Proceedings of the 35th Chinese Control Conference
July 27-29, 2016, Chengdu, China

Differential Evolution with Strategy of Improved Population Diversity

ZHAO Li[1], SUN Chao-jiao[2], HUANG Xian-chi[3], ZHOU Bing-xu[3]
1. Xixian New District Logistics Research Institute of Shaanxi International School of Trade & Commerce, Xianyang 712046, China
E-mail: zhaoli0108@163.com

2. College of Aeronautics and Astronautics Engineering, Air Force Engineering University, Xi'an 710038, China

3. PLA Air Force the 95503 unit, Zunyi 563127, China

Abstract: Differential Evolution (DE) algorithm is well known as a simple and efficient scheme for global optimization over continuous spaces. In order to ameliorate the population diversity, an improved differential evolution (IDE) algorithm is proposed in this paper. The idea is to vary the assembling positions of the premature individuals by mutation operation. It is proved in theory that the direction pointed to the center individual is the reasonable one to improve the diversity. Then the adaptive disturbance scheme after population premature is designed. The example based on the standard function shows that the IDE with good diversity has much better searching capacity and calculating precision in the whole evolution than that of DE.

Key Words: differential evolution, evolutionary algorithms, population diversity, global numerical optimization

1 Introduction

Differential Evolution (DE), which proposed by Storn and Price[1-2], is a simple yet powerful population-based stochastic search algorithm. Because its theory is very simple and few control variables need to optimize, DE algorithm can optimize globally, efficiently and effectively in real parameter optimization domain. Until now, DE is also particular in the intelligence optimization and has been applied to solve the numerous optimization problems.

Inspired by the natural evolution of species, DE is used to optimize the Chebyshev polynomial problems at the initial stage. And then the researchers found that the new algorithm has outstanding performance in the complicated global optimization with high dimensions. It brings some new method to solve the nonlinear combination problems. Now its application domains are very extensive, such as electromagnetics [3], engineering design [4], and optimal power flow [5], global numerical optimization [6]. Many studies show that the convergence velocity and stability performance are better than the Genetic Algorithm (GA) or the Particle Swarm Optimization (PSO) over several numerical benchmarks. Although the characteristic of differential evolution is so outstanding, its defects are also obvious in the theory and application. The one is that the computation cost is too high. The other is that the diversity decreases fast in the last stage of evolution. All these may cause the population to fall into the local optimum and the premature convergence.

Therefore, the experts and researchers have proposed many advanced DE schemes with better performance, by introducing the adaptive parameters or mixing with other evolutionary algorithms. In [7], the neighbor guided selection scheme for parents involved in mutation and the direction induced mutation strategy are proposed to fully exploit the neighborhood and direction information of population. The adaptive ranking mutation operator for the differential evolution is proposed in [8] when solving the constrained optimization problems. In [9], a two-stage strategy by combing two different algorithms is used to improve the overall search efficiency. A directional mutation operator is proposed in [10] to recognize good variation directions and increase the number of generations having fitness improvement. In [11], a learning-enhanced DE that promotes individuals to exchange information within the same cluster and between different clusters is proposed. However, all the algorithms mentioned above can't guarantee the addition of the population diversity in theory. It may lead to localization for some nonlinear optimization problems.

In this paper, we proposed a new differential evolutionary algorithm which improves the population diversity through individual evolution at the direction pointed to the population center. The conventional DE and related work are reviewed in Sections II. Section III describes the improved differential evolution (IDE) and the proving course is given in theory. Experimental results of the two typical functions demonstrate the performance of IDE in comparison with the conventional DE in Sections IV. Finally, some conclusions are drawn in Section V.

2 The DE Algorithm

Like any other evolutionary algorithm, DE starts with a population of D-dimensional parameter vectors representing the candidate solutions. The population, which encodes candidate solutions, i.e., $X^{(t)}=\{X_1^{(t)},X_2^{(t)},\bullet\ ,X_D^{(t)}\}$ has the different fitness $f(X^{(t)})$. The initial individuals of population should better cover the entire search space as much as possible, by uniformly randomizing individuals within the search space constrained by the prescribed minimum and maximum parameter bounds. Denote the variation constant by F and the cross probability by CR . The individual produced in the course of variation is $u_i^{(t)}$. Then the individual by exchanging randomly between the former individual and $u_i^{(t)}$ is $W_i^{(t)}$.

