±1.86E+01	1.32E+02±5.77E+01–	2.14E+02±1.12E+01	3.92E+01±1.27E+01+
	F_{11}	1.60E+01±5.77E+00	2.89E+01±3.29E+00–	3.95E+01±7.05E–01	2.01E+01±5.79E+00+
	F_{12}	1.29E+04±1.93E+04	1.49E+04±1.76E+04=	3.13E+03±3.54E+03	3.75E+03±4.87E+03=
Expanded multimodal	F_{13}	3.63E+00±1.02E+00	9.82E+00±2.99E+00 –	1.80E+01±1.11E+00	3.48E+00±7.88E–01+
functions	F_{14}	1.31E+01±2.33E–01	1.21E+01±3.90E–01+	1.34E+01±1.63E–01	1.30E+01±2.44E–01+
		+	3	+	11
		–	8	–	1
		=	3	=	2

Note: "Mean Error" and "Std Dev" indicate the average and standard deviation of the function error values obtained in 25 runs, respectively. Wilcoxon's rank sum test at a 0.05 significance level is performed between DE/current-to-best/1/bin and U-DE/current-to-best/1/bin, and between DE/current-to-rand/1/bin and U-DE/current-to-rand/1/bin. "+", "–", and "=" denote that the performance of DE without the restrained condition is better than, worse than, and similar to that of DE with the restrained condition, respectively

Table S4 Experimental results of JADE, U-JADE, jDE, and U-jDE over 51 independent runs on the 28 test functions with 30D from IEEE CEC2013 using 300,000 FEs

Test functions (30D)		JADE Mean Error±Std Dev	U-JADE Mean Error±Std Dev	jDE Mean Error±Std Dev	U-jDE Mean Error±Std Dev
	cf_1	0.00E+00±0.00E+00	0.00E+00±0.00E+00=	0.00E+00±0.00E+00	0.00E+00±0.00E+00=
	cf_2	7.80E+03±6.74E+03	6.88E+03±5.13E+03+	1.39E+05±8.26E+04	1.37E+05±9.15E+04=
Unimodal	cf_3	1.74E+06±5.71E+06	3.42E+05±1.46E+06+	1.25E+06±1.69E+06	1.35E+06±1.87E+06=
functions	cf_4	5.35E+03±1.36E+04	4.72E+03±1.12E+04+	5.22E+00±5.57E+00	5.25E+00±5.63E+00=
	cf_5	0.00E+00±0.00E+00	0.00E+00±0.00E+00=	0.00E+00±0.00E+00	0.00E+00±0.00E+00=
	cf_6	3.62E+00±9.17E+00	1.03E+00±5.17E+00+	1.33E+01±4.45E+00	1.27E+01±3.85E+00+
	cf_7	5.26E+00±8.25E+00	4.11E+00±5.01E+00+	3.56E+00±3.07E+00	3.35E+00±2.91E+00+
	cf_8	2.09E+01±4.33E–02	2.09E+01±1.31E–01=	2.09E+01±4.65E–02	2.09E+01±4.78E–02=
	cf_9	2.61E+01±1.83E+00	2.63E+01±1.42E+00=	2.82E+01±1.34E+00	2.51E+01±5.67E+00+
	cf_{10}	3.67E–02±2.45E–02	3.72E–02±2.53E–02=	4.01E–02±2.63E–02	3.66E–02±2.89E–02=
	cf_{11}	0.00E+00±0.00E+00	0.00E+00±0.00E+00=	0.00E+00±0.00E+00	0.00E+00±0.00E+00=
Basic	cf_{12}	2.38E+01±4.16E+00	2.37E+01±4.32E+00=	5.97E+01±1.01E+01	5.34E+01±8.96E+00+
multimodal	cf_{13}	4.43E+01±1.27E+01	4.50E+01±1.38E+01=	9.08E+01±1.67E+01	8.06E+01±1.78E+01+

(*Continued*)

Test functions (30D)		JADE	U-JADE	jDE	U-jDE
		Mean Error±Std Dev	Mean Error±Std Dev	Mean Error±Std Dev	Mean Error±Std Dev
functions	cf_{14}	3.31E–02±2.39E–02	3.32E–02±2.88E–02=	2.44E–03±6.78E–03	1.51E–02±1.67E–02–
	cf_{15}	3.31E+03±3.01E+02	3.28E+03±3.43E+02=	5.11E+03±3.59E+02	4.85E+03±3.65E+02+
	cf_{16}	1.70E+00±6.69E–01	1.76E+00±7.43E–01=	2.37E+00±2.77E–01	2.36E+00±3.32E–01=
	cf_{17}	3.04E+01±2.74E–14	3.04E+01±3.45E–14=	3.04E+01±3.82E–14	3.04E+01±4.08E–14=
	cf_{18}	7.69E+01±6.06E+00	7.67E+01±6.01E+00=	1.59E+02±1.56E+01	1.49E+02±1.63E+01+
	cf_{19}	1.45E+00±1.20E–01	1.47E+00±1.01E–01=	1.61E+00±1.55E–01	1.64E+00±1.41E–01=
	cf_{20}	1.05E+01±5.07E–01	1.06E+01±5.35E–01=	1.16E+01±3.63E–01	1.16E+01±3.50E–01=
Composition functions	cf_{21}	3.09E+02±7.22E+01	2.96E+02±6.02E+01+	2.76E+02±7.29E+01	2.76E+02±7.31E+01=
	cf_{22}	9.14E+01±3.49E+01	9.18E+01±2.80E+01=	1.31E+02±2.36E+01	1.09E+02±2.27E+01+
	cf_{23}	3.47E+03±4.74E+02	3.48E+03±4.57E+02=	5.49E+03±5.05E+02	5.48E+03±5.85E+02=
	cf_{24}	2.13E+02±1.23E+01	2.09E+02±7.13E+00+	2.13E+02±1.11E+01	2.14E+02±1.04E+01=
	cf_{25}	2.74E+02±1.17E+01	2.73E+02±9.89E+00=	2.49E+02±8.68E+00	2.49E+02±6.25E+00=
	cf_{26}	2.19E+02±4.92E+01	2.12E+02±3.78E+01+	2.05E+02±2.65E+01	2.02E+02±1.57E+01=
	cf_{27}	6.76E+02±2.33E+02	6.85E+02±2.33E+02=	6.86E+02±1.90E+02	6.23E+02±1.52E+02+
	cf_{28}	3.19E+02±1.39E+02	3.19E+02±1.39E+02=	3.00E+02±0.00E+00	3.00E+02±0.00E+00=
		+	8	+	9
		–	0	–	1
		=	20	=	18

