Supplementary Document for Multioutput Surrogate Assisted Evolutionary Algorithm for Expensive Multi-Modal Optimization Problems

1st Renzhi Chen
Defense Innovation Institute
Beijing, China
chenrenzhi1989@gmail.com

2nd Ke Li
Department of Computer Science
University of Exeter
Exeter, EX4 4QF, UK
k.li@exeter.ac.uk

I. Experimental Setup

A. Synthetic Benchmark Problem

We choose three classic multimodal optimization problems as the synthetic benchmark problem in our experiments.

- **Rastrigin Function** has several local minima. It is highly multimodal, but locations of the minima are regularly distributed. It is defined as:

$$f(\mathbf{x}) = an + \sum_{i=1}^{n}[x_i^2 - a\cos(cx_i)] \tag{1}$$

where $a = 10, c = 2\pi$ and the bounds is set $\mathbf{x} \in [-5.12, 5.12]^n$.

- **Ackley Function** [1] is widely used for testing optimization algorithms which is characterized by a nearly flat outer region, and a large hole at the centre. The function poses a risk for optimization algorithms to be trapped in one of its many local minima. It is defined as:

$$\begin{aligned} f(\mathbf{x}) = -a\exp\left(-b\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2}\right) \\ -\exp\left(\frac{1}{n}\sum_{i=1}^{n}\cos(cx_i)\right) + a + e \end{aligned} \tag{2}$$

where $a = 20, b = 0.2, c = 2\pi$, and the bounds is set $\mathbf{x} \in [-5, 5]^n$.

- **Griewank Function** has many widespread local minima, which are regularly distributed. It is defined as:

$$f(\mathbf{x}) = \frac{1}{4000}\sum_{i=1}^{n}x_i^2 - \prod_{i=1}^{n}\cos\left(\frac{x_i}{\sqrt{i}}\right) + 1 \tag{3}$$

where the bounds is set $\mathbf{x} \in [-10, 10]^n$.

B. Statistical Tests

To have a statistical interpretation of the significance of comparison results, we use the following three statistical measures in our empirical study.

- Wilcoxon signed-rank test [2]: This is a non-parametric statistical test that makes little assumption about the underlying distribution of the data and has been recommended in many empirical studies in the EA community [3]. In particular, the significance level is set to $p = 0.05$ in our experiments.
- Scott-Knott test [4]: We apply the Scott-Knott test to rank the performance of different peer techniques over 21 runs on each test problem. In a nutshell, the Scott-Knott test uses a statistical test and effect size to divide the performance of peer algorithms into several clusters according to their metric values. After that, each cluster can be assigned a rank according to the mean metric values achieved by the peer algorithms within the cluster. In particular, the smaller the rank is, the better performance of the algorithm achieves.
- A_{12} effect size [5]: To ensure the resulted differences are not generated from a trivial effect, we apply A_{12} as the effect size measure to evaluate the probability that one algorithm is better than another. Specifically, given a pair of peer algorithms, $A_{12} < 0.56$ means they are *equivalent*. $0.56 \leq A_{12} < 0.64$ indicates a *small* effect size while $0.64 \leq A_{12} < 0.71$ and $A_{12} \geq 0.71$ mean a *medium* and a *large* effect size, respectively.

II. Results and Analysis

A. Comparison with the Peer Algorithms for Different Dimensions

the trajectories of $\mathcal{L}(\mathbf{x})$ shown in Figs. 1.

B. Comparison with the Peer Algorithms for Different Number of Optimal

The statistical comparison results on $\bar{\epsilon}_{\mathrm{f}}$ and $\bar{\epsilon}_{\mathrm{t}}$ are given in Tables II and the trajectories of $\mathcal{L}(\mathbf{x})$ shown in Figs. 2.

References

[1] T. Bäck, *Evolutionary algorithms in theory and practice - evolution strategies, evolutionary programming, genetic algorithms*. Oxford University Press, 1996.