Like Genetic Algorithm, the conventional DE has the following four operations.

1) Initialization

*This work is supported by The Twelfth Five year Plan of Education Science in Shaanxi Province under Grant SGH140865.

The initial population $X^{(0)}$ is chosen arbitrarily in the solution space, and the variation constant F, the cross probability CR and the population scale number NP are set respectively.

2) Mutation operation

DE employs the mutation operation to produce a mutant vector with respect to each individual, in the current population. The evolutionary operation is described as follows

$$u_i^{(t)} = X_{r1}^{(t)} + F(X_{r2}^{(t)} - X_{r3}^{(t)}) \quad (1)$$

where $r_1, r_2, r_3 \in \text{rand}(1, n)$ represents a uniformly distributed random variable within the range [1, n], satisfied the relationship $i \neq r_1 \neq r_2 \neq r_3$.

3) Crossover operation

After the mutation between individuals, the crossover operation is applied to the target vector and the corresponding mutant vector to generate a trial vector. The evolutionary operation is described as follows

$$W_{ij}^{(t)} = \begin{cases} u_{ij}^{(t)} & \text{rand}[0,1] < CR \text{ or } j = j_{\text{rand}} \\ X_{ij}^{(t)} & \text{rand}[0,1] \geq CR \text{ or } j \neq j_{\text{rand}} \end{cases} \quad (2)$$

4) Selection operation

The selection operation is to compare the objective function value of each trial vector with the corresponding target vector in every population. The target vector is used to be the population for the next generation. The evolutionary operation is described as follows

$$X_i^{(t+1)} = \begin{cases} W_i^{(t)} & f(W_i^{(t)}) \leq f(X_i^{(t)}) \\ X_i^{(t)} & f(W_i^{(t)}) > f(X_i^{(t)}) \end{cases} \quad (3)$$

The above three steps, that is mutation operation, crossover operation and selection operation, are repeated generation after generation until some specific termination criteria are satisfied.

3 IDE Algorithm Design

When the conventional DE keeps convergence fast in the posterior of evolution, the difference information among individuals becomes small. Then the corresponding diversity of the population is deceased with the iterative number of evolution, which may lead to the premature convergence or stagnation. The population will fall into the early convergence and local optimum. The traditional second variation method with random disturbance sometimes can increase the diversity of the population, but the variation is blindfold. At many varying direction, the population diversity may decreases which can't improve the global searching capacity of the DE algorithm. Therefore, a new differential evolutionary algorithm is proposed, which improves the population diversity through individual evolution at the direction pointed to the population center. And the prove course is given as follows in theory.

Definite 1 The diversity of the population is

$$I = \frac{1}{NP}\sum_{i=1}^{NP}\sqrt{\sum_{j=1}^{D}(X_{ij} - \overline{X_j})^2} \quad (4)$$

where $\overline{X_j}$ is the mean of the individual in dimension j,

$$\overline{X_j} = \frac{1}{NP}\sum_{i=1}^{NP} X_{ij} \quad (5)$$

Definite 2 If the center of the population is stated as $\overline{X^{(t)}} = (\overline{X_1^{(t)}}, \overline{X_2^{(t)}}, \cdots, \overline{X_D^{(t)}})$, the vector to $X_i^{(t)}$ is

$$\vec{d_i} = X_i^{(t)} - \overline{X^{(t)}} = (X_{i1}^{(t)} - \overline{X_1^{(t)}}, \cdots, X_{iD}^{(t)} - \overline{X_D^{(t)}}) \quad (6)$$

and the corresponding distance is

$$\left\|\vec{d_i}\right\| = \left\|X_i^{(t)} - \overline{X^{(t)}}\right\| = \sqrt{\sum_{j=1}^{D}(X_{ij}^{(t)} - \overline{X_j^{(t)}})^2} \quad (7)$$

