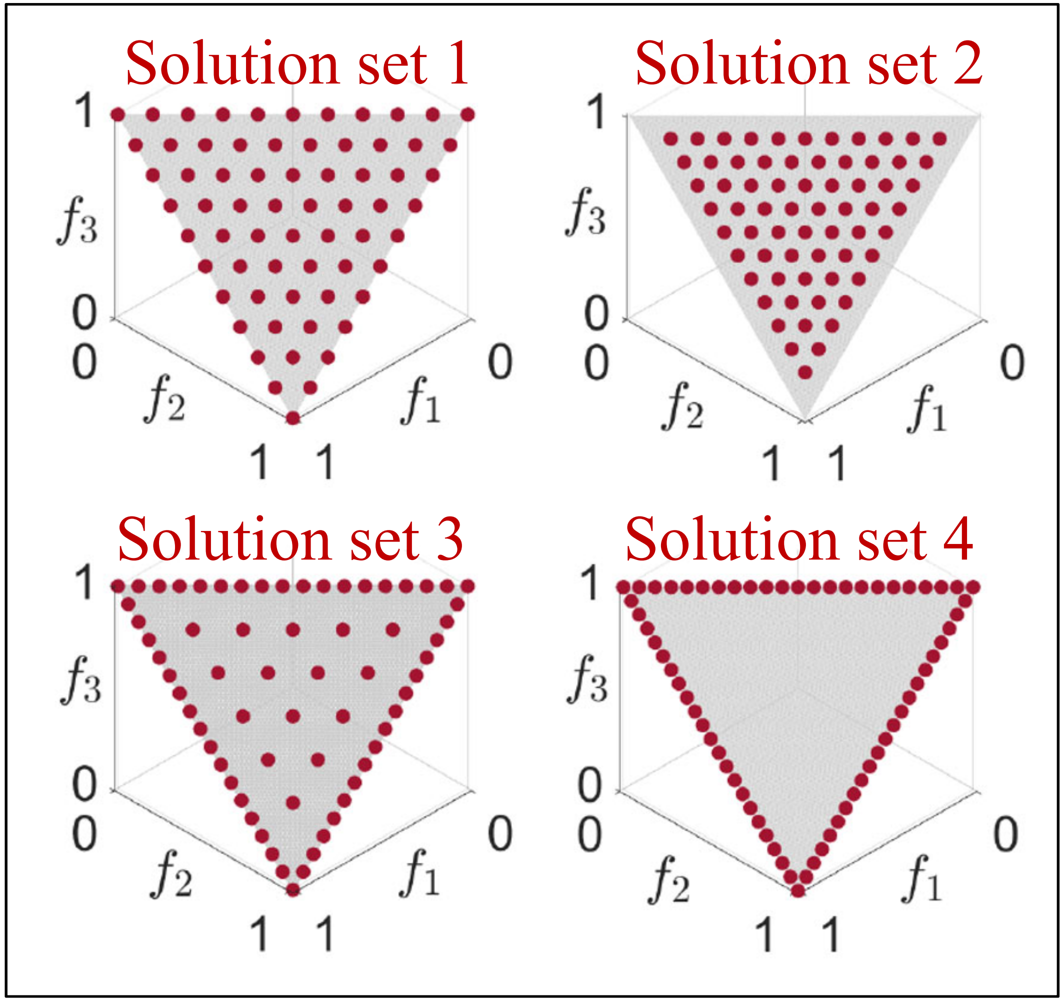
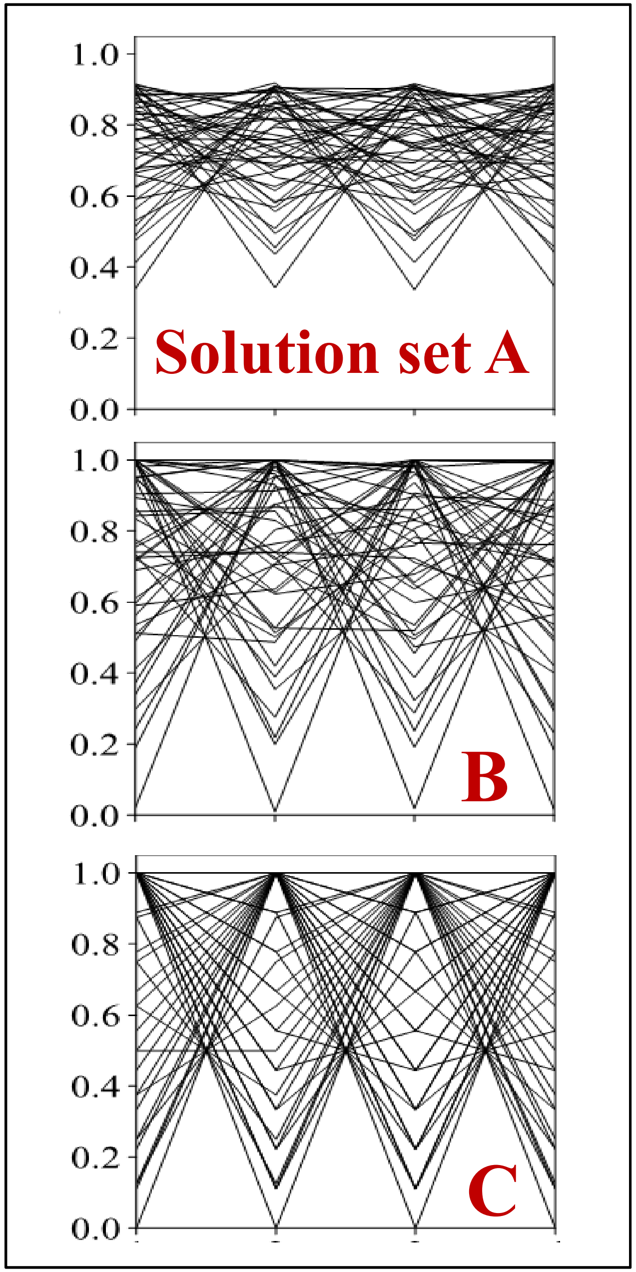
Fair Performance Comparison of Evolutionary Multi-Objective Algorithms

Question Which is the better solution set ?





Hisao Ishibuchi (SUSTech in Shenzhen, China)

Today talk is from Sapporo

My Hotel

Hokkaido University

Fair Performance Comparison of Evolutionary Multi-Objective Algorithms

Hisao Ishibuchi
Southern University of Science and Technology
Shenzhen, China

In this talk, I will explain the following:

In computational experiments, we can easily obtain clearly different comparison results depending on the settings of our computational experiments.

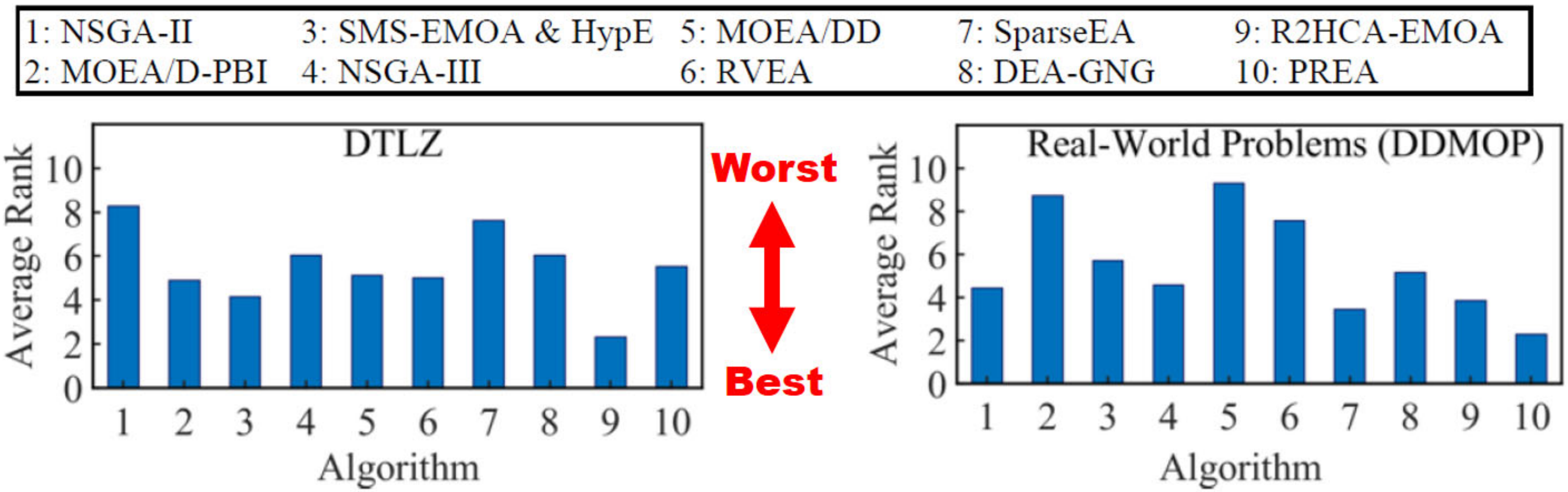


Fig. 1. Results on DTLZ

Fig. 2. Results on DDMOP

In Fig. 1, NSGA-II (ID 1) is worst, and PREA (ID 10) is not good.
In Fig. 1, NSGA-II (ID 1) is good, and PREA (ID 10) is the best.

Multi-Objective Optimization

Optimization Problems

Single-Objective Problem:

Maximize $f(\boldsymbol{x})$

Multi-Objective Problems:

Maximize $f_1(\boldsymbol{x}),\ f_2(\boldsymbol{x})$

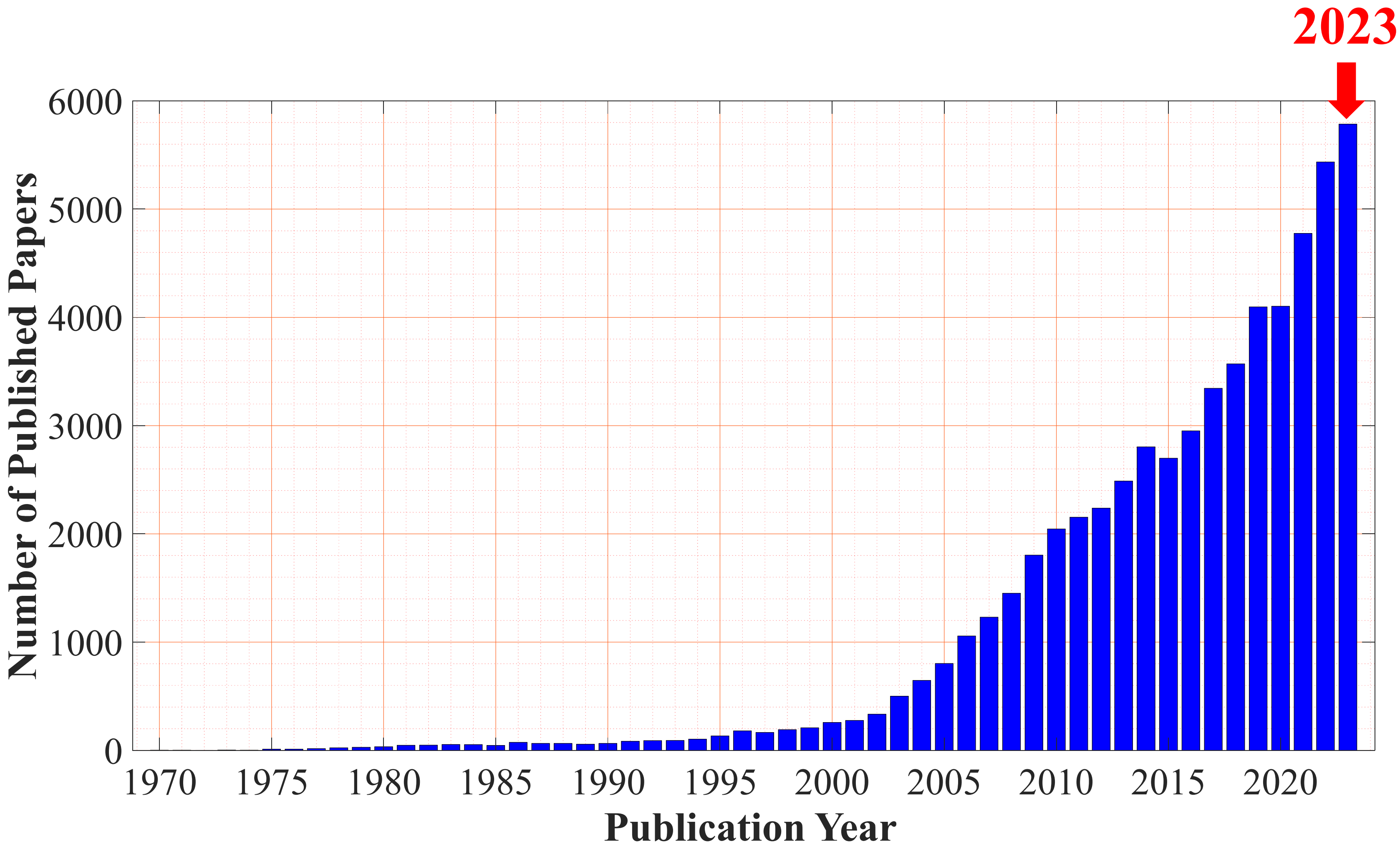
Maximize $f_1(\boldsymbol{x}),\ f_2(\boldsymbol{x}),\ f_3(\boldsymbol{x})$

Maximize $f_1(\boldsymbol{x}),\ f_2(\boldsymbol{x}),\ f_3(\boldsymbol{x}),\ f_4(\boldsymbol{x})$

Maximize $f_1(\boldsymbol{x}),\ f_2(\boldsymbol{x}),\ f_3(\boldsymbol{x}),\ f_4(\boldsymbol{x}),\ f_5(\boldsymbol{x}),\ \cdots$

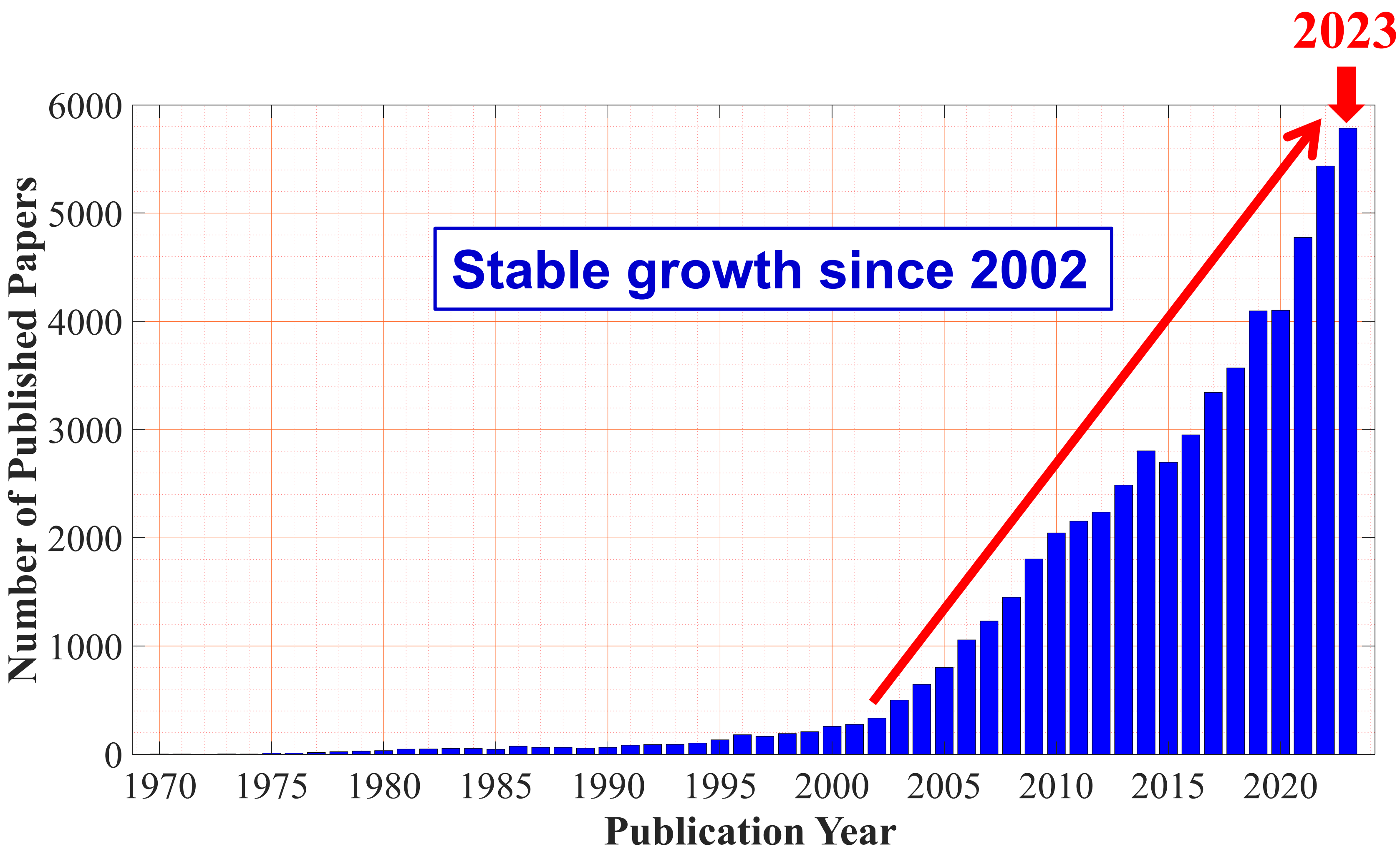
Popularity of Multi-Objective Optimization Research

The number of papers with “Multi-objective” or “Multiobjective” in the paper titles (Scopus Database: January 17, 2024)



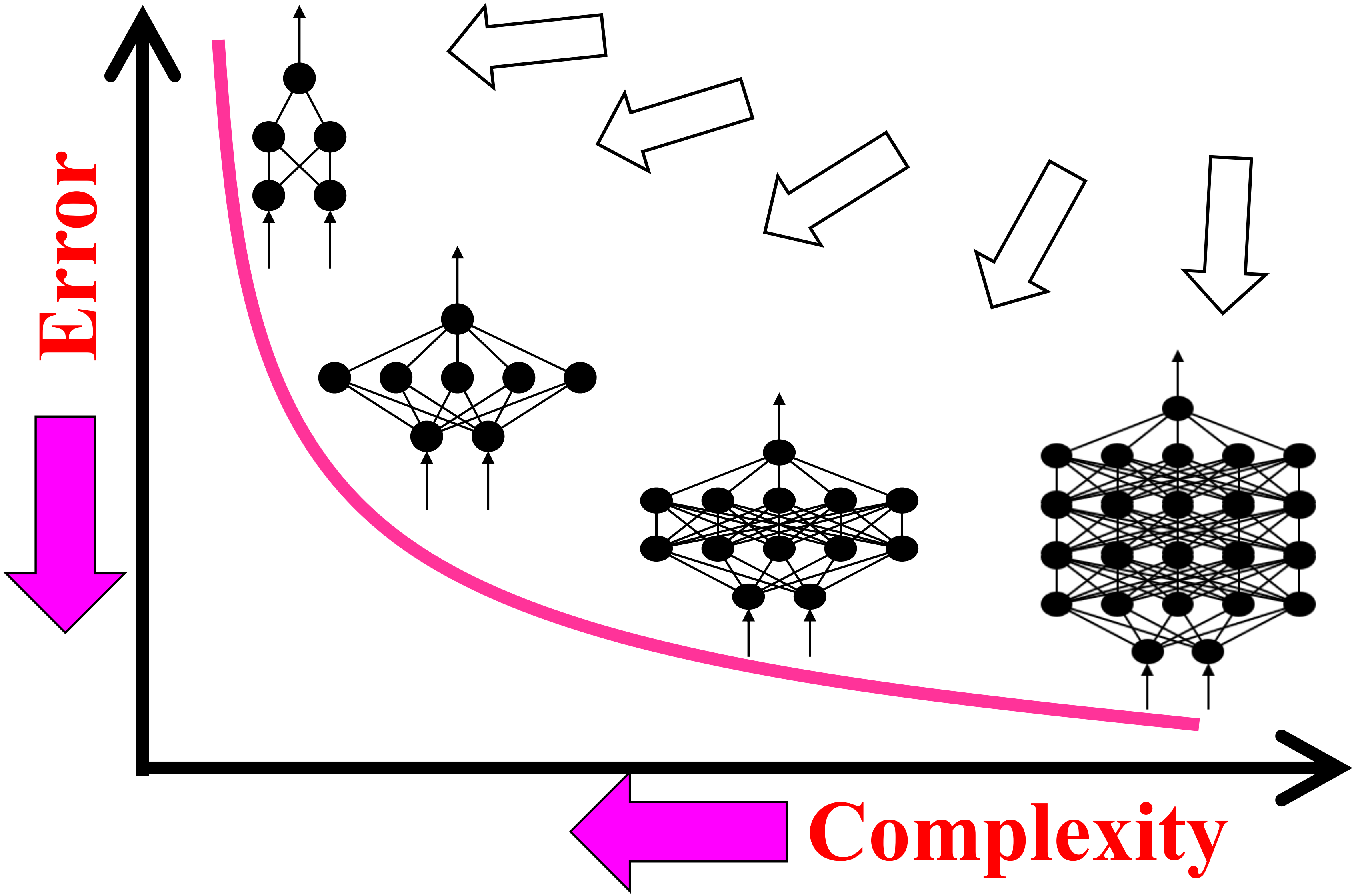
Popularity of Multi-Objective Optimization Research