[2] F. Wilcoxon, "Individual comparisons by ranking methods," 1945.

[3] J. Derrac, S. García, D. Molina, and F. Herrera, "A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms," *Swarm Evol. Comput.*, vol. 1, no. 1, pp. 3–18, 2011.

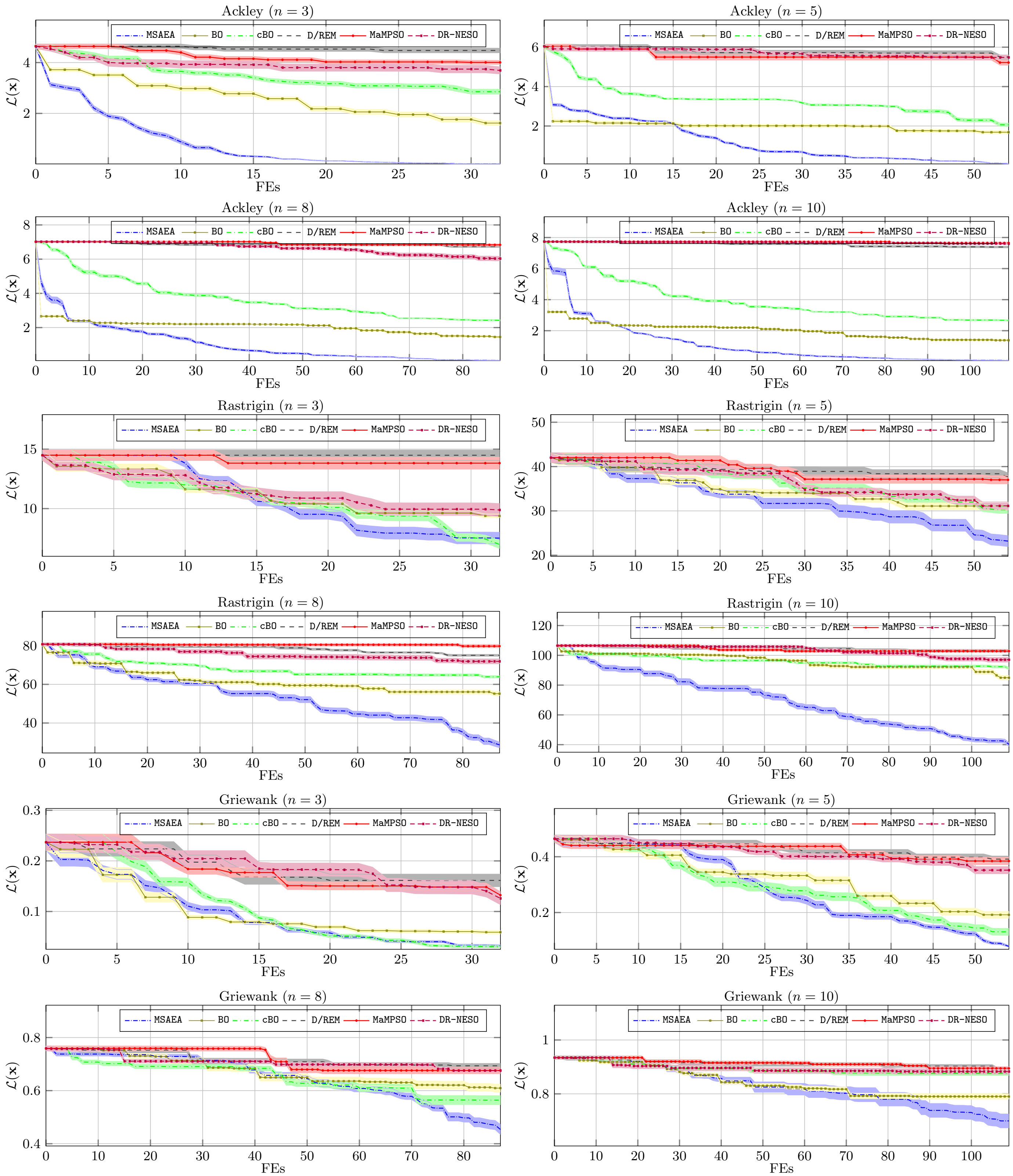

Fig. 1. Trajectories of $\mathcal{L}(\mathbf{x})$ with confidence bounds achieved by six algorithms on Ackley and Rastrigin functions.