Note: "Mean Error" and "Std Dev" indicate the average and standard deviation of the function error values obtained in 51 runs, respectively. Wilcoxon's rank sum test at a 0.05 significance level is performed between JADE and U-JADE, and between jDE and U-jDE. "+", "–", and "=" denote that the performance of DE without the restrained condition is better than, worse than, and similar to that of DE with the restrained condition, respectively

Table S5 Experimental results of SaDE, U-SaDE, EPSDE, and U-EPSDE over 51 independent runs on the 28 test functions with 30D from IEEE CEC2013 using 300,000 FEs

Test functions (30D)		SaDE	U-SaDE	EPSDE	U-EPSDE
		Mean Error±Std Dev	Mean Error±Std Dev	Mean Error±Std Dev	Mean Error±Std Dev
Unimodal functions	cf_1	0.00E+00±0.00E+00	0.00E+00±0.00E+00=	0.00E+00±0.00E+00	0.00E+00±0.00E+00=
	cf_2	4.03E+05±1.91E+05	4.26E+05±2.42E+05=	8.16E+05±5.00E+06	8.64E+05±3.50E+06=
	cf_3	1.72E+07±3.10E+07	1.38E+07±1.84E+07+	1.52E+08±4.11E+08	5.73E+07±3.25E+08+
	cf_4	3.28E+03±1.67E+03	3.43E+03±1.84E+03=	8.47E+03±2.78E+04	3.30E+03±9.79E+03+
	cf_5	0.00E+00±0.00E+00	0.00E+00±0.00E+00=	0.00E+00±0.00E+00	0.00E+00±0.00E+00=
Basic multimodal functions	cf_6	3.09E+01±2.92E+01	2.93E+01±2.80E+01+	9.31E+00±1.02E+00	9.07E+00±2.23E+00+
	cf_7	2.90E+01±1.41E+01	2.52E+01±1.12E+01+	6.56E+01±4.88E+01	6.01E+01±3.84E+01+
	cf_8	2.09E+01±5.08E–02	2.09E+01±5.52E–02=	2.09E+01±4.52E–02	2.09E+01±4.28E–02=
	cf_9	1.78E+01±2.09E+00	1.75E+01±2.91E+00=	3.36E+01±3.59E+00	3.38E+01±3.63E+00=
	cf_{10}	2.69E–01±1.51E–01	2.72E–01±1.34E–01=	9.78E–02±6.92E–02	9.95E–02±5.80E–02=
	cf_{11}	1.56E–01±4.61E–01	1.85E–01±5.47E–01=	0.00E+00±0.00E+00	0.00E+00±0.00E+00=
	cf_{12}	4.81E+01±1.14E+01	4.57E+01±9.52E+00+	4.86E+01±9.87E+00	4.89E+01±1.65E+01=
	cf_{13}	9.87E+01±2.45E+01	9.92E+01±1.92E+01=	7.94E+01±1.75E+01	7.90E+01±2.06E+01=
	cf_{14}	7.76E–01±1.04E+00	7.82E–01±1.08E+00=	3.47E–01±4.89E–01	2.88E–01±3.52E–01+
	cf_{15}	4.74E+03±1.02E+03	4.75E+03±1.03E+03=	6.65E+03±8.15E+02	6.57E+03±7.55E+02=
	cf_{16}	2.24E+00±2.62E–01	2.22E+00±2.63E–01=	2.49E+00±2.59E–01	2.44E+00±2.94E–01=
	cf_{17}	3.04E+01±4.62E–02	3.04E+01±4.34E–02=	3.04E+01±4.86E–02	3.04E+01±2.22E–03=
	cf_{18}	1.30E+02±4.36E+01	1.17E+02±4.37E+01+	1.36E+02±1.74E+01	1.34E+02±1.35E+01=
	cf_{19}	4.10E+00±8.18E–01	3.83E+00±9.04E–01+	1.86E+00±2.40E–01	1.85E+00±2.15E–01=
	cf_{20}	1.08E+01±6.55E–01	1.06E+01±7.11E–01=	1.32E+01±6.00E–01	1.30E+01±6.75E–01+
Composition functions	cf_{21}	3.14E+02±6.22E+01	3.20E+02±6.58E+01=	2.90E+02±7.61E+01	2.95E+02±8.23E+01=
	cf_{22}	1.26E+02±4.47E+01	1.19E+02±3.44E+01+	3.33E+02±1.58E+02	3.00E+02±1.22E+02+
	cf_{23}	4.68E+03±1.10E+03	4.36E+03±1.11E+03+	7.07E+03±7.79E+02	6.86E+03±8.09E+02+
	cf_{24}	2.26E+02±6.69E+00	2.25E+02±5.70E+00=	2.90E+02±6.65E+00	2.88E+02±7.83E+00=
	cf_{25}	2.64E+02±1.22E+01	2.65E+02±1.14E+01=	2.98E+02±2.94E+00	2.98E+02±2.85E+00=
	cf_{26}	2.10E+02±3.53E+01	2.05E+02±2.51E+01+	3.59E+02±6.55E+01	3.61E+02±4.89E+01=

(*Continued*)