Theorem 1 For the individual with D dimension, the population scale is $NP \geq 3$. When the individual $X_r^{(t)}$ varies along $\vec{d_r}$ to the reverse direction of the center of the population $\overline{X^{(t)}}$, in other words, the following relationship is satisfied.

$$X_r^{(t+1)} - X_r^{(t)} = k\vec{d_r} = k(X_r^{(t)} - \overline{X^{(t)}}) \quad (8)$$

where $0 < k < NP$. Then the diversity of the population after variance is increased.

Proof. The individual of the varied population can be expressed as

$$X_{ij}^{(t+1)} = \begin{cases} X_{ij}^{(t)} & i \neq r \\ X_{rj}^{(t)} + \Delta X_{rj} & i = r \end{cases} \quad (9)$$

The coordinate of the new population center is

$$\overline{X_j^{(t+1)}} = \frac{1}{NP}(\sum_{i=1}^{NP} X_{ij}^{(t)} + \Delta X_{rj}) = \overline{X_j^{(t)}} + \frac{1}{NP}\Delta X_{rj} \quad (10)$$

where $j = 1, 2, \cdots, G$.

If the individual varies in the direction showing in (8), the increment of the population diversity can be transform into the following form.

$$\begin{aligned}\Delta I &= \frac{1}{NP}\sum_{i=1}^{NP}\left(\sqrt{\sum_{j=1}^{D}(X_{ij}^{(t+1)} - \overline{X_j^{(t+1)}})^2} - \sqrt{\sum_{j=1}^{D}(X_{ij}^{(t)} - \overline{X_j^{(t)}})^2}\right) \\ &= \frac{1}{NP}\sum_{i=1}^{NP}\left[\frac{1}{NP}\left(\sqrt{\sum_{j=1}^{D}(X_{rj}^{(t+1)} - \overline{X_j^{(t+1)}})^2} - \sqrt{\sum_{j=1}^{D}(X_{rj}^{(t)} - \overline{X_j^{(t+1)}})^2}\right)\right. \\ &\left. + \sqrt{\sum_{j=1}^{D}(X_{ij}^{(t)} - \overline{X_j^{(t+1)}})^2} - \sqrt{\sum_{j=1}^{D}(X_{ij}^{(t)} - \overline{X_j^{(t)}})^2}\right]\end{aligned} \quad (11)$$

According to (8), we can obtain that

$$X_r^{(t+1)} - \overline{X^{(t+1)}} = \varpi(X_r^{(t)} - \overline{X^{(t+1)}}) \quad (12)$$

where $\varpi \in R$, which mean that the above two vectors are in the same line. Then we can further find out

$$X_r^{(t+1)} - \overline{X^{(t+1)}} = \zeta(X_r^{(t)} - \overline{X^{(t+1)}}) \quad (13)$$

where $\zeta \in R$.

Because $0 < k < NP$, it exists

$$\begin{aligned}X_{rj}^{(t+1)} - \zeta X_{rj}^{(t)} &= (1-\zeta)(\overline{X_j^{(t)}} + \frac{1}{NP}(X_{rj}^{(t+1)} - X_{rj}^{(t)})) \\ &= (1-\zeta)[(\frac{1}{NP} - \frac{1}{k})X_{rj}^{(t+1)} + (1 + \frac{1}{k} - \frac{1}{NP})X_{rj}^{(t)}]\end{aligned} \quad (14)$$