The number of papers with “Multi-objective” or “Multiobjective” in the paper titles (Scopus Database: January 17, 2024)

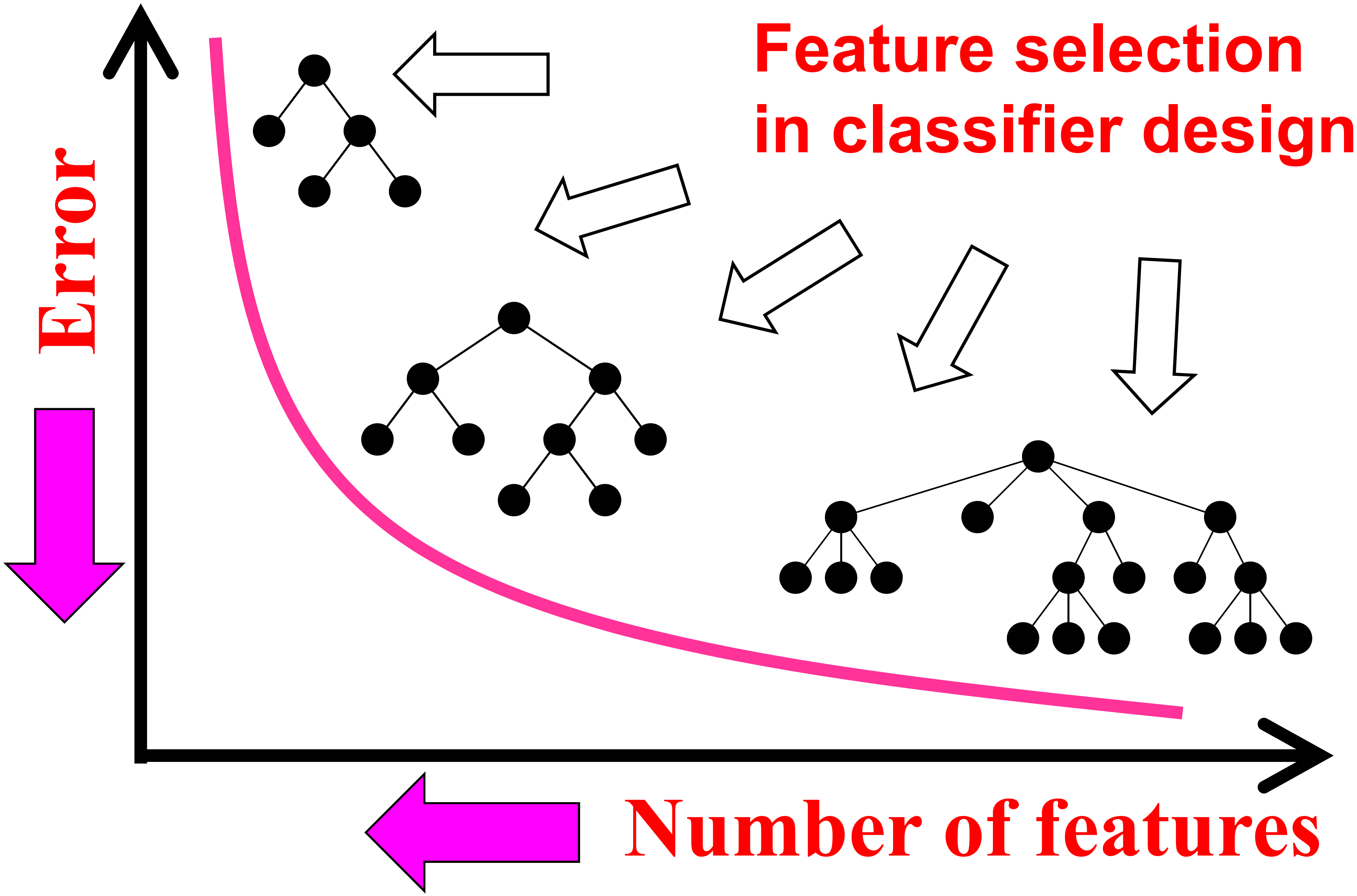


Almost all problems have multiple objectives

Example: Multi-objective neural architecture search

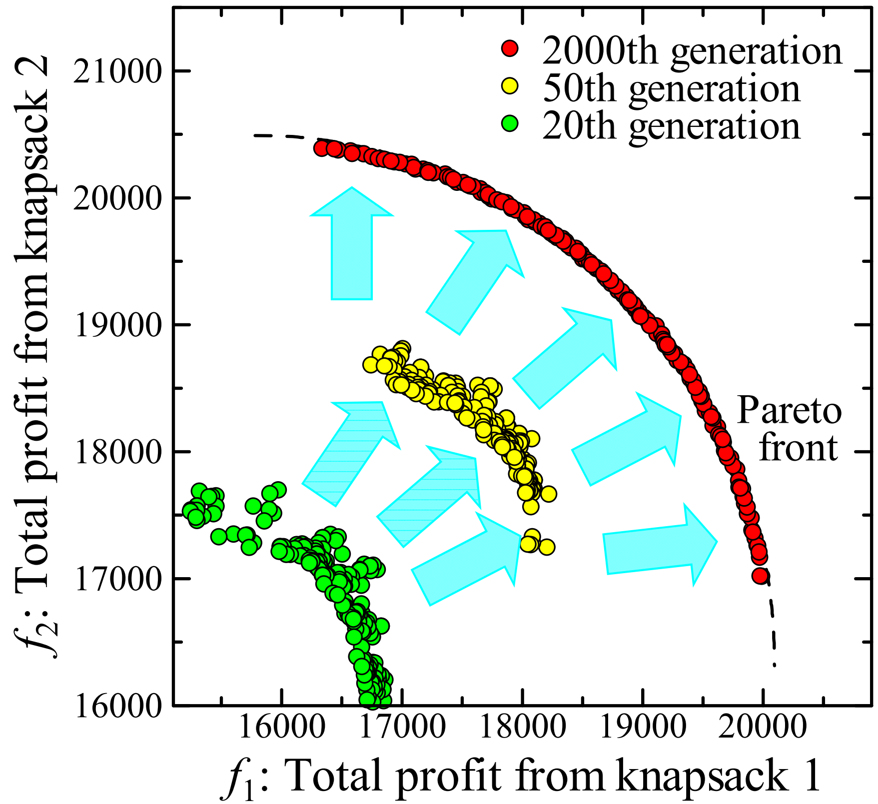


Almost all problems have multiple objectives



Evolutionary Multi-Objective Optimization (EMO)
= Evolutionary Search for Pareto Optimal Solutions

The Goal of EMO: To find a set of well-distributed solutions over the entire Pareto front.



Example of Good Search Behavior of an EMO algorithm

Difficulties in Fair Performance Comparison of Evolutionary Multi-Objective Optimization Algorithms

(0) Visual Comparison

(1) Specification of Termination Condition

(2) Specification of Population Size

(3) Choice of Performance Indicators (e.g., HV, IGD)

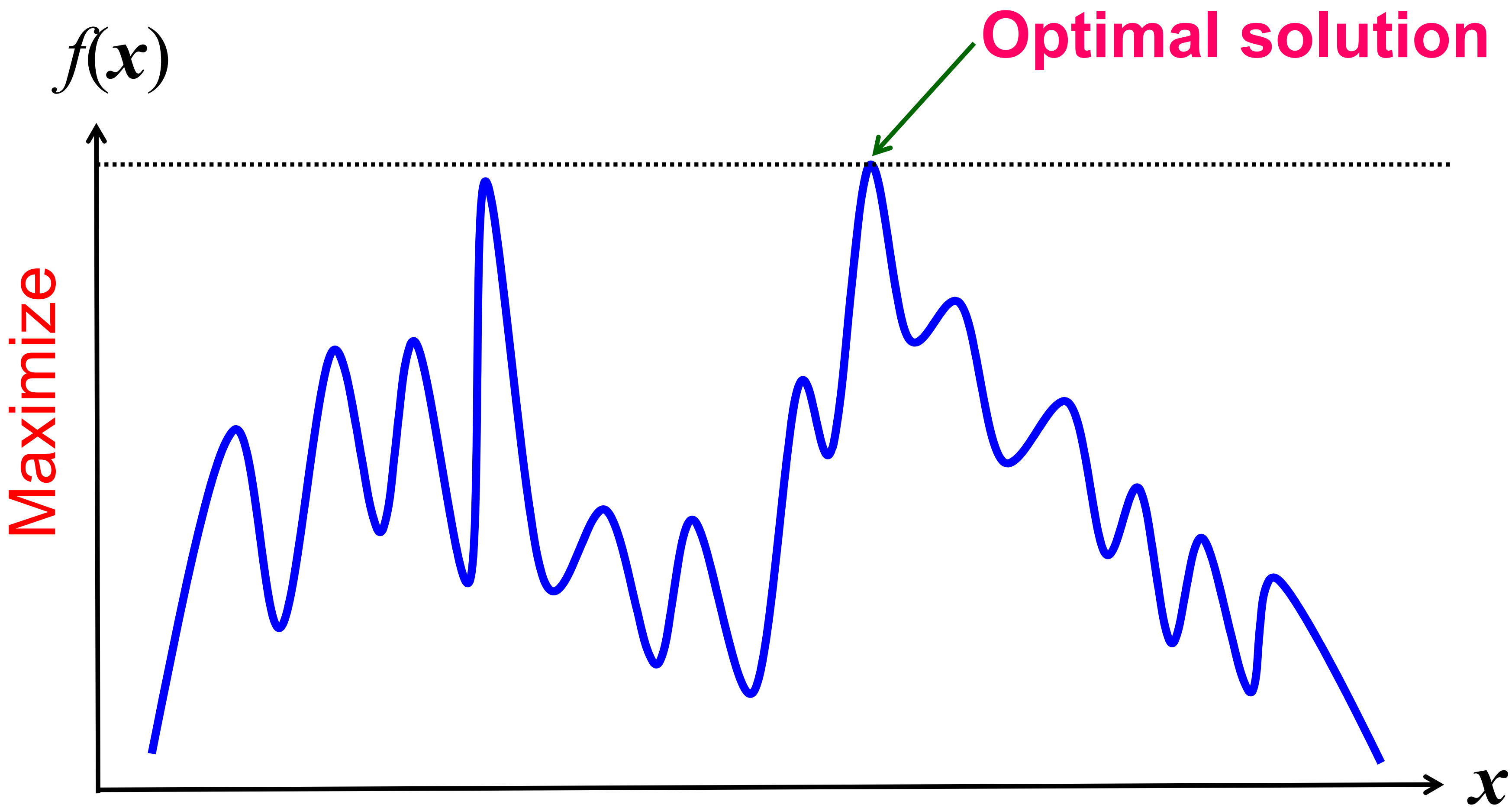
(4) Setting in Performance Indicators (e.g., reference point)

(5) Choice of Test Problems

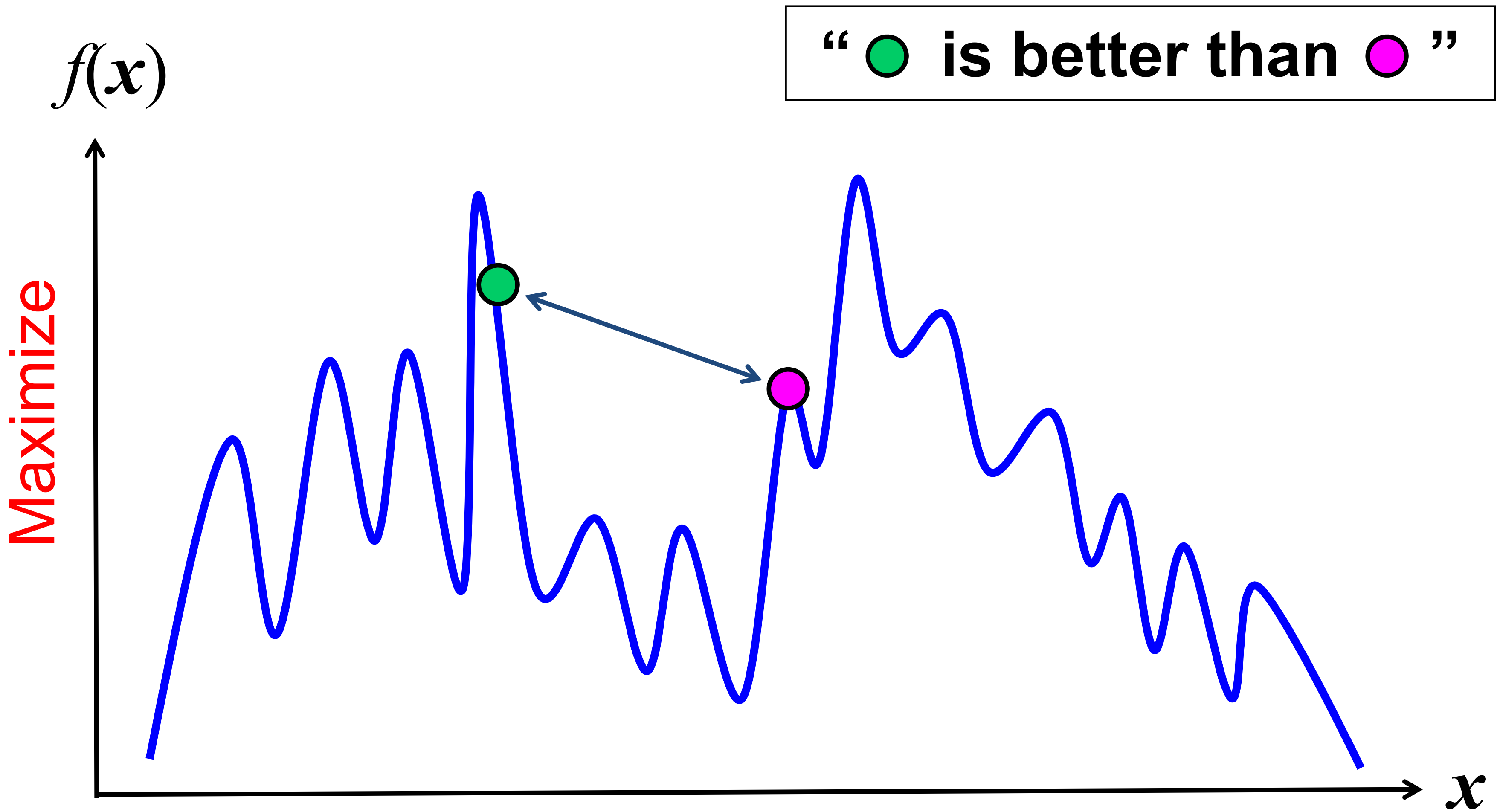
This Talk is mainly based on my recent paper:

Hisao Ishibuchi, Lie Meng Pang, and Ke Shang, "**Difficulties in Fair Performance Comparison of Multi-Objective Evolutionary Algorithms**" ***IEEE Computational Intelligence Magazine*** **(February 2022)**

Single-Objective Optimization: Maximize $f(x)$

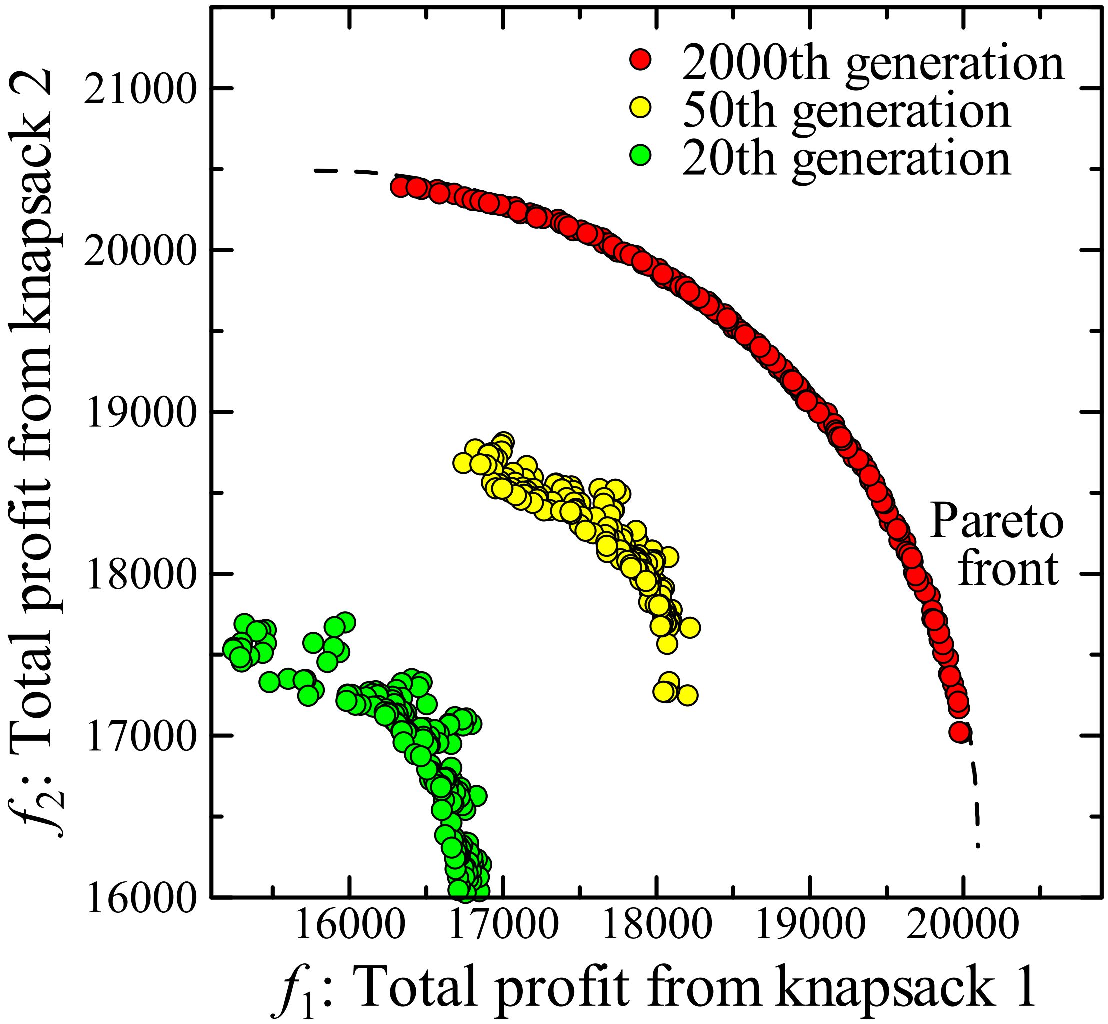


The final result of optimization is a single solution.
Comparison of solutions is easy.



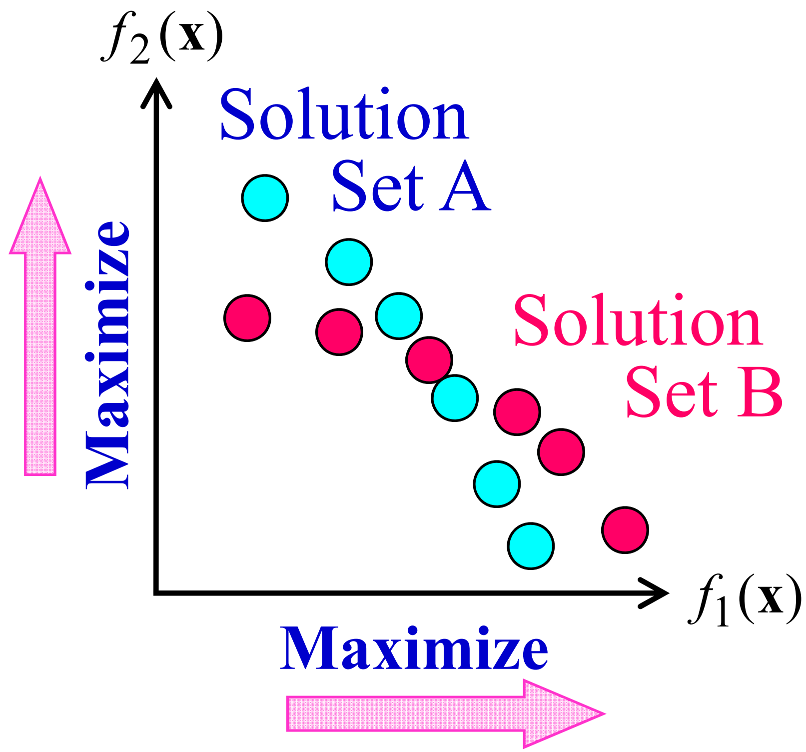
Two-Objective Optimization: Maximize $f_1(\boldsymbol{x}),\ f_2(\boldsymbol{x})$

The final result of optimization is a solution set.



