2013 IEEE Congress on Evolutionary Computation
June 20-23, Cancún, México

Differential Evolution on the CEC-2013 Single-Objective Continuous Optimization Testbed

A. K. Qin*†, Xiaodong Li*
*School of Computer Science and Information Technology, RMIT University, Melbourne, 3001, Victoria, Australia
Email:{kai.qin, xiaodong.li}@rmit.edu.au
†School of Automation, Southeast University, Nanjing, China, 210096

***Abstract*—Differential evolution (DE) is one of the most powerful continuous optimizers in the field of evolutionary computation. This work systematically benchmarks a classic DE algorithm (DE/rand/1/bin) on the CEC-2013 single-objective continuous optimization testbed. We report, for each test function at different problem dimensionality, the best achieved performance among a wide range of potentially effective parameter settings. It reflects the intrinsic optimization capability of DE/rand/1/bin on this testbed and can serve as a baseline for performance comparison in future research using this testbed. Furthermore, we conduct parameter sensitivity analysis using advanced non-parametric statistical tests to discover statistically significantly superior parameter settings. This analysis provides a statistically reliable rule of thumb for choosing the parameters of DE/rand/1/bin to solve unseen problems. Moreover, we report the performance of DE/rand/1/bin using one superior parameter setting advocated by parameter sensitivity analysis.**

I. Introduction

Differential evolution (DE) [1]–[3] is a very effective and powerful stochastic optimizer, which was proposed by Storn and Price in 1995 and has now developed into one of the most promising research areas in the field of evolutionary computation. Over the past decades, numerous studies have been carried out to improve DE's performance [4]–[12], to give a theoretical explanation of the behavior of DE [13], to apply DE and its variants to solve various scientific and engineering problems [3], as evidenced by a huge body of publications on DE in the forms of monographs, edited volumes and research articles. DE related algorithms have demonstrated the outstanding performance when solving many challenging tasks. It is worth noting that DE variants have always been one of the top performers in previous optimization competitions held at the *IEEE Congress on Evolutionary Computation (CEC)* such as CEC-2005 single-objective, CEC-2007 and CEC-2009 multi-objective and CEC-2008 large-scale continuous optimization.

This work systematically benchmarks the performance of a well-known classic DE algorithm (**DE/rand/1/bin**) on the newly proposed CEC-2013 single-objective continuous optimization testbed. This testbed, consisting of 28 test functions, expands its CEC-2005 counterpart by adding new test functions, introducing oscillation and symmetric breaking transformations and modifying the formula of composition functions. As the most widely used classic DE algorithm, DE/rand/1/bin has succeeded in solving various numerical and real-world problems and is often chosen as a baseline to gauge the effectiveness of newly proposed optimizers. However, its own performance can be significantly influenced by three involved control parameters, i.e., population size (NP), crossover rate (CR) and mutation scale factor (F). Consequently, using improper parameter settings will greatly degrade its efficacy as both a problem solver and a baseline method.

We evaluate DE/rand/1/bin, using a wide range of parameter settings, on all 28 test functions in the CEC-2013 testbed at three different problem dimensionality (10D, 30D and 50D) and present the best achieved results among all these parameter settings for each function at each dimensionality. The reported results reveal the intrinsic optimization capability of DE/rand/1/bin on this testbed and provide a handy reference for future research using this testbed by choosing DE/rand/bin as a baseline method for performance comparison. Furthermore, we conduct parameter sensitivity analysis with respect to all 28 functions at each dimensionality using the Iman and Davenport test and the Hochberg post-hoc procedure [14], [15]. The results indicate that medium NP values (e.g., 40, 50 and 60), large CR values (e.g., 0.9) and medium F values (e.g., 0.5) can lead to the statistically significantly better performance than the other parameter settings at any tested problem dimensionality. This finding provides a statistically reliable rule of thumb for choosing the parameters of DE/rand/1/bin for solving unseen problems. Moreover, we present the results corresponding to one of the superior parameter settings, i.e., NP: 50, CR: 0.9 and F: 0.5, suggested by parameter sensitivity analysis to comply with the protocol of the CEC-2013 testbed about using a single parameter setting to conduct all experiments [16].

The remaining paper proceeds with a brief review of DE in Section II. Section III presents experimental results on the CEC-2013 testbed and parameter sensitivity analysis. Section IV concludes the paper.

II. Differential Evolution

A general black-box continuous optimization task aims at finding the optimal real-valued decision variables to minimize (or maximize) one or more real-valued objective functions given no prior knowledge about function characteristics. For example, the single-objective continuous optimization considered in the scope of this paper can be formulated as:

$$\mathbf{x}^* = \arg\min_{\mathbf{x}\in\mathcal{R}^D} f(\mathbf{x}), f(\mathbf{x}) \in \mathcal{R}$$

where $\mathbf{x} = \{x^1, \dots, x^D\} \in \mathcal{R}^D$ is a decision vector composed of D real-valued decision variables. The real-valued objective function $f(\mathbf{x}) \in \mathcal{R}$ quantifies the quality of such a decision vector. DE is a population-based stochastic search algorithm very proficient in solving black-box continuous optimization problems. It has been widely applied in various scientific and engineering fields [3].

A. Algorithm Description

DE evolves a population of individuals (decision vectors) towards global optima via three operations: mutation, recombination and selection. Specifically, a fixed-size population is first randomly initialized within the solution space. Then, each individual in the population, so-called target vector, undergoes the following three operations in sequence:

- *Mutation*: a base vector is first generated using population members, which is the reference point of the mutation. Then, the vector difference of randomly sampled population members excluding the target vector under consideration is scaled and added to the base vector to produce a mutant vector. There exist different ways to create the base vector and vector difference, which correspond to different mutation strategies.
- *Recombination*: the above-generated mutant vector and the target vector under consideration are recombined to generate a trial vector. Discrete recombination, so-called crossover, is most often used in DE, e.g., binominal (uniform) crossover and exponential (circular two-point) crossover. Alternatively, line recombination can be used.
- *Selection*: if the trial vector has better quality than the target vector under consideration, it will replace the target vector and enter the population for the next generation. Otherwise, the target vector will remain in the population for the next generation.

The population is iteratively updated by applying these three operations until certain termination criteria are met, e.g., the maximum number of function evaluations is reached.

The success of DE is mainly attributed to its differential mutation scheme and its capability of exploiting contour matching (the population tends to distribute along functional level sets), which distinguish DE from other existing evolutionary algorithms. Scaled difference vectors with respect to all possible pairs of population members distributed along function level sets can well adapt to the property of the searching landscape currently explored, which thus provide promising mutation directions with adaptive mutation step-sizes balancing between the global and local search. Specifically, at the initial searching stage, population members spread over the entire solution space. Accordingly, lengths of difference vectors are large to favor the global search. As the evolution goes on, population members gradually converge to a sub-region of the solution space. Consequently, the local search is advocated by small lengths of differential vectors.

DE variants can be denoted using "DE/x/y/z" in which:

- "x" defines the scheme to generate the base vector, e.g., "rand" and "best" indicate the base vector is a randomly selected population member and the so-far best population member, respectively.
- "y" defines the number of pairs of population members used to construct the vector difference, e.g., 1 and 2 mean the vector difference is composed of one and two pairs of population members, respectively.
- "z" defines the recombination scheme, e.g., "bin" and "exp" stand for the binominal and exponential crossover (discrete recombination), respectively.

