

Benchmarking of the Indicator-based evolution algorithm using the Bi-Objective BBOB Test Suite

Final report ^{*}

Daro Heng

Mihaela Sorostinean

Karim Kouki

Aris Tritas

Ahmed Mazari

ABSTRACT

The objective of this project was to study, implement and benchmark the Indicator Based Evolutionary Algorithm (IBEA) using the Comparing Continuous Optimizer (COCO) platform. Firstly we provide a brief overview of the algorithm. Secondly, we describe our implementation and experimental setup. Finally, we discuss the obtained results and compare them to both baseline approaches and related work.

1. SETTING

In the context of a multi-objective optimization, the main goal is to find a good approximation of the set of Pareto-optimal solutions. An evolutionary algorithm is an exploration strategy of the domain space \mathbb{R}^n which seeks optimal solution vectors defined in the objective space \mathbb{R}^k (here $k = 2$) by promoting at each generation the fittest and/or the newest individuals from a population of decision vectors.

The performance measure used by IBEA for determining the relative quality of two solution sets in the objective space is a binary indicator function $I : \Omega \times \Omega \rightarrow \mathbb{R}$. In terms of two decision vectors \mathbf{x}^1 and \mathbf{x}^2 , the domination relation is defined as : $\mathbf{x}^1 > \mathbf{x}^2 \Leftrightarrow (f_i(\mathbf{x}^1) \leq f_i(\mathbf{x}^2) \forall i \in \{1, \dots, n\} \text{ and } f_j(\mathbf{x}^1) < f_j(\mathbf{x}^2) \text{ for at least one objective})$. The epsilon indicator, which is compliant with this domination relationship, is defined for pairwise comparisons as:

$$I_{\epsilon+}(\{\mathbf{x}^1\}, \{\mathbf{x}^2\}) = \min_{\epsilon} f_i(\mathbf{x}^1) - \epsilon \leq f_i(\mathbf{x}^2) \forall i \in \{1, \dots, n\}$$

. Its focus is to summarize the minimum distance needed to improve on the Pareto set approximation to a single scalar. Intuitively, the reduction to a single scalar may incur a loss of information regarding some dimensions of objective space. Indeed, the fact that the indicator chooses the minimum improvement across all objectives implies *conservative* rather than more *optimistic* updates of the fitness estimate. Clearly, we may miss out on potential improvement on some

of the targeted objectives. The fitness value is defined in the sequel as a measure of the usefulness of each individual with regards to the optimization goal. As such, the algorithm tries to maximize it.

$$F(\mathbf{x}^1) = \sum_{\mathbf{x}^2 \in P \setminus \{\mathbf{x}^1\}} -\exp \frac{I(\{\mathbf{x}^1\}, \{\mathbf{x}^2\})}{\kappa}$$

where P is the population set, and $\mathbf{x} \in \mathbb{R}^n$. The fitness function of an approximation set is defined as a *dominance preserving relation*.

Moreover, as the optimization functions are typically unnormalized, IBEA scales both the values taken by the objective function as well as the values taken by the indicator function. Consequently the need for parameter tuning in face of problem and indicator function diversity is decreased.

2. IMPLEMENTATION

The input of the algorithm is the size of the population (α), a maximum number of generations, a budget in function evaluations, and a fitness scaling factor (κ). The output of the algorithm is an approximation of the Pareto-set.

1. **Initialization:** generate an initial population P of given size uniformly between the lower and upper bound of the domain space.
2. **Fitness assignment:** compute and assign a fitness value to each individual in P .
3. **Environmental selection:** detect and remove individuals which have the smallest fitness values from the population until the current size of the population P does not exceed α . Update the fitness values for all remaining individuals.
4. **Termination:** after the environmental selection is performed, the termination criterion of the algorithm is checked; if the maximum number of generations is reached or another termination criterion is met, the algorithm returns the set of decision vectors A .
5. **Mating Selection:** consists in creating a temporary mating pool P' which is filled with individuals from P by performing binary tournament selection with replacement on P .
6. **Variation:** finally the variation step consists applying recombination and mutation operators to the previously created mating pool. The offspring resulting from variation is added to P and the generation counter

^{*}Submission deadline: October 21st.

is incremented. The algorithm is then performed again from step 2, until a termination criterion is met.