Algorithm comparison

⇒ **Comparison between solution sets**



Difficulties in Fair Performance Comparison of Evolutionary Multi-Objective Optimization Algorithms

(0) Visual Comparison

(1) Specification of Termination Condition

(2) Specification of Population Size

(3) Choice of Performance Indicators (e.g., HV, IGD)

(4) Setting in Performance Indicators (e.g., reference point)

(5) Choice of Test Problems

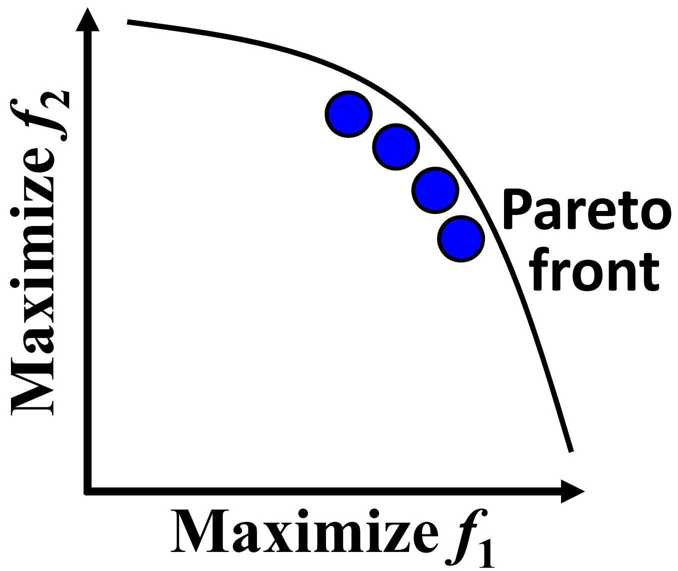
This Talk is mainly based on my recent paper:

Hisao Ishibuchi, Lie Meng Pang, and Ke Shang, "**Difficulties in Fair Performance Comparison of Multi-Objective Evolutionary Algorithms**" ***IEEE Computational Intelligence Magazine*** **(February 2022)**

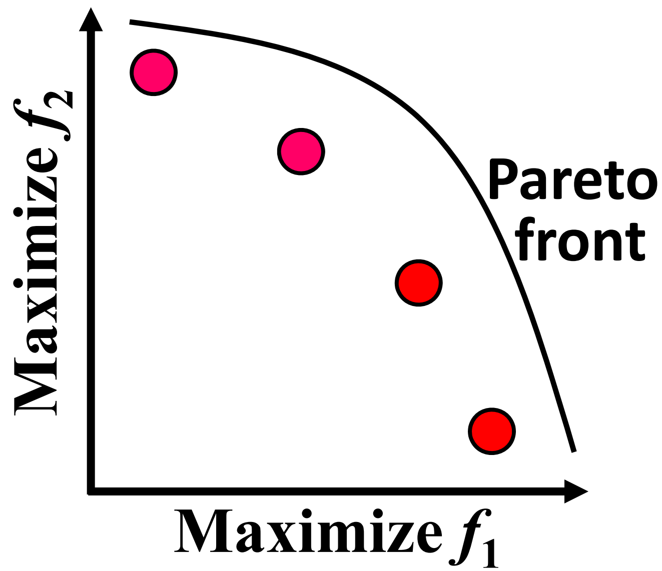
The final result of optimization is a solution set.

Comparison of solution sets is not easy.

Which is a better solution set?



Solution Set A



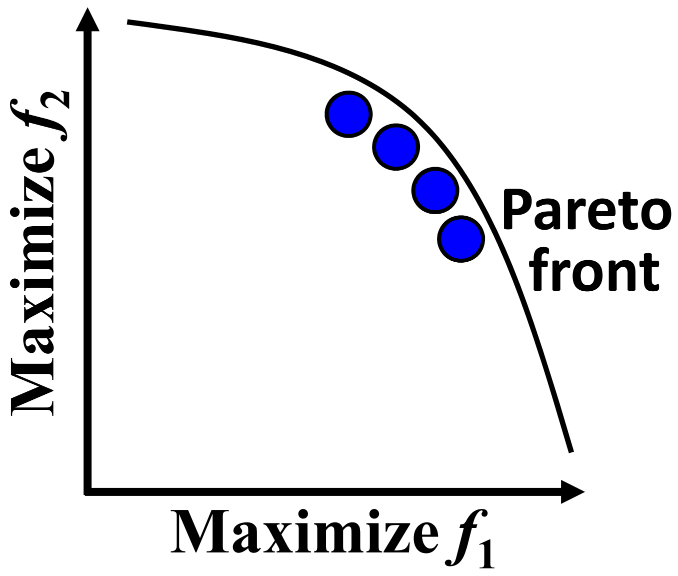
Solution Set B

The final result of optimization is a solution set.

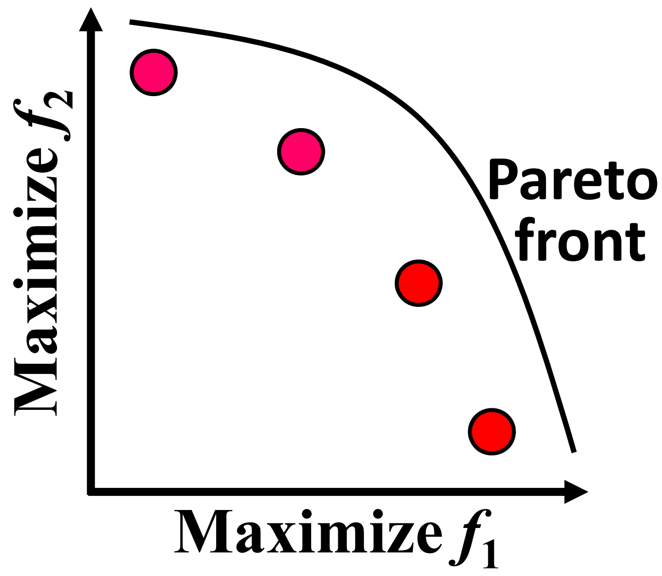
Comparison of solution sets is not easy.

In the case of two objectives, we can understand the quality of each solution set even if we cannot say which is better.

Which is a better solution set?



Solution Set A

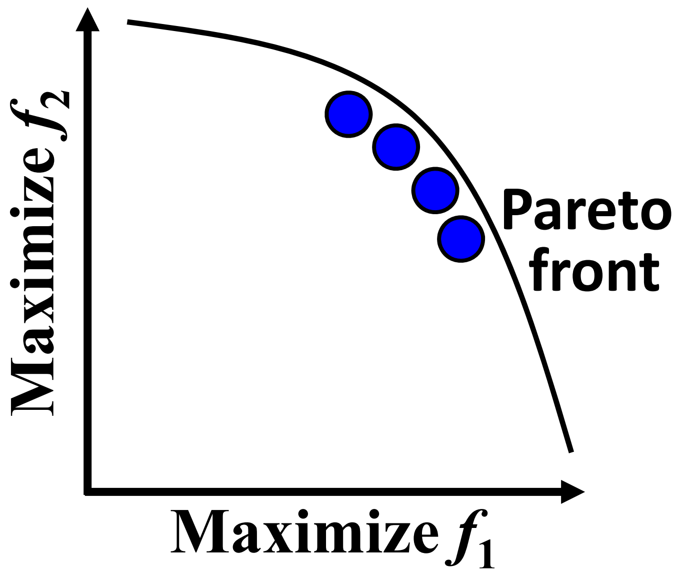


Solution Set B

The final result of optimization is a solution set.
Comparison of solution sets is not easy.

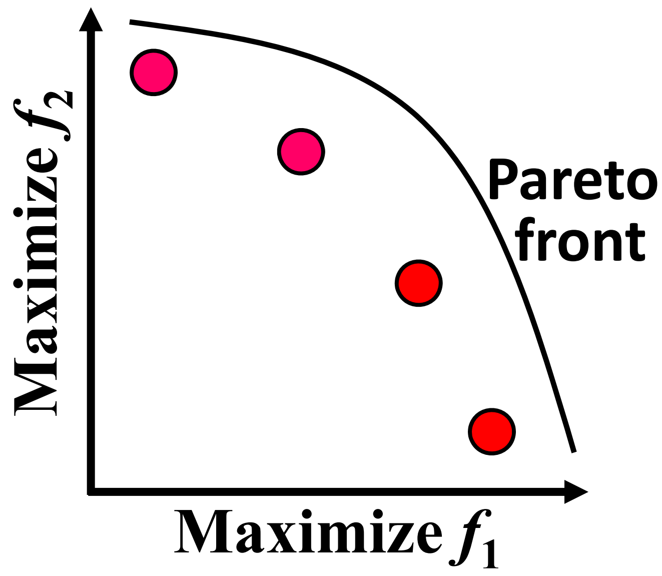
In the case of two objectives, we can understand the quality of each solution set even if we cannot say which is better.

Good Convergence



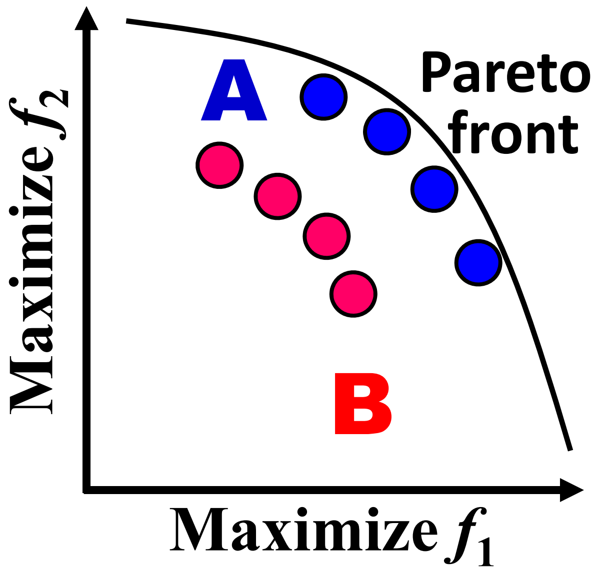
Solution Set A

Good Diversity

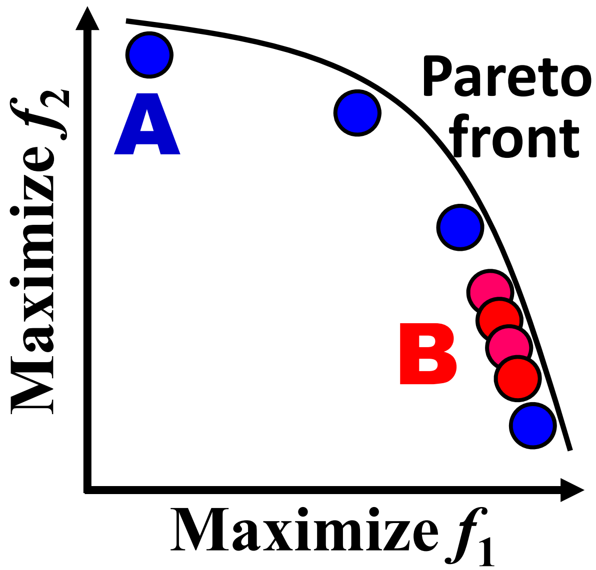


Solution Set B

Which is a better solution set between A and B?

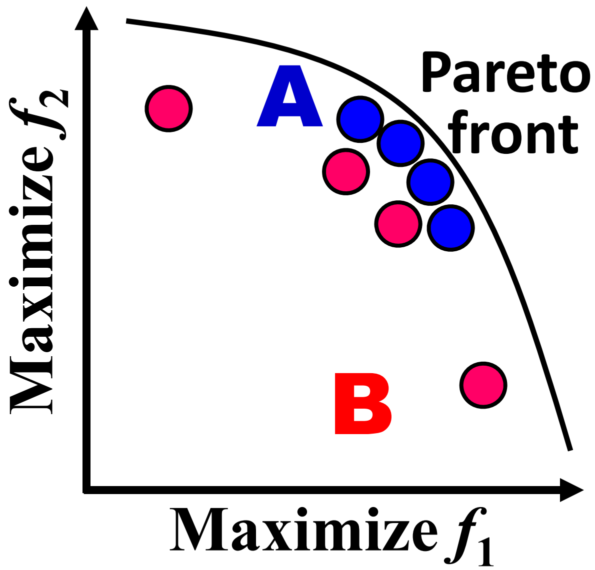


A is clearly better.

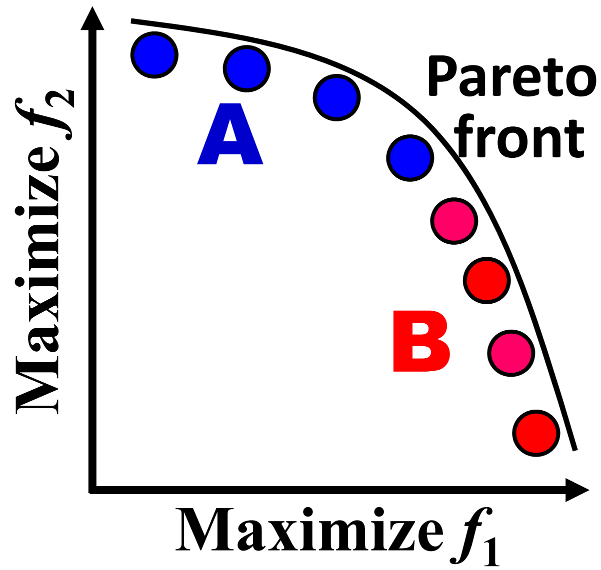


A is better.

Which is a better solution set between A and B?

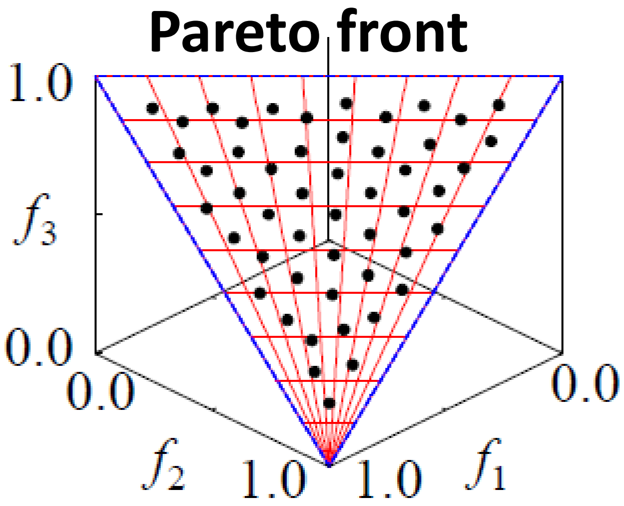


A seems to be better.

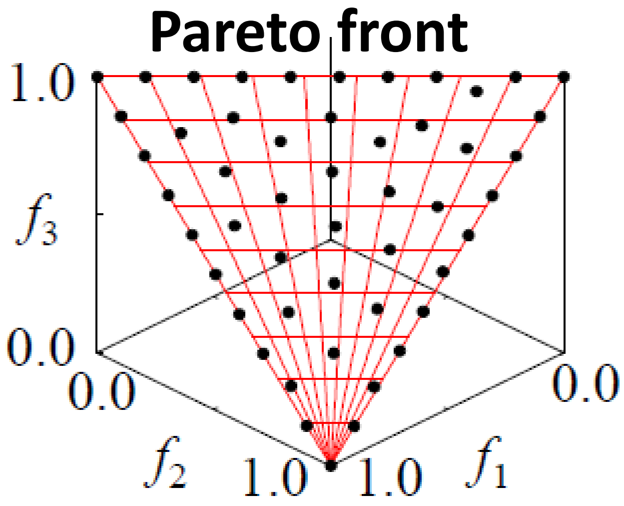


Which is better ?
(Not clear)

Which is a better solution set between A and B?

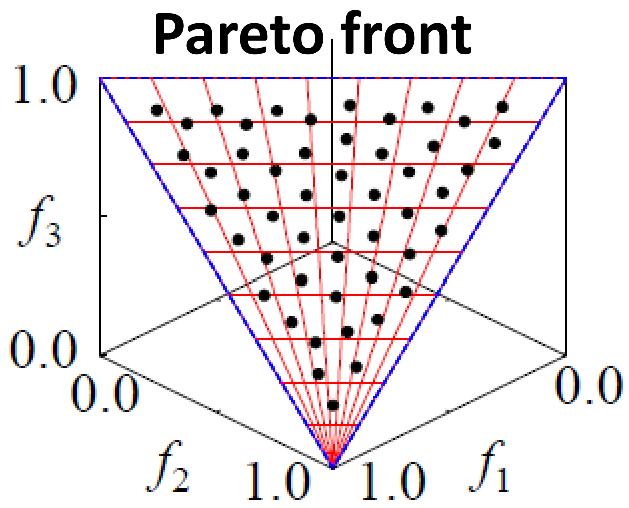


Solution Set A

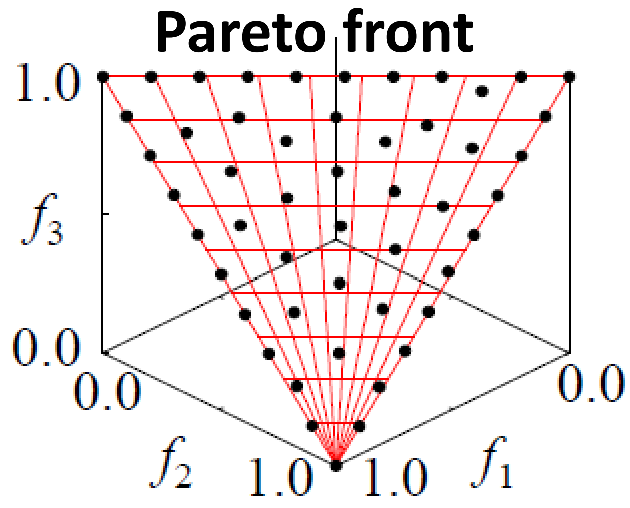


Solution Set B

Which is a better solution set between A and B?



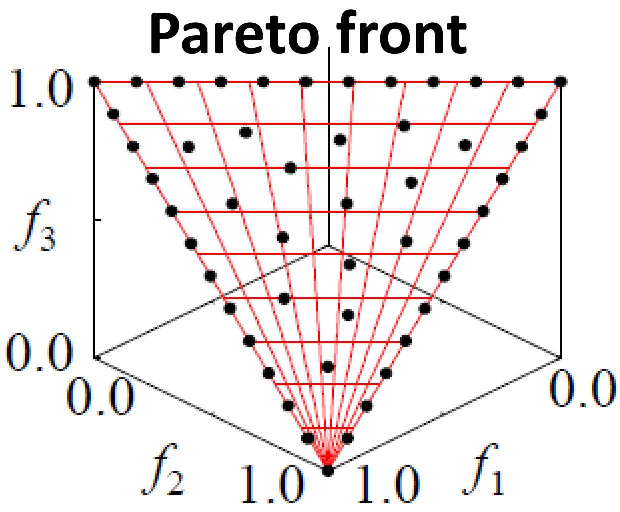
Solution Set A

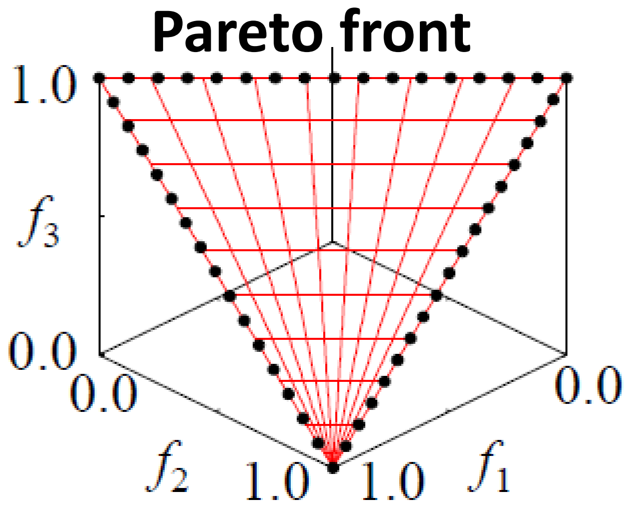


Solution Set B

B is better since B covers the entire Pareto Front.

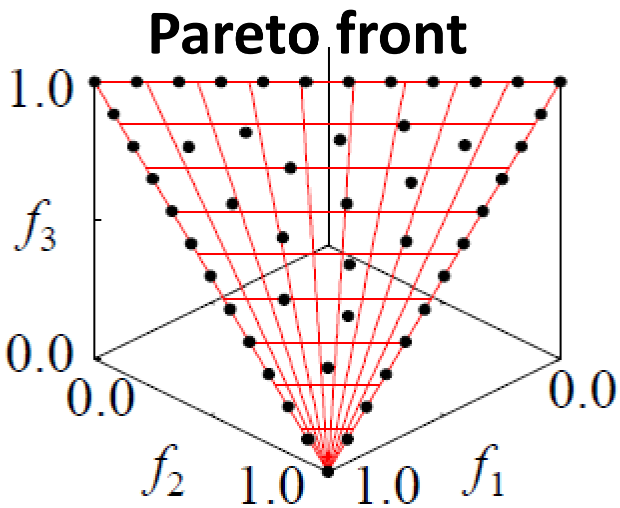
Which is a better solution set between A and B?



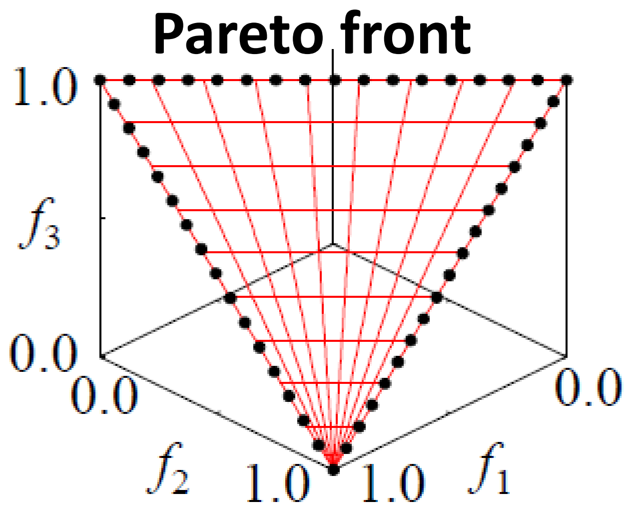


Solution Set A **Solution Set B**

Which is a better solution set between A and B?



