

A Hybrid Estimation of Distribution Algorithm with Differential Evolution for Global Optimization

Bing Dong

East China Normal University, Shanghai, China

November 30, 2016

- 1 Background
- 2 EDA/DE-EIG
- 3 Experiment results
- 4 Conclusions and future work

The box-constrained continuous global optimization can be stated in the following:

$$\begin{array}{ll} \min & f(x) \\ \text{s.t} & x \in [a_i, b_i]^n \end{array} \quad (1)$$

- $x = (x_1, x_2, \dots, x_n)^T \in R^n$ is a decision vector
- $[a_i, b_i]^n$ is the search space
- $f : R^n \rightarrow R$ is the objective function

Algorithm 1: DE

```

1 Initial the population  $P_0$  randomly :
   $P_0 = \{x_{1,D}, x_{2,D}, x_{3,D}, \dots, x_{N,D}\}$ 
2 while not terminate do
  // mutation
3    $v_{i,G} = x_{r1,G} + F \cdot (x_{r2,G} - x_{r3,G})$ 
  // crossover
4   if  $\text{rand}_j(0,1) \leq CR$  or  $j = j_{rand}$  then
5      $u_{i,j,G} = v_{i,j,G}$ 
6   else
7      $u_{i,j,G} = x_{i,j,G}$ 
8   end
  // selection
9   if  $f(u_{i,G}) \leq f(x_{i,G})$  then
10     $x_{i,G+1} = u_{i,G}$ 
11  else
12     $x_{i,G+1} = x_{i,G}$ 
13  end
14 end
  
```

Figure 1: The algorithm framework of DE

- N is the population size, and D is the dimension of the target vector.
- At each generation G , a mutant vector $v_{i,G}$ is obtained by mutation. F is the scaling factor, $r1, r2, r3$ are mutually different integers randomly selected from $[1, N]$.
- The trial vector $u_{i,j,G}$ is generated by combining $v_{i,G}$ and $x_{i,j,G}$. $\text{rand}_j(0,1) \in [0,1]$ is a uniformly distributed random number, and j_{rand} is a random integer between j and D . CR is the controlling parameter.
- $u_{i,G}$ and $x_{i,G}$ compete to enter the next generation in accordance with the objective function value.

Algorithm 2: EDA

```
1 Initialization: Initial the  $Pop(t)$  randomly, and  $t$  is the  
   generation.  
2 while not terminate do  
3   Modeling: Build a probabilistic model  $p(x)$   
   according to the statistical information of the  $Pop(t)$ .  
4   Sampling: Generate a new solution set  $Q$  by  
   sampling from the built probabilistic model  $p(x)$ .  
5   Selection: Select from  $Q \cup Pop(t)$  to construct the  
   next population  $Pop(t+1)$ . The selection criterion is  
   the objective function value.  
6    $t = t + 1$   
7 end
```

- t is the generation counter.
- $Pop(t) = \{x^1, x^2, \dots, x^N\}$ is the population at generation t .
- Q is a new solution set sampled from the probabilistic model.

Figure 2: The algorithm framework of DE

Algorithm 3: DE-EIG

```

1 Initial the population  $Pop(t) = \{x_1, x_2, x_3, \dots, x_N\}$  ( $N$ 
  is the size of the population)
2 while not terminate do
3    $v_{i,G} = x_{r1,G} + F \cdot (x_{r2,G} - x_{r3,G})$ 
4   if  $rand() < p$  then
5     if  $rand() \leq CR$  then
6        $u_{i,G} = v_{i,G}$ 
7     else
8        $u_{i,G} = x_{i,G}$ 
9     end
10  else
11    Compute the the eigenvector matrix  $E$  of  $x_{i,G}$ ,
      let  $E'$  be the inverse matrix.
12     $x'_{i,G} = E' \cdot x_{i,G}$ 
13     $v'_{i,G} = E' \cdot v_{i,G}$ 
14    if  $rand() \leq CR$  then
15       $u'_{i,G} = v'_{i,G}$ 
16    else
17       $u'_{i,G} = x'_{i,G}$ 
18    end
19     $u_{i,G} = E \cdot u'_{i,G}$ 
20  end
21  if  $f(u_{i,G}) \leq f(x_{i,G})$  then
22     $x_{i,G+1} = u_{i,G}$ 
23  else
24     $x_{i,G+1} = x_{i,G}$ 
25  end
26   $t = t + 1$ 
27 end

```

As traditional DE operates crossover in the original coordinate, it is inevitable to lose some statistics information. DE-EIG based on the eigenvectors makes the crossover rotationally invariant by transform the coordinate system of the individuals in the population.