Population-related data (i.e decision vectors, objective and fitness values) is stored in a hash table for timely access. All numerical computation is done with NumPy. The optimizer is an object which incurs some overhead in Python. Except for the recombination and mutation operators which were implemented in separate functions to achieve modularity during testing, the optimizer is essentially a single inlined function.

2.1 Recombination

At first, the crossover operators we experimented with were intermediate weighting and discrete recombination. Then we implemented the Simulated Binary Crossover [2]. The results reported here used SBX-5 for recombination.

2.2 Variation

For the mutation step, we began by simply adding isotropic Gaussian noise with fixed variance to the produced offspring. However, it is well known that fixed variance does not produce the fastest target values search.

Budget pitfall

Before the intermediate report we used an extremely small budget, which had the effect of delaying realistic benchmarking.

3. EXPERIMENTAL SETUP

Our idea was to evaluate the influence of hyper-parameters on the performance achieved for different function groups and in different number of dimensions. For a reasonably low budget (10^3) and 10^6 to 10^9 number of runs, we tried the following ranges for hyper-parameters (best values in bold):

- Population size $\in \{50, 80, \mathbf{100}, 150, 200\}$
- Number of offspring $\in \{20, 30, 35, 40, \mathbf{50}\}$
- Mutation probability: **low (0.1)**, high (>0.7)
- Crossover probability $\in [0.5, 1.0]$ (best was **0.7**)
- Initial mutation step-size $\in \{1.0, 2.5, \mathbf{5.0}, 10, 15\}$
- Distribution index of the SBX operator: 2 and 5

Our focus was mostly on crossover and mutation probabilities, as well as population and offspring size. Naturally, the total runtime of an experiment depends on the size of the population, the dimension of the problem and the budget. Carrying out experiments with all their combinations would be time consuming. As such, we set for a grid search approach: we fixed all parameters except one to reasonable defaults and tuned the last one by picking values from a range of interest and comparing the output of the test suite.

4. CPU TIMING

In order to evaluate the CPU timing of the algorithm, we have run the IBEA with restarts on the entire bbob-biobj test suite [5] for $10D$ function evaluations. The Python code was run on a Intel(R) Core(TM) i7 CPU. For a population size of 100, the time per function evaluation ranges between $2.2ms$ and $6.5ms$ depending on whether we use isotropic mutation or not.

5. RESULTS

Results of IBEA from experiments according to [4] and [1] on the benchmark functions given in [5] are presented in Figures 1, 2. The experiments were performed with COCO [3], version 15.4.

Presented here are comparative ECDFs on $2D$ with two baseline algorithms: Random Search and NSGA-II. Our best configuration is also compared with the implementation of IBEA- ϵ in the C language and the implementation of IBEA with Hypervolume indicator in Python.

Remark

In the aRT tables, IBEA corresponds to the following order: IBEA- ϵ in C, IBEA- ϵ in Python and IBEA-HV in Python.

5.1 Observations

In low dimension, we notice that the instances in which our algorithm performs moderately well is the **separable**, **moderate** and **weakly-structured** case. It fails however in the face of multimodal and ill-conditioned functions.

Generally speaking, derandomized step-size adaptation performs better than isotropic mutation. Moreover, extreme values of variance for the isotropic mutation negatively impact the performance of the algorithm.

In higher dimensions, the trends are the same. Clearly the supplementary degrees of freedom decrease the number of targets reached.

5.2 Comparison of isotropic and derandomized mutation operators

With no surprise that our algorithm has a runtime of $1e-3$ iff the variance is adapted. gets to 10^{-3} on f12

5.3 Related Work

To an extent the comparison as well as results of groups studying IBEA-HV is quite revealing. We can safely say that the ϵ indicator may be overly simplistic for the wide range of optimization functions benchmarked. We also observe that using a better indicator tends to find targets much faster.

Conclusion

Going forward, we may think that a more advanced step-size adaptation scheme such as CMA, and also importantly a different indicator function may function well for different problems.