Among classic DE algorithms, DE/rand/1/bin is most widely used. Its pseudo-code is described in Algorithm 1.

B. Control Parameters

DE has three control parameters:

- *Population size* (NP) determines the population diversity and accordingly influences the convergence speed. Its choice usually depends on the complexity and scale of the given problem. For example, large population sizes are preferred to solve highly multimodal problems or large-scale problems having hundreds (or thousands) of decision variables. However, limited computational budgets in practice may prevent DE with large population sizes from converging to a desirable solution.
- *Crossover rate* (CR), in the most often used discrete recombination, controls how many decision variables in a target vector interchange their values with those in the corresponding mutant vector to generate a trial vector. Its choice usually depends on the interaction of decision variables in a given problem. When fewer decision variables are interacted with each other, small CR values are more effective than large ones in terms of the convergence speed. However, when more decision variables are interrelated, large CR values are more effective.
- *Mutation scale factor* (F) adjusts the mutation step-size in the relative manner to control the exploration and exploitation power. Its choice should aim to prevent the undesirable convergence speed. Too small F values may lead to premature convergence while too large F values may much slow down the speed of convergence.

Among existing research works on DE, the ways of configuring parameters can be divided into three categories:

- *Fixed schemes* use a fixed parameter setting throughout the searching course, which is suggested on a basis of theoretical or empirical studies on some suites of test problems [1], [2], [13].
- *Control schemes* use some predefined rules to alter the parameter setting throughout the searching course [6].
- *Adaptive schemes* incessantly adapt the parameter setting by online learning its impact on the searching performance throughout the searching course [4], [5], [7]–[10], [12].

Recent works on investigating control parameters of DE mainly focus on adaptive schemes. In most of these works, the effectiveness of the newly proposed adaptive scheme is gauged in contrast with some fixed schemes. However, this comparison may become less reliable when the parameter setting in the chosen fixed scheme is based on test problems of different properties from the currently used ones. Furthermore, it is interesting to compare, for each test problem at certain dimensionality, the performance of a newly proposed adaptive

Algorithm 1 Classic DE Algorithm: DE/rand/1/bin

Input: *NP*, *CR*, *F*

Output: $\mathbf{x}^* = \arg\min_{\mathbf{x}\in\mathcal{R}^D} f(\mathbf{x})$

Initialize the generation counter g: $g = 0$
Initialize the population $\mathbf{P}_g$ of *NP* D-dimensional individuals: $\mathbf{P}_g = \{\mathbf{x}_{1,g}, \ldots, \mathbf{x}_{NP,g}\}$ with $\mathbf{x}_{i,g} = \{x^1_{i,g}, \ldots, x^D_{i,g}\}$
  The following uniformly random initialization is used in our experiments (Section III):
  for $j = 1 \to D$
    $x^j_{i,g} = rand_u(0,1)\cdot(b^j_{\mathrm{U}} - b^j_{\mathrm{L}}) + b^j_{\mathrm{L}}$, b^j_{L} and b^j_{U} are the upper and lower bounds of the j^{th} individual element
  end for
Evaluate the objective function value of each individual in $\mathbf{P}_g$, i.e., $f(\mathbf{x}_{i,g}), i = 1, \ldots, NP$
while the predefined termination criteria are not met **do**
  for $i = 1 \to NP$ **do**
    Randomly select in $\{1, \ldots, NP\}$ three mutually exclusive indices that are distinct from i:
      do $r_1 = ceil(rand_u(1, NP))$
      while $r_1 \neq i$
      do $r_2 = ceil(rand_u(1, NP))$
      while $r_2 \neq i$ and $r_2 \neq r_1$
      do $r_3 = ceil(rand_u(1, NP))$
      while $r_3 \neq i$ and $r_3 \neq r_1$ and $r_3 \neq r_2$
    Generate a mutant vector $\mathbf{v}_{i,g} = \{v^1_{i,g}, \ldots, v^D_{i,g}\}$:
      $\mathbf{v}_{i,g} = \mathbf{x}_{r_1,g} + F\cdot(\mathbf{x}_{r_2,g} - \mathbf{x}_{r_3,g})$
    Generate a trial vector $\mathbf{u}_{i,g} = \{u^1_{i,g}, \ldots, u^D_{i,g}\}$:
      $j_{rand} = ceil(rand_u(1, D))$
      for $j = 1 \to D$

$$u^j_{i,g} = \begin{cases} v^j_{i,g} & \text{if } rand_u(0,1) \le CR \text{ or } j = j_{rand} \\ x^j_{i,g} & \text{otherwise} \end{cases}$$

      end for
    Evaluate the objective function value of the generated trial vector $\mathbf{u}_{i,g}$
    Determine the i^{th} individual in the population for the next generation $\mathbf{P}_{g+1} = \{\mathbf{x}_{1,g+1}, \ldots, \mathbf{x}_{NP,g+1}\}$:

$$\mathbf{x}_{i,g+1} = \begin{cases} \mathbf{u}_{i,g} & \text{if} f(\mathbf{u}_{i,g}) \le f(\mathbf{x}_{i,g}) \\ \mathbf{x}_{i,g} & \text{otherwise} \end{cases}$$

  end for
  Increase the generation counter: $g = g + 1$
end while
NOTE: (1) $rand_u(a, b)$ is a uniform random number generator sampling in $[a, b]$.
    (2) $ceil(c)$ takes on the smallest integer larger than or equal to c.

scheme with the best achievable performance estimated over the entire parameter space. This helps to discover whether and when this new adaptive scheme can achieve or exceed the performance obtained using the best-calibrated parameter setting.

III. Experiments

We evaluate the performance of DE/rand/1/bin under each combination of 13 NP values (20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250 and 300), three CR values (0.1, 0.5 and 0.9) and two F values (0.5 and 0.8) on 28 numerical test functions in the CEC-2013 testbed at three problem dimensionality (10D, 30D and 50D). In addition to the performance measures described in the protocol of the CEC-2013 testbed [16], we also report the success rate, the expected running time to succeed (ERT) and the empirical cumulative distribution function (ECDF) of the number of function evaluations at success.

To reveal the intrinsic optimization ability of DE/rand/1/bin on the CEC-2013 testbed, we report, for each function at each of three problem dimensionality, the results corresponding to the best parameter setting (among 78 tested ones) that leads to the smallest average objective function error value (Section III-B) at execution termination over all execution runs. To provide a statistically reliable rule of thumb for the parameter choice, we perform parameter sensitive analysis using advanced non-parametric statistical tests to compare all 78 parameter settings over all 28 test functions at 10D, 30D and 50D, respectively. To comply with the protocol of the CEC-2013 testbed in regard to parameter configuration, we report the results corresponding to one parameter setting advocated by parameter sensitivity analysis.

A. CEC-2013 Testbed

The CEC-2013 testbed contains 28 numerical test functions of different characteristics, which are grouped into three categories: uni-modal functions (f_1-f_5), multi-modal functions (f_6-f_{20}) and composition functions (f_{21}-f_{28}). This testbed improves its CEC-2005 counterpart with additional test functions, oscillation and symmetric-breaking transforms as well as modification of the formula of composition functions. The complete description of this testbed is available in [16].