Test functions (30D)		SaDE	U-SaDE	EPSDE	U-EPSDE
		Mean Error±Std Dev	Mean Error±Std Dev	Mean Error±Std Dev	Mean Error±Std Dev
	cf_{27}	5.94E+02±6.53E+01	5.86E+02±7.75E+01+	1.21E+03±7.03E+01	1.20E+03±7.91E+01=
	cf_{28}	3.00E+02±0.00E+00	2.96E+02±2.85E+01+	3.20E+02±1.43E+02	3.00E+02±0.00E+00+
		+	11	+	9
		−	0	−	0
		=	17	=	19

Note: "Mean Error" and "Std Dev" indicate the average and standard deviation of the function error values obtained in 51 runs, respectively. Wilcoxon's rank sum test at a 0.05 significance level is performed between SaDE and U-SaDE, and between EPSDE and U-EPSDE. "+", "−", and "=" denote that the performance of DE without the restrained condition is better than, worse than, and similar to that of DE with the restrained condition, respectively

Table S6 Experimental results of CoDE, U-CoDE, LSHADE, U-LSHADE, JADE/eig, and U-JADE/eig over 51 independent runs on the 28 test functions with 30D from IEEE CEC2013 using 300,000 FEs

Test functions (30D)		CoDE	U-CoDE	LSHADE	U-LSHADE	JADE/eig	U-JADE/eig
		Mean Error±Std Dev	Mean Error±Std Dev	Mean Error±Std Dev	Mean Error±Std Dev	Mean Error±Std Dev	Mean Error±Std Dev
Unimodal functions	cf_1	0.00E+00±0.00E+00	0.00E+00±0.00E+00=	0.00E+00±0.00E+00	0.00E+00±0.00E+00=	0.00E+00±0.00E+00	0.00E+00±0.00E+00=
	cf_2	1.86E+05±9.59E+04	8.28E+04±4.98E+04+	0.00E+00±0.00E+00	0.00E+00±0.00E+00=	8.53E+03±7.30E+03	7.31E+03±5.50E+03+
	cf_3	1.13E+09±6.65E+08	9.43E+05±2.39E+06+	1.78E+00±1.05E+01	5.81E–02±4.01E–01+	4.97E+05±2.46E+06	1.16E+04±5.30E+04+
	cf_4	6.21E–01±7.05E–01	1.07E–01±2.29E–01+	0.00E+00±0.00E+00	0.00E+00±0.00E+00=	4.13E–07±7.45E–07	5.79E–07±6.28E–07=
	cf_5	0.00E+00±0.00E+00	0.00E+00±0.00E+00=	0.00E+00±0.00E+00	0.00E+00±0.00E+00=	0.00E+00±0.00E+00	0.00E+00±0.00E+00=
Basic multimodal functions	cf_6	7.21E+00±7.14E+00	4.11E+00±8.99E+00+	0.00E+00±0.00E+00	0.00E+00±0.00E+00=	3.10E+00±8.59E+00	1.11E+00±5.19E+00+
	cf_7	6.84E+02±2.09E+02	7.33E+01±4.00E+01+	6.45E–01±4.46E–01	5.60E–01±4.99E–01+	2.35E+00±2.46E+00	5.02E+00±7.65E+00–
	cf_8	2.09E+01±6.18E–02	2.08E+01±9.72E–02+	2.08E+01±1.12E–01	2.07E+01±1.87E–01=	2.09E+01±5.69E–02	2.09E+01±5.69E–02=
	cf_9	3.22E+01±1.42E+00	1.38E+01±3.26E+00+	2.63E+01±1.28E+00	2.58E+01±1.78E+00+	2.57E+01±1.87E+00	2.60E+01±1.77E+00=
	cf_{10}	9.05E–03±2.32E–02	3.74E–02±2.52E–02–	2.90E–04±1.44E–03	2.90E–04±1.44E–03=	3.11E–02±1.90E–02	2.56E–02±1.72E–02+
	cf_{11}	0.00E+00±0.00E+00	0.00E+00±0.00E+00=	0.00E+00±0.00E+00	0.00E+00±0.00E+00=	0.00E+00±0.00E+00	0.00E+00±0.00E+00=
	cf_{12}	1.89E+02±1.61E+01	3.68E+01±9.00E+00+	5.44E+01±1.32E+01	5.46E+01±1.35E+01=	2.51E+01±4.64E+00	2.49E+01±4.25E+00+
	cf_{13}	2.04E+02±1.50E+01	7.58E+01±2.40E+01+	5.95E+00±2.78E+00	5.42E+00±2.09E+00+	5.42E+01±1.32E+01	5.25E+01±1.33E+01+
	cf_{14}	1.19E+02±2.52E+01	3.28E+00±3.53E+00+	3.11E–02±2.54E–02	2.44E–02±2.19E–02+	2.45E+01±5.92E+00	2.50E+01±7.13E+00=
	cf_{15}	6.65E+03±3.57E+02	3.53E+03±5.76E+02+	2.69E+03±3.45E+02	2.70E+03±3.49E+02=	3.27E+03±3.56E+02	3.16E+03±3.54E+02+
	cf_{15}	2.46E+00±2.51E–01	3.40E–01±2.36E–01+	7.85E–01±1.75E–01	5.25E–01±3.09E–01+	1.81E+00±6.80E–01	1.53E+00±8.91E–01+
	cf_{17}	3.09E+01±1.86E–01	3.04E+01±3.29E–02+	3.04E+01±7.24E–12	3.04E+01±1.60E–11=	3.06E+01±9.09E–02	3.06E+01±7.13E–02=
	cf_{18}	2.49E+02±1.21E+01	6.42E+01±1.27E+01+	5.16E+01±3.21E+00	5.18E+01±2.64E+00=	7.75E+01±6.74E+00	7.53E+01±5.65E+00+
	cf_{19}	6.14E+00±5.49E–01	1.56E+00±2.82E–01+	1.17E+00±9.37E–01	1.18E+00±9.90E–01=	1.72E+00±1.36E–01	1.69E+00±1.54E–01=
	cf_{20}	1.25E+01±2.46E–01	1.06E+01±6.65E–01+	1.02E+01±1.45E+00	1.03E+01±1.49E+00=	1.04E+01±4.75E–01	1.03E+01±4.28E–01=
Composition functions	cf_{21}	3.17E+02±1.09E+02	3.03E+02±9.35E+01+	2.93E+02±3.69E+01	2.93E+02±3.44E+01=	3.14E+02±7.28E+01	2.91E+02±7.04E+01+
	cf_{22}	1.01E+03±2.18E+02	1.09E+02±2.72E+01+	1.08E+02±2.37E+00	1.08E+02±2.51E+00=	1.44E+02±2.10E+01	1.46E+02±3.25E+01=
	cf_{23}	6.93E+03±3.24E+02	3.54E+03±6.49E+02+	2.49E+03±2.90E+02	2.50E+03±3.37E+02=	3.24E+03±3.90E+02	3.29E+03±3.78E+02=
	cf_{24}	2.77E+02±5.19E+00	2.21E+02±8.24E+00+	2.00E+02±8.01E–01	2.00E+02±7.13E–01=	2.09E+02±1.18E+01	2.09E+02±1.23E+01=
	cf_{25}	3.01E+02±3.77E+00	2.84E+02±1.35E+01+	2.41E+02±4.45E+00	2.42E+02±6.80E+00=	2.63E+02±1.54E+01	2.61E+02±1.62E+01=
	cf_{26}	2.00E+02±6.39E–03	2.15E+02±4.19E+01–	2.00E+02±2.72E–14	2.00E+02±2.63E–14=	2.09E+02±3.22E+01	2.10E+02±3.45E+01=
	cf_{27}	1.10E+03±3.74E+01	6.05E+02±9.58E+01+	3.02E+02±5.64E+00	3.02E+02±6.35E+00=	5.52E+02±1.97E+02	5.09E+02±2.15E+02+
	cf_{28}	3.00E+02±4.66E–10	3.00E+02±0.00E–10=	3.00E+02±0.00E+00	3.00E+02±0.00E+00=	3.21E+02±1.51E+02	3.00E+02±0.00E+00+
		+	22	+	6	+	12
		–	2	–	0	–	1
		=	4	=	22	=	15