Solution Set A



Solution Set B

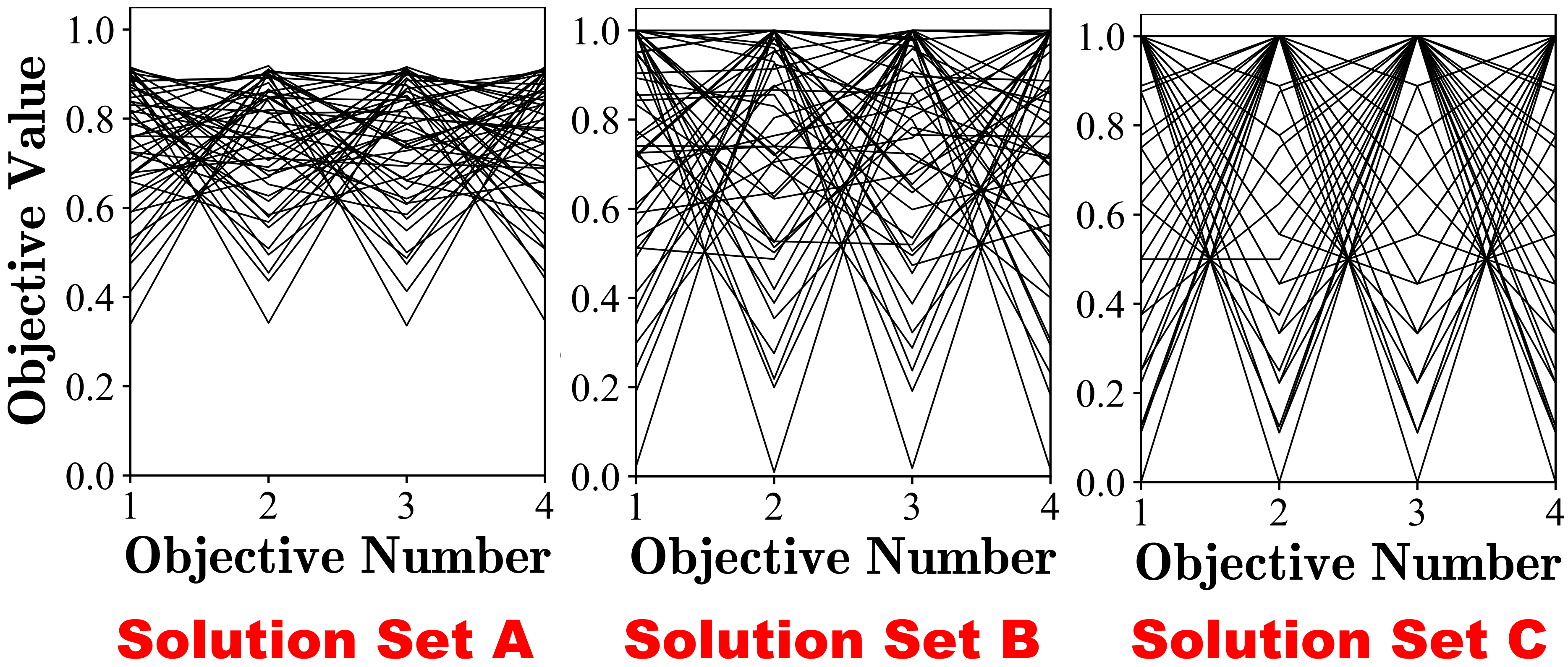
A is better since A covers the entire Pareto Front.

Visual comparison is easy for some cases.

Four-Objective Problems (Parallel Coordinates)

Which is the best solution set among A, B and C?

(Minimization. All solutions are on the Pareto front in $[0, 1]^4$)

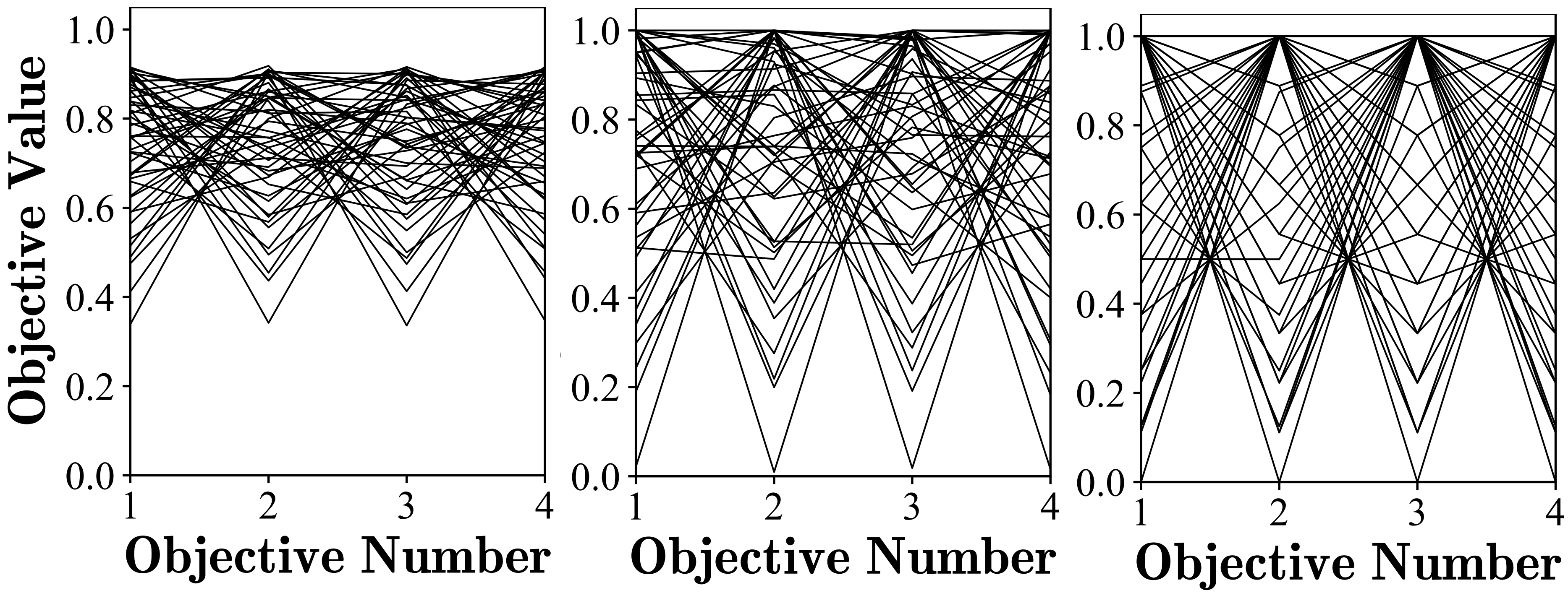


H. Ishibuchi et al., "Optimal distributions of solutions for hypervolume maximization on triangular and inverted triangular Pareto fronts of four-objective Problems," IEEE SSCI 2019, pp. 1857-1864.

Four-Objective Problems (Parallel Coordinates)

Which is the best solution set among A, B and C?

(Minimization. All solutions are on the Pareto front in $[0, 1]^4$)



Solution Set A **Solution Set B** **Solution Set C**

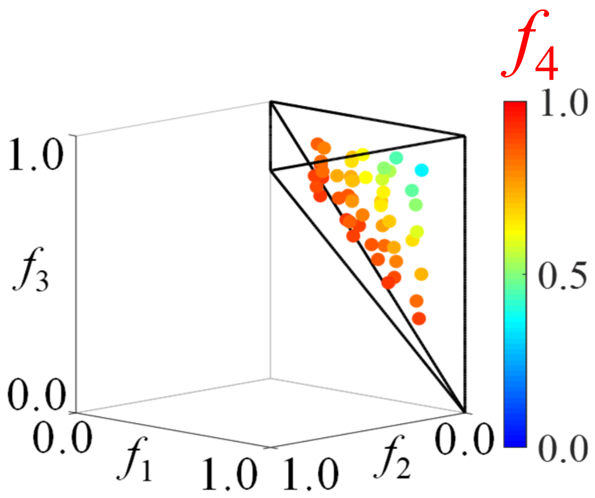
Solution Set A is clearly bad.

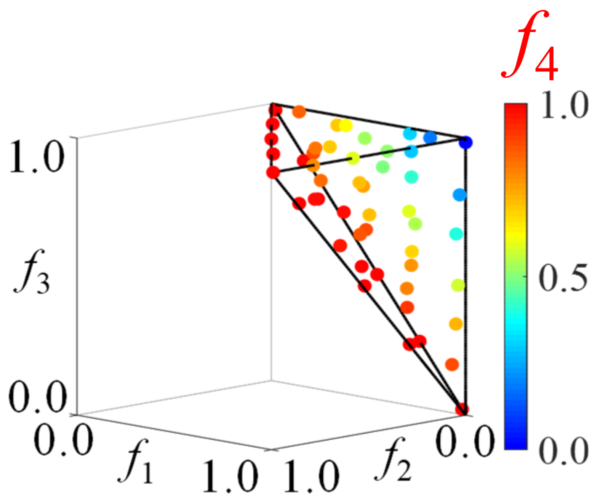
Solution set C looks better than B, or B is better than C ?

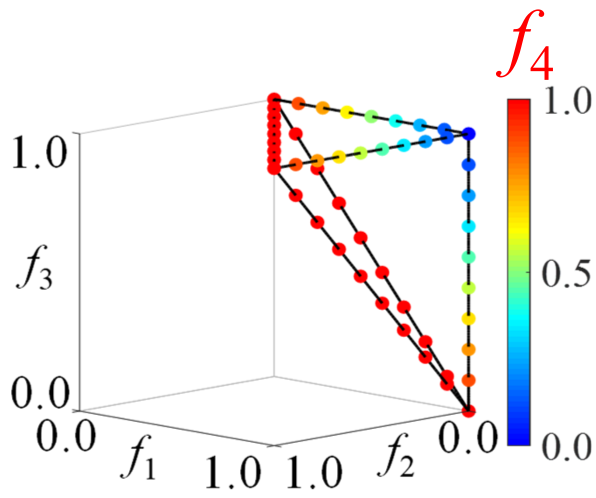
Four-Objective Problems (Parallel Coordinates)

Which is the best solution set among A, B and C?

(Minimization. All solutions are on the Pareto front in $[0, 1]^4$)





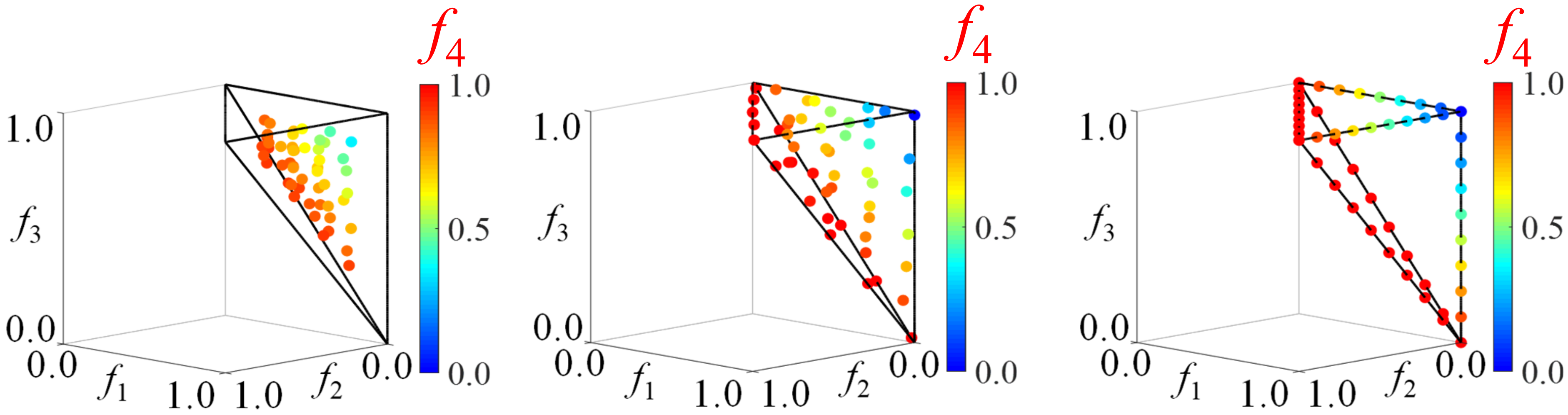


Solution Set A **Solution Set B** **Solution Set C**

Four-Objective Problems (Parallel Coordinates)

Which is the best solution set among A, B and C?

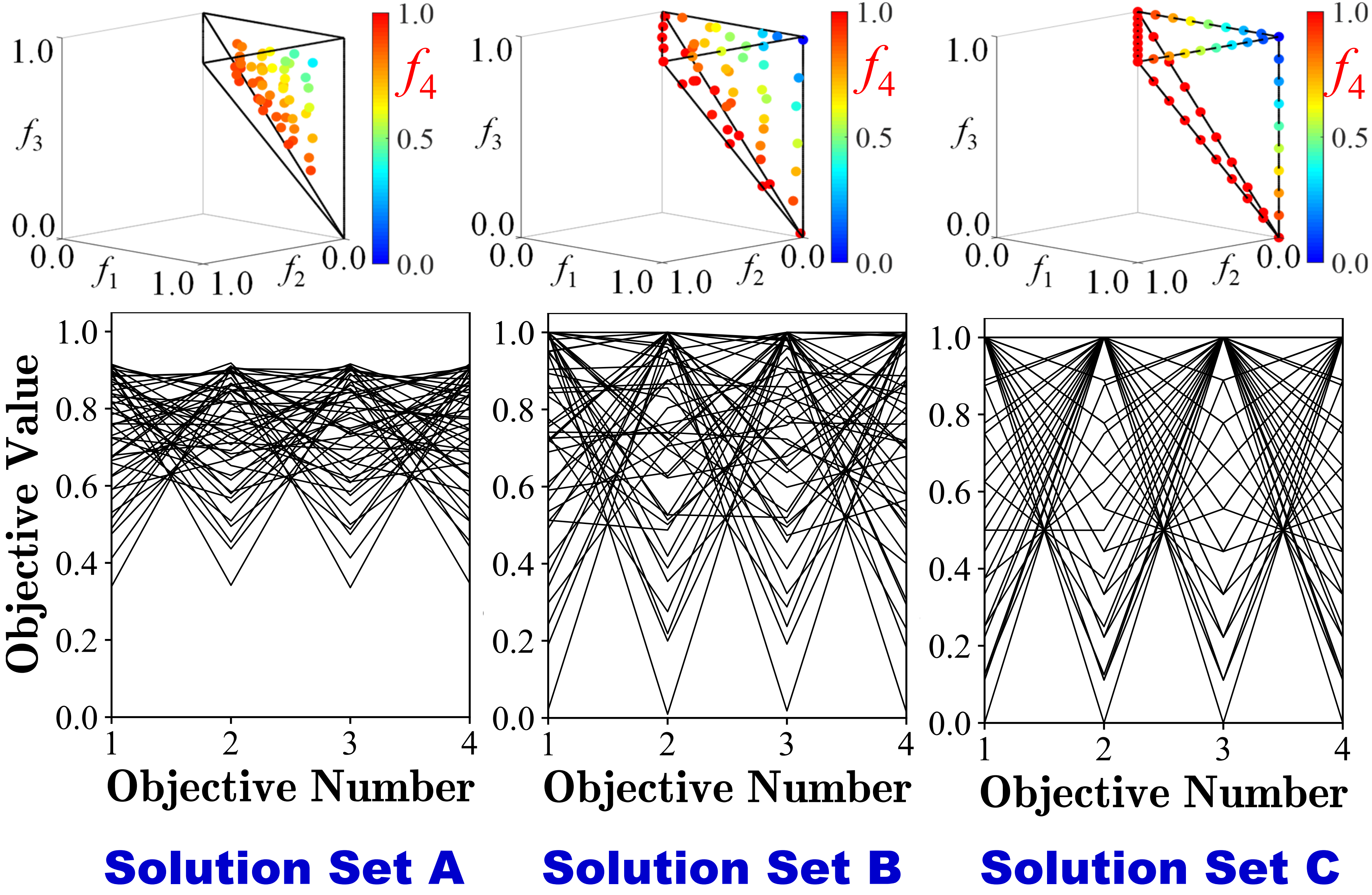
(Minimization. All solutions are on the Pareto front in $[0, 1]^4$)



Solution Set A **Solution Set B** **Solution Set C**

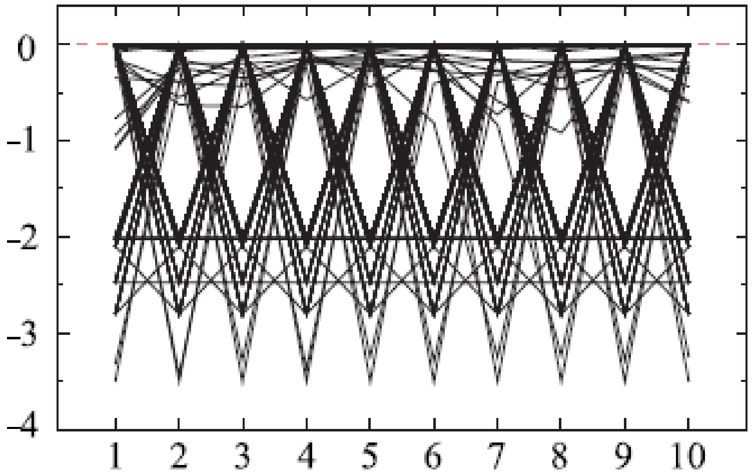
Solution Set B is clearly the best.

A is only inside, and C is only on the boundary.

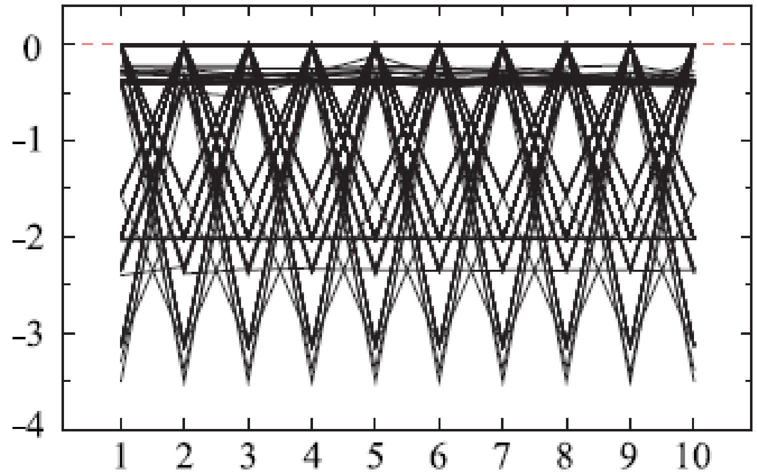


Visual comparison is difficult ➡ Indicators: HV, IGD

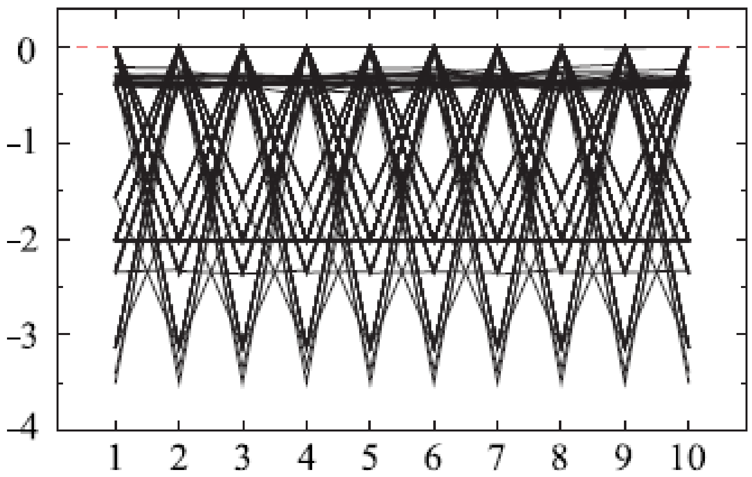
Visual comparison is very difficult for many-objective problems



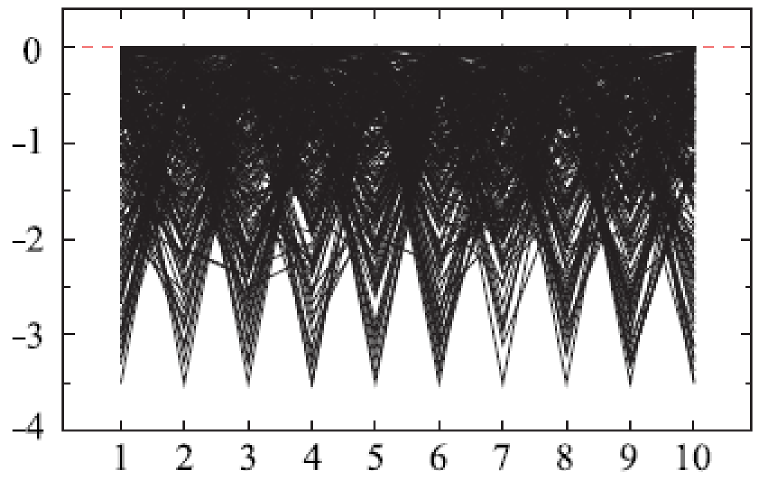
MOEA/D-Tch



MOEA/D-WS



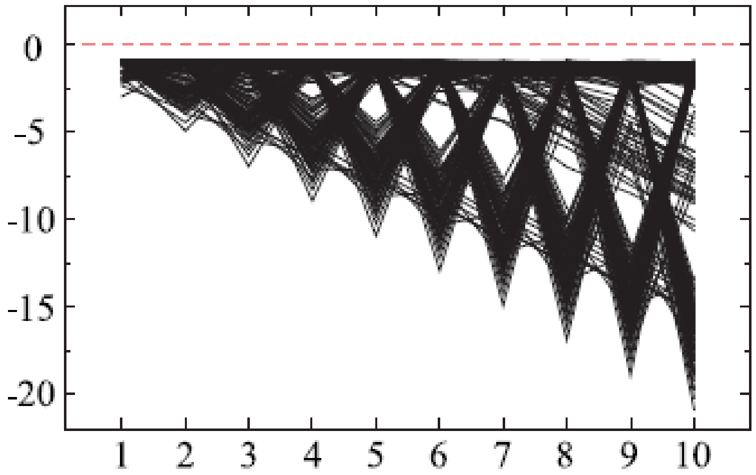
MOEA/D-IPBI (0.1)