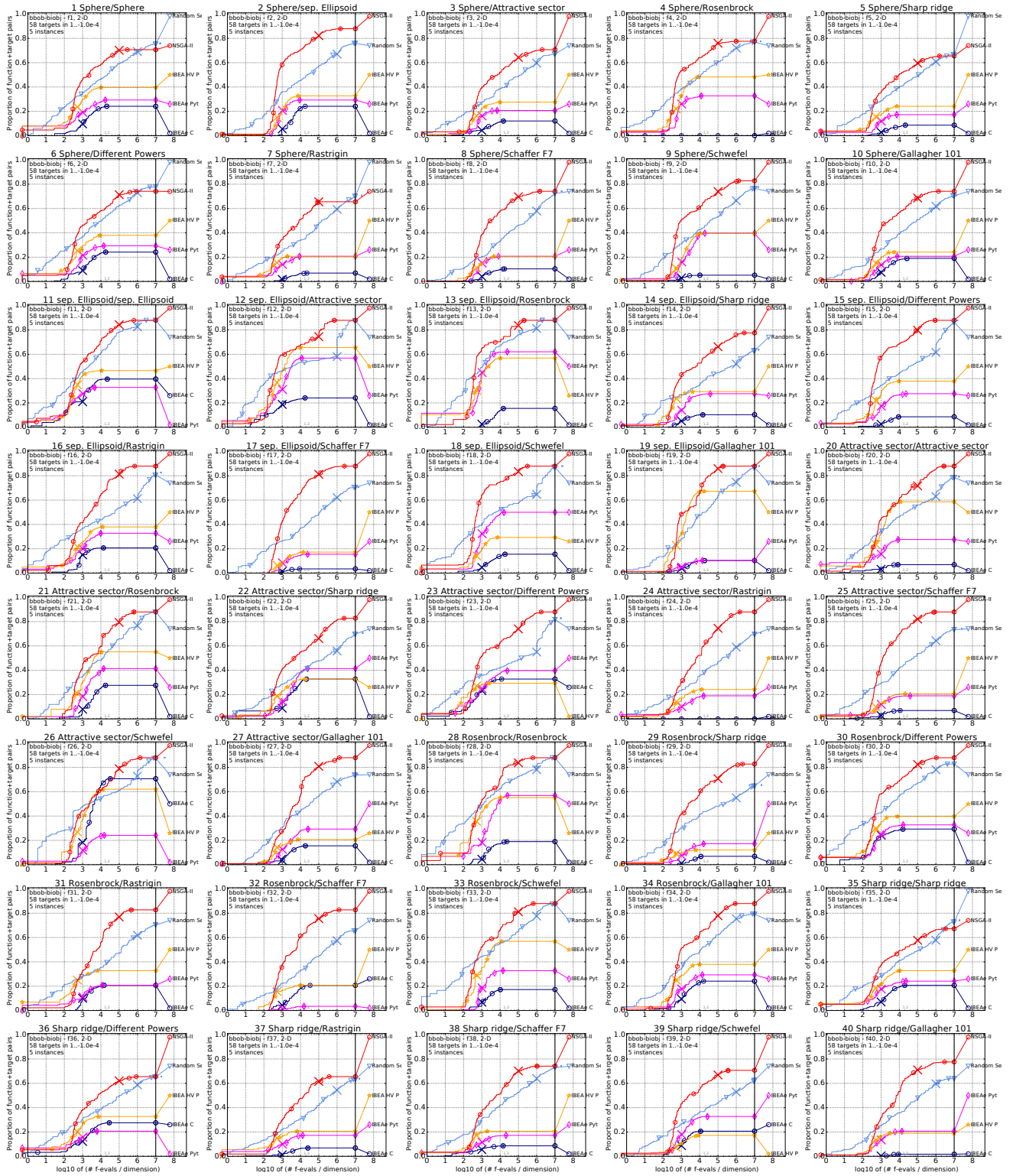


Figure 1: Bootstrapped empirical cumulative distribution of the number of objective function evaluations divided by dimension (FEvals/DIM) for 58 targets with target precision in $\{-10^{-4}, -10^{-4.2}, -10^{-4.4}, -10^{-4.6}, -10^{-4.8}, -10^{-5}, 0, 10^{-5}, 10^{-4.9}, 10^{-4.8}, \dots, 10^{-0.1}, 10^0\}$ for each single function f_1 to f_{40} in 10-D.