Note: "Mean Error" and "Std Dev" indicate the average and standard deviation of the function error values obtained in 51 runs, respectively. Wilcoxon's rank sum test at a 0.05 significance level is performed between CoDE and U-CoDE, between LSHADE and U-LSHADE, and between JADE/eig and U-JADE/eig. "+", "−", and "=" denote that the performance of DE without the restrained condition is better than, worse than, and similar to that of DE with the restrained condition, respectively

Table S7 Experimental results of JADE, U-JADE, jDE, and U-jDE over 51 independent runs on the 28 test functions with 50D from IEEE CEC2013 using 500,000 FEs

Test functions (50D)		JADE	U-JADE	jDE	U-jDE
		Mean Error±Std Dev	Mean Error±Std Dev	Mean Error±Std Dev	Mean Error±Std Dev
Unimodal functions	cf_1	0.00E+00±0.00E+00	0.00E+00±0.00E+00=	0.00E+00±0.00E+00	0.00E+00±0.00E+00=
	cf_2	2.35E+04±1.11E+04	2.42E+04±1.47E+04=	5.24E+05±2.44E+05	5.45E+05±2.11E+05=
	cf_3	4.87E+06±9.77E+06	3.83E+06±9.61E+06+	7.05E+06±1.58E+07	5.13E+06±9.38E+06+
	cf_4	9.78E+03±2.02E+04	5.36E+03±1.66E+04+	1.20E+01±9.97E+00	1.22E+01±1.60E+01=
	cf_5	0.00E+00±0.00E+00	0.00E+00±0.00E+00=	0.00E+00±0.00E+00	0.00E+00±0.00E+00=

(*Continued*)