B. Experimental Setup

We test DE/rand/1/bin under 78 parameter settings ([NP, CR, F] $\in$ [20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300] $\times$ [0.1, 0.5, 0.9] $\times$ [0.5, 0.8]) on each of 28 test functions at three problem dimensionality (10D, 30D and 50D), respectively. These parameter settings include most of the parameter setting advices in previous works.

For each test function at each problem dimensionality, DE/rand/1/bin under each of 78 parameter settings are executed 51 times with each run using different random seeds while all parameter settings share the same random seed with respect to any individual run.

According to the protocol of the CEC-2013 testbed [16], two stopping criteria are applied: (1) the maximum number of function evaluations (maxFEvals) is reached where maxFEvals is set to 10^4 times problem dimensionality. (2) The difference of objective function values between the best solution found so far and the global optimal solution, so-called object function error value (FEV), is smaller than or equal to 10^{-8}. In such a case, **the FEV is set to** 10^{-8}. Here, we do not follow the CEC-2013 protocol to set the FEV to zero since this may deliberately reduce the average FEV at termination if only a few runs reach 10^{-8}.

We implement all algorithms in OCTAVE, and execute them on a Linux PC having the AMD Opteron 2376 CPU at 2.3 GHz.

C. Performance measures

The optimization performance is measured by (1) the best, worse, median and mean (standard deviation) of the FEVs achieved when the algorithm terminates over 51 runs; (2) the success rate (SR) over 51 runs. An optimization algorithm is regarded as succeeding in solving the problem once it achieves the FEV smaller than 10^{-8}; (3) the expected running time to succeed (ERT) [17]. This measure estimates the expected number of function evaluations to succeed at the first time, which is calculated as the total number of function evaluations when the algorithm succeeds or terminates (if not succeeding) summed over 51 runs and divided by the total number of successful runs. If the success rate equals zero, this measure becomes invalid.

Practical optimization tasks often impose demanding requirements on the computation time of optimizers, which is usually proportional to the number of executed function evaluations. To clearly inspect an optimization algorithm's efficacy with respect to various computation budgets, i.e., the maximally allowed number of function evaluations, the empirical cumulative distribution function (ECDF) [17] of the number of executed function evaluations at success over all 51 runs of all 28 test functions are illustrated at three problem dimensionality (10D, 30D and 50D), respectively.

The computational complexity is measured by CPU-seconds with respect to each problem dimensionality according to the protocol of the CEC-2013 testbed [16].

D. Results

Table I reports, for each test function at each dimensionality, the performance of DE/rand/1/bin using the best-calibrated parameter setting, among 78 configurations: [NP, CR, F] $\in$ [20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300] $\times$ [0.1, 0.5, 0.9] $\times$ [0.5, 0.8], which leads to the smallest average FEV at execution termination over 51 runs. It reflects the intrinsic optimization capability (represented by the best achieved performance over all tested parameter settings) of DE/rand/1/bin on the CEC-2013 testbed. It is observed that the performance of DE/rand/1/bin decreases as problem dimensionality increases. For 10D problems, DE/rand/1/bin achieves the non-zero SRs on 10 (f_1, f_2, f_4, f_5, f_6, f_7, f_{10}, f_{11}, f_{13} and f_{14}) out of 28 test functions among which the SR on seven problems (f_1, f_2, f_4, f_5, f_6, f_{11} and f_{14}) hits 1. On five (f_3, f_9, f_{12}, f_{16} and f_{19}) out of the remaining 18 test functions, it achieves FEVs less than 1.00e+00 in at least one execution run. For both 30D and 50D problems, DE/rand/1/bin achieves the non-zero SRs (equal to 1) on three (f_1, f_5 and f_{11}) out of 28 test functions. For the remaining 25 test functions, it achieves FEVs less than 1.00e+00 in at least one execution run on five functions (f_3, f_6, f_7, f_{10}, f_{14}) at 30D and four functions (f_3, f_7, f_{10}, f_{14}) at 50D.

TABLE IV. COMPUTATIONAL COMPLEXITY MEASURED BY CPU SECONDS ([16]) AT 10D, 30D AND 50D, RESPECTIVELY. T_0 MEASURES THE COMPUTATION TIME OF BASIC OPERATIONS. T_1 MEASURES THE COMPUTATION TIME OF ONE EXECUTION RUN ON TEST FUNCTION f_{14}. $\widehat{T}_2$ TAKES INTO ACCOUNT THE VARIATION OF T_1 IN DIFFERENT EXECUTION RUNS.

DIM	T_0	T_1	$\widehat{T}_2$	$(\widehat{T}_2 - T_1)/T_0$
$D = 10$	18.49	62.30	61.97	-0.02
$D = 30$		71.54	70.68	-0.05
$D = 50$		75.08	73.44	-0.09

Figures 1a, 1c and 1e illustrate the empirical cumulative distribution function (ECDF) of the number of executed function evaluations when the algorithm succeeds in reaching some pre-specified FEV divided by problem dimensionality, which is accumulated over all 51 runs of all CEC-2013 test functions. It is observed that, for all three problem dimensionality, given smaller termination FEVs, e.g., 10^{-1}, 10^{-4} and 10^{-8}, the proportion of successful execution runs starts to increase from zero after around $10^{2.5} \cdot D$ ($316 \cdot D$) function evaluations. Such increasing tendency is still fairly strong even when the maxFEvals ($10^4 \cdot D$) is reached, which implies that the larger maxFEvals might lead to the improved performance of DE/rand/1/bin.

E. Parameter sensitivity analysis

Choosing suitable parameter settings of DE/rand/1/bin for solving numerical or practical optimization problems in the black-box manner can be time-consuming. The trial-and-error scheme, as used to produce the results in Table III, may reveal the best achievable performance over the parameter space at the expense of computational resources. In many real-world applications, one single objective function evaluation may take seconds or minutes, far many times slower than the evaluation of any CEC-2013 function. As a result, one execution run on even a 10D problem may takes a couple of days or months, which makes the trial-and-error scheme infeasible for the parameter choice. One alternative, so-called fixed schemes for parameter configuration (Section II-B), is to carry out extensive empirical studies on a considerable number of test functions of diverse properties, and deduce some rules of thumb of choosing parameters that can consistently lead to

TABLE I. PERFORMANCE (PFM) OF DE/RAND/1/BIN USING THE BEST PARAMETER SETTING (AMONG 78 CONFIGURATIONS: [NP, CR, F] ∈ [20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300] × [0.1, 0.5, 0.9] × [0.5, 0.8]) THAT LEADS TO THE SMALLEST AVERAGE FEV AT EXECUTION TERMINATION OVER 51 RUNS WITH RESPECT TO EACH OF 28 CEC-2013 TEST FUNCTIONS AT PROBLEM DIMENSIONALITY (DIM) 10D, 30D AND 50D, RESPECTIVELY. **BEST**, **WORST**, **MEDIAN**, **MEAN (STD)** STAND FOR THE BEST, WORST, MEDIAN, MEAN (STANDARD DEVIATION) OF THE FEVS AT EXECUTION TERMINATION OVER 51 RUNS, RESPECTIVELY. **SR** AND **ERT** REPRESENT THE SUCCESS RATE AND THE EXPECTED RUNNING TIME TO SUCCEED, RESPECTIVELY. **ERT** IS DENOTED BY "-" (INVALID) WHEN **SR** EQUALS ZERO.