Figure 3: The algorithm framework of DE-EIG

Algorithm 4: DE/EDA

```

1 Generate population  $Pop(t)$  randomly consists of N
  solutions  $x_1, x_2, \dots, x_N$  from the feasible search space.
  while not terminate do
2   Construct the probabilistic model:
3    $p_k(x) = \prod_{i=1}^n \mathcal{N}(x_i; \mu_i, \sigma_i)$ 
4   For all  $j = 1, 2, \dots, n$ , produce a trial solution
      $u = (u_1, u_2, \dots, u_n)$ 
5   if  $rand() < CRP$  then
6      $u_j = \frac{(x_i)_j + (x_d)_j}{2} + F \cdot [(x_d)_j - (x_i)_j + (x_b)_j - (x_c)_j]$ 
7   else
8      $u_j$  is sampled according to  $\mathcal{N}(x_i; \mu_i, \sigma_i)$ 
9   end
10  where  $CRP$  is the controlling parameter.
11  if  $f(u) < f(x_i)$  then
12     $x_i^{t+1} = u$ 
13  else
14     $x_i^{t+1} = x_i^t$ 
15  end
16   $t = t + 1$ 
17 end
  
```

CRP is utilized to control the offspring generation by DE or EDA. DE/EDA utilized the global information extracted by EDA and the differential information exploited by DE to obtain promising solutions.

Figure 4: The algorithm framework of DE/EDA

Algorithm 5: EDA/DE-EIG

```

1 Initial the population  $Pop(t) = \{x_1, x_2, x_3, \dots, x_N\}$  ( $N$ 
  is the size of the population)
2 while not terminate do
3   Construct the probabilistic model:
4    $p(x) = \prod_{i=1}^n \mathcal{N}(x_i; \mu_i, \sigma_i)$ 
5   Generate a trial solution  $u_{i,G}$  as follows:
6   if  $rand() < CRP$  then
7      $u_{i,G}$  is produced by DE-EIG.
8   else
9      $u_{i,G}$  is sampled from the probabilistic model
       $p(x)$ .
10  end
11  if  $f(u_{i,G}) < f(x_{i,G})$  then
12     $x_{i,G+1} = u_{i,G}$ 
13  else
14     $x_{i,G+1} = x_{i,G}$ 
15  end
16  if  $Converge(\theta, G, G_e)$  then
17    Operate the expensive local search.
18  end
19   $t = t + 1$ 
20 end

```

Based on the framework of DE/EDA, DE-EIG is imported to improve the sampling in the algorithm. For $Converge(\theta, G, G_e)$ at line 16, it is essential to judge whether to operate expensive local search.

Figure 5: The algorithm framework of DE/EDA-EIG

Compared algorithms and Experimental setting

In this paper, EDA/DE-EIG is compared with JADE and DE/EDA on the first 13 test instances from YYL test instances. The global minimum objective value is 0 for all the test instances.

- The dimension of the population is 30. All algorithms are run independently 50 times and stopped after 450,000 function evaluations.
- JADE: The parameters $N = 150$, $p = 0.05$, $c = 0.1$, $F = 0.5$ and $CR = 0.9$.
- DE/EDA: The parameters are set as: $N = 150$, $F = 0.5$ and $CRP = 0.9$.
- EDA/DE-EIG: The CRP is 0.5; F is set to be 0.5; CR is set to be 0.6; the parameter p to control the probability to operate crossover two coordinate systems is 0.5; the convergence threshold $\theta = 0.1$; the size of the population N is 150. For the parameter setting of expensive local search, it is same as that in EDA/LS.

TABLE I
 STATISTICAL RESULTS (*mean* \pm *std*) FOR THE THREE ALGORITHMS ON INSTANCES $f1 - f13$.