Test functions (50D)		JADE	U-JADE	jDE	U-jDE
		Mean Error±Std Dev	Mean Error±Std Dev	Mean Error±Std Dev	Mean Error±Std Dev
Basic multimodal functions	cf_6	4.37E+01±1.11E+00	4.28E+01±5.66E+00+	4.38E+01±4.43E–01	4.39E+01±7.65E–01=
	cf_7	2.29E+01±1.04E+01	2.32E+01±1.34E+01=	1.81E+01±6.92E+00	1.92E+01±7.07E+00=
	cf_8	2.11E+01±1.03E–01	2.11E+01±8.98E–02=	2.11E+01±4.65E–02	2.12E+01±3.60E–02=
	cf_9	5.46E+01±2.01E+00	5.41E+01±2.36E+00=	5.50E+01±2.61E+00	5.38E+01±4.86E+00+
	cf_{10}	3.09E–02±2.09E–02	3.11E–02±2.24E–02=	5.30E–02±3.55E–02	5.35E–02±4.06E–02=
	cf_{11}	0.00E+00±0.00E+00	0.00E+00±0.00E+00=	0.00E+00±0.00E+00	0.00E+00±0.00E+00=
	cf_{12}	5.68E+01±9.48E+00	5.61E+01±9.39E+00=	1.06E+02±1.64E+01	9.25E+01±2.17E+01+
	cf_{13}	1.28E+02±2.52E+01	1.27E+02±2.33E+01=	1.81E+02±2.80E+01	1.70E+02±2.49E+01+
	cf_{14}	4.28E–02±2.61E–02	4.25E–02±2.70E–02=	6.33E–03±1.69E–02	4.64E–03±1.03E–02=
	cf_{15}	6.97E+03±4.67E+02	6.97E+03±4.72E+02=	9.87E+03±4.40E+02	9.59E+03±6.84E+02+
	cf_{16}	2.00E+00±7.86E–01	2.01E+00±7.91E–01=	3.01E+00±3.58E–01	2.97E+00±4.40E–01+
	cf_{17}	5.08E+01±3.52E–14	5.08E+01±3.61E–14=	5.08E+01±6.97E–14	5.08E+01±7.89E–14=
	cf_{18}	1.40E+02±1.11E+01	1.41E+02±1.06E+01=	2.79E+02±2.54E+01	2.59E+02±1.91E+01+
	cf_{19}	2.75E+00±1.82E–01	2.69E+00±2.01E–01+	2.91E+00±2.06E–01	2.88E+00±2.10E–01+
	cf_{20}	1.96E+01±6.07E–01	1.97E+01±5.72E–01=	2.14E+01±4.47E–01	2.11E+01±5.10E–01+
Composition functions	cf_{21}	8.06E+02±4.07E+02	7.69E+02±4.19E+02+	5.79E+02±4.58E+02	5.83E+02±4.65E+02=
	cf_{22}	2.43E+01±4.63E+01	1.28E+01±5.55E+00+	1.03E+02±5.13E+01	2.05E+01±1.23E+01+
	cf_{23}	7.32E+03±8.42E+02	7.28E+03±5.69E+02+	1.08E+04±7.47E+02	1.06E+04±6.76E+02=
	cf_{24}	2.49E+02±2.10E+01	2.47E+02±2.04E+01=	2.55E+02±1.49E+01	2.53E+02±1.20E+01=
	cf_{25}	3.52E+02±2.62E+01	3.52E+02±1.86E+01=	3.08E+02±2.11E+01	3.01E+02±1.04E+01+
	cf_{26}	3.40E+02±1.05E+02	3.42E+02±1.01E+02=	2.32E+02±6.69E+01	2.34E+02±7.05E+01=
	cf_{27}	1.40E+03±3.22E+02	1.32E+03±3.01E+02+	1.09E+03±2.23E+02	1.03E+03±1.55E+02=
	cf_{28}	5.73E+02±7.00E+02	5.73E+02±6.98E+02=	4.57E+02±4.12E+02	4.00E+02±2.41E–14+
		+	8	+	12
		–	0	–	0
		=	20	=	16

Note: "Mean Error" and "Std Dev" indicate the average and standard deviation of the function error values obtained in 51 runs, respectively. Wilcoxon's rank sum test at a 0.05 significance level is performed between JADE and U-JADE, and between jDE and U-jDE. "+", "–", and "=" denote that the performance of DE without the restrained condition is better than, worse than, and similar to that of DE with the restrained condition, respectively

Table S8 Experimental results of SaDE, U-SaDE, EPSDE, and U-EPSDE over 51 independent runs on the 28 test functions with 50D from IEEE CEC2013 using 500,000 FEs

Test functions (50D)		SaDE	U-SaDE	EPSDE	U-EPSDE
		Mean Error±Std Dev	Mean Error±Std Dev	Mean Error±Std Dev	Mean Error±Std Dev
Unimodal functions	cf_1	0.00E+00±0.00E+00	0.00E+00±0.00E+00=	0.00E+00±0.00E+00	0.00E+00±0.00E+00=
	cf_2	8.65E+05±3.14E+05	8.78E+05±3.25E+05=	1.41E+07±2.86E+07	3.62E+06±1.41E+07+
	cf_3	8.62E+07±1.13E+08	7.95E+07±7.20E+07+	2.75E+09±8.50E+09	4.23E+08±1.31E+09+
	cf_4	5.11E+03±1.92E+03	5.23E+03±1.85E+03=	1.12E+04±3.54E+04	5.25E+03±2.13E+04+
	cf_5	0.00E+00±0.00E+00	0.00E+00±0.00E+00=	0.00E+00±0.00E+00	0.00E+00±0.00E+00=
Basic multimodal functions	cf_6	5.31E+01±1.95E+01	5.33E+01±1.99E+01=	3.62E+01±1.87E+00	3.57E+01±4.71E+00+
	cf_7	4.97E+01±8.99E+00	4.99E+01±9.13E+00=	8.44E+01±3.31E+01	7.62E+01±3.19E+01+
	cf_8	2.11E+01±4.06E–02	2.11E+01±3.39E–02=	2.11E+01±3.67E–02	2.11E+01±4.47E–02=
	cf_9	3.96E+01±4.61E+00	3.83E+01±3.44E+00+	7.05E+01±3.47E+00	6.96E+01±4.02E+00+
	cf_{10}	2.69E–01±1.68E–01	2.78E–01±1.55E–01=	1.32E–01±7.41E–02	1.24E–01±6.97E–02+
	cf_{11}	2.04E+00±1.78E+00	2.04E+00±1.71E+00=	9.75E–02±3.59E–01	9.53E–02±2.78E–01+
	cf_{12}	1.25E+02±2.45E+01	1.18E+02±2.14E+01+	1.62E+02±2.87E+01	1.66E+02±3.57E+01=
	cf_{13}	2.56E+02±3.96E+01	2.47E+02±3.97E+01+	2.46E+02±4.99E+01	2.47E+02±4.17E+01=
	cf_{14}	7.22E+00±5.81E+00	6.95E+00±3.62E+00+	9.96E+02±8.15E+02	8.73E+02±8.16E+02+
	cf_{15}	8.54E+03±2.11E+03	8.61E+03±2.23E+03=	1.40E+04±5.84E+02	1.38E+04±5.97E+02+
	cf_{16}	3.12E+00±3.09E–01	3.03E+00±2.56E–01=	3.35E+00±3.20E–01	3.27E+00±2.98E–01+
	cf_{17}	5.12E+01±4.33E–01	5.13E+01±3.93E–01=	5.08E+01±3.02E–01	5.10E+01±1.15E+00=
	cf_{18}	1.58E+02±7.35E+01	1.59E+02±7.41E+01=	3.37E+02±2.63E+01	3.41E+02±2.67E+01=

(*Continued*)

