

Tutorial Proposal: Difficulties in Fair Performance Comparison of Multiobjective Evolutionary Algorithms

Hisao Ishibuchi, Lie Meng Pang, Ke Shang
Southern University of Science and Technology
Shenzhen, China

hisao@sustech.edu.cn, panglm@sustech.edu.cn, kshang@foxmail.com

<http://gecco-2022.sigevo.org/>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

GECCO '22 Companion, Boston, USA

© 2022 Copyright held by the owner/author(s). 978-x-xxxx-xxxx-

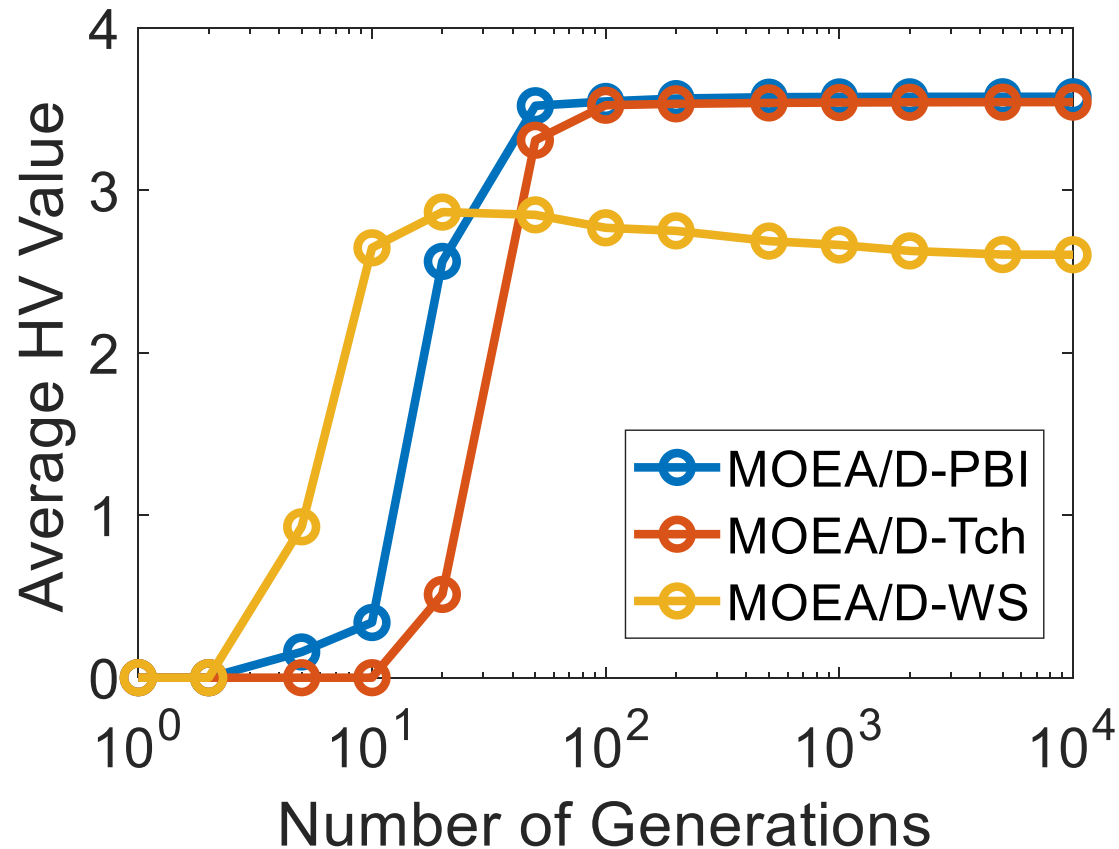
x/YY/MM...\$15.00

Add the DOI URL assigned to your paper



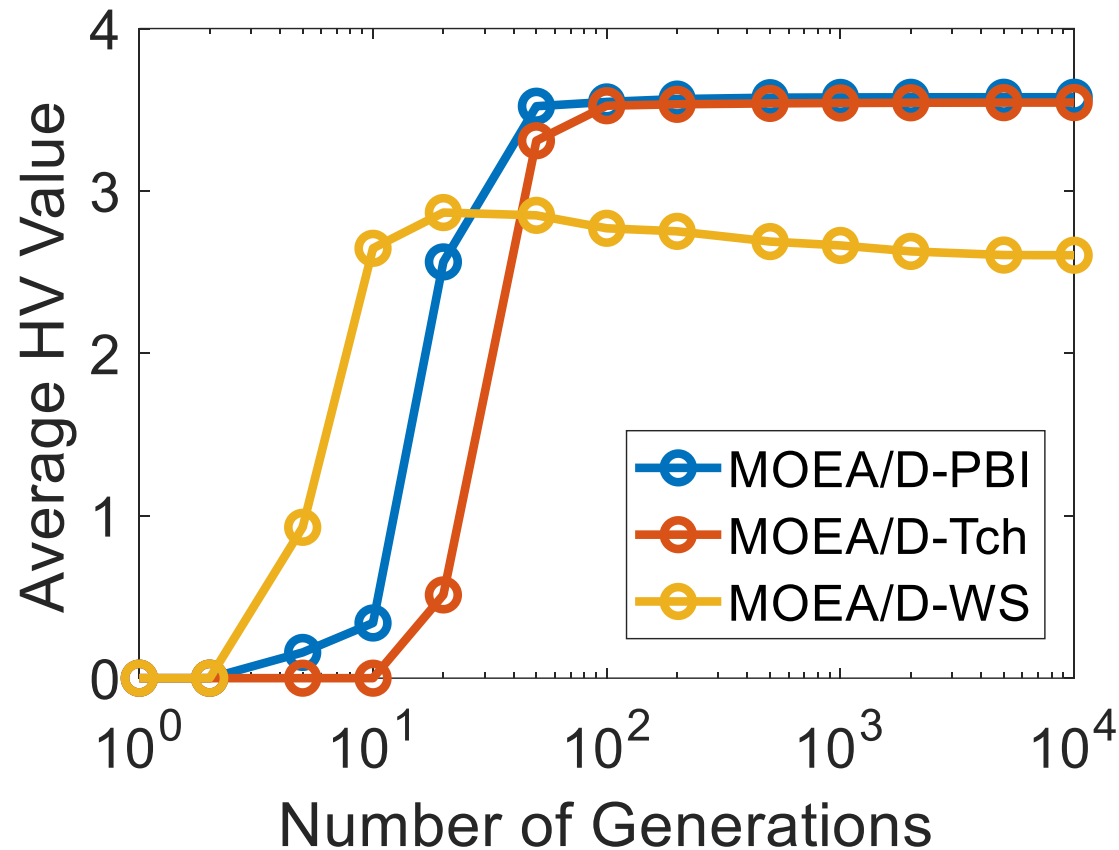
Difficulties in Fair Performance Comparison of EMO Algorithms