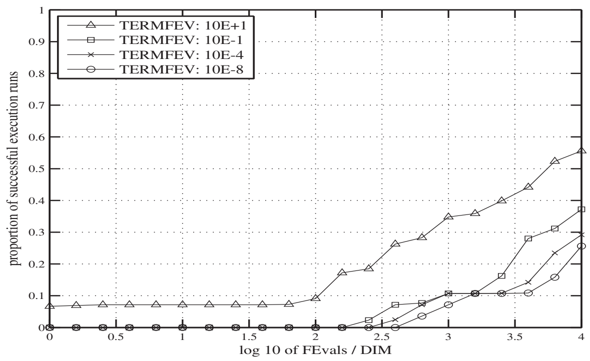
DIM	PFM	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	f12	f13	f14
10D	Best	1.00e-08	1.00e-08	3.51e-05	1.00e-08	1.00e-08	1.00e-08	1.00e-08	2.02e+01	1.47e-06	1.00e-08	1.00e-08	9.95e-01	1.00e-08	1.00e-08
	Worst	1.00e-08	1.00e-08	1.42e+00	1.00e-08	1.00e-08	1.00e-08	1.48e-04	2.05e+01	2.68e+00	1.18e-01	1.00e-08	2.19e+01	2.81e+01	1.00e-08
	Median	1.00e-08	1.00e-08	5.30e-02	1.00e-08	1.00e-08	1.00e-08	8.85e-07	2.03e+01	2.41e-01	4.92e-02	1.00e-08	6.96e+00	1.15e+01	1.00e-08
	Mean	1.00e-08	1.00e-08	1.06e-01	1.00e-08	1.00e-08	1.00e-08	1.47e-05	2.03e+01	5.05e-01	4.92e-02	1.00e-08	8.24e+00	1.20e+01	1.00e-08
	Std	0.00e-00	0.00e-00	2.25e-01	0.00e-00	0.00e-00	0.00e-00	3.05e-05	7.66e-02	6.05e-01	2.58e-02	0.00e-00	5.62e+00	6.01e+00	0.00e-00
	SR	1.00	1.00	0.00	1.00	1.00	1.00	0.10	0.00	0.00	0.06	1.00	0.00	0.02	1.00
	ERT	5.81e+03	6.76e+04	-	5.94e+04	7.38e+03	6.48e+04	1.02e+06	-	-	1.63e+06	1.15e+04	-	5.04e+06	8.76e+04
		f15	**f16**	**f17**	**f18**	**f19**	**f20**	**f21**	**f22**	**f23**	**f24**	**f25**	**f26**	**f27**	**f28**
	Best	2.90e+02	5.11e-01	1.45e+00	2.14e+01	1.27e-02	1.68e+00	1.67e+02	5.13e+00	4.17e+01	1.12e+02	1.41e+02	1.05e+02	3.00e+02	1.03e+02
	Worst	1.13e+03	1.19e+00	1.02e+01	3.55e+01	3.31e-01	3.50e+00	2.86e+02	4.56e+01	1.92e+03	1.94e+02	2.17e+02	2.00e+02	3.34e+02	3.16e+02
	Median	7.30e+02	9.05e-01	1.01e+01	2.83e+01	1.43e-01	2.28e+00	2.22e+02	1.47e+01	8.58e+02	1.41e+02	1.88e+02	1.23e+02	3.00e+02	1.06e+02
	Mean	7.25e+02	9.08e-01	8.57e+00	2.80e+01	1.68e-01	2.35e+00	2.25e+02	1.47e+01	8.62e+02	1.43e+02	1.89e+02	1.28e+02	3.01e+02	2.02e+02
	Std	1.80e+02	1.72e-01	2.91e+00	3.41e+00	8.86e-02	3.01e-01	2.31e+01	6.05e+00	3.99e+02	1.47e+01	1.61e+01	1.92e+01	4.77e+00	1.04e+02
	SR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	ERT	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DIM	**PFM**	**f1**	**f2**	**f3**	**f4**	**f5**	**f6**	**f7**	**f8**	**f9**	**f10**	**f11**	**f12**	**f13**	**f14**
30D	Best	1.00e-08	2.38e+04	4.21e-07	1.32e+01	1.00e-08	9.33e-02	3.22e-03	2.08e+01	4.49e+00	2.93e-08	1.00e-08	2.34e+01	3.50e+01	1.25e-01
	Worst	1.00e-08	1.84e+05	8.17e+00	1.69e+03	1.00e-08	2.64e+01	1.21e+00	2.10e+01	1.64e+01	1.23e-02	1.00e-08	9.95e+01	2.02e+02	2.91e-01
	Median	1.00e-08	7.60e+04	1.06e-03	9.44e+01	1.00e-08	6.19e+00	5.48e-02	2.09e+01	8.85e+00	2.54e-07	1.00e-08	3.78e+01	9.48e+01	2.29e-01
	Mean	1.00e-08	8.49e+04	1.78e-01	1.59e+02	1.00e-08	8.32e+00	1.50e-01	2.09e+01	9.17e+00	1.02e-03	1.00e-08	4.41e+01	1.02e+02	2.18e-01
	Std	0.00e-00	4.18e+04	1.14e+00	2.42e+02	0.00e-00	6.83e+00	2.50e-01	5.66e-02	2.24e+00	2.88e-03	0.00e-00	1.88e+01	3.96e+01	4.41e-02
	SR	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
	ERT	1.65e+04	-	-	-	2.15e+04	-	-	-	-	-	4.74e+04	-	-	-
		f15	**f16**	**f17**	**f18**	**f19**	**f20**	**f21**	**f22**	**f23**	**f24**	**f25**	**f26**	**f27**	**f28**
	Best	4.26e+03	1.49e+00	3.04e+01	1.20e+02	1.19e+00	8.94e+00	2.00e+02	1.90e+01	2.07e+03	2.00e+02	2.27e+02	2.00e+02	3.00e+02	3.00e+02
	Worst	6.41e+03	2.75e+00	3.04e+01	2.19e+02	2.34e+00	1.40e+01	3.00e+02	1.31e+02	7.71e+03	2.07e+02	2.52e+02	2.00e+02	5.76e+02	3.00e+02
	Median	5.62e+03	2.18e+00	3.04e+01	1.94e+02	1.83e+00	1.15e+01	2.04e+02	6.75e+01	6.53e+03	2.01e+02	2.39e+02	2.00e+02	3.02e+02	3.00e+02
	Mean	5.59e+03	2.19e+00	3.04e+01	1.89e+02	1.80e+00	1.15e+01	2.16e+02	7.26e+01	5.97e+03	2.01e+02	2.39e+02	2.00e+02	3.62e+02	3.00e+02
	Std	3.78e+02	2.95e-01	0.00e-00	2.41e+01	2.63e-01	7.97e-01	2.80e+01	4.08e+01	1.46e+03	1.31e+00	5.48e+00	1.49e-02	9.20e+01	0.00e-00
	SR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	ERT	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DIM	**PFM**	**f1**	**f2**	**f3**	**f4**	**f5**	**f6**	**f7**	**f8**	**f9**	**f10**	**f11**	**f12**	**f13**	**f14**
50D	Best	1.00e-08	2.07e+05	9.01e-01	3.23e+02	1.00e-08	4.34e+01	6.73e-02	2.10e+01	1.87e+01	1.93e-07	1.00e-08	6.18e+01	1.43e+02	4.00e-01
	Worst	1.00e-08	9.74e+05	1.99e+04	3.66e+03	1.00e-08	4.34e+01	9.09e+00	2.12e+01	7.11e+01	4.44e-02	1.00e-08	1.67e+02	4.41e+02	1.17e+01
	Median	1.00e-08	4.70e+05	7.27e+01	1.27e+03	1.00e-08	4.34e+01	9.96e-01	2.11e+01	2.62e+01	1.48e-02	1.00e-08	1.02e+02	2.95e+02	2.40e+00
	Mean	1.00e-08	4.72e+05	7.88e+02	1.38e+03	1.00e-08	4.34e+01	1.80e+00	2.11e+01	2.73e+01	1.70e-02	1.00e-08	1.08e+02	2.91e+02	2.80e+00
	Std	0.00e-00	1.63e+05	2.83e+03	8.14e+02	0.00e-00	0.00e+00	1.94e+00	4.15e-02	7.55e+00	1.14e-02	0.00e-00	2.59e+01	8.10e+01	1.94e+00
	SR	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
	ERT	2.65e+04	-	-	-	3.56e+04	-	-	-	-	-	2.73e+05	-	-	-
		f15	**f16**	**f17**	**f18**	**f19**	**f20**	**f21**	**f22**	**f23**	**f24**	**f25**	**f26**	**f27**	**f28**
	Best	1.17e+04	2.27e+00	5.08e+01	3.54e+02	3.40e+00	1.99e+01	2.00e+02	2.37e+01	4.32e+03	2.01e+02	2.62e+02	2.04e+02	3.09e+02	4.00e+02
	Worst	1.38e+04	3.72e+00	5.08e+01	4.18e+02	1.91e+01	2.35e+01	3.19e+02	7.01e+02	1.42e+04	2.56e+02	2.87e+02	2.24e+02	7.92e+02	4.00e+02
	Median	1.29e+04	3.23e+00	5.08e+01	3.99e+02	5.15e+00	2.12e+01	2.00e+02	5.61e+01	1.35e+04	2.02e+02	2.74e+02	2.11e+02	6.16e+02	4.00e+02
	Mean	1.28e+04	3.18e+00	5.08e+01	3.95e+02	5.89e+00	2.13e+01	2.13e+02	9.85e+01	1.27e+04	2.14e+02	2.74e+02	2.11e+02	6.02e+02	4.00e+02
	Std	4.75e+02	3.23e-01	2.09e-03	1.51e+01	2.73e+00	6.31e-01	3.13e+01	1.12e+02	2.24e+03	1.70e+01	5.37e+00	3.23e+00	1.09e+02	0.00e-00
	SR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	ERT	-	-	-	-	-	-	-	-	-	-	-	-	-	-

