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

Bing Dong

East China Normal University, Shanghai, China

December 3, 2016

Outline

Outline for Section 1

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

Definition

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

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

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

DE

Algorithm 1: DE

Initial the population P_0 randomly : $P_0 = \{x_{1,D}, x_{2,D}, x_{3,D}, \cdots, x_{N,D}\}$
while *not terminate* **do**
  // mutation
  $v_{i,G} = x_{r1,G} + F \cdot (x_{r2,G} - x_{r3,G})$
  // crossover
  if $rand_j(0,1) \leq CR$ *or* $j = j_{rand}$ **then**
    $u_{i,j,G} = v_{i,j,G}$
  else
    $u_{i,j,G} = x_{i,j,G}$
  end
  // selection
  if $f(u_{i,G}) \leq f(x_{i,G})$ **then**
    $x_{i,G+1} = u_{i,G}$
  else
    $x_{i,G+1} = x_{i,G}$
  end
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 combing $v_{i,G}$ and $x_{i,j,G}$. $rand_j(0,1) \in [0,1]$ 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.

EDA

Algorithm 2: EDA

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

Figure 2: The algorithm framework of DE

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

Outline for Section 2

DE-EIG

Algorithm 3: DE-EIG

Initial the population $Pop(t) = \{x_1, x_2, x_3, \cdots, x_N\}$ (N is the size of the population)
while *not terminate* **do**
  $v_{i,G} = x_{r1,G} + F \cdot (x_{r2,G} - x_{r3,G})$
  if $rand() < p$ **then**
    if $rand() \leq CR$ **then**
      $u_{i,G} = v_{i,G}$
    else
      $u_{i,G} = x_{i,G}$
    end
  else
    Compute the the eigenvector matrix E of $x_{i,G}$, let E' be the inverse matrix.
    $x'_{i,G} = E' \cdot x_{i,G}$
    $v'_{i,G} = E' \cdot v_{i,G}$
    if $rand() \leq CR$ **then**
      $u'_{i,G} = v'_{i,G}$
    else
      $u'_{i,G} = x'_{i,G}$
    end
    $u_{i,G} = E \cdot u'_{i,G}$
  end
  if $f(u_{i,G}) \leq f(x_{i,G})$ **then**
    $x_{i,G+1} = u_{i,G}$
  else
    $x_{i,G+1} = x_{i,G}$
  end
  $t = t + 1$
end

Figure 3: The algorithm framework of DE-EIG

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.

DE/EDA

Algorithm 4: DE/EDA

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

Figure 4: The algorithm framework of DE/EDA

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.

DE/EDA-EIG

Algorithm 5: EDA/DE-EIG

1 Initial the population $Pop(t) = \{x_1, x_2, x_3, \cdots, x_N\}$ (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** $Converage(\theta, G, G_e)$ **then**
17 Operate the expensive local search.
18 **end**
19 $t = t + 1$
20 **end**

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

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

Outline for Section 3

Compared algorithms and experimental setting

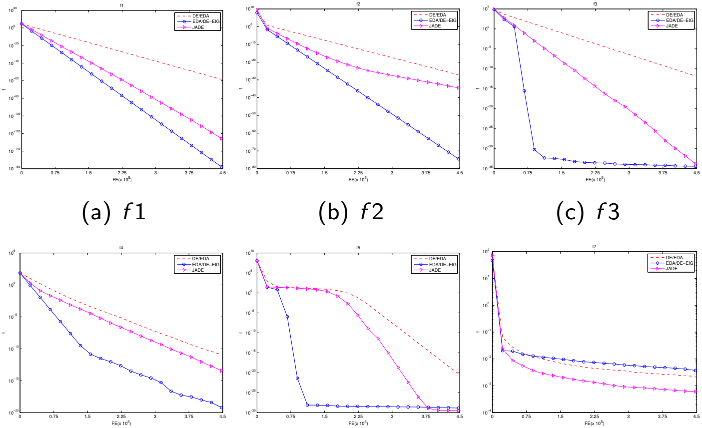
In this paper, EDA/DE-EIG is compared with JADE and DE/EDA on the first 13 test instances form YYL test instances. The global minimum objective value is 0 for all the test instances.

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

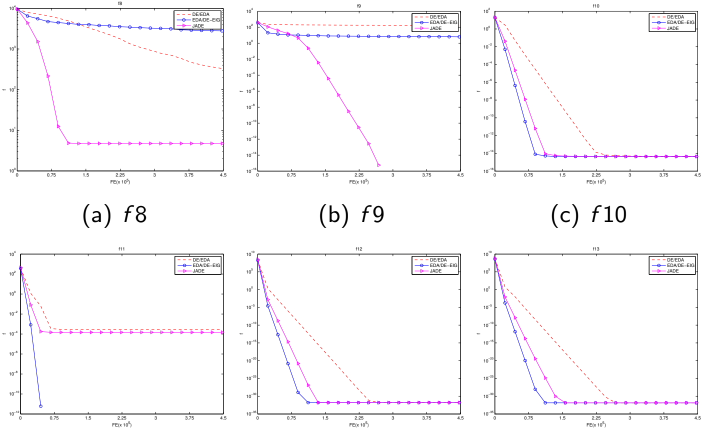
[1] The bold ones mean the best.



(a) $f1$ (b) $f2$ (c) $f3$

(d) $f4$ (e) $f5$ (f) $f7$

Figure 6: The mean function value versus on $f1 - f7$ except $f6$.



(a) $f8$ (b) $f9$ (c) $f10$

(d) $f11$ (e) $f12$ (f) $f13$

Figure 7: The mean function value versus on $f8 - f13$

Outline for Section 4

Conclusions

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.

Future work

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.