Difficulty 1: Performance comparison results depend on the termination condition. Different comparison results are obtained from different specifications of the termination condition.

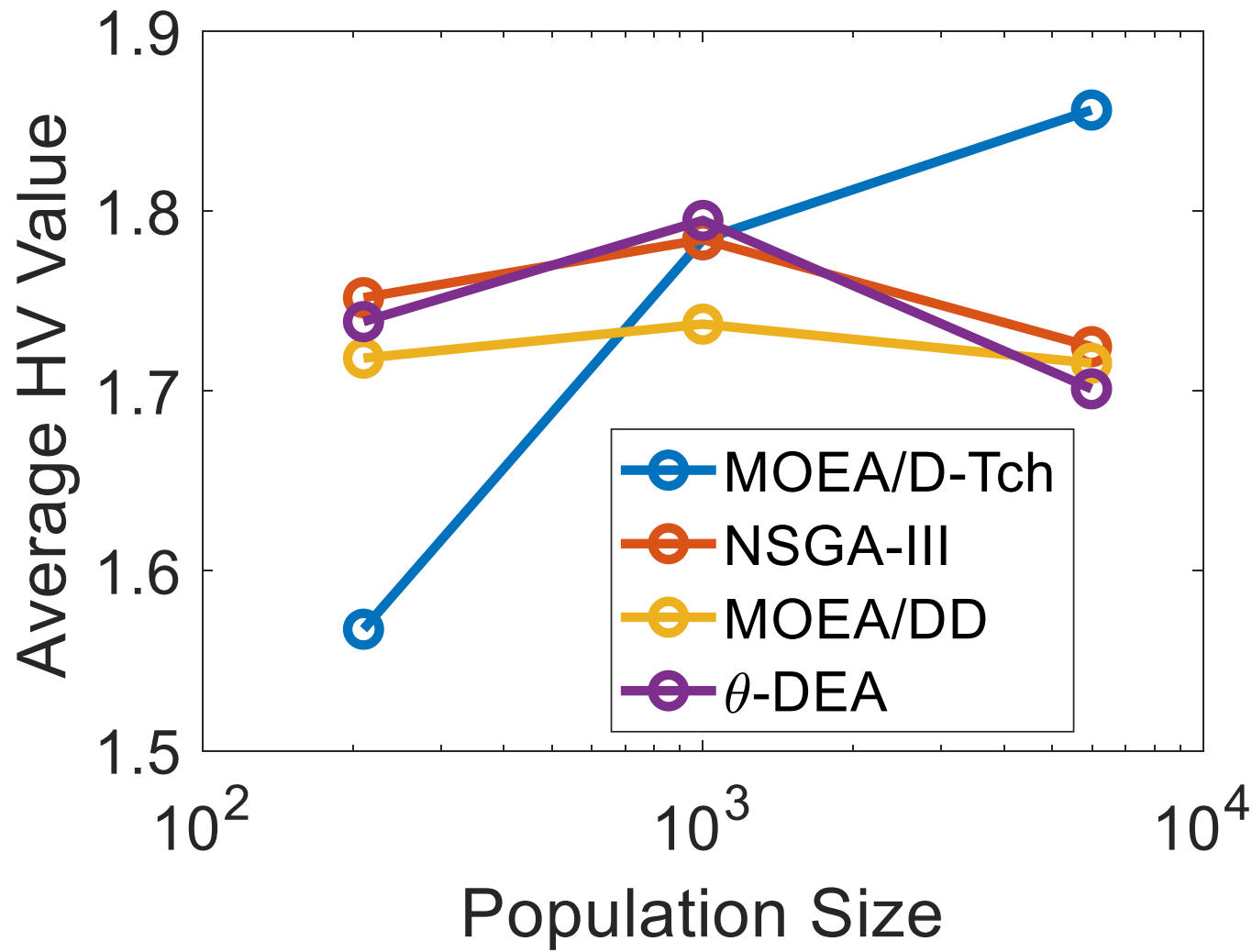


Difficulties in Fair Performance Comparison of EMO Algorithms

Difficulty 1: Performance comparison results depend on the termination condition.
==> Such dependency indicates the importance of performance comparison under the anytime algorithm framework.