Now, by using the underdetermined coefficients method, we can obtain

$$\zeta = 1 + \frac{NP \times k}{NP - k} \quad (15)$$

It can be transformed into

$$|\frac{NP - k}{k}|(|\zeta| - 1) = NP \quad (16)$$

By integrated the (16) and (11), it can be simplified as

$$\Delta I = \frac{1}{NP}\sum_{i=1}^{NP}\Bigg[\frac{|\zeta|-1}{NP}\sqrt{\sum_{j=1}^{D}(X_{rj}^{(t)}-\overline{X_j^{(t+1)}})^2} + \sqrt{\sum_{j=1}^{D}(X_{ij}^{(t)}-\overline{X_j^{(t+1)}})^2} - \sqrt{\sum_{j=1}^{D}(X_{ij}^{(t)}-\overline{X_j^{(t)}})^2}\Bigg]$$
$$= \frac{1}{NP}\sum_{i=1}^{NP}\Bigg[\sqrt{\sum_{j=1}^{D}(\overline{X_j^{(t+1)}}-\overline{X_j^{(t)}})^2} + \sqrt{\sum_{j=1}^{D}(X_{ij}^{(t)}-\overline{X_j^{(t+1)}})^2} - \sqrt{\sum_{j=1}^{D}(X_{ij}^{(t)}-\overline{X_j^{(t)}})^2}\Bigg] \tag{17}$$

Until now, we use the Minkowski's inequality and the expression

$$\Delta I > 0 \tag{18}$$

is easily obtained. □

According to the conclusion of theorem 1, we can design the following disturbance scheme for the conventional DE when the population premature and convergence too early.

$$X_q^{(t)} = X_p^{(t)} + (X_p^{(t)} - \overline{X^{(t)}})(1+\xi) \tag{19}$$

where $X_p^{(t)}$ and $X_q^{(t)}$ are the individual before disturbance, ξ is the random number in $[0,1]$. Remarkably, the designed disturbance scheme can ensure the increase of population diversity and improve the diversity performance after second disturbance.

How to determine the premature time is very important to the optimization performance. If the average fitness of the current population is f_{avg}, then we define fitness errors

$$\delta^2 = \sum_{i=1}^{NP}\left|\frac{f(X_i^{(t)})-f_{avg}}{f}\right| \tag{20}$$

where

$$f = \min\{\max\{|f(X_i^{(t)})-f_{avg}|\},1\} \tag{21}$$

When $\delta^2 \le \Delta$ and the fitness of the best individual is larger than ε, the population falls into the premature. $\Delta > 0$ is the variance of the fitness and ε is the precision.

Now the steps are list as follows:

Step 1 Initialize the population. Set the scale number NP, the variation constant F, the cross probability CR, the variance of the fitness Δ and the precision ε.

Step 2 Calculate the fitness of the individual $f(X_i^{(t)})$, then find out the best fitness in the population $f(X_{gbest}^{(t)})$ and the best individual $X_{gbest}^{(t)}$.

Step 3 If the condition, which $f(X_{gbest}^{(t)})$ is less than ε or the maximum iteration is reached, is satisfied. Then the optimization is finished. Otherwise continue.

Step 4 Calculate the variance of the fitness as in (20). If $\delta^2 < \Delta$ and $f(X_{gbest}^{(t)}) > \varepsilon$, the mutation operation is implemented for the $X_{gbest}^{(t)}$ and some random individuals. Otherwise, come to step 5.

Step 5 Do mutation operation, crossover operation and selection operation for all the individual in the population to form the new population. Come back to step 2.

4 Experiment and Analysis

To validate the effectiveness of the IDE algorithm, the standard test function is chosen to do some experiment. All the simulation and analysis is compared to the conventional DE algorithm in [12]. The standard test function includes:

1) Sphere function

$$f_1(x) = \sum_{i=1}^{D} x_i^2 \tag{22}$$

where $x_i \in [-10,10]$, $i = 1,2,3,\cdots,D$. It is easy to find out that the minimum of the Sphere function is 0 at the position $x_i = 0$.

2) Rosenbrock function

$$f_3(x) = \sum_{i=1}^{D-1}\{100(x_i^2 - x_{i+1}) + (x_i - 1)^2\} \tag{23}$$

where $x_i \in [-10,10]$, $i = 1,2,3,\cdots,D$. The minimum of the Rosenbrock function is 0 but is difficult to locate.

The parameters are set as following. The dimension of the function D is 10. The scale number NP is 100. The variation constant F is 0.85. The variance of the fitness Δ is 0.001. The precision ε is 0.001. The experiments are carried out for twenty times respectively, then the results are listed in Table 1.