[4] N. Mittas and L. Angelis, "Ranking and clustering software cost estimation models through a multiple comparisons algorithm," *IEEE Trans. Software Eng.*, vol. 39, no. 4, pp. 537–551, 2013.

[5] A. Vargha and H. D. Delaney, "A critique and improvement of the CL

TABLE I

PERFORMANCE COMPARISON RESULTS OF $\overline{\epsilon}_f$ AND $\overline{\epsilon}_t$ BETWEEN MSAEA AND THE OTHER 5 PEER ALGORITHMS

$\overline{\epsilon}_t$	n	MSAEA	BO	cBO	MaMPSO	D/REM	DR-NESO
Ackley	3	**3.670E-2(7.69E-2)**	1.249E+0(8.95E-1)†	2.589E+0(8.37E-1)†	4.617E+0(8.49E-1)†	4.120E+0(6.61E-1)†	3.741E+0(1.12E+0)†
	5	**1.419E-1(2.38E-1)**	1.806E+0(6.57E-1)†	2.359E+0(9.40E-1)†	5.562E+0(8.07E-1)†	5.389E+0(1.06E+0)†	5.268E+0(1.20E+0)†
	8	**8.114E-2(8.96E-2)**	1.408E+0(4.02E-1)†	2.472E+0(3.73E-1)†	6.880E+0(4.52E-1)†	7.035E+0(5.75E-1)†	6.456E+0(1.12E+0)†
	10	**9.649E-2(8.75E-2)**	1.287E+0(4.18E-1)†	2.717E+0(4.08E-1)†	7.206E+0(6.57E-1)†	7.487E+0(5.38E-1)†	7.427E+0(6.34E-1)†
Rastrigin	3	7.781E+0(5.99E+0)	1.017E+1(4.53E+0)	**7.455E+0(4.03E+0)**	1.735E+1(5.31E+0)†	1.558E+1(5.96E+0)†	1.374E+1(6.68E+0)†
	5	**2.047E+1(1.24E+1)**	3.180E+1(7.13E+0)†	3.066E+1(6.89E+0)†	3.685E+1(8.00E+0)†	3.641E+1(8.35E+0)†	3.320E+1(8.75E+0)†
	8	**3.058E+1(1.57E+1)**	5.809E+1(1.24E+1)†	6.409E+1(6.20E+0)†	7.698E+1(5.94E+0)†	8.085E+1(7.87E+0)†	7.250E+1(1.32E+1)†
	10	**3.800E+1(1.59E+1)**	8.136E+1(1.42E+1)†	9.052E+1(1.05E+1)†	1.003E+2(1.24E+1)†	1.031E+2(1.12E+1)†	9.456E+1(1.68E+1)†
Griewank	3	3.282E-2(1.50E-2)	5.907E-2(3.47E-2)†	**3.062E-2(2.53E-2)**	1.615E-1(1.24E-1)†	1.324E-1(6.73E-2)†	1.258E-1(1.22E-1)†
	5	**7.886E-2(3.01E-2)**	1.918E-1(1.25E-1)†	1.308E-1(1.25E-1)	3.915E-1(1.10E-1)†	3.846E-1(1.29E-1)†	3.519E-1(1.29E-1)†
	8	**4.537E-1(1.87E-1)**	6.095E-1(1.47E-1)	5.643E-1(1.70E-1)	6.935E-1(8.91E-2)†	6.760E-1(1.22E-1)†	6.756E-1(5.37E-2)†