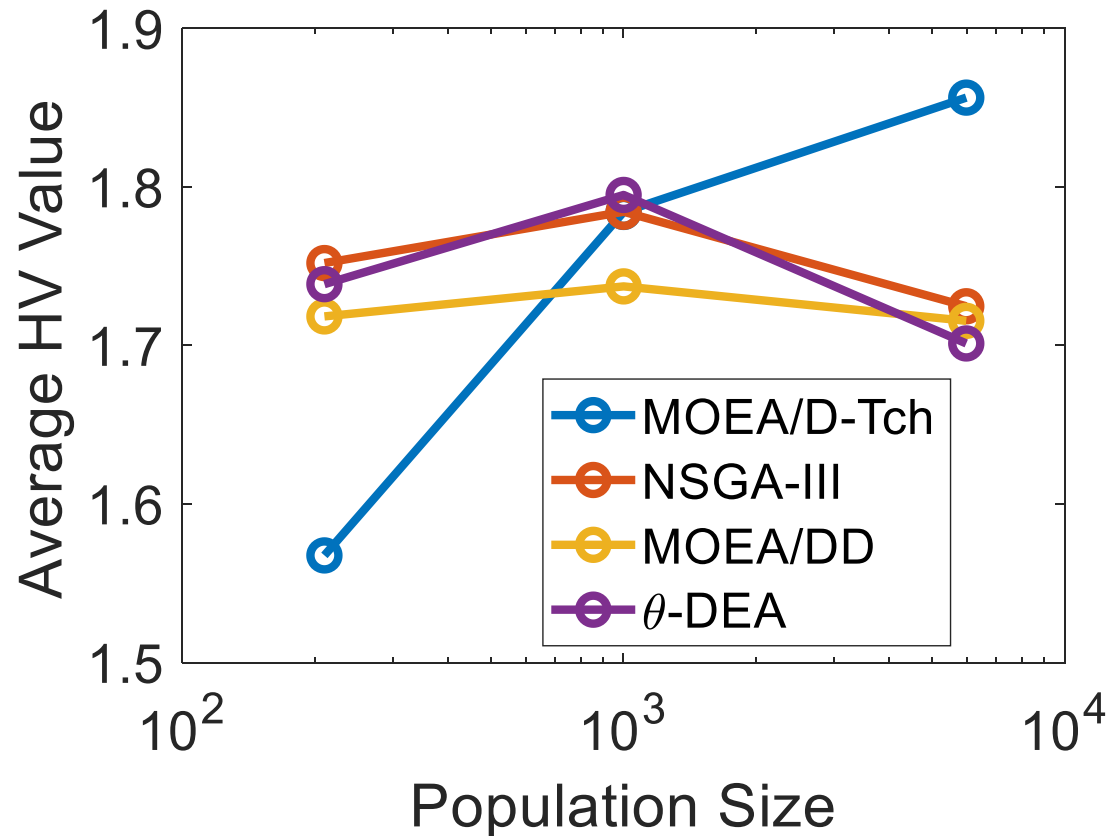
Difficulty 2: Performance comparison results depend on the population size.



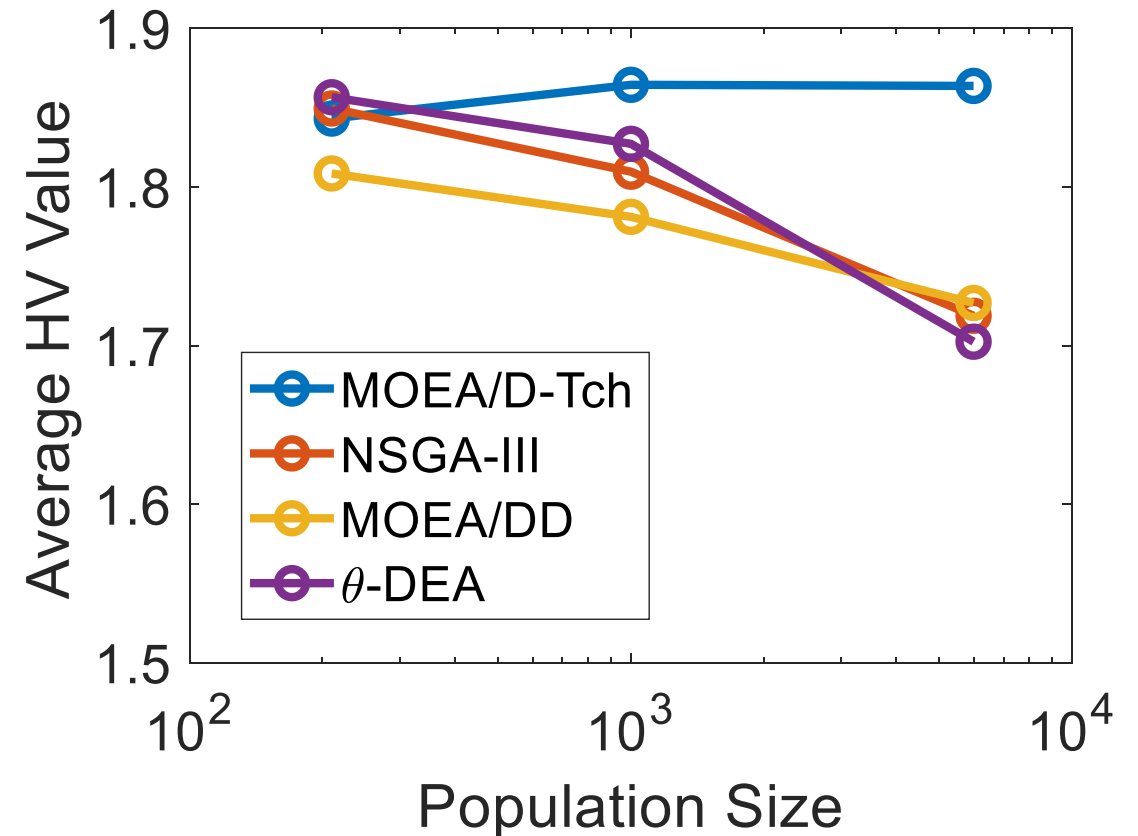
HV of the final population.