TABLE II. PARAMETER SENSITIVITY ANALYSIS RESULTS BY USING THE IMAN AND DAVENPORT TEST WITH THE HOCHBERG POST-HOC PROCEDURE ON 28 CEC-2013 TEST FUNCTIONS AT PROBLEM DIMENSIONALITY 10D, 30D AND 50D, RESPECTIVELY. AMONG 78 TESTED PARAMETER SETTINGS ([NP, CR, F] ∈ [20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300] × [0.1, 0.5, 0.9] × [0.5, 0.8]), THOSE LEADING TO THE STATISTICALLY SIGNIFICANTLY BETTER PERFORMANCE (AT THE SIGNIFICANCE LEVEL 0.05) OVER OTHERS WITH RESPECT TO 10D, 30D AND 50D PROBLEMS ARE DENOTED BY 10, 30 AND 50 AND ORDERLY SEPARATED BY "/" IN THEIR CORRESPONDING TABLE CELLS. "-" MEANS THE CORRESPONDING PARAMETER SETTING IS STATISTICALLY SIGNIFICANTLY WORSE THAN SOME OTHER PARAMETER SETTINGS FOR SOLVING ALL 28 TEST FUNCTIONS AT CERTAIN PROBLEM DIMENSIONALITY.

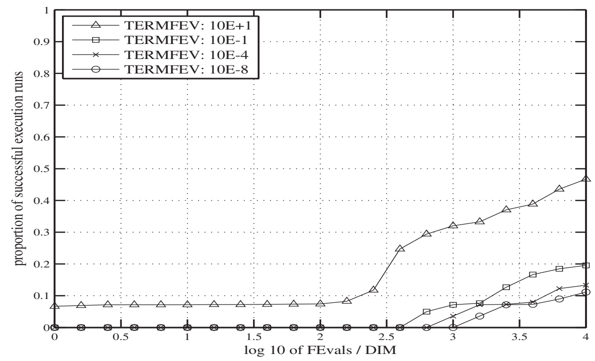
Parameter		Population Size												
CR	**F**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**150**	**200**	**250**	**300**
0.1	**0.5**	-/30/50	-/30/-	-/30/-	-/30/-	-/30/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-
	0.8	-/30/-	-/30/-	-/30/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-
0.5	**0.5**	10/-/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-
	0.8	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-
0.9	**0.5**	-/-/-	-/-/-	10/30/50	10/30/50	10/30/50	-/30/-	-/30/-	-/30/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-
	0.8	-/30/-	-/30/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-

the satisfactory performance on most of test functions. Since consideration was given to a wide range of problems, the generally good performance can be expected by applying such derived rules of thumb to solve unseen problems.

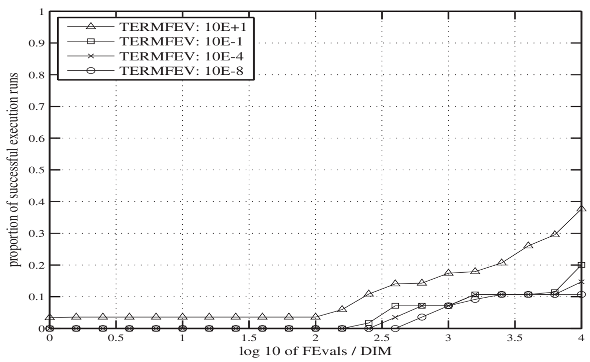
Although previous works had already given many rules of thumb for choosing the parameters of DE/rand/1/bin [1]–[3], most of them were not derived in a statistically rigorous way. In this paper, we employ advanced non-parametric statistical