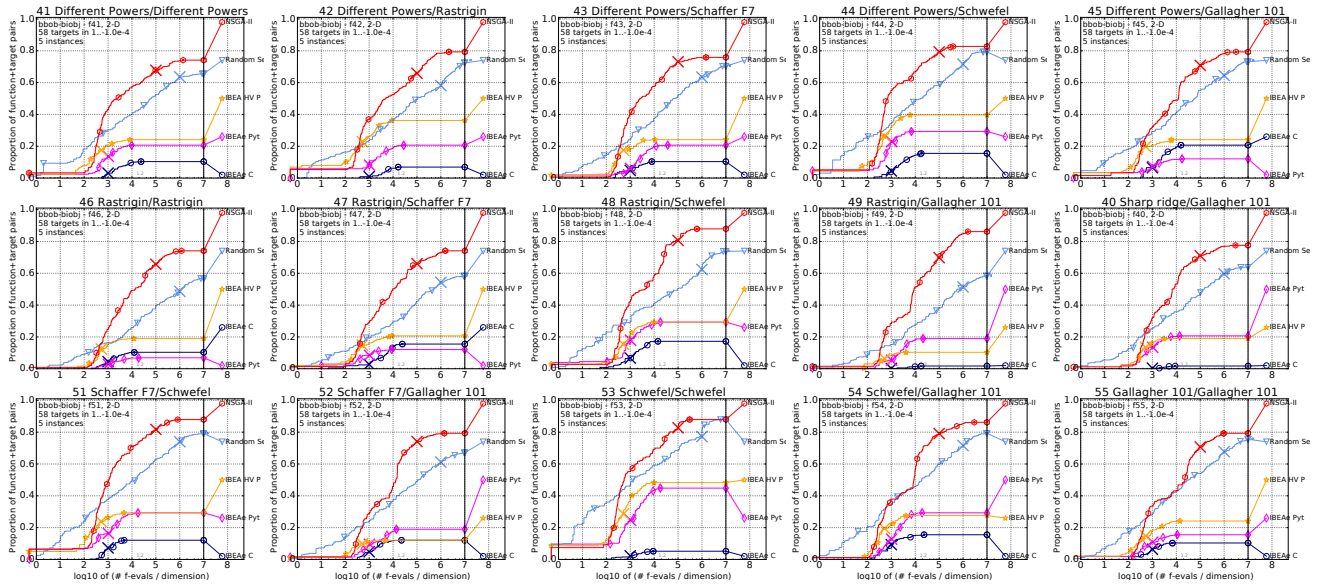


Figure 2: Bootstrapped empirical cumulative distribution of the number of objective function evaluations divided by dimension (FEvals/DIM) as in Fig. 1 but for functions f_{41} to f_{55} in 10-D.