Table 1: The Result of the optimization

Function	Algorithm	Average	Variance	Optimum	Worst-case
Sphere	DE	2.975e^{-4}	3.718e^{-4}	9.595e^{-5}	7.649e^{-2}
	IDE	1.193e^{-7}	3.253e^{-8}	7.193e^{-8}	6.816e^{-7}
Rosenbrock	DE	0.12389	0.054205	0.037124	0.33430
	IDE	4.080e^{-5}	2.263e^{-6}	5.871e^{-6}	9.764e^{-5}

According to the Table 1, it is obvious that the worst-case data of IDE algorithm is better than that of the conventional DE. It is because that the IDE algorithm can guarantee the diversity after the mutation operation, but the DE is very weak to fall into the premature and reduces to the local optimum. For the Rosenbrock function which is much difficult to optimize than others, the DE algorithm is neither falling into the local optimum nor turning into the stop flags. However, the IDE algorithm can obtain the reasonable good result because of its global optimal performance.

In order to dramatically draw the diversity variation in the evolutional course of the optimal solution, we take the Rosenbrock function as example to discuss at detail. The individuals are scattered both in the 20th and 117th generation with the designed adaptive disturbance scheme (19). The first and second variables denoted by the population, which is compared between the population before scattering and the population after scattering, are shown in Fig. 1 and Fig. 2 respectively.

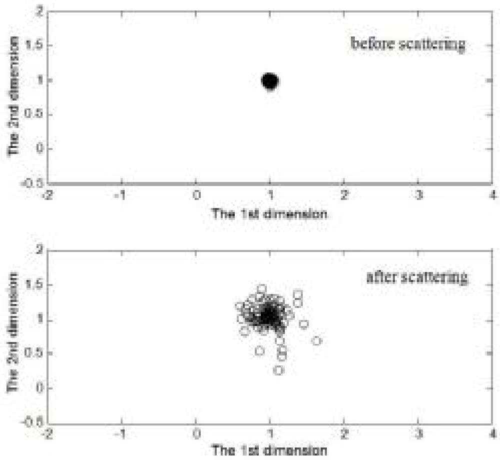

Fig. 1: The population distribution of IDE in 20th generation

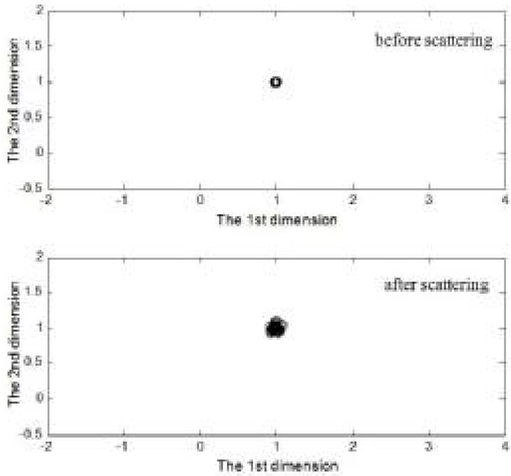

Fig. 2: The population distribution of IDE in 117th generation

It is illustrated in Fig. 1 that the population is falling into premature in the early stage of the evolution, which isn't very propitious to solve the global optimum. After scattering, the individuals are widely distributed in the area of the local optimum. This operation will increase the probability to obtain the global solution. Also we can see in Fig. 2 that in the later stage of the evolution, the IDE algorithm can keep good diversity through the adaptive mutation with the disturbance function. All the above analysis illuminates that the diversity of the DE algorithm is sharply declined in the later of the evolution and the population falls into a stagnant state. Therefore, the result is much worse than that of IDE algorithm.

The diversity variation curves of the IDE algorithm are shown in Fig. 3. The IDE algorithm can keep better diversity of population through the designed adaptive disturbance scheme, especially at the time of the declined diversity in the latter evolution. The population of DE algorithm generates the assembled phenomenon, which reduces to the falling of local optimum. However, the diversity of the IDE algorithm is increased and is much larger than that of DE when the population assembling. And the population is converged fast after the scattering operation.