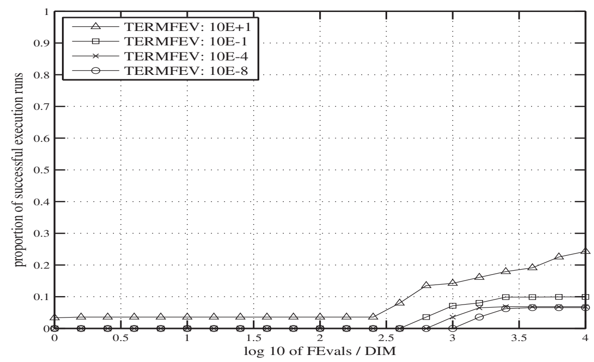
(a) Problem dimensionality: 10D



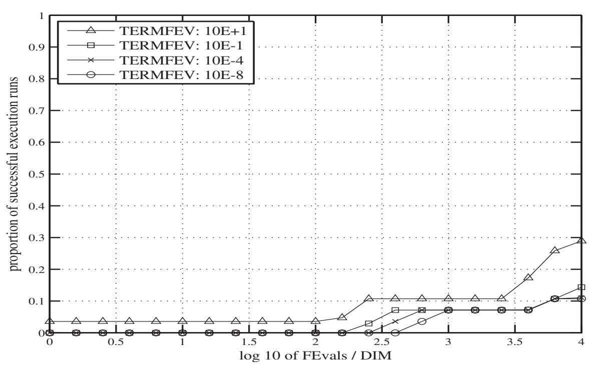
(b) Problem dimensionality: 10D



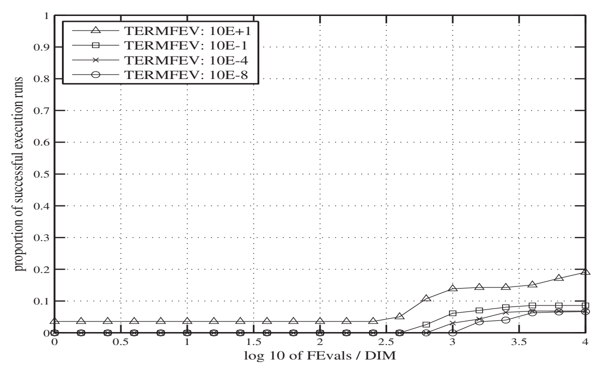
(c) Problem dimensionality: 30D



(d) Problem dimensionality: 30D



(e) Problem dimensionality: 50D



(f) Problem dimensionality: 50D

Fig. 1. Empirical cumulative distribution function (ECDF) of the number of executed function evaluations (FEvals) when the algorithm succeeds in reaching certain function error values (denoted in the legend as TERMFEV: 10e+1, 10e-1, 10e-4 and 10e-8) divided by problem dimensionality, which is accumulated over all 51 execution runs of 28 CEC-2013 test functions at problem dimensionality 10D, 30D and 50D, respectively. **(a)(c)(e)**: DE/rand/1/bin using the best parameter setting (among 78 tested cases: [NP, CR, F] $\in$ [20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300] $\times$ [0.1, 0.5, 0.9] $\times$ [0.5, 0.8]) that leads to the smallest mean FEV at execution termination over 51 runs with respect to each of 28 test functions. **(b)(d)(f)**: DE/rand/1/bin using one superior parameter setting (NP:50, CR:0.9, F:0.5) advocated by parameter sensitivity analysis (Section III-D) on 28 test functions.

tests to conduct parameter sensitivity analysis in order to discover some parameter settings (out of a wide range of potentially effective parameter settings) that can lead to the statistically significantly better performance than the other settings. Specifically, we apply the Iman and Davenport test [14], [15] to compare 78 parameter settings over all 28 test functions to determine whether at least two parameter settings out of 78 tested cases can lead to the statistically significantly different (at the significance level 0.05) performance. If so, the Hochberg post-hoc procedure [14], [15] is used to demarcate statistically significantly superior parameter settings.

The Iman and Davenport test is a derivative of the famous Friedman two-way analysis of variances by ranks [18], which is a statistical hypothesis testing method to detecting the existence of the significant difference between the behavior of two or more algorithms. It is a non-parametric test, which converts original results to ranks before calculating the test statistic. For example, given several algorithms in comparison, their numerical performance indices will be first ranked with respect to each test problem where smaller ranks indicate the better performance. Then, the average rank over all test problems will be used to calculate the test statistic. Finally, the calculated test statistic will be converted to the p-value

TABLE III. PERFORMANCE (PFM) OF DE/RAND/1/BIN USING ONE SUPERIOR PARAMETER SETTING (NP:50, CR:0.9, F:0.5) ADVOCATED BY PARAMETER SENSITIVITY ANALYSIS (SECTION III-D) ON ALL 28 CEC-2013 TEST FUNCTIONS AT PROBLEM DIMENSIONALITY (DIM) 10D, 30D AND 50D, RESPECTIVELY. **BEST**, **WORST**, **MEDIAN**, **MEAN** (**STD**) STAND FOR THE BEST, WORST, MEDIAN, MEAN (STANDARD DEVIATION) OF THE FEVS AT EXECUTION TERMINATION OVER 51 RUNS, RESPECTIVELY. **SR** AND **ERT** REPRESENT THE SUCCESS RATE AND THE EXPECTED RUNNING TIME TO SUCCEED. **ERT** IS DENOTED BY "-" (INVALID) WHEN **SR** EQUALS ZERO.