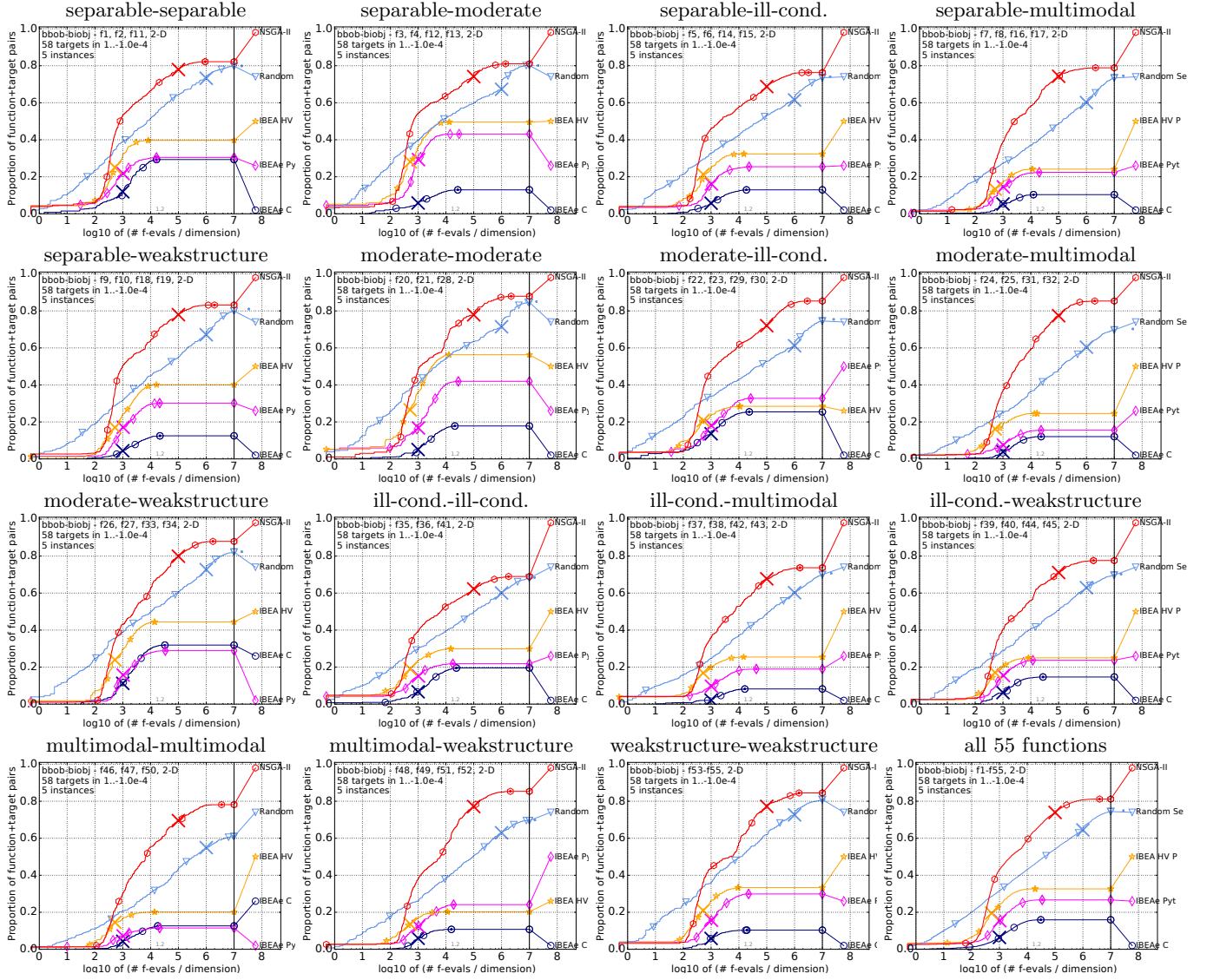


Figure 3: Bootstrapped empirical cumulative distribution of the number of objective function evaluations divided by dimension (FEvals/DIM) for 58 targets with target precision in $\{-10^{-4}, -10^{-4.2}, -10^{-4.4}, -10^{-4.6}, -10^{-4.8}, -10^{-5}, 0, 10^{-5}, 10^{-4.9}, 10^{-4.8}, \dots, 10^{-0.1}, 10^0\}$ for all functions and subgroups in 4-D.

6. REFERENCES

- [1] D. Brockhoff, T. Tušar, D. Tušar, T. Wagner, N. Hansen, and A. Auger. Biobjective performance assessment with the coco platform. *arXiv preprint arXiv:1605.01746*, 2016.
- [2] K. Deb and R. B. Agrawal. Simulated binary crossover for continuous search space. *Complex Systems*, 9(3):1–15, 1994.
- [3] N. Hansen, A. Auger, O. Mersmann, T. Tušar, and D. Brockhoff. COCO: A platform for comparing continuous optimizers in a black-box setting. *ArXiv e-prints*, arXiv:1603.08785, 2016.
- [4] N. Hansen, T. Tutar, O. Mersmann, A. Auger, and D. Brockhoff. Coco: The experimental procedure. *arXiv preprint arXiv:1603.08776*, 2016.
- [5] T. Tutar, D. Brockhoff, N. Hansen, and A. Auger. Coco: The bi-objective black box optimization benchmarking (bbob-biobj) test suite. *arXiv preprint arXiv:1604.00359*, 2016.