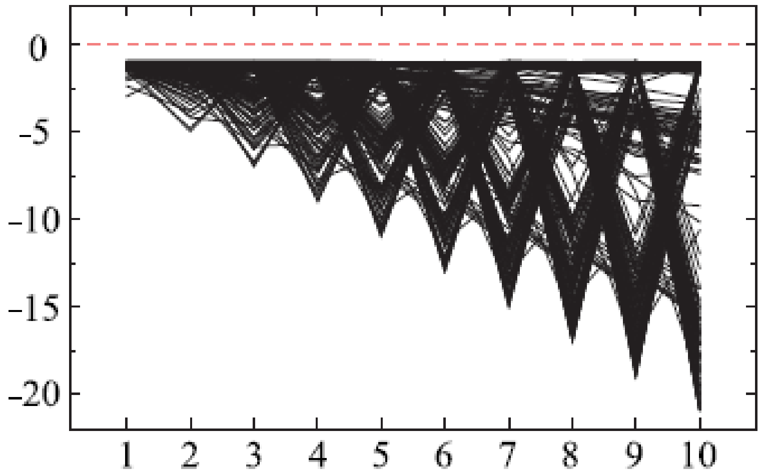
NSGA-II

Minus-DTLZ2

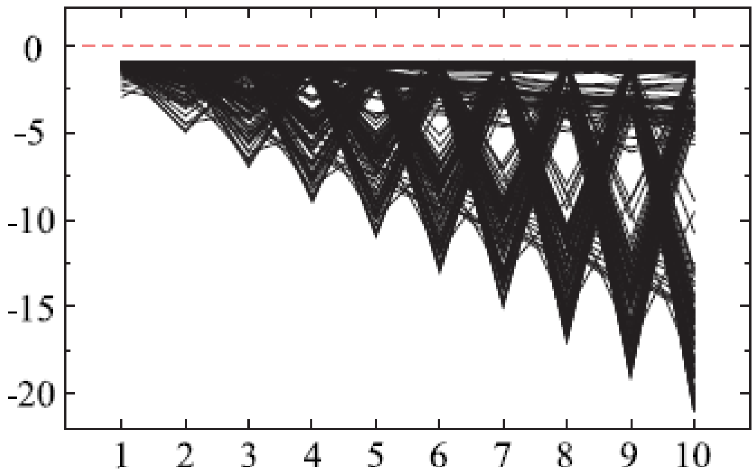
Visual comparison is very difficult for many-objective problems



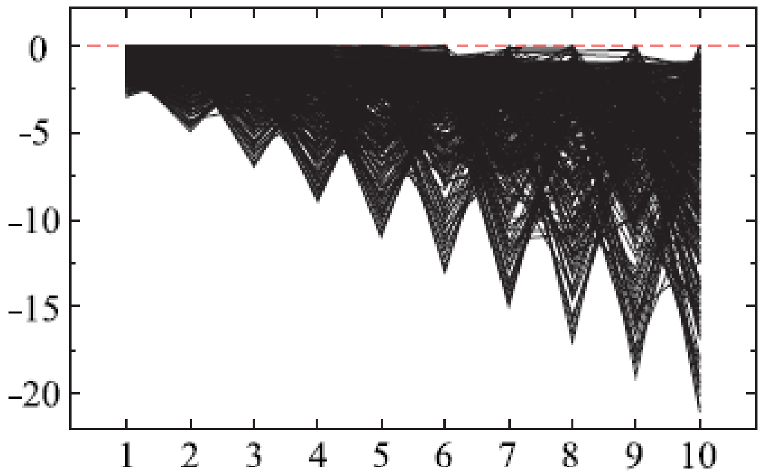
MOEA/D-Tch



MOEA/D-WS

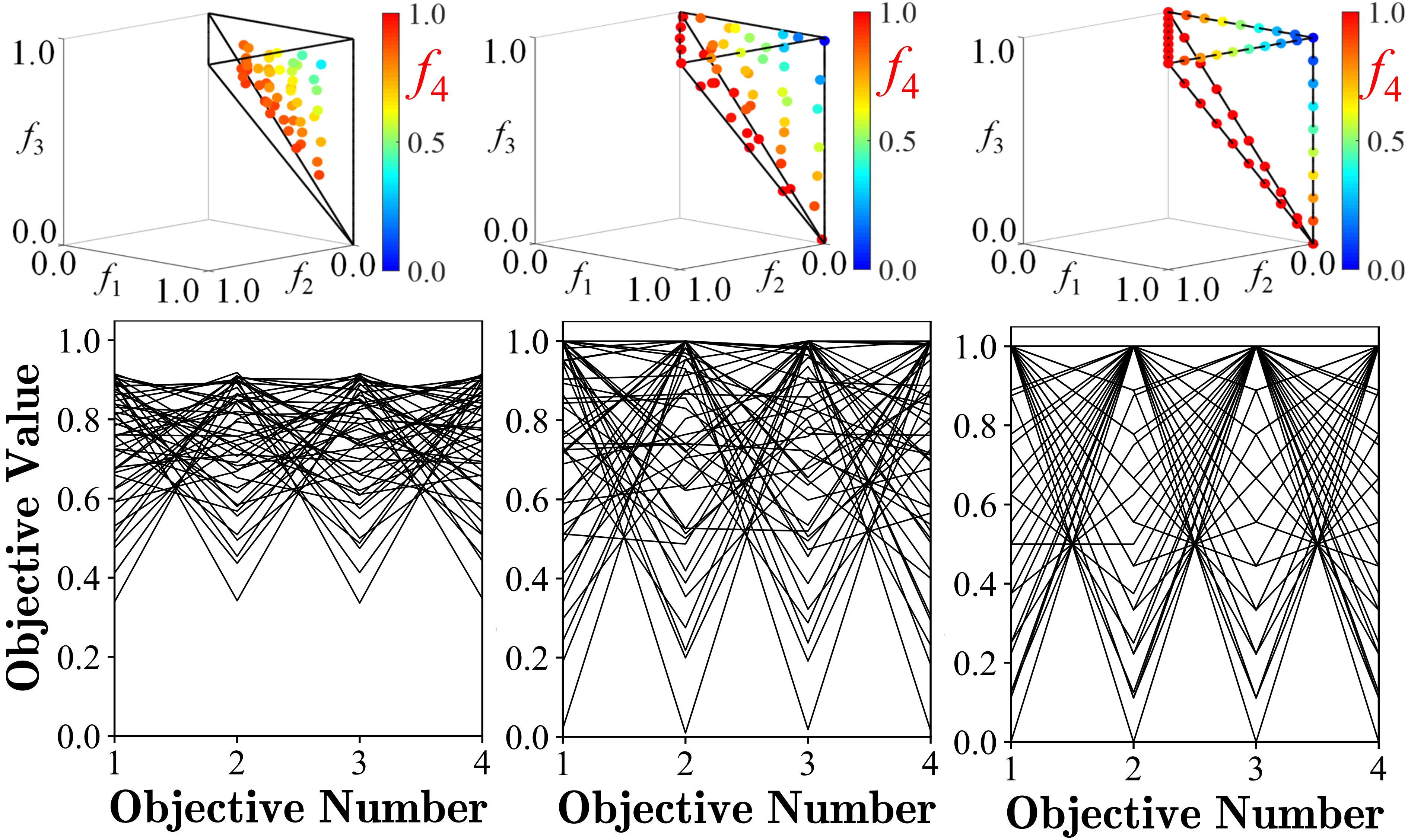


MOEA/D-IPBI (0.1)



NSGA-II

Minus-WFG2

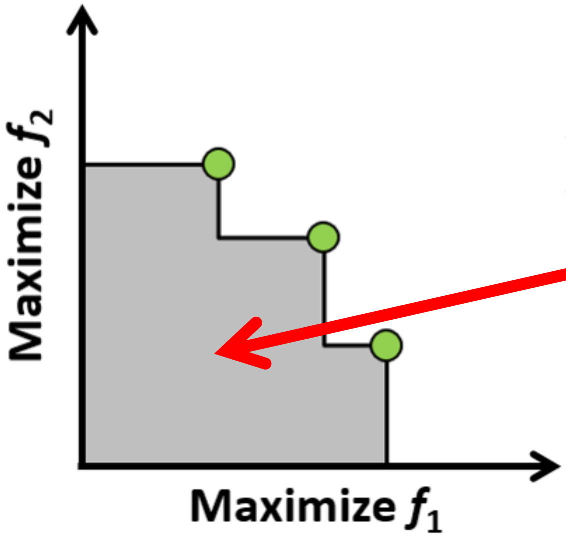


Solution Set A **Solution Set B** **Solution Set C**

Visual comparison is difficult → **Indicators: HV, IGD**

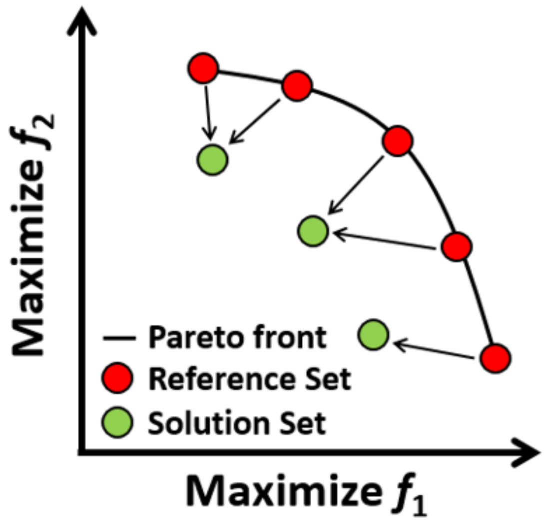
Frequently-Used Performance Indicators

- Hypervolume (HV) Indicator
- Inverted Generational Distance (IGD) Indicator.



HV Indicator

Size of this dominate region



IGD Indicator

Average distance from each reference point to the nearest solution (Average arrow length)

Difficulties in Fair Performance Comparison of Evolutionary Multi-Objective Optimization Algorithms

(0) Visual Comparison

(1) Specification of Termination Condition

(2) Specification of Population Size

(3) Choice of Performance Indicators (e.g., HV, IGD)

(4) Setting in Performance Indicators (e.g., reference point)

(5) Choice of Test Problems

This Talk is mainly based on my recent paper:

Hisao Ishibuchi, Lie Meng Pang, and Ke Shang, "**Difficulties in Fair Performance Comparison of Multi-Objective Evolutionary Algorithms**" ***IEEE Computational Intelligence Magazine*** **(February 2022)**

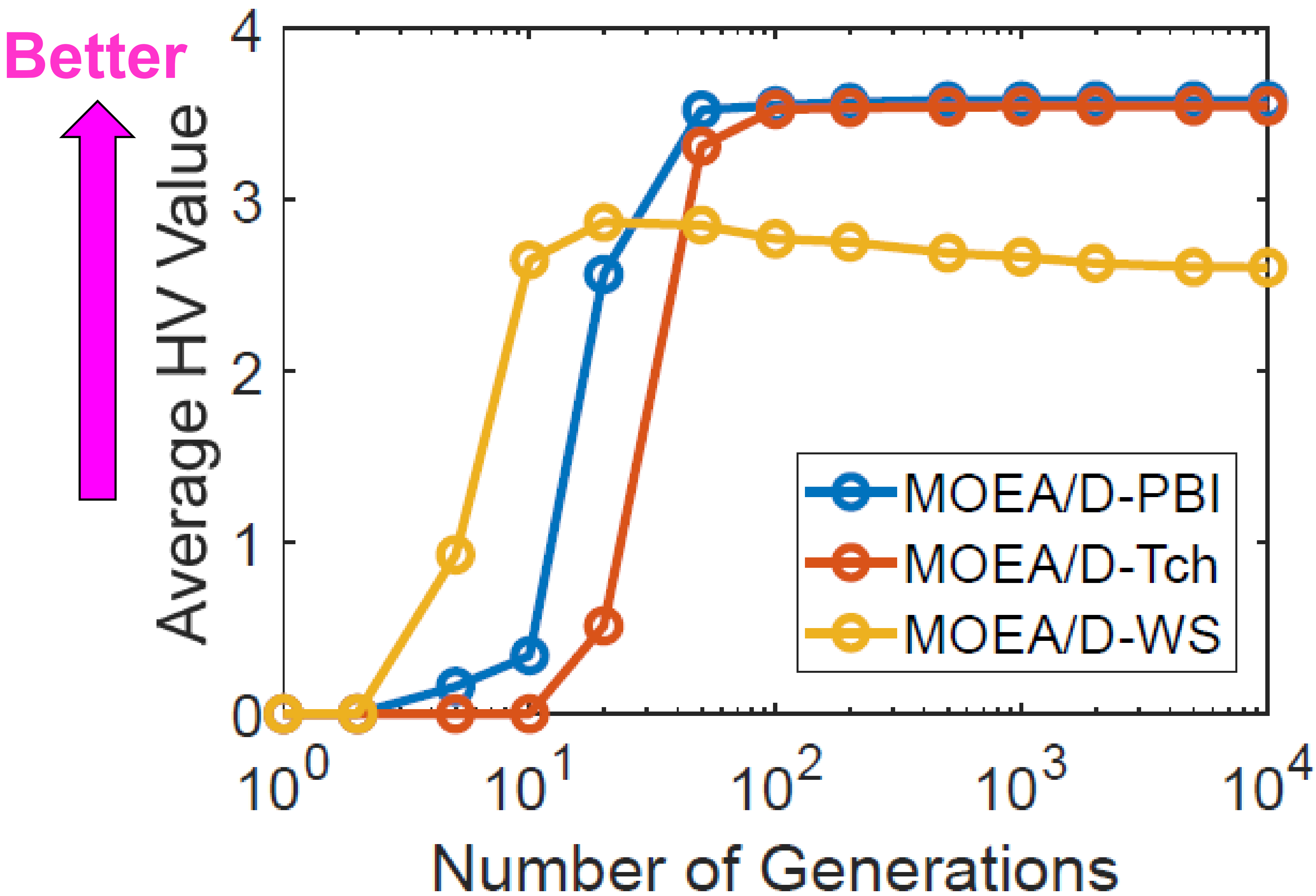


Fig. 5. Average HV results for 7-objective HTNY19.

[Ishibuchi et al. IEEE CIM 2022]

Which is the best algorithm ?

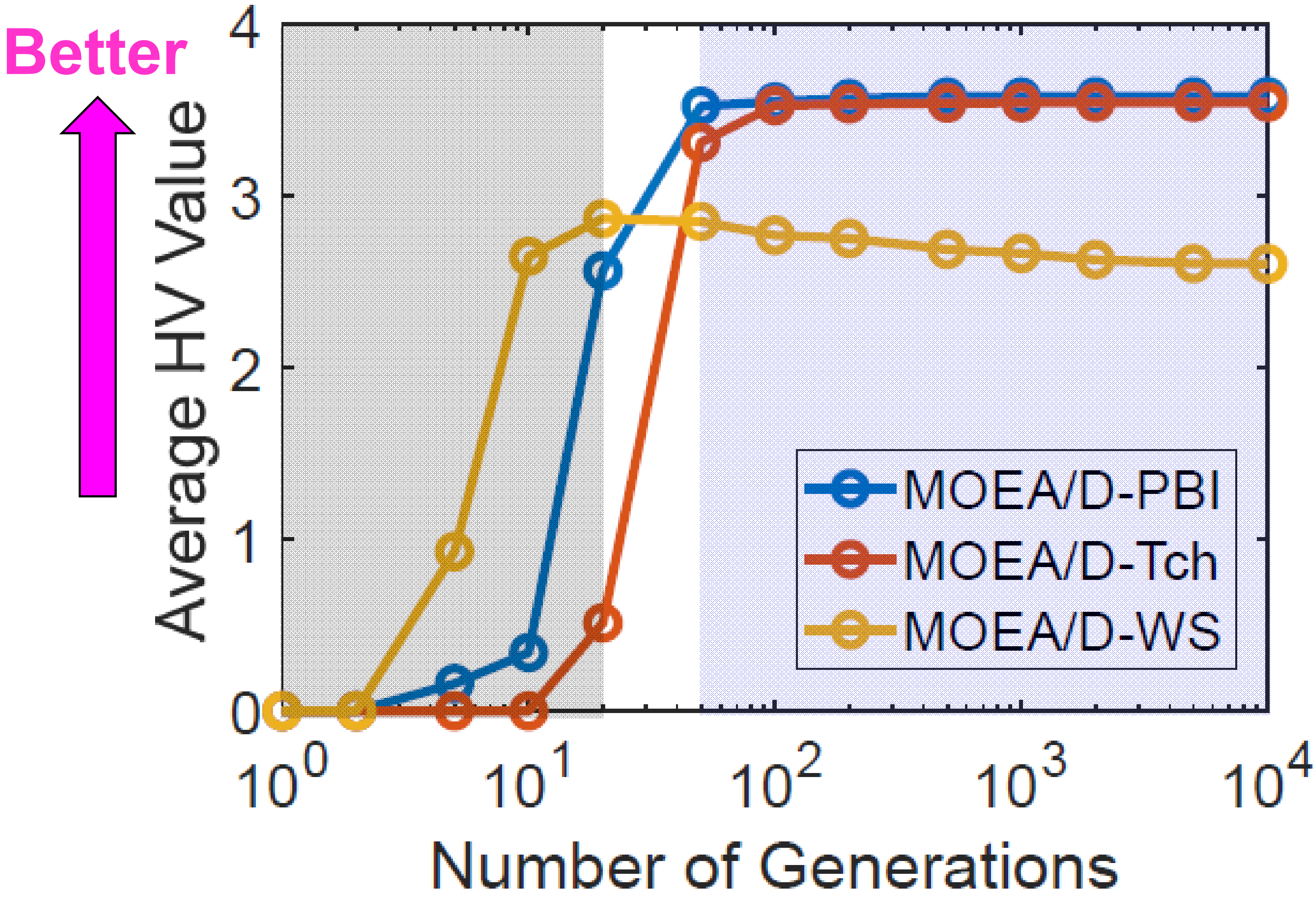


Fig. 5. Average HV results for 7-objective HTNY19.

Which is the best algorithm ?

[Ishibuchi et al. IEEE CIM 2022]

Totally different termination conditions have been used in the literature in the EMO community:

ParEGO Paper [1]: 100 - 260 solution evaluations **($\sim 10^2$)**
Expensive problems for expensive multi-objective optimization

NSGA-III Paper [2]: 22.75 - 405 thousand solution evaluations
Standard problems for many-objective optimization **($\sim 10^2$ K)**

LMEA Paper [3]: 1 - 230 million solution evaluations **($\sim 10^2$ M)**
Large-scale problems for large-scale many-objective optimization.

[1] J. Knowles, "ParEGO: A hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems", IEEE Transactions on Evolutionary Computation (2005).

[2] K. Deb and H. Jain, "An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, Part I: Solving problems with box constraints," IEEE Transactions on Evolutionary Computation (2014).

[3] X. Zhang, Y. Tian, R. Cheng, and Y. Jin, "A decision variable clustering-based evolutionary algorithm for large-scale many-objective optimization," IEEE Transactions on Evolutionary Computation (2018).

Observation.

Our experimental results strongly depend on the termination condition.

Question.

How to specify the termination condition?

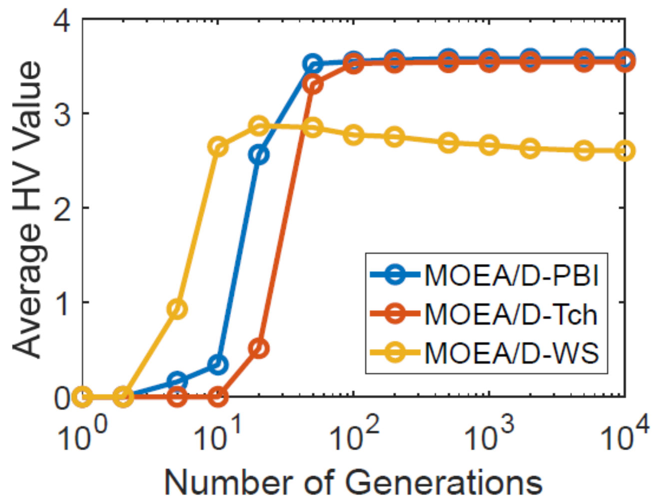


Fig. 5. Average HV results for 7-objective HTNY19.

Observation.

Our experimental results strongly depend on the termination condition.

Question.

How to specify the termination condition?