DIM	PFM	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	f12	f13	f14
10D	Best	1.00e-08	1.52e-03	1.00e-08	2.72e-07	1.00e-08	7.59e-03	1.00e-08	2.02e+01	1.00e-08	1.00e-08	1.00e-08	9.95e-01	1.99e+00	3.66e+00
	Worse	1.00e-08	3.06e+04	1.09e+01	5.48e+02	1.00e-08	9.81e+00	6.80e-02	2.05e+01	3.59e+00	1.18e-01	1.09e+01	2.19e+01	3.12e+01	1.12e+03
	Median	1.00e-08	2.85e+02	3.11e-01	1.92e+00	1.00e-08	7.10e-01	8.41e-06	2.04e+01	1.03e+00	4.92e-02	9.95e-01	6.96e+00	1.18e+01	7.48e+01
	Mean	1.00e-08	2.42e+03	1.41e+00	2.71e+01	1.00e-08	3.29e+00	1.44e-03	2.04e+01	1.14e+00	4.92e-02	1.14e+00	8.24e+00	1.21e+01	2.20e+02
	Std	0.00e-00	5.27e+03	2.51e+00	8.34e+01	0.00e-00	4.30e+00	9.50e-03	6.64e-02	9.90e-01	2.58e-02	2.13e+00	5.62e+00	6.31e+00	2.80e+02
	SR	1.00	0.00	0.16	0.00	1.00	0.00	0.24	0.00	0.22	0.06	0.45	0.00	0.00	0.00
	ERT	1.38e+04	-	5.96e+05	-	1.74e+04	-	4.05e+05	-	4.27e+05	1.63e+06	1.89e+05	-	-	-
		f15	**f16**	**f17**	**f18**	**f19**	**f20**	**f21**	**f22**	**f23**	**f24**	**f25**	**f26**	**f27**	**f28**
	Best	2.83e+02	3.73e-01	6.68e+00	2.03e+01	4.98e-01	1.74e+00	2.00e+02	1.84e+01	1.10e+02	1.03e+02	2.00e+02	1.02e+02	3.00e+02	1.00e+02
	Worse	1.49e+03	1.54e+00	2.71e+01	4.36e+01	2.21e+00	3.47e+00	4.00e+02	8.69e+02	1.61e+03	2.11e+02	2.10e+02	3.08e+02	5.20e+02	3.00e+02
	Median	1.19e+03	1.03e+00	1.67e+01	3.12e+01	9.74e-01	2.33e+00	4.00e+02	1.16e+02	1.01e+03	2.05e+02	2.00e+02	2.00e+02	3.00e+02	3.00e+02
	Mean	1.13e+03	1.01e+00	1.78e+01	3.16e+01	1.07e+00	2.36e+00	3.73e+02	2.23e+02	9.77e+02	2.02e+02	2.02e+02	1.67e+02	3.37e+02	2.92e+02
	Std	2.69e+02	2.42e-01	5.36e+00	5.58e+00	4.98e-01	3.58e-01	6.96e+01	2.36e+02	3.34e+02	1.46e+01	2.76e+00	4.86e+01	7.58e+01	3.92e+01
	SR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	ERT	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DIM	**PFM**	**f1**	**f2**	**f3**	**f4**	**f5**	**f6**	**f7**	**f8**	**f9**	**f10**	**f11**	**f12**	**f13**	**f14**
30D	Best	1.00e-08	4.56e+04	3.04e+00	4.42e+01	1.00e-08	1.37e+01	7.71e-02	2.08e+01	4.49e+00	1.97e-02	3.98e+00	1.69e+01	3.30e+01	5.99e+01
	Worst	1.00e-08	3.31e+05	2.44e+07	1.94e+03	5.99e-04	7.06e+01	7.27e+00	2.10e+01	1.64e+01	2.37e-01	2.49e+01	1.85e+02	1.84e+02	3.28e+03
	Median	1.00e-08	1.46e+05	4.96e+05	3.46e+02	1.00e-08	1.41e+01	1.17e+00	2.10e+01	8.85e+00	6.41e-02	1.39e+01	1.57e+02	1.65e+02	4.79e+02
	Mean	1.00e-08	1.54e+05	3.65e+06	4.62e+02	3.09e-05	1.98e+01	1.58e+00	2.09e+01	9.17e+00	7.62e-02	1.42e+01	1.14e+02	1.53e+02	5.72e+02
	Std	0.00e-00	6.58e+04	5.84e+06	4.15e+02	1.17e-04	1.37e+01	1.60e+00	4.99e-02	2.24e+00	4.92e-02	4.56e+00	6.69e+01	3.90e+01	5.46e+02
	SR	1.00	0.00	0.00	0.00	0.84	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	ERT	3.94e+04	-	-	-	1.23e+05	-	-	-	-	-	-	-	-	-
		f15	**f16**	**f17**	**f18**	**f19**	**f20**	**f21**	**f22**	**f23**	**f24**	**f25**	**f26**	**f27**	**f28**
	Best	4.39e+03	1.28e+00	4.08e+01	1.73e+02	1.33e+00	1.09e+01	2.00e+02	1.44e+02	6.56e+03	2.01e+02	2.39e+02	2.00e+02	3.13e+02	3.00e+02
	Worst	7.58e+03	2.88e+00	1.55e+02	2.14e+02	1.19e+01	1.26e+01	4.44e+02	9.65e+02	7.69e+03	2.44e+02	2.61e+02	3.39e+02	6.53e+02	3.00e+02
	Median	7.11e+03	2.52e+00	5.20e+01	2.00e+02	2.76e+00	1.19e+01	3.00e+02	4.23e+02	7.12e+03	2.16e+02	2.48e+02	2.00e+02	5.09e+02	3.00e+02
	Mean	7.01e+03	2.45e+00	5.62e+01	1.99e+02	3.93e+00	1.19e+01	3.07e+02	4.44e+02	7.11e+03	2.17e+02	2.48e+02	2.37e+02	4.95e+02	3.00e+02
	Std	5.42e+02	3.02e-01	1.96e+01	9.87e+00	2.87e+00	3.41e-01	7.92e+01	2.01e+02	2.64e+02	1.09e+01	4.38e+00	5.52e+01	9.30e+01	0.00e-00
	SR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	ERT	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DIM	**PFM**	**f1**	**f2**	**f3**	**f4**	**f5**	**f6**	**f7**	**f8**	**f9**	**f10**	**f11**	**f12**	**f13**	**f14**
50D	Best	1.00e-08	1.30e+05	5.13e+05	7.22e+02	1.00e-08	4.34e+01	2.32e+00	2.10e+01	1.48e+01	4.19e-02	1.89e+01	4.48e+01	3.14e+02	3.45e+02
	Worst	1.00e-08	1.16e+06	7.63e+07	5.17e+03	3.07e-03	8.56e+01	2.73e+01	2.12e+01	7.35e+01	2.88e-01	6.57e+01	3.70e+02	3.90e+02	2.38e+03
	Median	1.00e-08	4.55e+05	7.75e+06	1.71e+03	1.00e-08	4.34e+01	1.29e+01	2.11e+01	2.32e+01	1.21e-01	3.58e+01	3.44e+02	3.52e+02	1.17e+03
	Mean	1.00e-08	4.91e+05	1.97e+07	1.83e+03	1.34e-04	4.46e+01	1.24e+01	2.11e+01	2.94e+01	1.37e-01	3.64e+01	2.94e+02	3.50e+02	1.25e+03
	Std	0.00e-00	1.89e+05	2.30e+07	8.34e+02	5.52e-04	5.93e+00	5.80e+00	4.11e-02	1.67e+01	6.44e-02	9.09e+00	1.12e+02	1.34e+01	4.82e+02
	SR	1.00	0.00	0.00	0.00	0.86	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	ERT	6.92e+04	-	-	-	2.33e+05	-	-	-	-	-	-	-	-	-
		f15	**f16**	**f17**	**f18**	**f19**	**f20**	**f21**	**f22**	**f23**	**f24**	**f25**	**f26**	**f27**	**f28**
	Best	1.24e+04	2.18e+00	7.50e+01	3.60e+02	3.40e+00	2.11e+01	2.00e+02	6.27e+02	1.20e+04	2.23e+02	2.71e+02	2.00e+02	5.46e+02	4.00e+02
	Worst	1.45e+04	3.76e+00	1.26e+02	4.29e+02	1.91e+01	2.24e+01	1.12e+03	2.28e+03	1.45e+04	2.74e+02	3.11e+02	3.68e+02	9.94e+02	3.40e+03
	Median	1.38e+04	3.35e+00	9.45e+01	3.97e+02	5.15e+00	2.19e+01	8.36e+02	1.12e+03	1.37e+04	2.53e+02	2.92e+02	3.49e+02	8.39e+02	4.00e+02
	Mean	1.38e+04	3.25e+00	9.52e+01	3.95e+02	5.89e+00	2.18e+01	7.05e+02	1.23e+03	1.36e+04	2.50e+02	2.92e+02	3.21e+02	8.34e+02	5.16e+02
	Std	3.75e+02	3.33e-01	1.15e+01	1.31e+01	2.73e+00	2.72e-01	4.23e+02	4.20e+02	4.45e+02	1.26e+01	9.50e+00	6.10e+01	8.68e+01	5.82e+02
	SR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	ERT	-	-	-	-	-	-	-	-	-	-	-	-	-	-

according to the F distribution. If the obtained p-value is smaller or equal to a pre-specified significance level (e.g., 0.05 used in our experiments), we can draw a conclusion that the significant difference exists between the performance of two or more algorithms in comparison. However, the Iman and Davenport test cannot discover which algorithms are significantly different from others. To resolve this issue, post-hoc procedures can be applied, which perform pairwise comparisons under the control of the family-wise error rate (FWER) [14]. Here, we choose the Hochberg procedure, which had been recommended in [15] for demarcating algorithms having the statistically significantly better performance than others. This post-hoc procedure first determines the control method as the algorithm with the lowest rank (algorithm's ranks are obtained from the Iman and Davenport test in our work). Then, the test statistic is calculated for each algorithm except for the control method based on the ranks of that algorithm and the control method, and converted to the p-value according to the standard normal distribution. After that, the obtained p-values are sorted. Finally, in the descending order, each p-value is compared with an adjusted significant level (for controlling the FWER). Once a p-value is smaller or equal to its adjusted significant level, the comparison is terminated. Those algorithms with p-values larger than this p-value that terminates the comparison together with the control method are claimed to be statistically significantly better than other algorithms. The detailed description of the Iman and Davenport test and the Hochberg post-hoc procedure is available in [14], [15], [18] .