	10	**6.997E-1(2.57E-1)**	7.903E-1(9.96E-2)	8.759E-1(7.32E-2)†	8.947E-1(5.25E-2)†	8.948E-1(5.99E-2)†	8.829E-1(9.82E-2)†
$\overline{\epsilon}_f$	n	MSAEA	BO	cBO	MaMPSO	D/REM	DR-NESO
Ackley	3	**3.369E+0(4.53E-1)**	4.167E+0(5.11E-1)†	4.665E+0(5.01E-1)†	5.304E+0(6.44E-1)†	5.127E+0(5.34E-1)†	5.015E+0(7.63E-1)†
	5	**4.113E+0(3.74E-1)**	4.580E+0(4.55E-1)†	5.217E+0(3.74E-1)†	6.408E+0(7.03E-1)†	6.351E+0(7.33E-1)†	6.339E+0(7.12E-1)†
	8	**4.346E+0(3.29E-1)**	4.837E+0(3.28E-1)†	5.744E+0(3.68E-1)†	7.339E+0(4.61E-1)†	7.392E+0(5.24E-1)†	7.298E+0(5.56E-1)†
	10	**4.614E+0(2.88E-1)**	5.040E+0(2.33E-1)†	6.108E+0(2.87E-1)†	7.733E+0(4.66E-1)†	7.812E+0(4.35E-1)†	7.805E+0(4.48E-1)†
Rastrigin	3	**1.924E+1(4.32E+0)**	2.025E+1(4.43E+0)†	1.974E+1(3.98E+0)	2.183E+1(4.82E+0)†	2.158E+1(4.66E+0)†	2.071E+1(5.28E+0)†
	5	**4.094E+1(6.02E+0)**	4.362E+1(4.77E+0)†	4.363E+1(5.22E+0)†	4.471E+1(6.01E+0)†	4.508E+1(6.09E+0)†	4.431E+1(6.03E+0)†
	8	**7.633E+1(7.10E+0)**	8.024E+1(7.20E+0)†	8.248E+1(4.99E+0)†	8.778E+1(6.11E+0)†	8.910E+1(6.80E+0)†	8.696E+1(8.01E+0)†
	10	**9.376E+1(8.30E+0)**	1.053E+2(8.65E+0)†	1.072E+2(8.78E+0)†	1.107E+2(1.10E+1)†	1.119E+2(1.08E+1)†	1.107E+2(1.14E+1)†
GriewankS	3	**2.385E-1(9.57E-2)**	2.422E-1(1.02E-1)	2.455E-1(1.05E-1)	2.851E-1(1.47E-1)†	2.816E-1(1.30E-1)†	2.863E-1(1.52E-1)†
	5	**4.475E-1(1.16E-1)**	4.663E-1(1.25E-1)	4.508E-1(1.14E-1)	5.190E-1(1.29E-1)†	5.165E-1(1.19E-1)†	5.125E-1(1.23E-1)†
	8	**7.597E-1(5.64E-2)**	7.731E-1(8.33E-2)	7.607E-1(7.37E-2)	7.949E-1(6.05E-2)†	7.947E-1(7.37E-2)	7.901E-1(6.33E-2)
	10	**9.044E-1(7.33E-2)**	9.109E-1(4.16E-2)	9.357E-1(4.81E-2)	9.457E-1(5.09E-2)	9.466E-1(4.91E-2)	9.372E-1(6.01E-2)

† indicates that MSAEA is significantly better than the corresponding peer algorithm according to the Wilcoxon signed-rank test at the 5% significance level.