Difficulty 2: Performance comparison results depend on the population size.

==> Comparison of EMO algorithms using selected solution subsets of the same size from all the examined solutions.

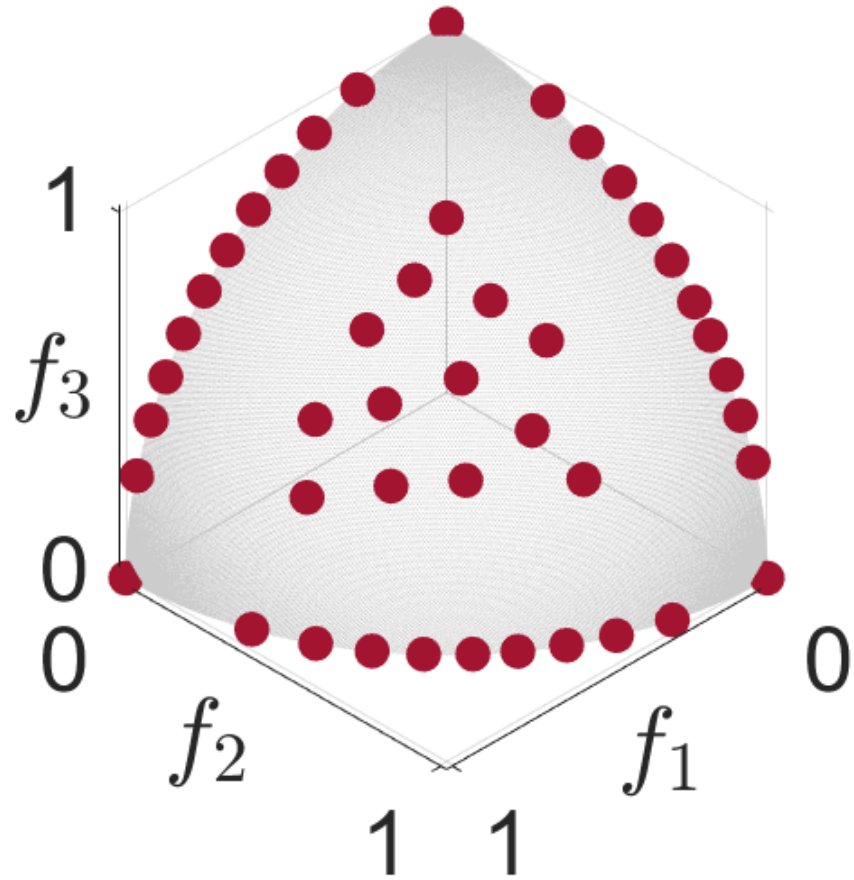


HV of the final population.