Test functions (50D)		SaDE	U-SaDE	EPSDE	U-EPSDE
		Mean Error±Std Dev	Mean Error±Std Dev	Mean Error±Std Dev	Mean Error±Std Dev
	cf_{19}	1.09E+01±2.67E+00	1.09E+01±2.33E+00=	6.17E+00±8.21E–01	6.18E+00±9.64E–01=
	cf_{20}	1.99E+01±9.52E–01	2.00E+01±1.03E+00=	2.25E+01±9.77E–01	2.24E+01±9.60E–01=
Composition functions	cf_{21}	8.32E+02±3.73E+02	8.34E+02±3.60E+02=	7.71E+02±4.05E+02	7.78E+02±4.02E+02=
	cf_{22}	9.30E+01±2.18E+02	3.36E+01±5.52E+01+	2.04E+03±5.57E+02	2.07E+03±5.50E+02=
	cf_{23}	8.40E+03±2.11E+03	8.42E+03±2.26E+03=	1.41E+04±6.16E+02	1.40E+04±8.00E+02=
	cf_{24}	2.78E+02±1.01E+01	2.78E+02±1.03E+01=	3.81E+02±5.38E+00	3.79E+02±5.19E+00=
	cf_{25}	3.45E+02±9.28E+00	3.43E+02±1.01E+01=	3.83E+02±4.15E+00	3.82E+02±3.83E+00=
	cf_{26}	2.95E+02±9.08E+01	2.69E+02±9.05E+01+	4.73E+02±8.59E+00	4.67E+02±3.30E+01=
	cf_{27}	1.18E+03±1.20E+02	1.18E+03±1.05E+02=	2.11E+03±4.34E+01	2.10E+03±5.01E+01+
	cf_{28}	5.34E+02±6.71E+02	4.00E+02±1.79E–14+	7.65E+02±1.01E+03	7.68E+02±1.05E+03=
		+	8	+	12
		–	0	–	0
		=	20	=	16

Note: "Mean Error" and "Std Dev" indicate the average and standard deviation of the function error values obtained in 51 runs, respectively. Wilcoxon's rank sum test at a 0.05 significance level is performed between SaDE and U-SaDE, and between EPSDE and U-EPSDE. "+", "–", and "=" denote that the performance of DE without the restrained condition is better than, worse than, and similar to that of DE with the restrained condition, respectively

Table S9 Experimental results of CoDE, U-CoDE, LSHADE, U-LSHADE, JADE/eig, and U-JADE/eig over 51 independent runs on the 28 test functions with 50D from IEEE CEC2013 using 500,000 FEs