TABLE II

PERFORMANCE COMPARISON RESULTS OF $\overline{\epsilon}_f$ AND $\overline{\epsilon}_t$ BETWEEN MSAEA AND THE OTHER 5 PEER ALGORITHMS OVER ACKLEY ($c = \pi, 2\pi, 4\pi$)

$\overline{\epsilon}_t$	n	MSAEA	BO	cBO	MaMPSO	D/REM	DR-NESO
Ackley ($c = \pi$)	3	**2.336E-2(1.46E-2)**	2.818E-1(1.32E-1)†	1.697E+0(1.22E+0)†	4.464E+0(9.10E-1)†	4.189E+0(8.30E-1)†	3.032E+0(1.45E+0)†
	5	**4.995E-2(2.80E-2)**	1.074E+0(6.76E-1)†	1.593E+0(8.59E-1)†	5.675E+0(8.03E-1)†	5.444E+0(6.48E-1)†	4.953E+0(1.64E+0)†
	8	**6.883E-2(3.09E-2)**	5.565E-1(1.64E-1)†	1.551E+0(5.59E-1)†	6.690E+0(3.65E-1)†	7.028E+0(6.16E-1)†	7.039E+0(6.01E-1)†
	10	**6.408E-2(3.07E-2)**	4.562E-1(1.12E-1)†	2.308E+0(4.96E-1)†	7.440E+0(4.40E-1)†	7.681E+0(5.33E-1)†	7.373E+0(8.42E-1)†
Ackley ($c = 2\pi$)	3	**1.761E-2(9.66E-3)**	1.620E+0(9.19E-1)†	2.856E+0(9.57E-1)†	4.475E+0(8.97E-1)†	4.007E+0(7.89E-1)†	3.691E+0(1.41E+0)†
	5	**9.075E-2(5.53E-2)**	1.685E+0(7.70E-1)†	2.064E+0(8.23E-1)†	5.501E+0(1.02E+0)†	5.220E+0(1.34E+0)†	5.480E+0(1.17E+0)†
	8	**8.403E-2(7.63E-2)**	1.451E+0(4.85E-1)†	2.426E+0(4.20E-1)†	6.726E+0(3.84E-1)†	6.837E+0(5.85E-1)†	6.033E+0(1.34E+0)†
	10	**9.096E-2(6.49E-2)**	1.385E+0(5.19E-1)†	2.664E+0(4.36E-1)†	7.387E+0(5.05E-1)†	7.653E+0(4.76E-1)†	7.597E+0(5.53E-1)†
Ackley ($c = 4\pi$)	3	**2.949E-1(4.07E-1)**	2.131E+0(6.41E-1)†	2.620E+0(5.87E-1)†	4.425E+0(1.03E+0)†	4.105E+0(9.96E-1)†	2.957E+0(7.18E-1)†
	5	**6.206E-1(4.74E-1)**	2.510E+0(3.99E-1)†	2.678E+0(4.54E-1)†	5.530E+0(7.58E-1)†	5.689E+0(7.35E-1)†	5.263E+0(1.13E+0)†
	8	**9.770E-1(7.23E-1)**	2.125E+0(4.05E-1)†	2.876E+0(5.95E-1)†	6.769E+0(3.48E-1)†	7.014E+0(3.98E-1)†	6.681E+0(4.69E-1)†