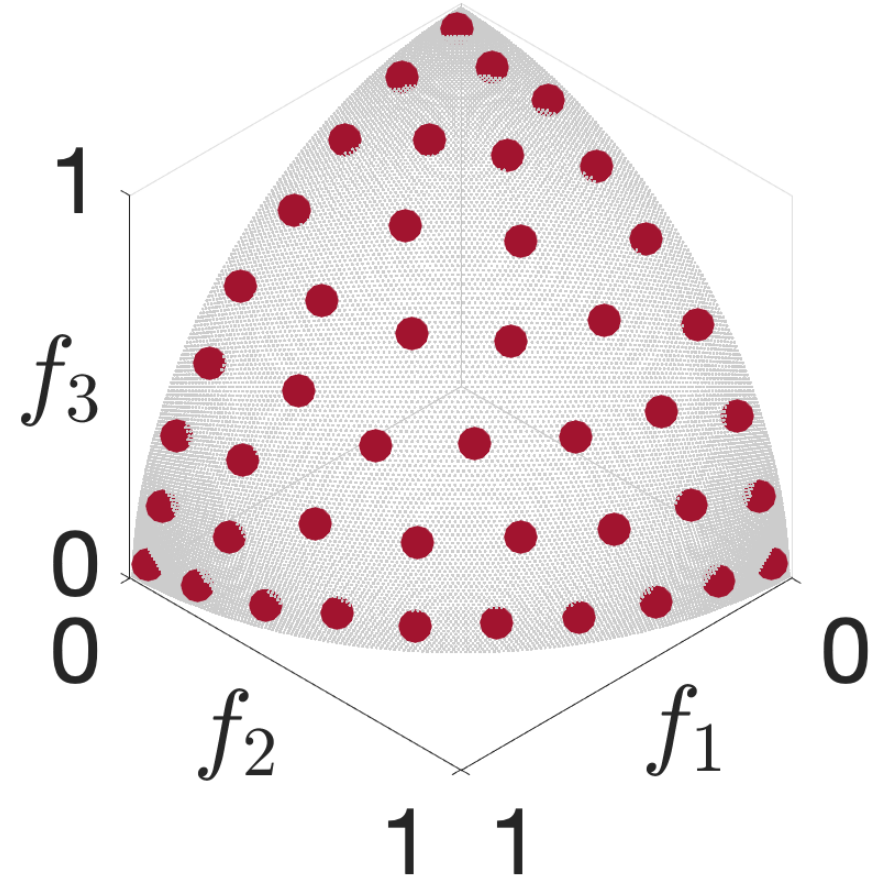


HV of the selected solution subset.
(210 solutions are selected)

Difficulty 3: Performance comparison results depend on the performance indicator.