Test functions (50D)		CoDE	U-CoDE	LSHADE	U-LSHADE	JADE/eig	U-JADE/eig
		Mean Error±Std Dev	Mean Error±Std Dev	Mean Error±Std Dev	Mean Error±Std Dev	Mean Error±Std Dev	Mean Error±Std Dev
Unimodal functions	cf_1	6.03E–07±1.91E–07	0.00E+00±0.00E+00+	0.00E+00±0.00E+00	0.00E+00±0.00E+00=	0.00E+00±0.00E+00	0.00E+00±0.00E+00=
	cf_2	2.32E+06±9.91E+05	5.40E+05±2.11E+05+	7.47E+02±1.12E+03	7.58E+02±1.34E+03=	4.05E+04±2.42E+04	3.88E+04±1.98E+04+
	cf_3	3.31E+10±4.98E+09	1.43E+06±2.33E+06+	4.67E+03±1.30E+04	3.09E+03±1.23E+04+	3.21E+06±7.47E+06	2.50E+06±5.09E+06+
	cf_4	2.35E+03±4.57E+03	3.96E–01±3.50E–01+	0.00E+00±0.00E+00	0.00E+00±0.00E+00=	8.72E–03±4.76E–02	6.23E–03±4.17E–02=
	cf_5	4.01E–04±7.37E–05	0.00E+00±0.00E+00+	0.00E+00±0.00E+00	0.00E+00±0.00E+00=	0.00E+00±0.00E+00	0.00E+00±0.00E+00=
Basic multimodal functions	cf_6	4.40E+01±3.38E–01	4.34E+01±4.50E–14+	4.34E+01±2.17E–14	4.34E+01±2.17E–14=	4.21E+01±8.69E+00	4.23E+01±8.65E+00=
	cf_7	1.17E+03±3.13E+02	5.06E+01±1.92E+01+	2.51E+00±1.25E+00	2.11E+00±1.34E+00+	2.19E+01±1.01E+01	2.65E+01±1.08E+01–
	cf_8	2.11E+01±4.24E–02	2.11E+01±4.17E–02=	2.11E+01±1.00E–01	2.09E+01±1.70E–01=	2.11E+01±8.87E–02	2.11E+01±8.42E–02=
	cf_9	6.50E+01±1.71E+00	2.44E+01±5.17E+00+	5.28E+01±2.45E+00	5.26E+01±1.92E+00=	5.32E+01±2.01E+00	5.27E+01±2.42E+00+
	cf_{10}	1.97E+00±4.55E–01	3.02E–02±2.11E–02+	9.08E–03±9.93E–03	9.75E–03±9.51E–03=	2.82E–02±1.84E–02	2.63E–02±2.08E–02+
	cf_{11}	4.93E+01±3.51E+00	1.26E+01±6.08E+00+	0.00E+00±0.00E+00	0.00E+00±0.00E+00=	0.00E+00±0.00E+00	0.00E+00±0.00E+00=
	cf_{12}	4.80E+02±1.74E+01	6.78E+01±1.53E+01+	1.44E+01±2.59E+00	1.45E+01±2.48E+00=	6.88E+01±1.12E+01	6.57E+01±1.11E+01+
	cf_{13}	4.83E+02±1.91E+01	1.39E+02±3.77E+01+	2.09E+01±8.05E+00	2.09E+01±7.89E+00=	1.51E+02±2.53E+01	1.49E+02±2.68E+01=
	cf_{14}	3.03E+03±2.26E+02	1.30E+03±2.75E+02+	2.13E–01±4.60E–02	2.14E–01±5.12E–02=	6.85E+01±1.10E+01	6.41E+01±1.02E+01+
	cf_{15}	1.39E+04±4.14E+02	6.60E+03±8.33E+02+	6.31E+03±3.35E+02	6.29E+03±3.82E+02=	6.91E+03±4.44E+02	6.73E+03±4.63E+02+
	cf_{16}	3.22E+00±3.55E–01	2.88E+00±5.83E–01+	1.25E+00±2.16E–01	1.05E+00±3.92E–01+	1.92E+00±8.91E–01	1.74E+00±8.62E–01+
	cf_{17}	1.16E+02±4.48E+00	8.89E+01±3.65E+00+	5.08E+01±1.60E–03	5.08E+01±2.45E–03=	5.15E+01±1.82E–01	5.14E+01±1.76E–01=
	cf_{18}	5.46E+02±2.09E+01	1.10E+02±2.85E+01+	1.04E+02±5.80E+00	1.02E+02±6.04E+00=	1.47E+02±9.21E+00	1.48E+02±1.02E+01=
	cf_{19}	2.17E+01±1.33E+00	1.26E+01±2.08E+00+	2.50E+00±1.30E–01	2.52E+00±1.46E–01=	3.48E+00±3.15E–01	3.49E+00±3.29E–01=
	cf_{20}	2.26E+01±2.11E–01	2.08E+01±1.09E–01+	1.82E+01±4.22E–01	1.82E+01±6.46E–01=	1.97E+01±5.28E–01	1.96E+01±5.65E–01=
Composition functions	cf_{21}	2.18E+02±1.29E+02	4.78E+02±4.18E+02–	8.51E+02±4.24E+02	8.03E+02±4.34E+02+	7.86E+02±3.67E+02	7.88E+02±3.75E+02=
	cf_{22}	4.74E+03±3.76E+02	1.86E+03±5.16E+02+	1.37E+01±1.38E+00	1.38E+01±1.54E+00=	1.36E+02±9.72E+01	1.17E+02±6.54E+01+
	cf_{23}	1.40E+04±3.82E+02	6.89E+03±9.71E+02+	5.78E+03±4.14E+02	5.79E+03±4.65E+02=	7.14E+03±6.44E+02	7.11E+03±5.64E+02=
	cf_{24}	3.60E+02±5.81E+00	2.37E+02±1.12E+01+	2.11E+02±5.89E+00	2.11E+02±5.02E+00=	2.43E+02±1.74E+01	2.44E+02±1.72E+01=
	cf_{25}	3.85E+02±3.56E+00	3.83E+02±4.25E+00+	2.78E+02±6.55E+01	2.77E+02±6.22E+01=	3.31E+02±2.72E+01	3.29E+02±2.94E+01=
	cf_{26}	2.32E+02±8.54E+01	2.59E+02±7.80E+01–	2.47E+02±5.29E+01	2.48E+02±5.40E+01=	3.47E+02±9.52E+01	3.49E+02±9.27E+01=
	cf_{27}	1.96E+03±4.82E+01	8.83E+02±1.48E+01+	4.04E+02±5.24E+01	3.84E+02±4.45E+01+	1.15E+03±3.56E+02	1.18E+03±3.34E+02=
	cf_{28}	4.00E+02±1.81E–03	4.00E+02±2.27E–14+	4.00E+02±4.95E–14	4.00E+02±3.50E–14=	5.17E+02±5.88E+02	4.00E+02±5.74E–14+
		+	25	+	5	+	10
		–	2	–	0	–	1
		=	1	=	23	=	17

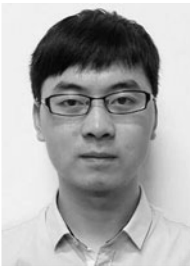
Note: "Mean Error" and "Std Dev" indicate the average and standard deviation of the function error values obtained in 51 runs, respectively. Wilcoxon's Rank sum test at a 0.05 significance level is performed between CoDE and U-CoDE, between LSHADE and U-LSHADE, and between JADE/eig and U-JADE/eig. "+", "–", and "=" denote that the performance of DE without the restrained condition is better than, worse than, and similar to that of DE with the restrained condition, respectively