The power and reliability of the above-described statistic tests can be influenced by the number of compared algorithms and the number of tested problems. It was suggested that the number of tested problems should be both not smaller than two times and not larger than eight times the number of compared algorithms [15]. Since we are investigating a population-based stochastic algorithm, different random seeds may cause the population to explore different parts of the solution space and consequently result in the distinct performance even when optimizing the same problem. Therefore, it is reasonable to treat applying different random seeds to the same test function as different test problems. Having considered this, for each test function at any tested problem dimensionality, we choose 10 execution runs corresponding to 10 distinct random seeds to serve as different test problems, which lead to 280 test problems in total for each tested problem dimensionality. This makes the number of test problems around 3.5 times the number of tested parameter settings, which comply with the suggestions in [15].

Table II reports the results of parameter sensitivity analysis. It can be observed that medium NP values (e.g., 40, 50 and 60), large CR values (e.g., 0.9) and medium F values (e.g., 0.5) can lead to the statistically significantly better performance than the other parameter settings at any tested problem dimensionality.

To comply with the protocol of the CEC-2013 testbed [16] about using a single parameter setting to conduct all experiments, Table III reports the performance of DE/rand/1/bin using the parameter setting (NP:50, CR:0.9, F:0.5) advocated by the above parameter sensitivity analysis. The computational complexity and ECDF corresponding to such a parameter setting are reported in Table IV and illustrated in Figures 1b, 1d and 1f, respectively.

IV. Conclusions

This work systematically benchmarks a classic DE algorithm (DE/rand/1/bin) on the CEC-2013 single-objective continuous optimization testbed. We report, for each of 28 CEC-2013 test functions at different problem dimensionality (10D, 30D and 50D), the best achieved performance among 78 potentially effective parameter settings ([NP, CR, F] $\in$ [20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300] $\times$ [0.1, 0.5, 0.9] $\times$ [0.5, 0.8]). It reflects the intrinsic optimization capability of DE/rand/1/bin on this testbed and provides a handy reference for future research using this testbed by choosing DE/rand/bin as a baseline method for performance comparison. Furthermore, we conduct parameter sensitivity analysis using the Iman and Davenport test followed by the Hochberg procedure to compare all 78 parameter settings over all 28 test functions at 10D, 30D and 50D, respectively. The results indicate that medium NP values (e.g., 40, 50 and 60), large CR values (e.g., 0.9) and medium F values (e.g., 0.5) can lead to the statistically significantly better performance than the other parameter settings at any tested problem dimensionality. It provides a statistically reliable rule of thumb for choosing the parameters of DE/rand/1/bin for solving unseen problems. Moreover, to comply with the protocol of the CEC-2013 testbed about using a single parameter setting to conduct all experiments, we report the performance of DE/rand/1/bin using one of the superior parameter settings (NP:50, CR:0.9, F:0.5) advocated by parameter sensitivity analysis.

Acknowledgment

This work was supported by NSFC under Grant No. 61005051, SRFDP under Grant No. 20100092120027 and ARC Discovery Grant (DP120102205). We would like to thank Kenneth V. Price for his valuable comments and suggestions on improving the quality of this paper.

References

[1] R. M. Storn and K. V. Price, "Differential evolution - a simple and efficient adaptive scheme for global optimization over continuous spaces," International Computer Science Institute, Berkeley, CA, USA, ICSI Technical Report 95-012, 1995.

[2] K. V. Price, R. M. Storn, and J. A. Lampinen, *Differential Evolution - A Practical Approach to Global Optimization*. Springer, 2005.

[3] S. Das and P. N. Suganthan, "Differential evolution: A survey of the state-of-the-art," *IEEE Transactions on Evolutionary Computation*, vol. 15, no. 1, pp. 4–31, Feb. 2011.

[4] H. A. Abbass, "The self-adaptive pareto differential evolution algorithm," in *Proc.of the 2002 IEEE Congress on Evolutionary Computation (CEC'02)*, vol. 1, Honolulu, Hawaii, USA, 2002, pp. 831–836.

[5] J. Liu and J. A. Lampinen, "A fuzzy adaptive differential evolution algorithm," *Soft Computing*, vol. 9, no. 6, pp. 448–462, Jun. 2004.

[6] S. Das, A. Konar, and U. K. Chakraborty, "Two improved differential evolution schemes for faster global search," in *Proc. of the 2005 conference on Genetic and evolutionary computation (GECCO'05)*. ACM, 2005, pp. 991–998.

[7] M. Omran, A. Salman, and A. Engelbrecht, "Self-adaptive differential evolution," *Computational intelligence and security*, pp. 192–199, 2005.

[8] J. Brest, S. Greiner, B. Bošković, M. Mernik, and V. Žumer, "Self-adapting control parameters in differential evolution: A comparative study on numerical benchmark problems," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 6, pp. 646–657, Dec. 2006.

[9] A. K. Qin, V. L. Huang, and P. N. Suganthan, "Differential evolution algorithm with strategy adaptation for global numerical optimization," *IEEE Transactions on Evolutionary Computation*, vol. 13, no. 2, pp. 398–417, 2009.

[10] Z. Yang, K. Tang, and X. Yao, "Scalability of generalized adaptive differential evolution for large-scale continuous optimization," *Soft Computing*, vol. 15, no. 11, pp. 2141–2155, Sep. 2010.

[11] Y. Wang, Z. Cai, and Q. Zhang, "Differential evolution with composite trial vector generation strategies and control parameters," *IEEE Transactions on Evolutionary Computation*, vol. 15, no. 1, pp. 55–66, Feb. 2011.

[12] R. Mallipeddi, P. N. Suganthan, Q. K. Pan, and M. F. Tasgetiren, "Differential evolution algorithm with ensemble of parameters and mutation strategies," *Applied Soft Computing*, vol. 11, no. 2, pp. 1679–1696, Mar. 2011.

[13] D. Zaharie, "Critical values for the control parameters of differential evolution algorithms," in *Proc. of MENDEL 2002: the 8th International Conference on Soft Computing*, Jun. 2002, pp. 62–67.

[14] S. García, D. Molina, M. Lozano, and F. Herrera, "A study on the use of non-parametric tests for analyzing the evolutionary algorithms behaviour: a case study on the cec2005 special session on real parameter optimization," *Journal of Heuristics*, vol. 15, no. 6, pp. 617–644, 2009.

[15] 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 and Evolutionary Computation*, vol. 1, no. 1, pp. 3–18, 2011.

[16] J. J. Liang, B.-Y. Qu, P. N. Suganthan, and A. G. Hernández-Díaz, "Problem definitions and evaluation criteria for the cec 2013 special session on real-parameter optimization," Computational Intelligence Laboratory, Zhengzhou University, Zhengzhou, China and Nanyang Technological University, Singapore, Technical Report 201212, 2013.

[17] N. Hansen, A. Auger, S. Finck, and R. Ros, "Real-parameter black-box optimization benchmarking: Experimental setup," INRIA, France, Technical Report, 2013.

[18] D. Sheskin, *Handbook of parametric and nonparametric statistical procedures*. CRC PressI Llc, 2004.