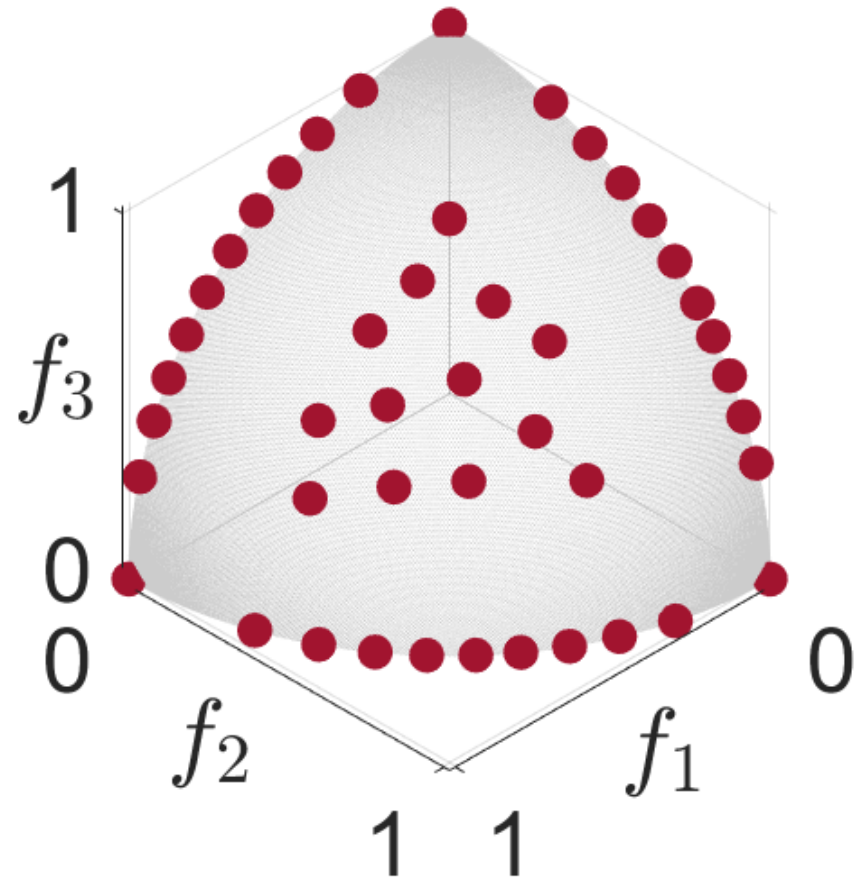


HV-Optimal Distribution

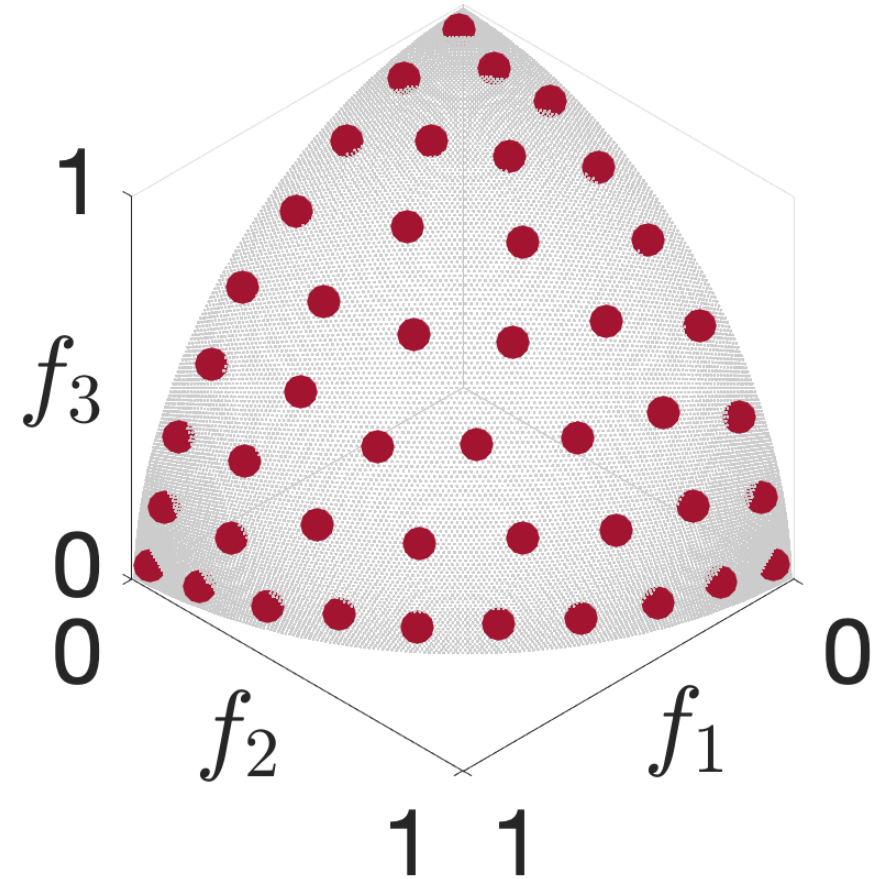


IGD-Optimal Distribution.

Difficulty 3: Performance comparison results depend on the performance indicator.
==> Use of at least two performance indicators.

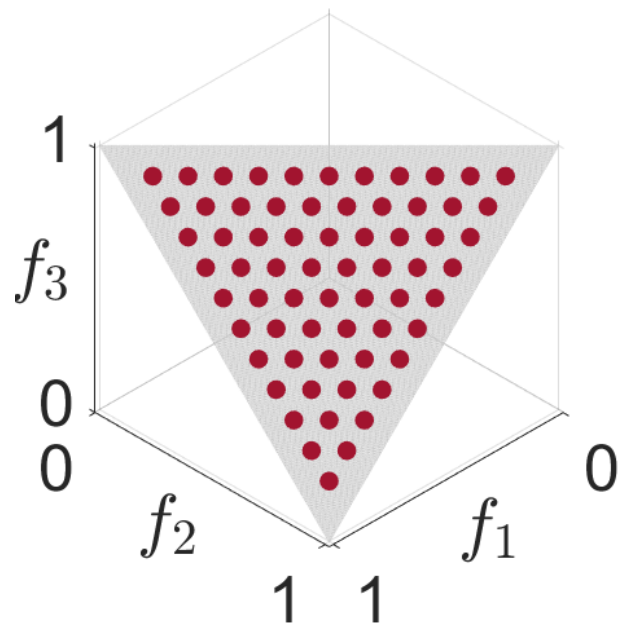


HV-Optimal Distribution

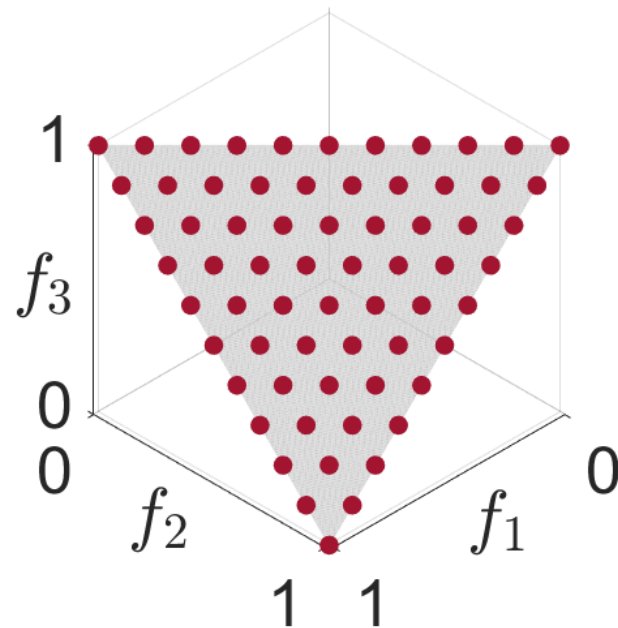


IGD-Optimal Distribution.

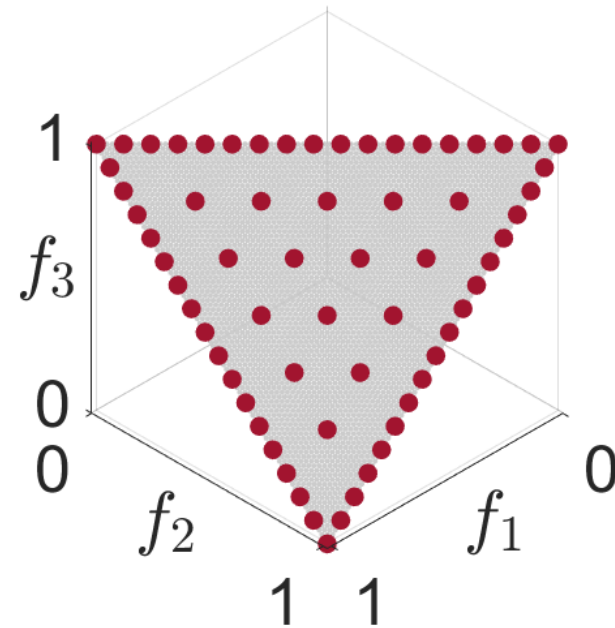
Difficulty 4: Performance comparison results depend on the parameter specification in each indicator. For example, any of the following four solution set can be shown as the best solution set depending on the reference point setting in the hypervolume indicator.