Δf	1e0	1e-1	1e-2	1e-3	#succ	Δf	1e0	1e-1	1e-2	1e-3	#succ	Δf	1e0	1e-1	1e-2	1e-3	#succ
f1						f20						f38					
IBEA	1226(457)	8440(5000)	∞	∞ 2000	0/5	IBEA	3025(4508)	∞	∞	∞ 2000	0/5	IBEA	1382(1515)	∞	∞	∞ 2000	0/5
IBEA	1(0)	1650(554)	∞	∞ 2063	0/5	IBEA	517(516)	3435(4812)	∞	∞ 2063	0/5	IBEA	182(234)	∞	∞	∞ 2063	0/5
IBEA	1(0)	1097(1126)	4827(2602)	∞ 1041	0/5	IBEA	404(520)	1688(3060)	4305(4164)	4777(8328)	0/5	IBEA	287(781)	1403(1401)	∞	∞ 1041	0/5
Rand	4.4(6)	76(44)	2202(1709)	6.9e4(5e4)	0/5	Rand	134(140)	537(2225)	1.2e5(3e5)	8.9e5(1e6)	0/5	Rand	1.6(0.5)	379(394)	8465(6593)	3.3e5(3e5)	0/5
NSGA	90(28)	547(58)	847(156)	5343(363)	0/5	NSGA	371(431)	693(537)	1228(311)	1.1e4(1e4)	1/5	NSGA	93(122)	676(22)	1668(378)	2.1e4(1e4)	0/5
f2						f21						f39					
IBEA	4518(2000)	9879(6000)	∞	∞ 2000	0/5	IBEA	3694(5588)	9894(9500)	∞	∞ 2000	0/5	IBEA	1808(1207)	3972(3244)	∞	∞ 2000	0/5
IBEA	653(172)	2509(4674)	∞	∞ 2063	0/5	IBEA	1713(3094)	2505(2047)	9554(2e4)	∞ 2063	0/5	IBEA	837(1119)	2397(1762)	∞	∞ 2063	0/5
IBEA	781(642)	1475(1301)	∞	∞ 1041	0/5	IBEA	309(190)	1988(1822)	4752(9109)	4969(3644)	0/5	IBEA	805(603)	∞	∞	∞ 1041	0/5
Rand	18(28)	409(572)	5282(5659)	1.6e5(3e5)	0/5	Rand	8.6(6)	227(112)	7747(8586)	1.8e5(2e5)	2/5	Rand	18(28)	420(299)	2.9e4(2e4)	3.8e6(4e6)	0/5
NSGA	191(180)	596(56)	882(155)	1526(616)	2/5	NSGA	338(214)	5220(1e4)	5763(1e4)	1.5e4(3e4)	2/5	NSGA	264(268)	690(78)	1431(702)	2.5e4(3e4)	0/5
f3						f22						f40					
IBEA	3128(4042)	∞	∞	∞ 2000	0/5	IBEA	1738(1064)	8006(1e4)	∞	∞ 2000	0/5	IBEA	9877(7000)	∞	∞	∞ 2000	0/5
IBEA	620(1755)	2880(2668)	∞	∞ 2063	0/5	IBEA	424(756)	9508(9799)	9881(1e4)	∞ 2063	0/5	IBEA	677(454)	8943(7736)	∞	∞ 2063	0/5
IBEA	314(312)	1311(2358)	∞	∞ 1041	0/5	IBEA	89(80)	2262(2181)	∞	∞ 1041	0/5	IBEA	177(61)	2406(2042)	∞	∞ 1041	0/5
Rand	663(826)	1.1e4(3e4)	3.8e5(1683)	6.8e5(2e6)	0/5	Rand	3.6(2)	660(431)	3.0e4(3e4)	3.2e6(4e6)	0/5	Rand	2.0(0.5)	761(539)	1.8e4(2e4)	5.6e5(6e5)	0/5
NSGA	191(214)	704(202)	1365(345)	2.2e4(7e3)	0/5	NSGA	242(189)	670(145)	1697(1170)	3.6e4(4e4)	1/5	NSGA	279(245)	906(378)	6671(1e4)	1.9e4(1e4)	0/5
f4						f23						f41					
IBEA	∞	∞	∞	∞ 2000	0/5	IBEA	556(1944)	1328(858)	∞	∞ 2000	0/5	IBEA	4242(3606)	∞	∞	∞ 2000	0/5
IBEA	367(411)	1326(295)	∞	∞ 2063	0/5	IBEA	29(36)	1046(617)	9641(1e4)	∞ 2063	0/5	IBEA	522(930)	4158(4658)	∞	∞ 2063	0/5