	10	**1.054E+0(5.62E-1)**	2.324E+0(2.92E-1)†	2.879E+0(3.12E-1)†	7.385E+0(4.61E-1)†	7.725E+0(4.26E-1)†	7.214E+0(7.68E-1)†
$\overline{\epsilon}_f$	n	MSAEA	BO	cBO	MaMPSO	D/REM	DR-NESO
Ackley ($c = \pi$)	3	**3.151E+0(5.18E-1)**	3.636E+0(4.76E-1)†	4.504E+0(5.77E-1)†	5.184E+0(7.27E-1)†	5.080E+0(7.00E-1)†	4.769E+0(8.08E-1)†
	5	**3.881E+0(2.71E-1)**	4.496E+0(4.85E-1)†	5.048E+0(4.62E-1)†	6.589E+0(6.42E-1)†	6.525E+0(5.42E-1)†	6.433E+0(7.49E-1)†
	8	**4.203E+0(2.83E-1)**	4.431E+0(2.46E-1)†	5.460E+0(3.67E-1)†	7.258E+0(3.84E-1)†	7.382E+0(5.37E-1)†	7.384E+0(5.33E-1)†
	10	**4.476E+0(2.47E-1)**	4.642E+0(1.63E-1)†	6.139E+0(2.94E-1)†	7.885E+0(2.96E-1)†	7.951E+0(4.23E-1)†	7.954E+0(3.59E-1)†
Ackley ($c = 2\pi$)	3	**3.294E+0(3.44E-1)**	4.220E+0(5.54E-1)†	4.661E+0(4.87E-1)†	5.199E+0(5.73E-1)†	5.035E+0(5.21E-1)†	4.878E+0(8.44E-1)†
	5	**4.079E+0(2.56E-1)**	4.500E+0(4.52E-1)†	5.158E+0(2.48E-1)†	6.401E+0(7.71E-1)†	6.302E+0(8.17E-1)†	6.365E+0(7.13E-1)†
	8	**4.344E+0(1.96E-1)**	4.865E+0(3.44E-1)†	5.687E+0(3.03E-1)†	7.267E+0(3.98E-1)†	7.306E+0(4.98E-1)†	7.168E+0(5.32E-1)†
	10	**4.641E+0(1.66E-1)**	5.089E+0(2.20E-1)†	6.050E+0(2.22E-1)†	7.864E+0(3.41E-1)†	7.936E+0(3.78E-1)†	7.920E+0(3.91E-1)†
Ackley ($c = 4\pi$)	3	**3.451E+0(4.98E-1)**	4.238E+0(6.59E-1)†	4.471E+0(6.61E-1)†	5.080E+0(9.25E-1)†	5.031E+0(9.31E-1)†	4.601E+0(7.34E-1)†
	5	**4.350E+0(2.39E-1)**	4.844E+0(3.53E-1)†	5.347E+0(3.66E-1)†	6.448E+0(5.87E-1)†	6.443E+0(5.21E-1)†	6.468E+0(5.89E-1)†
	8	**4.697E+0(3.10E-1)**	5.015E+0(2.93E-1)†	5.783E+0(3.47E-1)†	7.246E+0(3.93E-1)†	7.321E+0(4.49E-1)†	7.291E+0(4.31E-1)†
	10	**5.043E+0(2.52E-1)**	5.377E+0(2.02E-1)†	6.282E+0(3.04E-1)†	7.878E+0(3.42E-1)†	7.961E+0(3.78E-1)†	7.858E+0(3.67E-1)†