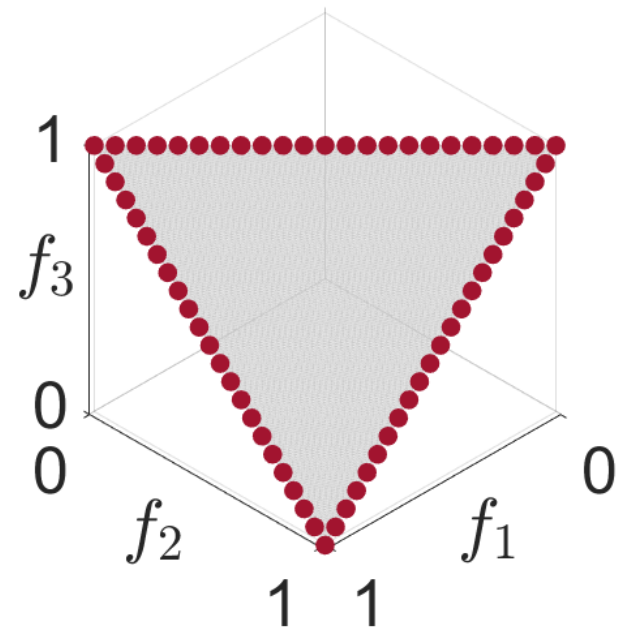
$r = 1.01$
[1, 1.05]



$r = 1.1$
[1.06, 1.29]



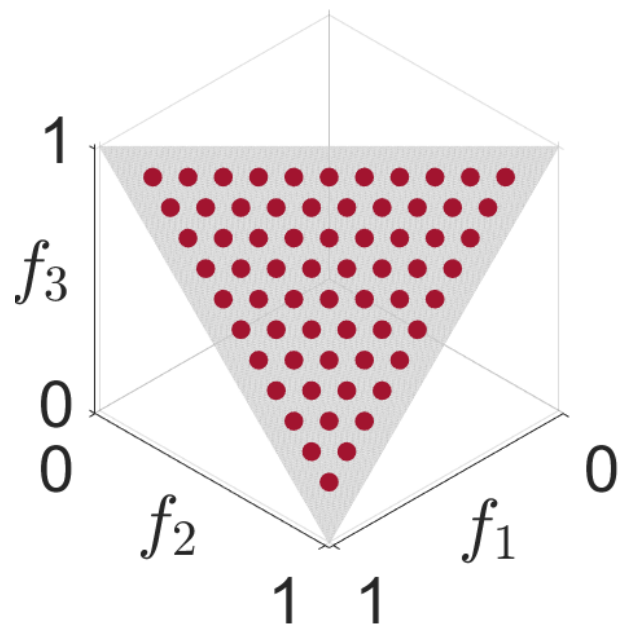
$r = 1.5$
[1.30, 6.08]



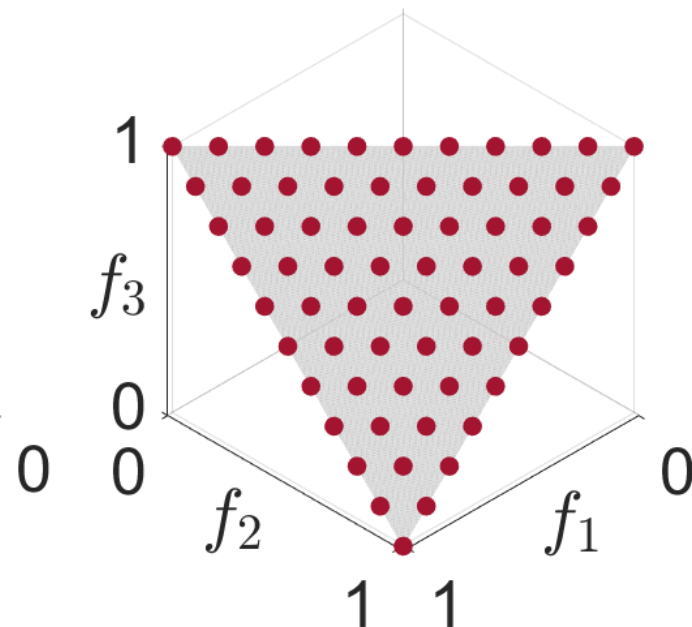
$r = 10$
[6.09, 100]

Difficulty 4: Performance comparison results depend on the parameter specification in each indicator. For example, any of the following four solution set can be shown as the best solution set depending on the reference point setting in the hypervolume indicator.

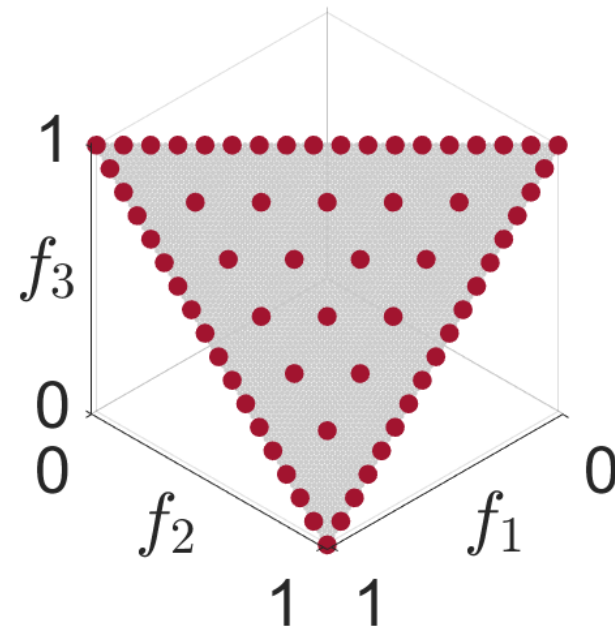
==> An appropriate specification of the reference point is needed for fair performance comparison.