IBEA	312(520)	1160(1404)	2119(1590)	∞ 1041	0/5	IBEA	40(48)	310(386)	∞	∞ 1041	0/5	IBEA	198(298)	852(643)	∞	∞ 1041	0/5
Rand	31(0)	194(420)	1490(1372)	5.2e4(7e4)	0/5	Rand	2.6(2)	57(88)	1.3e6(2e6)	1.4e6(4e6)	0/5	Rand	3.0(0.5)	219(58)	6677(2794)	2.4e5(1e5)	0/5
NSGA	152(189)	562(132)	808(113)	1795(1003)	0/5	NSGA	50(61)	423(160)	857(268)	7093(7127)	1/5	NSGA	246(208)	598(39)	1090(57)	8817(5514)	0/5
f5						f24						f42					
IBEA	2483(3144)	∞	∞	∞ 2000	0/5	IBEA	615(1032)	9773(7736)	∞	∞ 2000	0/5	IBEA	8755(7000)	∞	∞	∞ 2000	0/5
IBEA	378(628)	∞	∞	∞ 2063	0/5	IBEA	136(195)	5079(4164)	∞	∞ 1041	0/5	IBEA	517(1547)	9776(2e4)	∞	∞ 2063	0/5
IBEA	36(44)	4774(4945)	∞	∞ 1041	0/5	IBEA	136(195)	5079(4164)	∞	∞ 1041	0/5	IBEA	68(166)	1140(1272)	5135(6506)	∞ 1041	0/5
Rand	5.6(4)	213(182)	1.1e4(2714)	8.8e5(2e6)	0/5	Rand	9.0(14)	1264(1210)	2.5e4(2e4)	1.2e6(6e6)	0/5	Rand	6.2(2)	2497(3062)	2.3e4(2e4)	8.8e5(3e5)	0/5
NSGA	148(216)	595(80)	1832(1406)	5.3e4(3e4)	0/5	NSGA	80(190)	856(204)	4719(4220)	2.0e4(2e4)	1/5	NSGA	107(152)	677(174)	1857(1102)	2.5e4(2e4)	0/5
f6						f25						f43					
IBEA	1135(1646)	8695(7500)	∞	∞ 2000	0/5	IBEA	1878(866)	∞	∞	∞ 2000	0/5	IBEA	3510(5968)	∞	∞	∞ 2000	0/5
IBEA	1(0)	2510(2162)	∞	∞ 2063	0/5	IBEA	742(1042)	9595(7736)	∞	∞ 2063	0/5	IBEA	1376(4642)	9800(6189)	∞	∞ 2063	0/5
IBEA	1(0)	853(449)	5086(7547)	∞ 1041	0/5	IBEA	281(380)	5197(6506)	∞	∞ 1041	0/5	IBEA	101(209)	4610(6766)	∞	∞ 1041	0/5
Rand	2.8(4)	62(52)	1440(1233)	4.4e4(6e4)	0/5	Rand	6.0(4)	403(394)	2.8e4(3e4)	5.7e5(9e5)	0/5	Rand	2.8(2)	573(465)	1.6e4(1e4)	2.3e5(2e5)	0/5
NSGA	71(133)	493(180)	868(308)	7131(3863)	0/5	NSGA	190(220)	687(182)	1350(322)	3026(1604)	1/5	NSGA	185(154)	715(123)	1523(440)	9809(9991)	0/5
f7						f26						f44					
IBEA	3763(5362)	∞	∞	∞ 2000	0/5	IBEA	3287(5723)	3848(5500)	9136(1e4)	9136(6000)	0/5	IBEA	2168(3372)	∞	∞	∞ 2000	0/5
IBEA	178(442)	4024(2012)	∞	∞ 2063	0/5	IBEA	1979(754)	4832(4568)	∞	∞ 2063	0/5	IBEA	619(872)	2246(3249)	∞	∞ 2063	0/5
IBEA	261(520)	1422(1390)	∞	∞ 1041	0/5	IBEA	377(316)	1837(2478)	2162(2478)	4729(2863)	0/5	IBEA	239(339)	614(855)	2312(3676)	∞ 1041	0/5
Rand	5.2(5)	990(159)	4.5e4(4e4)	9.5e5(2e6)	0/5	Rand	12(8)	274(329)	1.8e4(4e4)	3.0e5(4e5)	1/5	Rand	4.6(3)	357(834)	3447(5184)	2.3e5(4e5)	0/5