References

1. Storn R, Price K. Differential evolution — a simple and efficient adaptive scheme for global optimization over continuous spaces. Berkeley, CA: International Computer Science Institute. Technical Report TR-95-012. 1995
2. Storn R, Price K. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. Journal of Global Optimization, 1997, 11(4): 341–359
3. Fan H, Lampinen J. A trigonometric mutation operation to differential evolution. Journal of Global Optimization, 2003, 27(1): 105–129
4. Zhang J, Sanderson A. JADE: adaptive differential evolution with optional external archive. IEEE Transactions on Evolutionary Computation, 2009, 13(5): 945–958
5. Das S, Abraham A. Differential evolution using a neighborhood-based mutation operator. IEEE Transactions on Evolutionary Computation, 2009, 13(3): 526–553
6. Wang Y, Cai Z X, Zhang Q F. Enhancing the search ability of differential evolution through orthogonal crossover. Information Sciences, 2012, 185(1): 153–177
7. Guo S M, Yang C C. Enhancing differential evolution utilizing eigenvector-based crossover operator. IEEE Transactions on Evolutionary Computation, 2015, 19(1): 31–49
8. Wang Y, Li H X, Huang T W, Li L. Differential evolution based on covariance matrix learning and bimodal distribution parameter setting. Applied Soft Computing, 2014, 18: 232–247
9. Liu J H, Lampinen J. A fuzzy adaptive differential evolution algorithm. Soft Computing, 2005, 9(6): 448–462
10. Brest J, Greiner S, Boskovic B, Mernik M, Zumer V. Self-adapting control parameters in differential evolution: a comparative study on numerical benchmark problems. IEEE Transactions on Evolutionary Computation, 2006, 10(6): 646–657
11. Noman N, Iba H. Accelerating differential evolution using an adaptive local search. IEEE Transactions on Evolutionary Computation, 2008, 12(1): 107–125
12. Rahnamayan S, Tizhoosh H R, Salama M M A. Opposition-based differential evolution. IEEE Transactions on Evolutionary Computation, 2008, 12(1): 64–79
13. Sun J Y, Zhang Q F, Tsang E P K. DE/EDA: a new evolutionary algorithm for global optimization. Information Sciences, 2003, 169(3–4): 249–262
14. Qin A K, Huang V L, Suganthan P N. Differential evolution algorithm with strategy adaptation for global numerical optimization. IEEE Transactions on Evolutionary Computation, 2009, 13(2): 398–417
15. Mallipeddi R, Suganthan P N, Pan Q, Tasgetiren M. Differential evolution algorithm with ensemble of parameters and mutation strategies. Applied Soft Computing, 2011, 11(2): 1679–1696
16. Wang Y, Cai Z X, Zhang Q F. Differential evolution with composite trial vector generation strategies and control parameters. IEEE Transactions on Evolutionary Computation, 2011, 15(1): 55–66
17. Das S, Suganthan P N. Differential evolution: a survey of the state-of-the-art. IEEE Transactions on Evolutionary Computation, 2011, 15(1): 4–31
18. Wang Y, Wang B C, Li H X, Yen G G. Incorporating objective function information into the feasibility rule for constrained evolutionary optimization. IEEE Transactions on Cybernetics, 2016, doi: 10.1109/TCYB.2015.2493239
19. Wang Y, Li H X, Yen G G, Song W. MOMMOP: multiobjective optimization for locating multiple optimal solutions of multimodal optimization problems. IEEE Transactions on Cybernetics, 2015, 45(4): 830–843
20. Tvrdík J. Modifications of differential evolution with composite trial vector generation strategies. In: Snášel V, Abraham A, Corchado E S, eds. Soft Computing Models in Industrial and Environmental Applications. Advances in Intelligent Systems and Computing, Vol 188. Berlin: Springer, 2013, 113–122
21. Price K, Storn R, Lampinen J. Differential Evolution-A Practical Approach to Global Optimization. Berlin: Springer-Verlag, 2005
22. Liu H, Huang H, Liu S S. Explore influence of differential operator in DE mutation with unrestrained method to generate mutant vector. In: Rutkowski L, Korytkowski M, Scherer R, et al. eds. Swarm and Evolutionary Computation. Lecture Notes in Computer Science, Vol 7269. Berlin: Springer, 2012, 292–300
23. Suganthan P N, Hansen N, Liang J, Deb K, Chen Y P, Auger A, Tiwari S. Problem definitions and evaluation criteria for the CEC 2005 special session on real-parameter optimization. KanGAL Report 2005005. 2005
24. Liang J J, Qu B Y, Suganthan P N, Hernández-Diaz A. Problem definitions and evaluation criteria for the CEC 2013 special session and competition on real-parameter optimization. Zhengzhou: Zhengzhou University. Technical Report. 2013
25. Tanabe R, Fukunaga A. Improving the search performance of shade using linear population size reduction. In: Proceeding of IEEE Congress on Evolutionary Computation. 2014, 1658–1665
26. Yang Z Y, Tang K, Yao X. Large scale evolutionary optimization using cooperative coevolution. Information Sciences, 2008, 87(15): 2985–2999

Yong Wang received the BS degree in automation from the Wuhan Institute of Technology, China in 2003, and the MS degree in pattern recognition and intelligent systems and the PhD degree in control science and engineering both from the Central South University (CSU), China in 2006 and 2011, respectively. He is currently an associate professor with the School of Information Science and Engineering, CSU. His current research interests include the theory, algorithm design, and application of computational intelligence. Dr. Wang was a recipient of the Hong Kong Scholar by the Mainland - Hong Kong Joint Postdoctoral Fellows Program, China in 2013, the Excellent Doctoral Dissertation by Hunan Province, China in 2013, the New Century Excellent Talents in University by the Ministry of Education, China in 2013, the 2015 IEEE Computational Intelligence Society Outstanding PhD Dissertation Award, the Hunan

Provincial Natural Science Fund for Distinguished Young Scholars, China in 2016, and EU Horizon 2020 Marie Sklodowska-Curie Fellowship, European Union in 2016. He is currently serving as an associate editor for the Swarm and Evolutionary Computation.



Zhi-Zhong Liu received the BS degree in automation from Central South University, China in 2013, where he is currently pursuing the PhD degree in control science and engineering. His current research interests include evolutionary computation, swarm intelligence, nonlinear equation systems, and multimodal optimization.

Jianbin Li is a professor with the Institute of Information Security and Big Data, Central South University, China. He received the BS degree from Tsinghua University, China. His research interests include network security and big data analysis.

Han-Xiong Li received the BE degree in aerospace engineering from the National University of Defense Technology, China in 1982, the ME degree in electrical engineering from Delft University of Technology, The Netherlands in 1991, and the PhD degree in electrical engineering from the University of Auckland, New Zealand in 1997. He is currently a professor with the Department of Systems Engineering and Engineering Management, the City University of Hong Kong, China. His current research interests include system intelligence and control, process design and control integration, and distributed parameter systems with applications to electronics packaging. Dr. Li was a recipient of the Distinguished Young Scholar (overseas) by the China National Science Foundation in 2004, the Chang Jiang Professorship by the Ministry of Education, China in 2006, and the National Professorship in China Thousand Talents Program in 2010. He serves as an associate editor of IEEE Transactions on Cybernetics and IEEE Transactions on Industrial Electronics (2009–2015). He is a fellow of the IEEE.

Jiahai Wang received the PhD degree from Toyama University, Japan in 2005. In 2005, he joined Sun Yat-sen University, China, where he is currently an associate professor with the School of Data and Computer Science. His current research interests include computational intelligence and its applications.