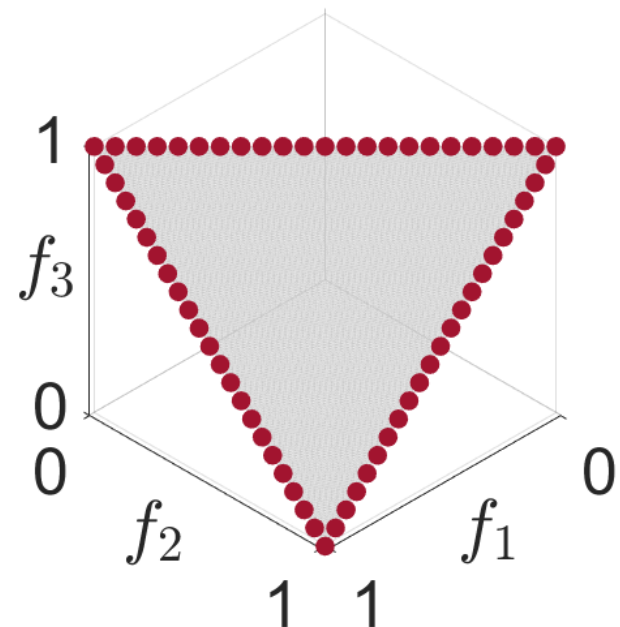
$r = 1.01$
[1, 1.05]



$r = 1.1$
[1.06, 1.29]



$r = 1.5$
[1.30, 6.08]



$r = 10$
[6.09, 100]

Difficulty 5: Performance comparison results depend on the choice of test problems.

Best Algorithm Percentage for Each Test Problem Set with 3, 5, 8 and 10 objectives:

Percentage of test problems on which each algorithm has the best average IGD value among the eight algorithms in each of the four test problem suites.

| Algorithm | DTLZ1-4 | WFG1-9 | Minus-DTLZ1-4 | Minus-WFG1-9 |
|---------------|---------|--------|---------------|--------------|
| NSGA-III | 6.25 | 16.67 | 12.50 | 5.56 |
| θ -DEA | 18.75 | 44.44 | 0.00 | 0.00 |
| MOEA/DD | 62.50 | 11.11 | 0.00 | 0.00 |
| MOEA/D-PBI | 12.50 | 5.56 | 0.00 | 0.00 |
| MOEA/D-Tch | 0.00 | 5.56 | 0.00 | 8.33 |
| MOEA/D-WS | 0.00 | 0.00 | 6.25 | 0.00 |
| MOEA/D-IPBI | 0.00 | 0.00 | 12.50 | 16.67 |
| NSGA-II | 0.00 | 16.67 | 68.75 | 69.44 |

Difficulty 5: Performance comparison results depend on the choice of test problems. Moreover, the frequently-used test problems DTLZ and WFG are unrealistic.

Best Algorithm Percentage for Each Test Problem Set with 3, 5, 8 and 10 objectives:

Percentage of test problems on which each algorithm has the best average IGD value among the eight algorithms in each of the four test problem suites.

| Algorithm | DTLZ1-4 | WFG1-9 | Minus-DTLZ1-4 | Minus-WFG1-9 |
|---------------|---------|--------|---------------|--------------|
| NSGA-III | 6.25 | 16.67 | 12.50 | 5.56 |
| θ -DEA | 18.75 | 44.44 | 0.00 | 0.00 |
| MOEA/DD | 62.50 | 11.11 | 0.00 | 0.00 |
| MOEA/D-PBI | 12.50 | 5.56 | 0.00 | 0.00 |
| MOEA/D-Tch | 0.00 | 5.56 | 0.00 | 8.33 |
| MOEA/D-WS | 0.00 | 0.00 | 6.25 | 0.00 |
| MOEA/D-IPBI | 0.00 | 0.00 | 12.50 | 16.67 |
| NSGA-II | 0.00 | 16.67 | 68.75 | 69.44 |

Difficulty 5: Performance comparison results depend on the choice of test problems. Moreover, the frequently-used test problems DTLZ and WFG are unrealistic.

==> Realistic test problems are needed.

Best Algorithm Percentage for Each Test Problem Set with 3, 5, 8 and 10 objectives:

Percentage of test problems on which each algorithm has the best average IGD value among the eight algorithms in each of the four test problem suites.

| Algorithm | DTLZ1-4 | WFG1-9 | Minus-DTLZ1-4 | Minus-WFG1-9 |
|---------------|---------|--------|---------------|--------------|
| NSGA-III | 6.25 | 16.67 | 12.50 | 5.56 |
| θ -DEA | 18.75 | 44.44 | 0.00 | 0.00 |
| MOEA/DD | 62.50 | 11.11 | 0.00 | 0.00 |
| MOEA/D-PBI | 12.50 | 5.56 | 0.00 | 0.00 |
| MOEA/D-Tch | 0.00 | 5.56 | 0.00 | 8.33 |
| MOEA/D-WS | 0.00 | 0.00 | 6.25 | 0.00 |
| MOEA/D-IPBI | 0.00 | 0.00 | 12.50 | 16.67 |
| NSGA-II | 0.00 | 16.67 | 68.75 | 69.44 |