NSGA	170(274)	726(317)	3942(5600)	2.5e4(1e4)	0/5	NSGA	327(252)	782(547)	6305(1e4)	1.3e4(3e4)	1/5	NSGA	195(268)	532(138)	849(188)	1627(690)	0/5
f8						f27						f45					
IBEA	3988(3500)	∞	∞	∞ 2000	0/5	IBEA	3084(3504)	∞	∞	∞ 2000	0/5	IBEA	1394(1215)	9132(1e4)	∞	∞ 2000	0/5
IBEA	583(572)	4764(2020)	∞	∞ 2063	0/5	IBEA	606(716)	4558(5244)	∞	∞ 2063	0/5	IBEA	245(360)	∞	∞	∞ 2063	0/5
IBEA	512(580)	2225(781)	∞	∞ 1041	0/5	IBEA	166(184)	4968(5465)	∞	∞ 1041	0/5	IBEA	52(78)	1371(1080)	∞	∞ 1041	0/5
Rand	50(96)	4980(8919)	9.4e4(1e5)	1.7e6(1e6)	0/5	Rand	18(15)	299(250)	5428(6275)	1.2e5(9e4)	0/5	Rand	7.8(5)	217(151)	1.1e4(1e4)	3.3e5(4e5)	0/5
NSGA	382(228)	825(173)	1648(183)	9484(3741)	0/5	NSGA	215(212)	1094(956)	7403(9920)	1.1e4(1e4)	1/5	NSGA	152(171)	652(110)	4602(4401)	2.1e4(2e4)	0/5
f9						f28						f46					
IBEA	3940(2788)	∞	∞	∞ 2000	0/5	IBEA	3400(2500)	8707(4500)	∞	∞ 2000	0/5	IBEA	1996(4602)	∞	∞	∞ 2000	0/5
IBEA	837(768)	4448(2855)	9453(1e4)	∞ 2063	0/5	IBEA	1991(3610)	3192(6608)	8752(9799)	8949(1e4)	0/5	IBEA	1643(2184)	∞	∞	∞ 2063	0/5
IBEA	605(687)	4561(8328)	5049(2863)	∞ 1041	0/5	IBEA	221(296)	352(404)	1062(1220)	4822(2342)	0/5	IBEA	332(190)	5197(4424)	∞	∞ 1041	0/5
Rand	21(24)	344(296)	8842(1e4)	3.5e5(3e5)	0/5	Rand	5.4(8)	64(142)	1256(1684)	3.2e4(5e4)	3/5	Rand	30(11)	7120(1e4)	2.1e5(3e5)	3.8e6(6e6)	0/5
NSGA	311(354)	678(192)	923(126)	3555(1952)	0/5	NSGA	291(154)	502(214)	851(495)	4970(4814)	3/5	NSGA	248(225)	1072(325)	6613(4863)	5.9e4(6e4)	0/5
f10						f29						f47					
IBEA	880(2118)	9578(6500)	∞	∞ 2000	0/5	IBEA	9253(5500)	∞	∞	∞ 2000	0/5	IBEA	3931(3892)	∞	∞	∞ 2000	0/5
IBEA	847(359)	4266(5331)	∞	∞ 2063	0/5	IBEA	693(1724)	∞	∞	∞ 2063	0/5	IBEA	675(1036)	∞	∞	∞ 2063	0/5
IBEA	375(110)	1539(521)	∞	∞ 1041	0/5	IBEA	191(426)	∞	∞	∞ 1041	0/5	IBEA	197(260)	4806(3904)	∞	∞ 1041	0/5
Rand	7.2(10)	491(452)	1.3e4(9389)	4.0e5(4e5)	0/5	Rand	34(80)	219(169)	1.5e4(2e4)	1.5e6(1e6)	0/5	Rand	5.2(6)	3813(8204)	7.8e4(9e4)	2.2e6(2e6)	0/5
NSGA	190(260)	606(166)	1376(455)	9039(6929)	0/5	NSGA	197(114)	616(171)	1193(354)	1.5e4(1e4)	0/5	NSGA	112(172)	862(200)	5473(6272)	3.6e4(3e4)	0/5
f11						f30						f48					
IBEA	583(562)	1581(2000)	8628(3000)	∞ 2000	0/5	IBEA	263(292)	832(968)	∞	∞ 2000	0/5	IBEA	3360(4252)	∞	∞	∞ 2000	0/5
IBEA	162(381)	726(651)	∞														