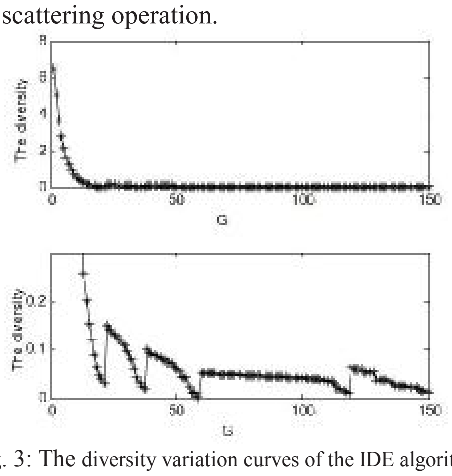

Fig. 3: The diversity variation curves of the IDE algorithm

It is shown in the test experiments that the IDE algorithm keep good diversity in the whole evolution. By comparing with the conventional DE algorithm, it has much better searching capacity and calculating precision. The optimization performance is increased greatly through the designed adaptive disturbance scheme.

5 Conclusions

In this study we proposed an improved direction pointed to the center individual to ameliorate the population diversity of the conventional differential evolution algorithm. Because the difference between the individuals becomes very small when DE keeps convergence fast, the corresponding diversity of the population is deceased with the iterative number of evolution. It leads to the premature convergence or stagnation. The adaptive disturbance scheme after population premature is designed based on the mutation operation at the direction to the center individual, which is proved that the diversity is increased in theory. At last, the standard test functions, Sphere function and Rosenbrock function are used to validate the effectiveness of the IDE algorithm. The simulation results show that the diversity IDE is much larger and the searching capacity is improved than DE algorithm.

References

[1] R. Storm, K. Price. *Differential evolution: a simple and efficient adaptive scheme for global optimization over continuous spaces*. Berkeley: International Computer Science Institute, TR-95-012, 1995.

[2] K. Price, Differential evolution: a fast and simple numerical optimizer, in *Biennial Conference of the North American Fuzzy Information Processing Society*, 1996: 524-527.

[3] P. Rocca, G. Oliveri, A. Massa, Differential evolution as applied to electromagnetic, *Antennas & Propagation Magazine IEEE*, 2011, 53(1): 38-49.

[4] T. Liao, Two hybrid differential evolution algorithms for engineering design optimization, *Applied Soft Computing*, 2010, 10(4): 1188-1199.

[5] M. Abido, N. Al-Ali, Multi-objective optimal power flow using differential evolution, *Arabian Journal for Science & Engineering*, 2012, 37(4): 991-1005.

[6] A. Qin, V. Huang, P. Suganthan, Differential evolution algorithm with strategy adaption for global numerical optimization, *IEEE Transactions on Evolutionary Computation*, 2009, 13(2): 398-417.

[7] Y. Cai, J. Wang, Differential evolution with neighborhood and direction information for numerical optimization, *IEEE Transactions on Cybernetics*, 2013, 43(6): 2202-2215.

[8] W. Gong, Z. Cai, D. Liang, Adaptive ranking mutation operator based differential evolution for constrained optimization, *IEEE Transactions on Cybernetics*, 2015, 45(4): 716-727.

[9] X.S Yang, S Deb, Two-stage eagle strategy with differential evolution, *Journal International Journal of Bio-Inspired Computation*, 2012, 4(1): 1-5.

[10] X. Zhang, S.Y. Yuen, A directional mutation operator for differential evolution algorithms, *Applied Soft Computing*, 2015, 30(C): 529-548.

[11] Y. Cai, J. Wang, J. Yin, Learning-enhanced differential evolution for numerical optimization, *Soft Computing*, 2012, 16(2): 303-330.

[12] R. Storm, K. Price, *Differential evolution: a simple and efficient heuristic for global optimization over continuous spaces*, Journal of Global Optimization, 1997, 11(4): 341-359.