† indicates that MSAEA is significantly better than the corresponding peer algorithm according to the Wilcoxon signed-rank test at the 5% significance level.

common language effect size statistics of McGraw and Wong," *J. Educ. Behav. Stat.*, vol. 25, no. 2, pp. 101–132, 2000.

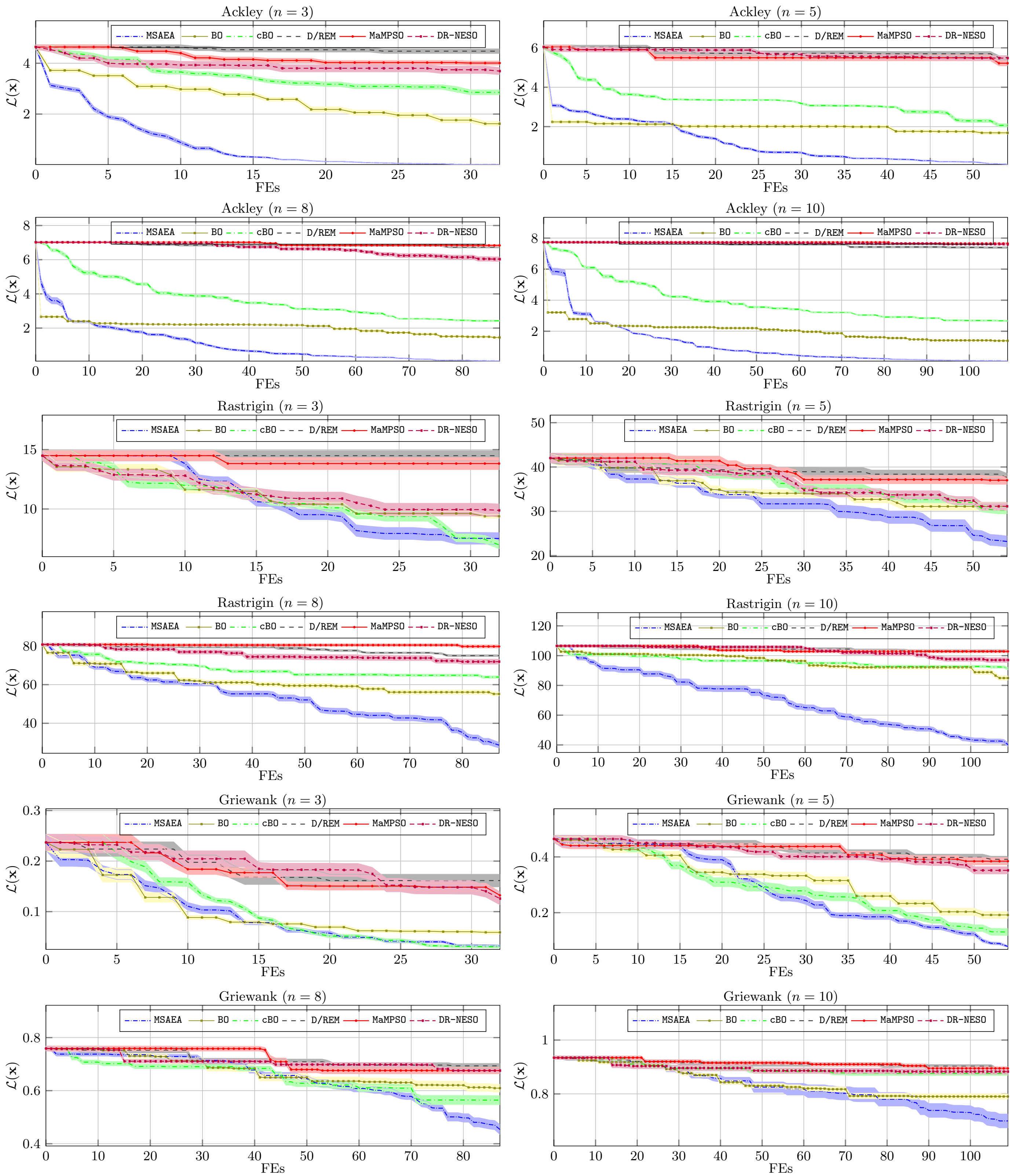

Fig. 2. Trajectories of $\mathcal{L}(\mathbf{x})$ with confidence bounds achieved by six algorithms on Ackley, Rastrigin and Griewank functions.

TABLE III
PERFORMANCE COMPARISON RESULTS OF $\overline{\epsilon}_f$ AND $\overline{\epsilon}_t$ BETWEEN MSAEA AND THE OTHER THREE VARIANTS ON ACKLEY, RASTRIGIN AND GRIEWANK FUNCTIONS