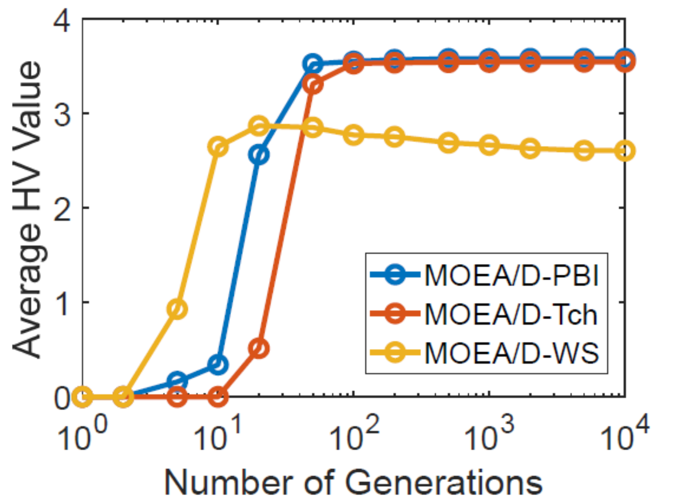


Fig. 5. Average HV results for 7-objective HTNY19.

Suggestion 1:

To use the same termination condition as in many other papers especially in well-known papers (e.g., 300 solution evaluations in expensive multi-objective optimization).

Observation.

Our experimental results strongly depend on the termination condition.

Question.

How to specify the termination condition?

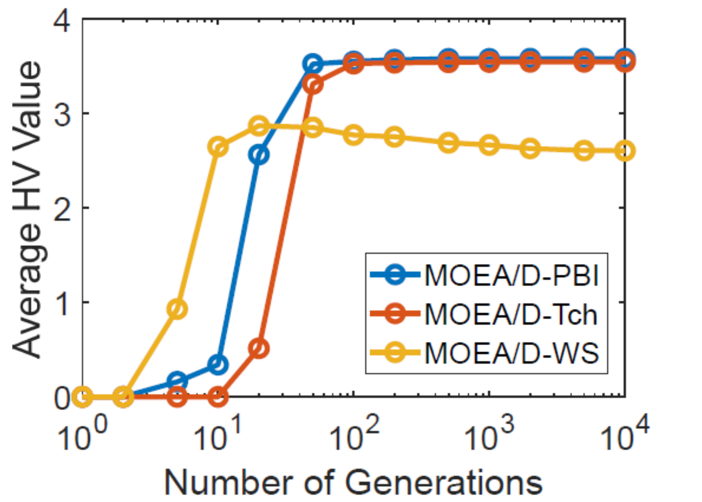


Fig. 5. Average HV results for 7-objective HTNY19.

Suggestion 1:

To use the same termination condition as in many other papers especially in well-known papers (e.g., 300 solution evaluations in expensive multi-objective optimization).

Suggestion 2:

To use various termination conditions, and compare different algorithms under the anytime algorithm framework.

Difficulties in Fair Performance Comparison of Evolutionary Multi-Objective Optimization Algorithms

(0) Visual Comparison

(1) Specification of Termination Condition

(2) Specification of Population Size

(3) Choice of Performance Indicators (e.g., HV, IGD)

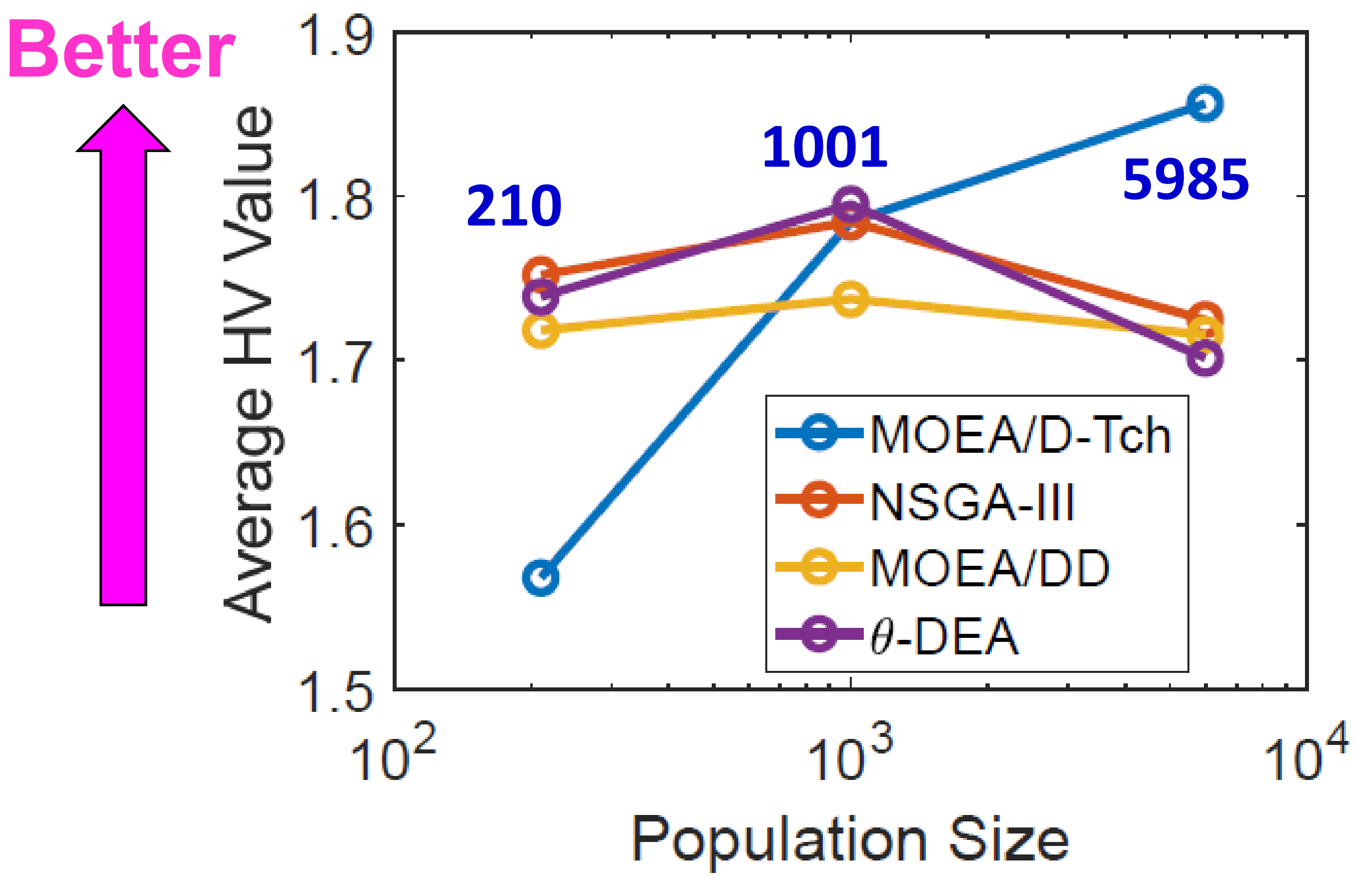
(4) Setting in Performance Indicators (e.g., reference point)

(5) Choice of Test Problems

This Talk is mainly based on my recent paper:

Hisao Ishibuchi, Lie Meng Pang, and Ke Shang, "**Difficulties in Fair Performance Comparison of Multi-Objective Evolutionary Algorithms**" ***IEEE Computational Intelligence Magazine*** **(February 2022)**

Experimental Results



Population size
x Generations
= Constant

Fig. 10. Average HV results of the final population for five-objective WFG3. [Ishibuchi et al. IEEE CIM 2022]

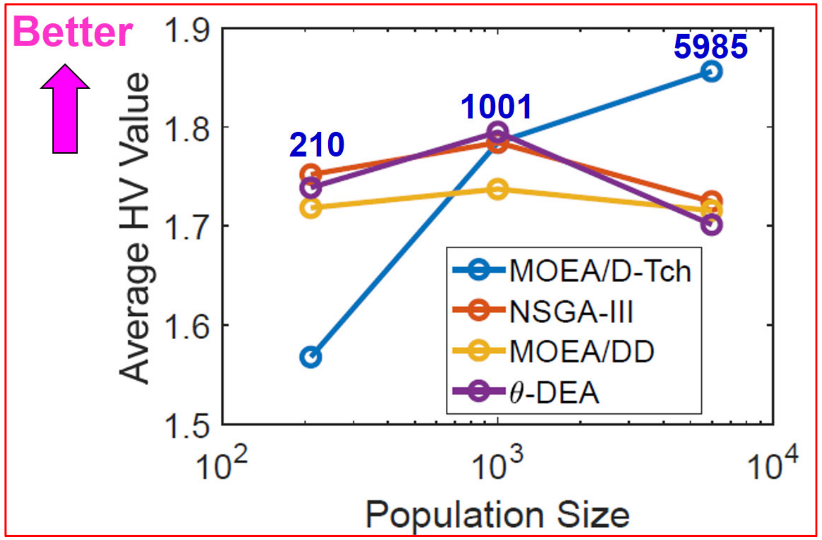
When the four algorithms are compared under the population size 210, MOEA/D is clearly the worst. However, MOEA/D is the best when the population size is very large.

Observation.

Our experimental results strongly depend on the population size.

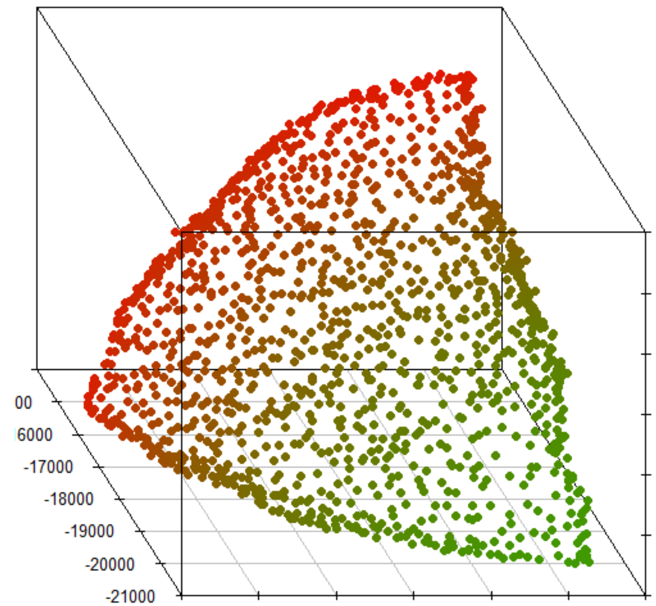
Question.

How to specify the population size?

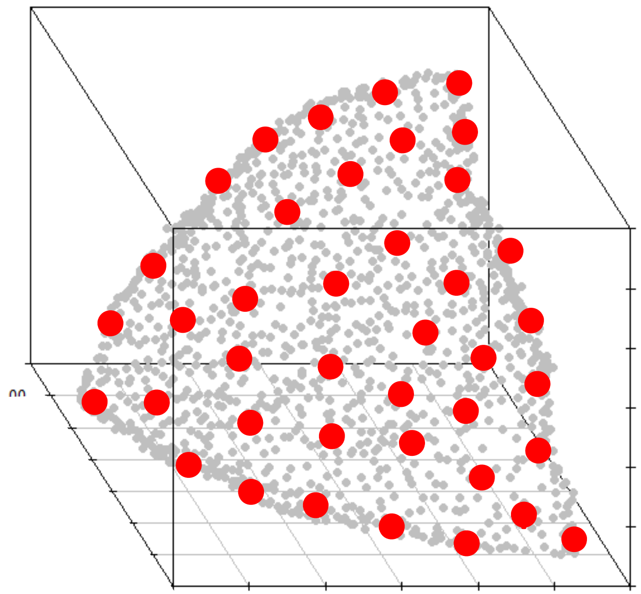


Population size specification is important !

Different specifications for different algorithms are unfair.



Algorithm A
(5000 solutions)

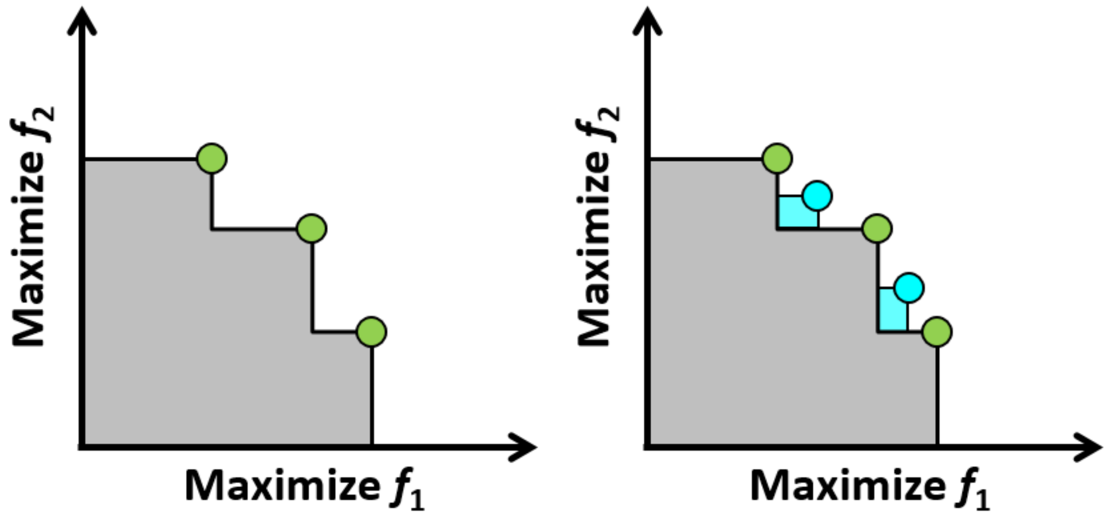


Algorithm B
(50 solutions)

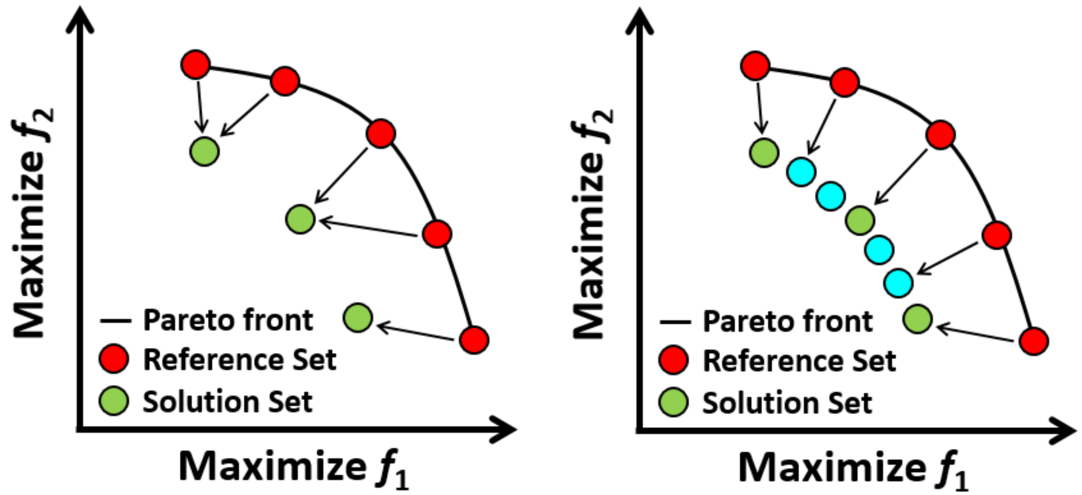
Comparison is unfair since frequently-used performance indicators such as HV and IGD can be improved by increasing the number of non-dominated solutions. Algorithm A will be evaluated as better.

Performance indicator values are usually improved by increasing the number of solutions in the examined solution set.

HV Indicator



IGD Indicator

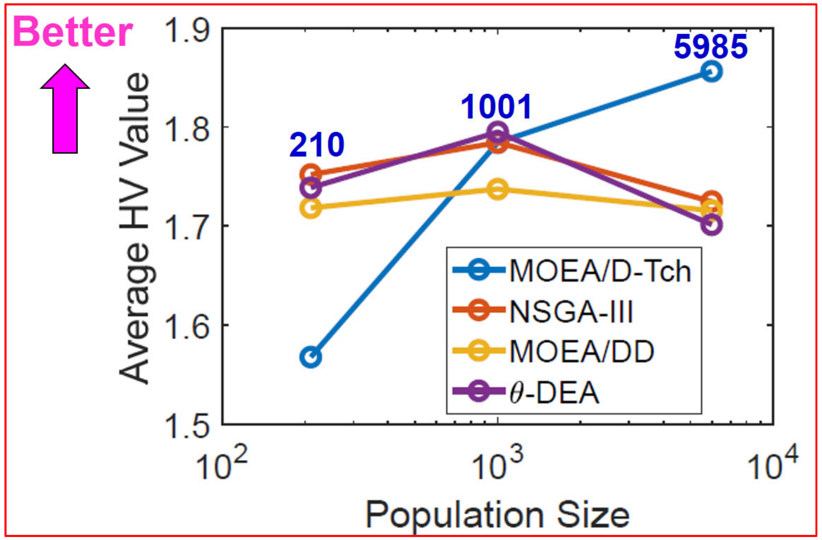


Observation.

Our experimental results strongly depend on the population size.

Question.

How to specify the population size?



Suggestion 1:

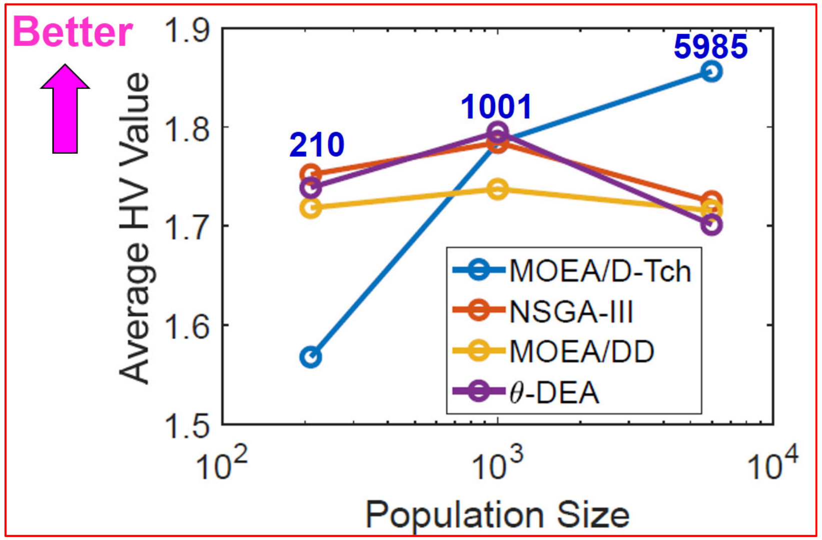
To use the same population size specification as in many other papers especially in well-known papers (e.g., 210).

Observation.

Our experimental results strongly depend on the population size.

Question.

How to specify the population size?



Suggestion 1:

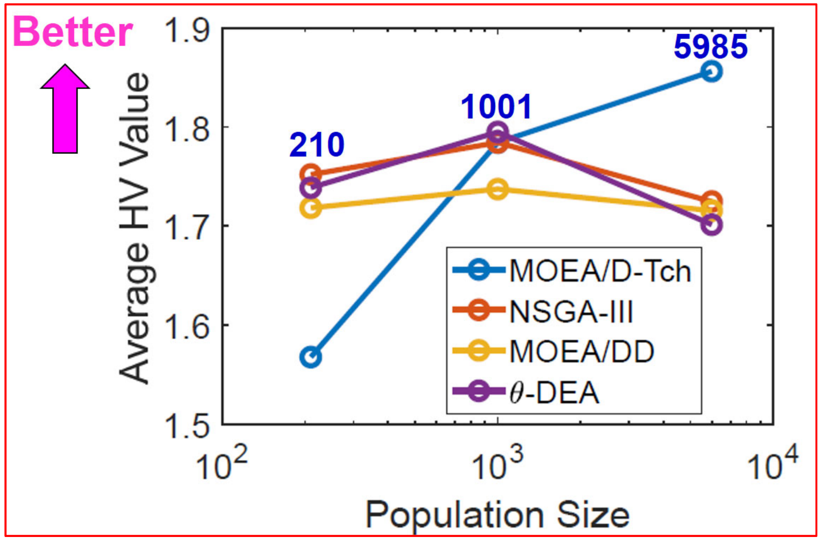
To use the same population size specification as in many other papers especially in well-known papers (e.g., 210). Other settings: 100, 500, ...

Observation.

Our experimental results strongly depend on the population size.

Question.

How to specify the population size?