| instances | EDA/DE-EIG | JADE | DE/EDA |
|-----------|---|---|---|
| $f1$ | 1.54e-159 \pm 5.11e-159 | $3.90e - 127 \pm 2.74e - 126(+)$ | $1.39e - 59 \pm 2.58e - 59(+)$ |
| $f2$ | 1.02e-75 \pm 7.46e-76 | $2.60e - 35 \pm 1.64e - 34(+)$ | $5.15e - 28 \pm 4.68e - 28(+)$ |
| $f3$ | 4.01e-35 \pm 8.47e-35 | $7.79e - 35 \pm 2.51e - 34(\sim)$ | $1.23e - 12 \pm 1.20e - 12(+)$ |
| $f4$ | 5.01e-20 \pm 3.06e-19 | $3.15e - 14 \pm 6.42e - 14(+)$ | $9.90e - 12 \pm 2.69e - 11(+)$ |
| $f5$ | $1.46e - 29 \pm 2.62e - 29$ | 3.85e-30 \pm 9.58e-30(-) | $3.37e - 21 \pm 8.66e - 21(+)$ |
| $f6$ | 0.00e+00 \pm 0.00e+00 | 0.00e+00 \pm 0.00e+00(\sim) | 0.00e+00 \pm 0.00e+00(\sim) |
| $f7$ | $3.60e - 03 \pm 1.00e - 03$ | 6.01e-04 \pm 2.23e-04(-) | $2.20e - 03 \pm 5.59e - 04(-)$ |
| $f8$ | $2.79e + 03 \pm 5.02e + 02$ | 4.74e+00 \pm 2.34e+01(-) | $1.82e + 03 \pm 6.72e + 02(-)$ |
| $f9$ | $6.23e + 00 \pm 2.21e + 00$ | 0.00e+00 \pm 0.00e+00(-) | $1.54e + 02 \pm 1.96e + 01(+)$ |
| $f10$ | 4.44e-15 \pm 0.00e+00 | 4.44e-15 \pm 0.00e+00(\sim) | 4.44e-15 \pm 0.00e+00(\sim) |
| $f11$ | 0.00e+00 \pm 0.00e+00 | $1.48e - 04 \pm 1.05e - 03(\sim)$ | $2.96e - 04 \pm 1.46e - 03(\sim)$ |
| $f12$ | 1.57e-32 \pm 5.53e-48 | 1.57e-32 \pm 5.53e-48(\sim) | 1.57e-32 \pm 5.53e-48(\sim) |
| $f13$ | 1.35e-32 \pm 1.11e-47 | 1.35e-32 \pm 1.11e-47(\sim) 3(+)6(\sim)4(-) | 1.35e-32 \pm 1.11e-47(\sim) 6(+)5(\sim)2(-) |

¹ The bold ones mean the best.

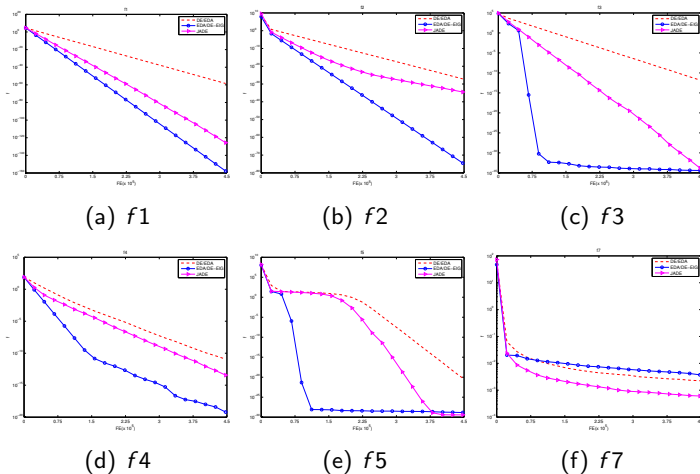
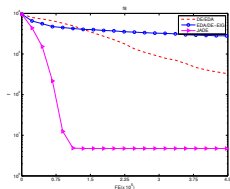
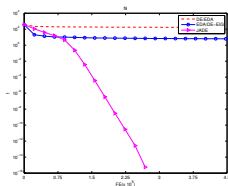


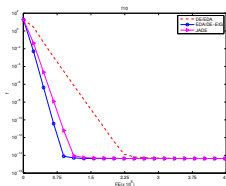
Figure 6: The mean function value versus on $f_1 - f_7$ except f_6 .



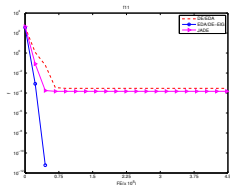
(a) f_8



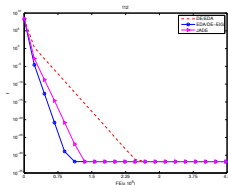
(b) f_9



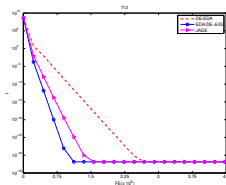
(c) f_{10}



(d) f_{11}



(e) f_{12}



(f) f_{13}

Figure 7: The mean function value versus on $f_8 - f_{13}$

DE/EDA is a promising method that utilizes both the the global and local information for global optimization. However, the potential improvement of the performance of this algorithm has not been exploited furthermore. In this paper, an improved DE, DE-EIG, is imported to combine with EDA, bringing an impressive improvement on the performance. DE-EIG is beneficial to utilize the statistics information of the population to accelerate the convergence. And expensive local search is applied to improve the performance further.

The experimental results have shown the distinct advantages of the proposed method, named as EDA/DE-EIG, in comparison with two state-of-art algorithms JADE and DE/EDA.

The results reported in this paper is preliminary and there are several ways to improve the algorithm performance. Firstly, the algorithm framework of EDA/DE-EIG can be simplified. Secondly, it is worth to investigate how to allocate the computational resources to both DE-EIG and EDA.

Thanks!

- B. Dong, A. Zhou, and G. Zhang, A Hybrid Estimation of Distribution Algorithm with Differential Evolution for Global Optimization, 2016 IEEE Symposium Series on Computational Intelligence (SSCI), 2016.