$\overline{\epsilon}_t$	n	MSAEA	MSAEA-v1	MSAEA-v2	MSAEA-v3
Ackley	3	**3.670E-2(7.69E-2)**	2.818E+0(1.08E+0)†	1.920E+0(9.02E-1)†	4.089E+0(7.89E-1)†
	5	**1.419E-1(2.38E-1)**	1.990E+0(8.33E-1)†	1.815E+0(6.30E-1)†	5.257E+0(1.51E+0)†
	8	**8.114E-2(8.96E-2)**	2.419E+0(4.47E-1)†	1.327E+0(4.91E-1)†	6.686E+0(6.09E-1)†
	10	**9.649E-2(8.75E-2)**	2.746E+0(4.44E-1)†	1.296E+0(4.19E-1)†	7.680E+0(5.06E-1)†
Rastrigin	3	7.781E+0(5.99E+0)	**7.244E+0(3.39E+0)**	9.047E+0(1.66E+0)†	1.515E+1(5.37E+0)†
	5	**2.047E+1(1.24E+1)**	3.212E+1(7.88E+0)	3.212E+1(9.02E+0)†	3.737E+1(1.02E+1)†
	8	**3.058E+1(1.57E+1)**	6.545E+1(5.74E+0)†	5.880E+1(9.92E+0)†	8.013E+1(8.53E+0)†
	10	**3.800E+1(1.59E+1)**	9.072E+1(8.45E+0)†	8.721E+1(1.33E+1)†	1.060E+2(8.37E+0)†
Griewank	3	**3.282E-2(1.50E-2)**	3.431E-2(2.84E-2)	5.972E-2(3.72E-2)†	1.539E-1(6.32E-2)†
	5	**7.886E-2(3.01E-2)**	1.419E-1(1.42E-1)†	1.860E-1(1.19E-1)†	3.835E-1(1.47E-1)†
	8	**4.537E-1(1.87E-1)**	5.674E-1(1.63E-1)†	6.044E-1(1.54E-1)†	6.431E-1(1.20E-1)†
	10	**6.997E-1(2.57E-1)**	8.758E-1(8.33E-2)†	8.071E-1(1.04E-1)†	9.210E-1(4.84E-2)†
$\overline{\epsilon}_f$	n	MSAEA	MSAEA-v1	MSAEA-v2	MSAEA-v3
Ackley	3	**3.369E+0(4.53E-1)**	4.522E+0(4.96E-1)†	4.145E+0(6.09E-1)†	4.986E+0(5.90E-1)†
	5	**4.113E+0(3.74E-1)**	5.221E+0(2.42E-1)†	4.580E+0(3.33E-1)†	6.469E+0(8.40E-1)†
	8	**4.346E+0(3.29E-1)**	5.592E+0(2.51E-1)†	4.741E+0(3.24E-1)†	7.162E+0(4.83E-1)†
	10	**4.614E+0(2.88E-1)**	6.107E+0(2.29E-1)†	5.070E+0(2.27E-1)†	7.974E+0(3.81E-1)†
Rastrigin	3	1.924E+1(4.32E+0)	1.755E+1(4.26E+0)	**1.751E+1(3.61E+0)**	1.950E+1(5.13E+0)†
	5	**4.094E+1(6.02E+0)**	4.396E+1(7.24E+0)†	4.329E+1(6.23E+0)†	4.506E+1(7.80E+0)†
	8	**7.633E+1(7.10E+0)**	8.279E+1(4.84E+0)†	8.015E+1(7.22E+0)†	8.829E+1(6.00E+0)†
	10	**9.376E+1(8.30E+0)**	1.085E+2(6.97E+0)†	1.088E+2(6.94E+0)†	1.137E+2(7.75E+0)†
Griewank	3	**2.385E-1(9.57E-2)**	2.812E-1(1.00E-1)†	2.758E-1(9.72E-2)†	3.272E-1(1.22E-1)†
	5	**4.475E-1(1.16E-1)**	4.477E-1(1.14E-1)†	4.603E-1(1.37E-1)†	5.159E-1(1.36E-1)†
	8	**7.597E-1(5.64E-2)**	7.977E-1(6.26E-2)†	7.717E-1(8.39E-2)†	7.868E-1(7.53E-2)†
	10	**9.044E-1(7.33E-2)**	9.403E-1(5.15E-2)†	9.232E-1(4.18E-2)†	9.609E-1(4.85E-2)†

† indicates that MSAEA is significantly better than the corresponding peer algorithm according to the Wilcoxon signed-rank test at the 5% significance level.