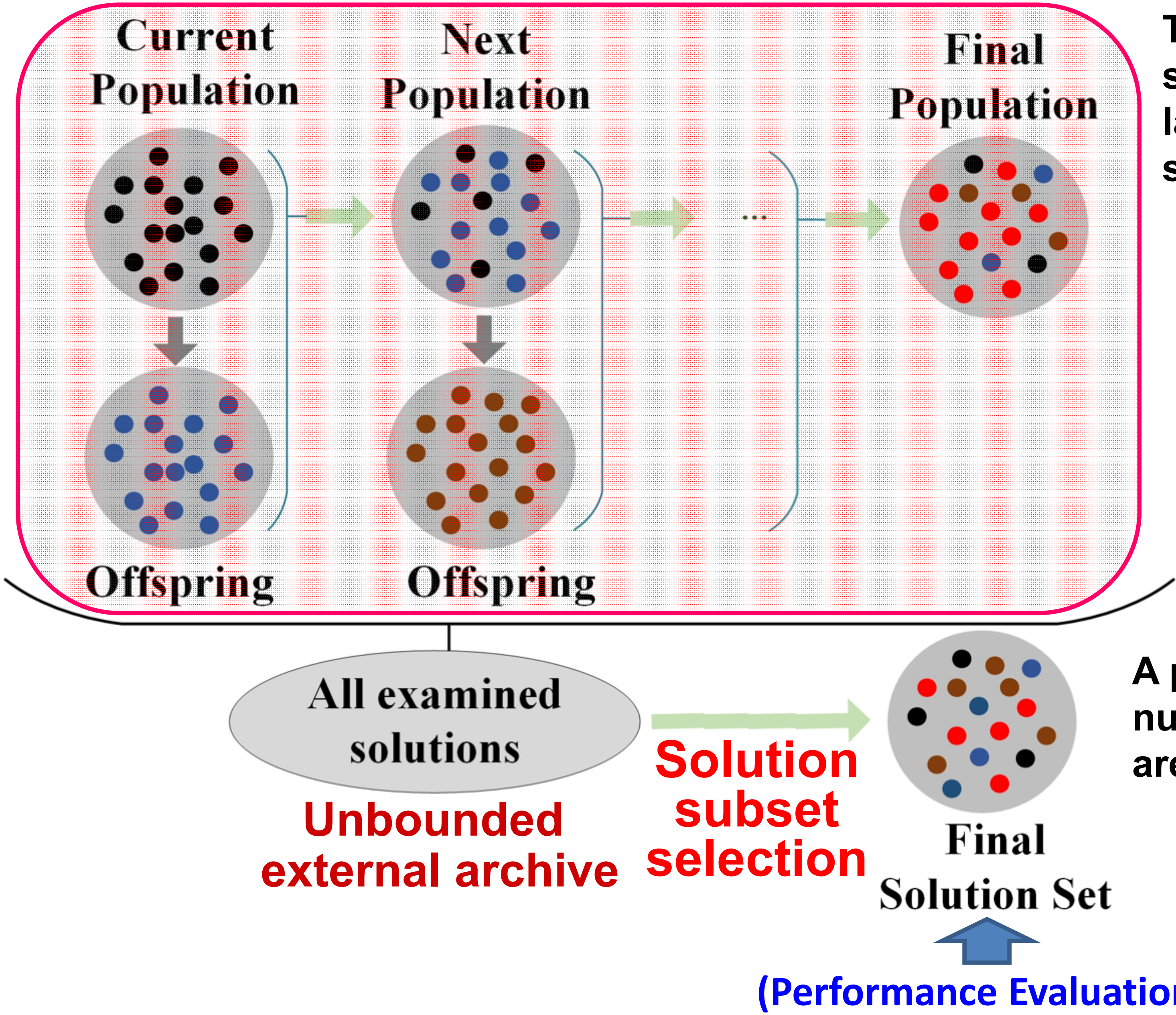
Suggestion 1:

To use the same population size specification as in many other papers especially in well-known papers (e.g., 210). Other settings: 100, 500, ...

Suggestion 2:

To use the best specification for each algorithm, and to select a pre-specified number of solutions from all the examined solutions for performance comparison.

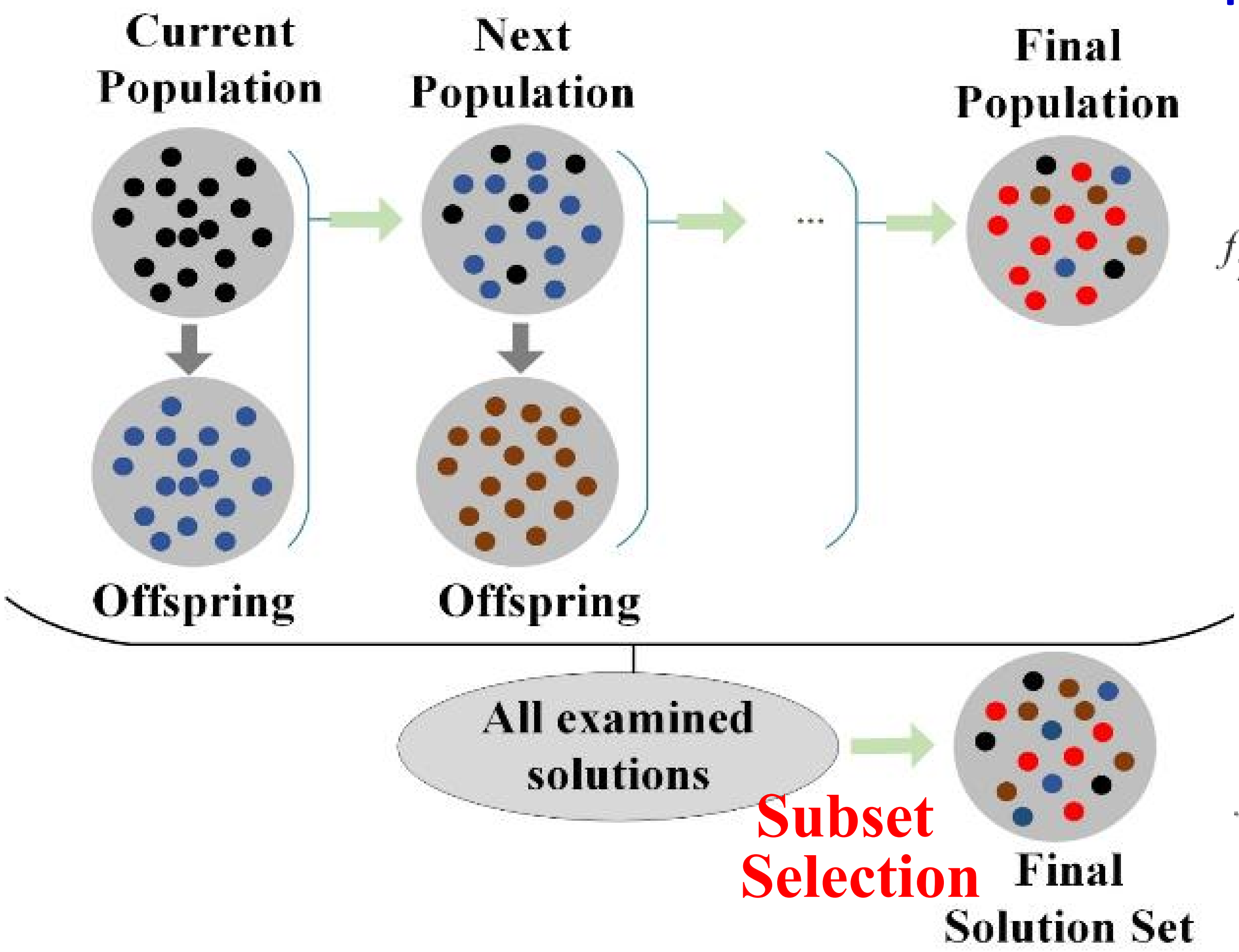
Framework with an Unbounded External Archive



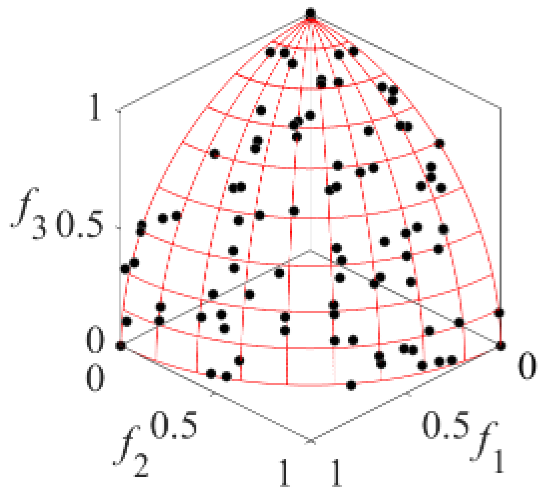
The population size can be very large and very small (if needed).

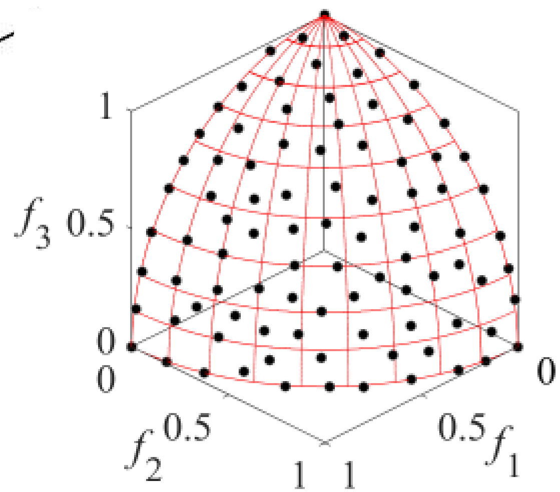
A pre-specified number of solutions are selected.

Final Population vs. Selected Solution Set

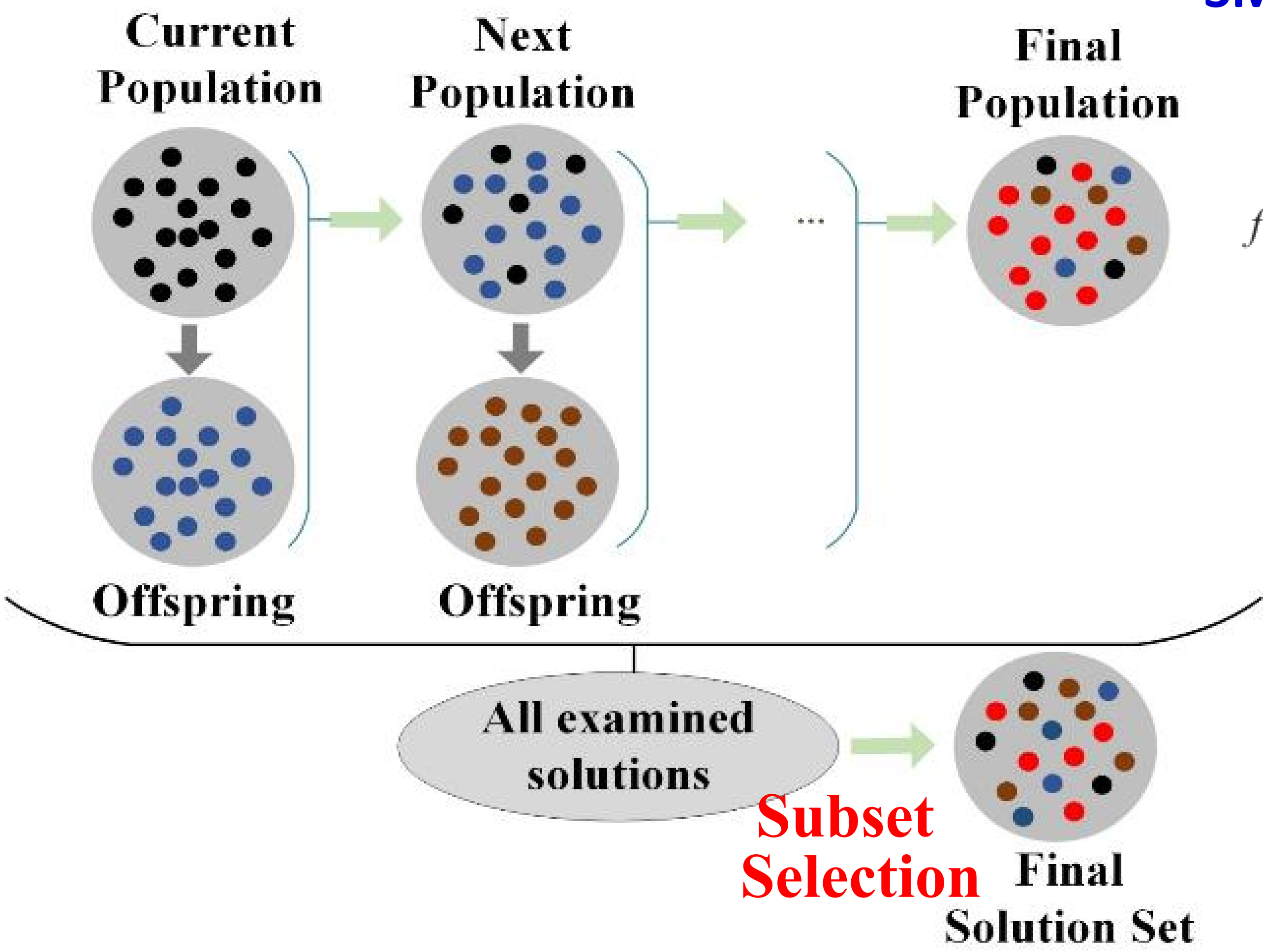


NSGA-II on DTLZ2

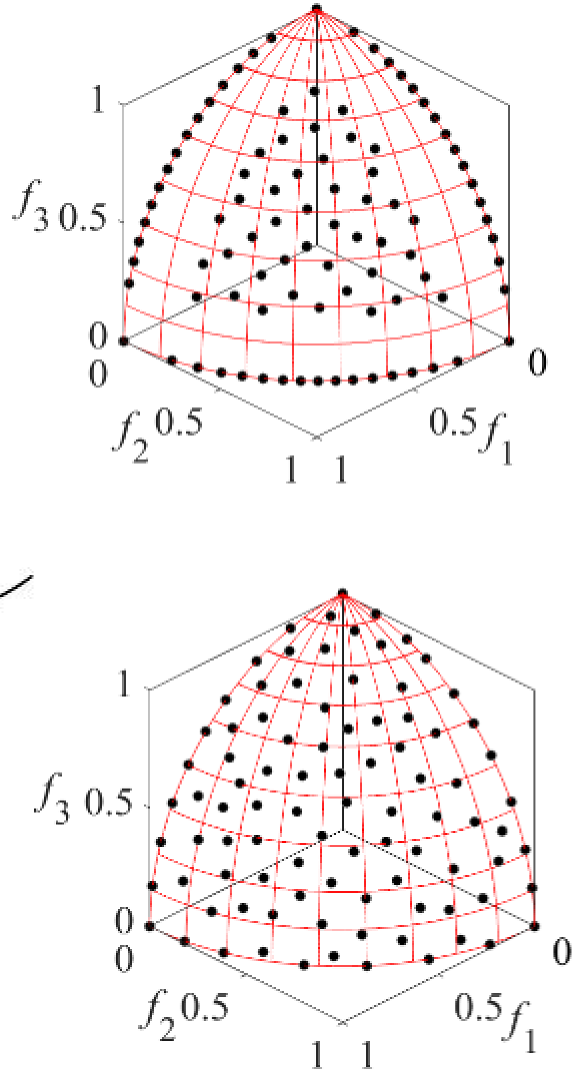




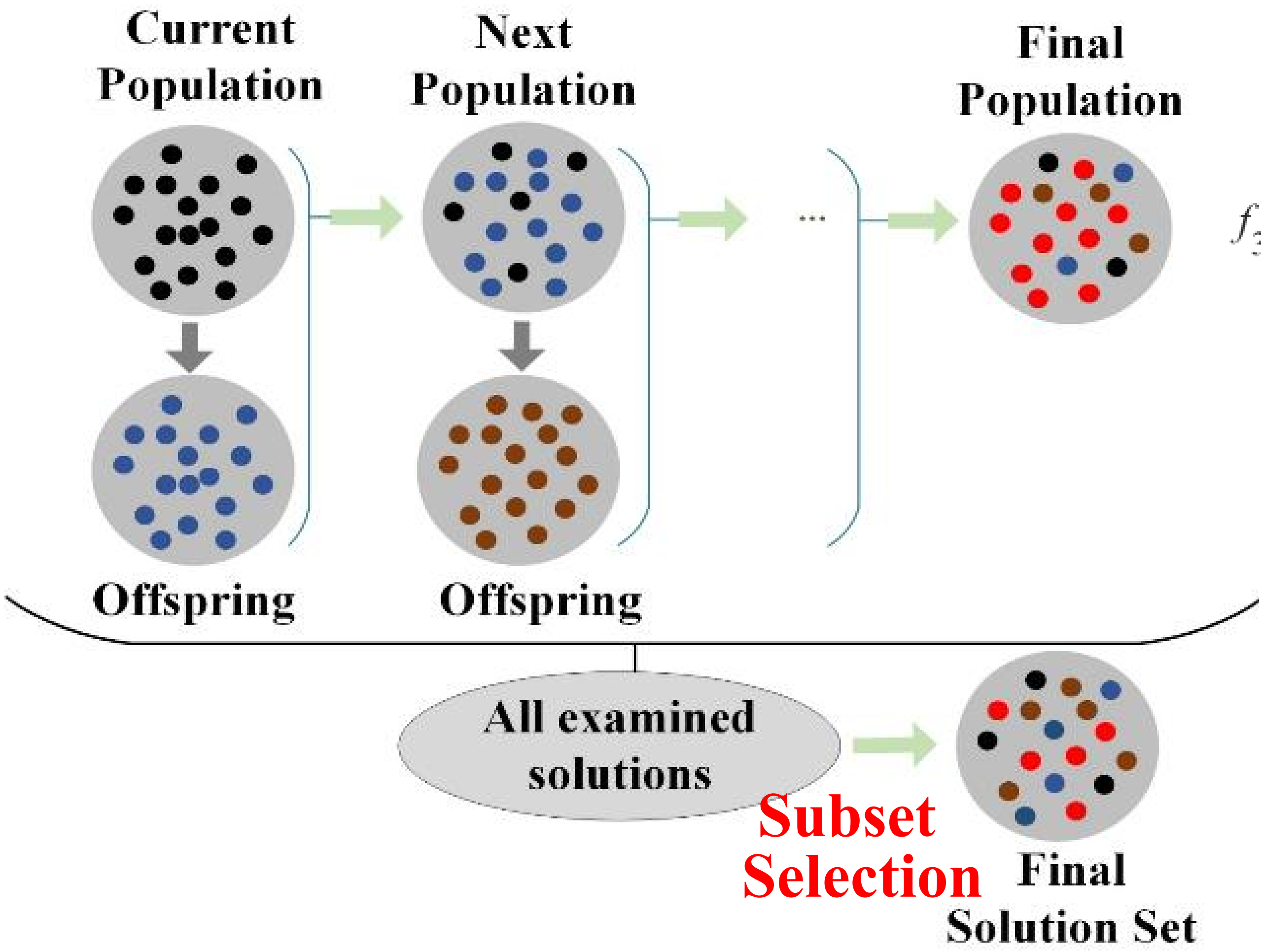
Final Population vs. Selected Solution Set



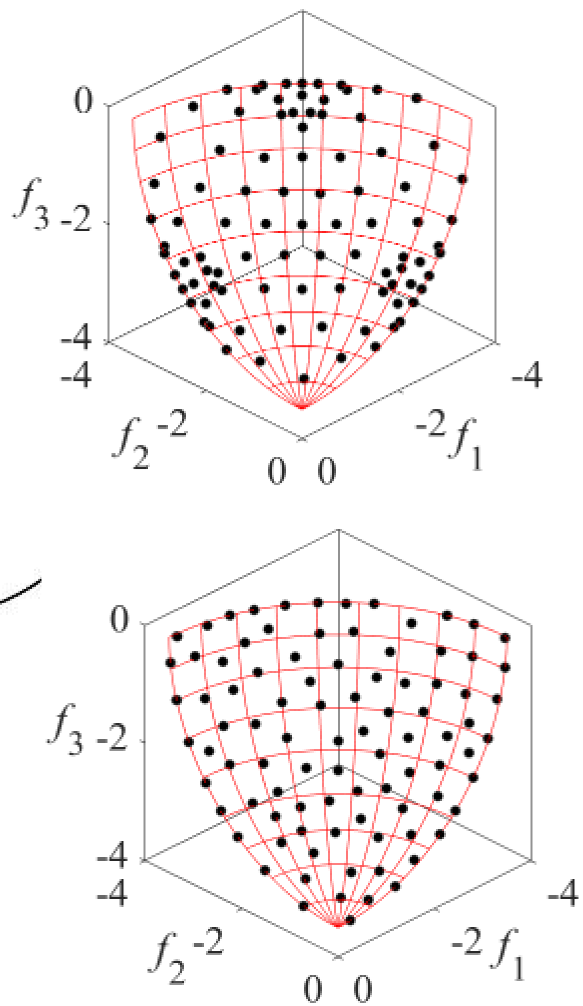
SMS-EMOA on DTLZ2



Final Population vs. Selected Solution Set

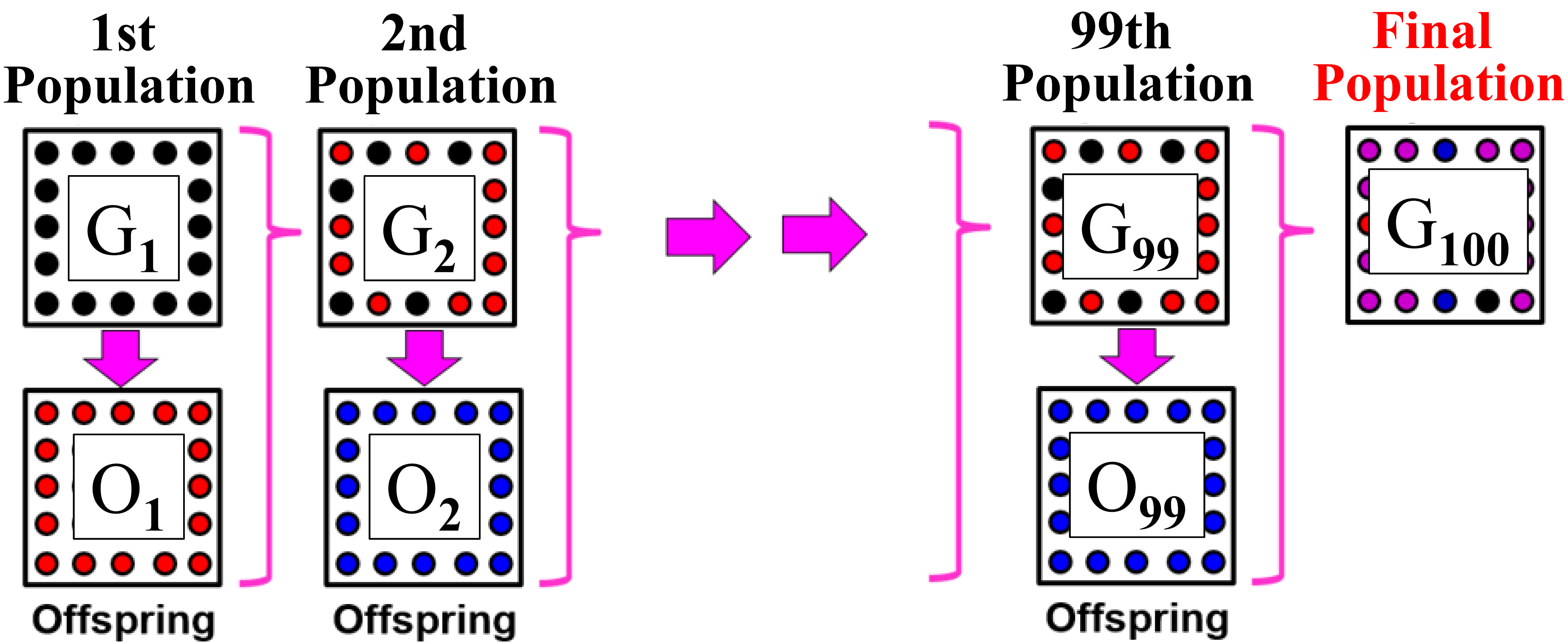


MOEA/D- PBI on Minus-DTLZ2



Improvement by Subset selection

Reason: The final population is not the best subset of all the examined solutions.



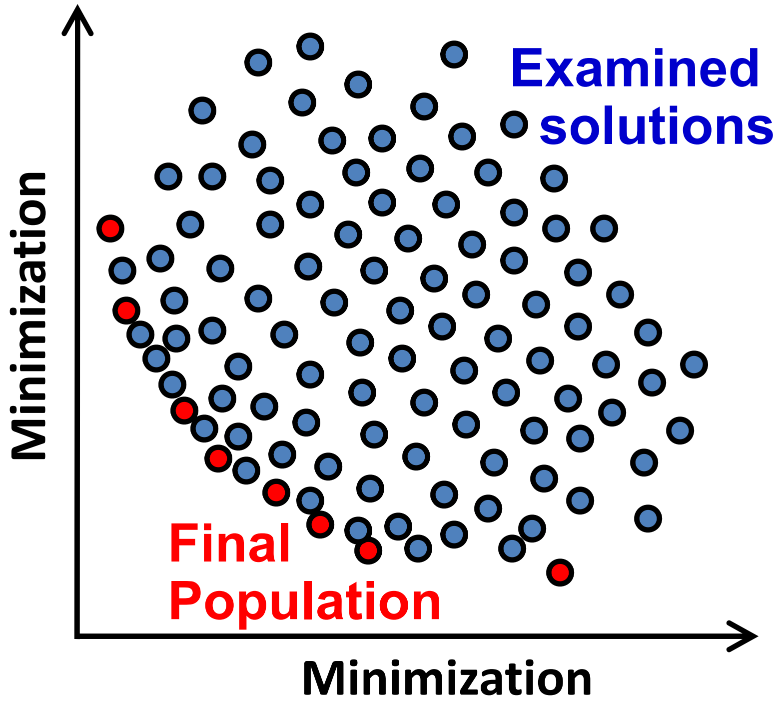
G100 is the best subset of $G_{99} \cup O_{99}$

G100 is **not** the best subset of $G_1 \cup O_1 \cup G_2 \cup O_2 \cup \ldots \cup G_{99} \cup O_{99}$

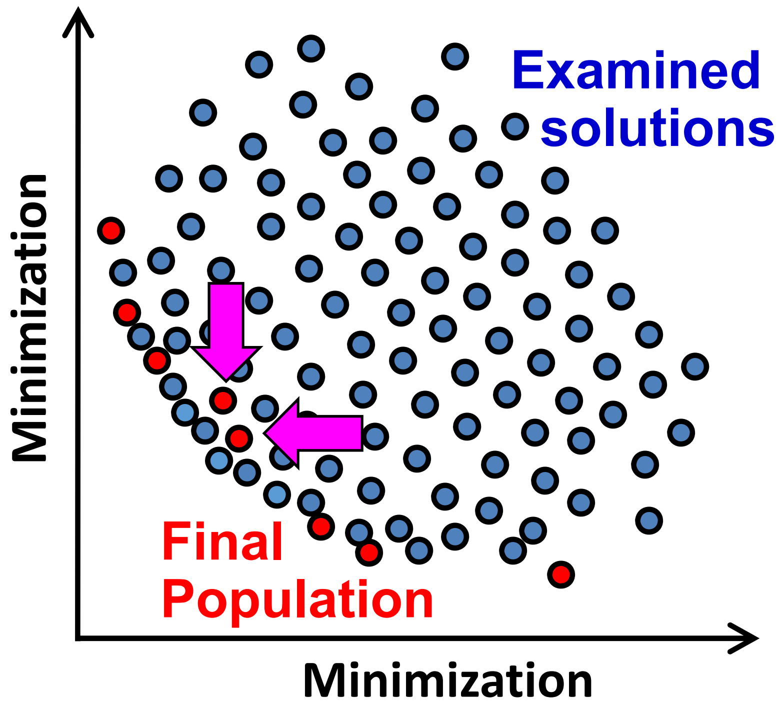
Improvement by Subset selection

Reason: The final population is not the best subset of all the examined solutions.

[7] M. Li, X. Yao, "An empirical investigation of the optimality and monotonicity proper-ties of multiobjective archiving methods" **EMO 2019**.



Desirable Results



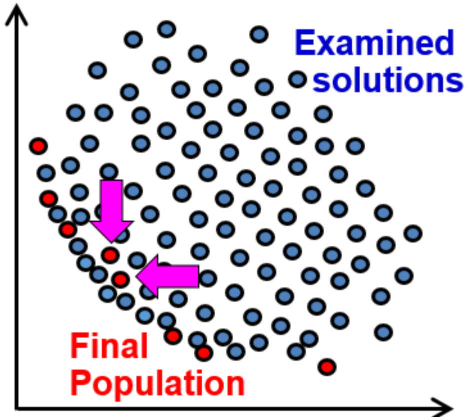
Many Cases

Table 2. The average deterioration ratio (DR) of 30 runs of the eight algorithms. The lower the better; 0.00% (in boldface) means that there is no archived solution which is dominated by the solutions eliminated in the previous archiving process.

Problem	NSGA-II	NSGA-II+ϵ	SPEA2	SPEA2+SDE	IBEA	SMS-EMOA	MOEA/D	NSGA-III
SCH1	**0.00%**	0.10%	**0.00%**	0.08%	0.20%	**0.00%**	0.25%	**0.00%**
SCH2	**0.00%**	0.22%	**0.00%**	0.14%	0.36%	**0.00%**	0.50%	**0.00%**
FON	10.21%	10.66%	4.05%	5.13%	1.02%	0.02%	2.99%	0.66%
KUR	5.06%	6.12%	1.83%	2.05%	0.49%	0.05%	5.86%	3.74%
ZDT1	3.79%	2.78%	1.90%	0.94%	0.05%	0.67%	21.65%	0.68%
ZDT2	2.42%	1.74%	1.23%	0.47%	0.11%	0.42%	29.26%	0.77%
ZDT3	3.46%	2.42%	1.92%	1.48%	0.06%	0.85%	22.98%	2.64%
ZDT4	0.58%	0.41%	0.40%	0.21%	0.26%	0.73%	12.25%	0.53%
ZDT6	0.97%	0.67%	0.62%	0.20%	**0.00%**	0.72%	19.00%	0.60%
WFG1	1.11%	0.60%	1.13%	0.30%	**0.00%**	3.96%	16.16%	2.11%
WFG2	1.81%	1.37%	1.20%	0.75%	0.03%	1.02%	17.45%	2.64%
WFG3	7.48%	4.75%	4.22%	1.71%	0.05%	0.45%	16.89%	1.72%
WFG4	8.34%	5.25%	4.70%	2.50%	0.08%	0.45%	15.81%	3.81%
WFG5	10.46%	8.00%	5.06%	4.12%	0.09%	0.31%	18.16%	2.13%
• • •								
DTLZ2-5	12.55%	2.40%	18.23%	1.14%	0.17%	0.02%	1.13%	2.69%
Average	5.424%	3.586%	4.240%	1.561%	0.421%	0.461%	9.366%	3.176%

[1] M. Li and X. Yao, "An empirical investigation of the optimality and monotonicity properties of multiobjective archiving methods," *EMO 2019*.

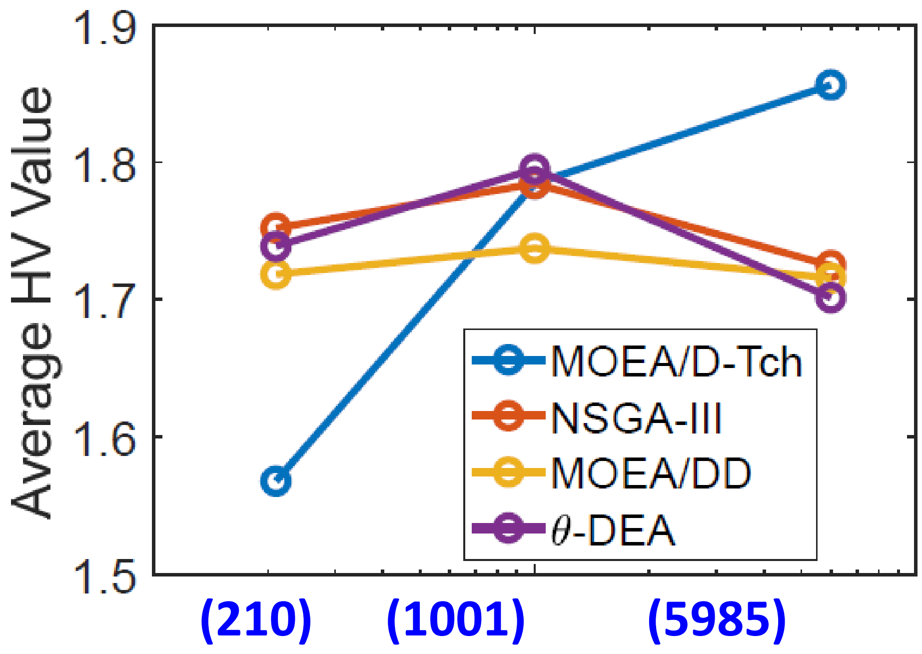
Table 2: Percentage of dominated solutions



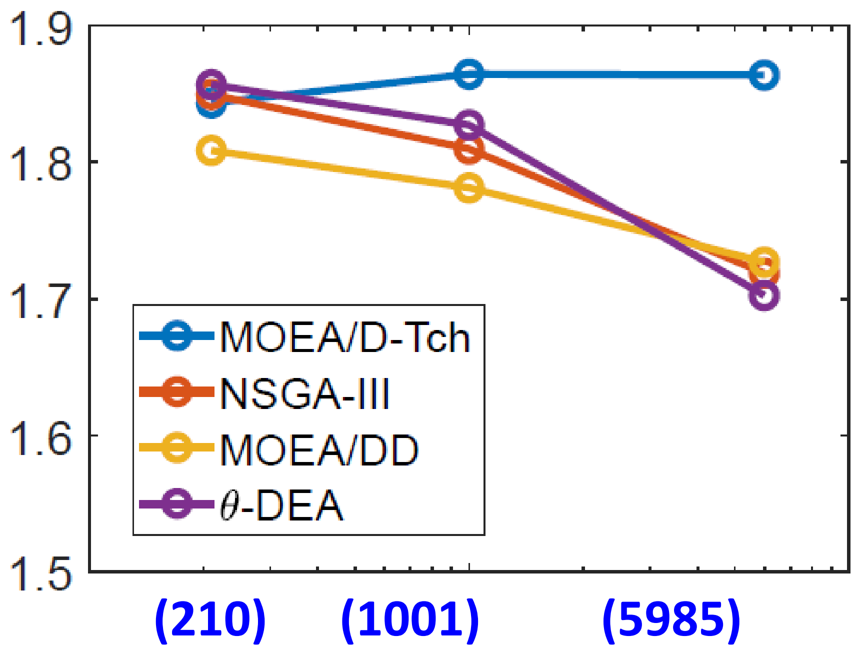
Final Population vs. Selected Solutions

Using a HV-based greedy subset selection algorithm,
210 solutions are selected from all the examined solutions.

[Ishibuchi et al. IEEE CIM 2022]



Final Population Results



Selected Solutions Results

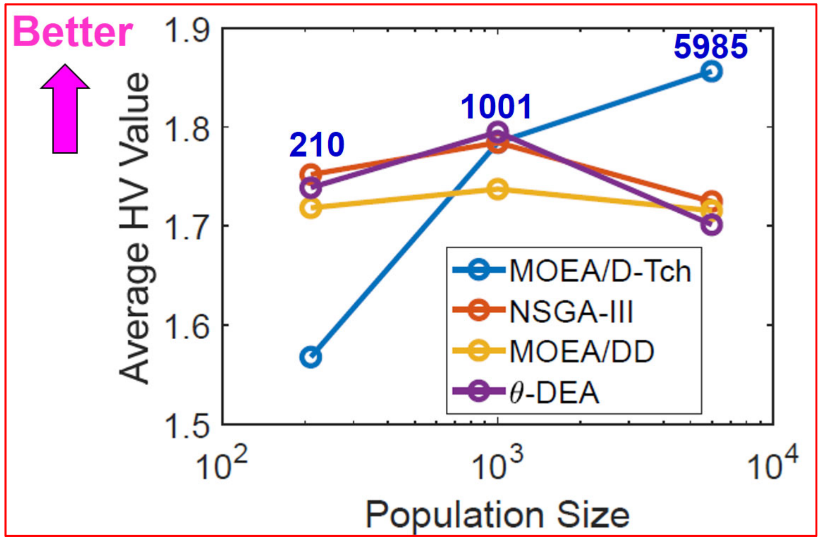
When the four algorithms are compared using the final population with the standard population size 210, MOEA/D is the worst. When they are compared using the selected 210 solutions, MOEA/D is the best.

Observation.

Our experimental results strongly depend on the population size.

Question.

How to specify the population size?



Suggestion 1:

To use the same population size specification as in many other papers especially in well-known papers (e.g., 210). Other settings: 100, 500, ...

Suggestion 2:

To use the best specification for each algorithm, and to select a pre-specified number of solutions from all the examined solutions for performance comparison.

Difficulties in Fair Performance Comparison of Evolutionary Multi-Objective Optimization Algorithms

(0) Visual Comparison

(1) Specification of Termination Condition

(2) Specification of Population Size

(3) Choice of Performance Indicators (e.g., HV, IGD)

(4) Setting in Performance Indicators (e.g., reference point)

(5) Choice of Test Problems

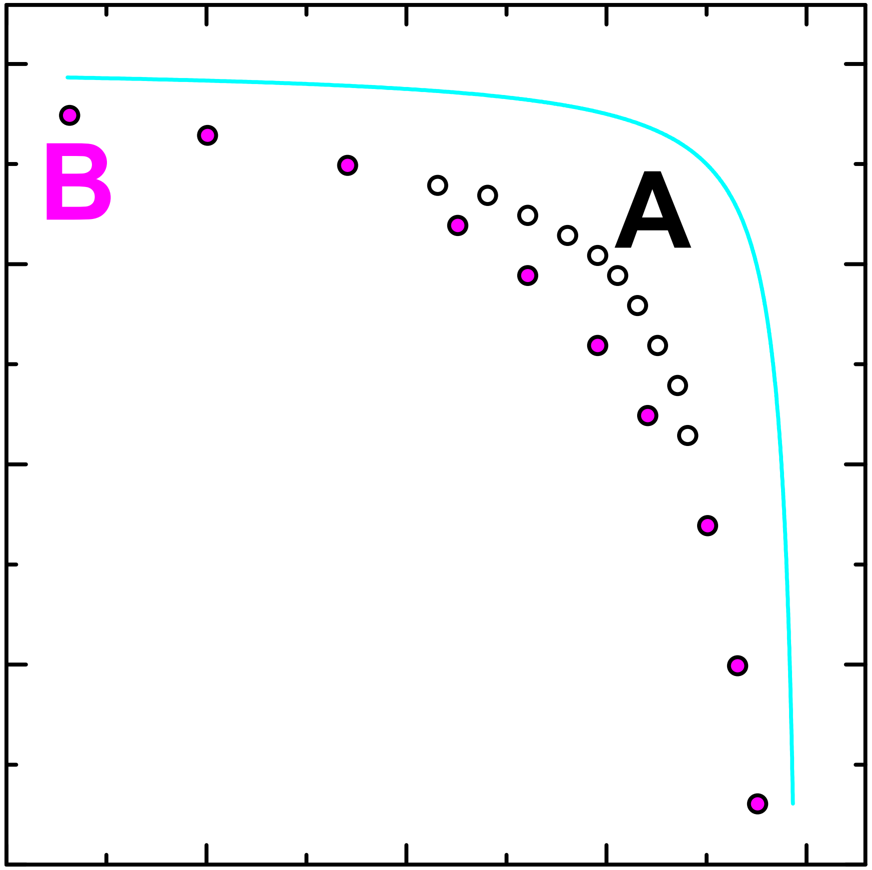
This Talk is mainly based on my recent paper:

Hisao Ishibuchi, Lie Meng Pang, and Ke Shang, "**Difficulties in Fair Performance Comparison of Multi-Objective Evolutionary Algorithms**" ***IEEE Computational Intelligence Magazine*** **(February 2022)**

Difficulties in the use of the HV indicator

HV-based comparison results depend on the reference point.

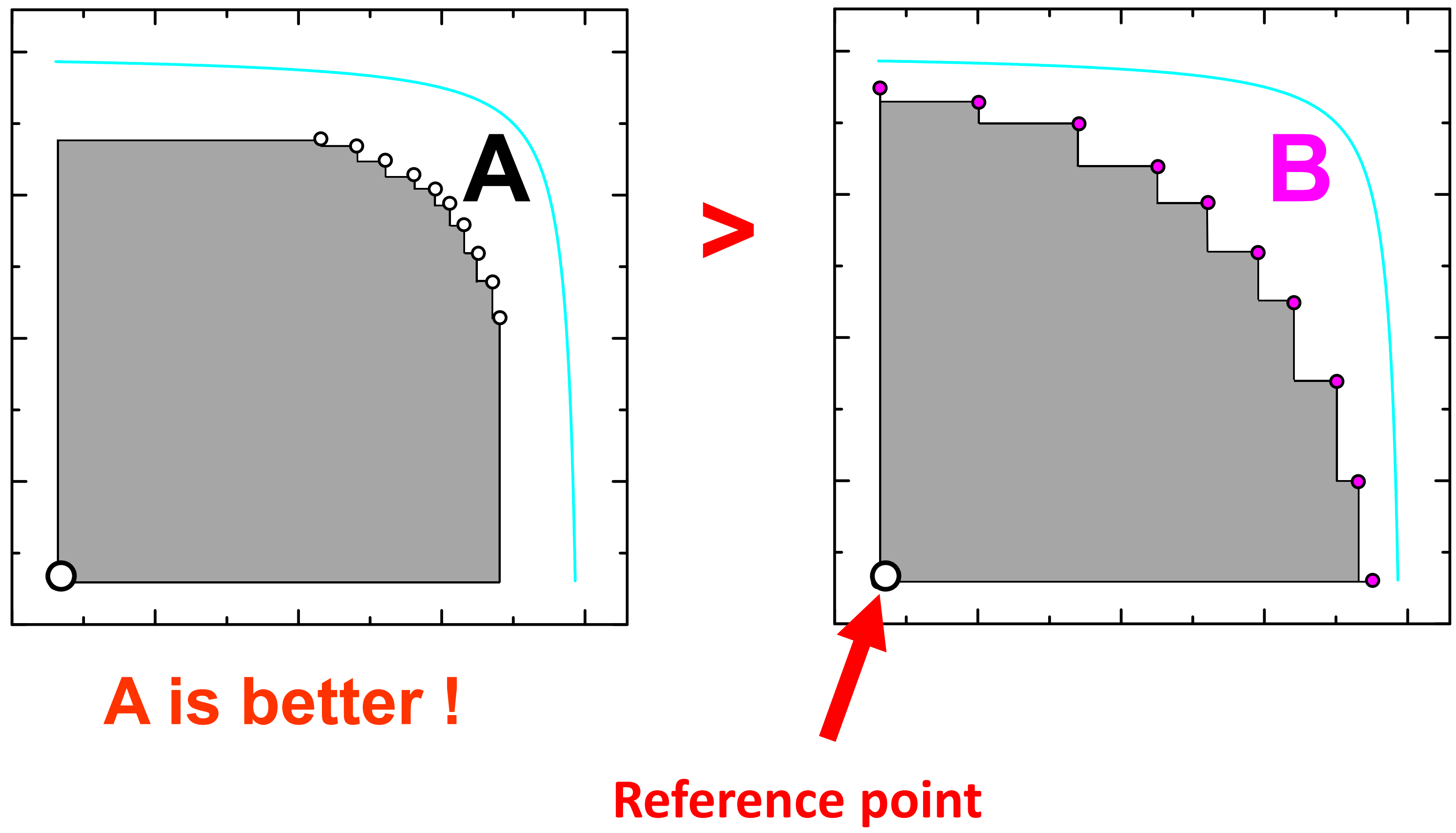
Q. Which is the better solution set between A and B (Two-Objective Maximization Problem)



Difficulties in the use of the HV indicator

HV-based comparison results depend on the reference point.

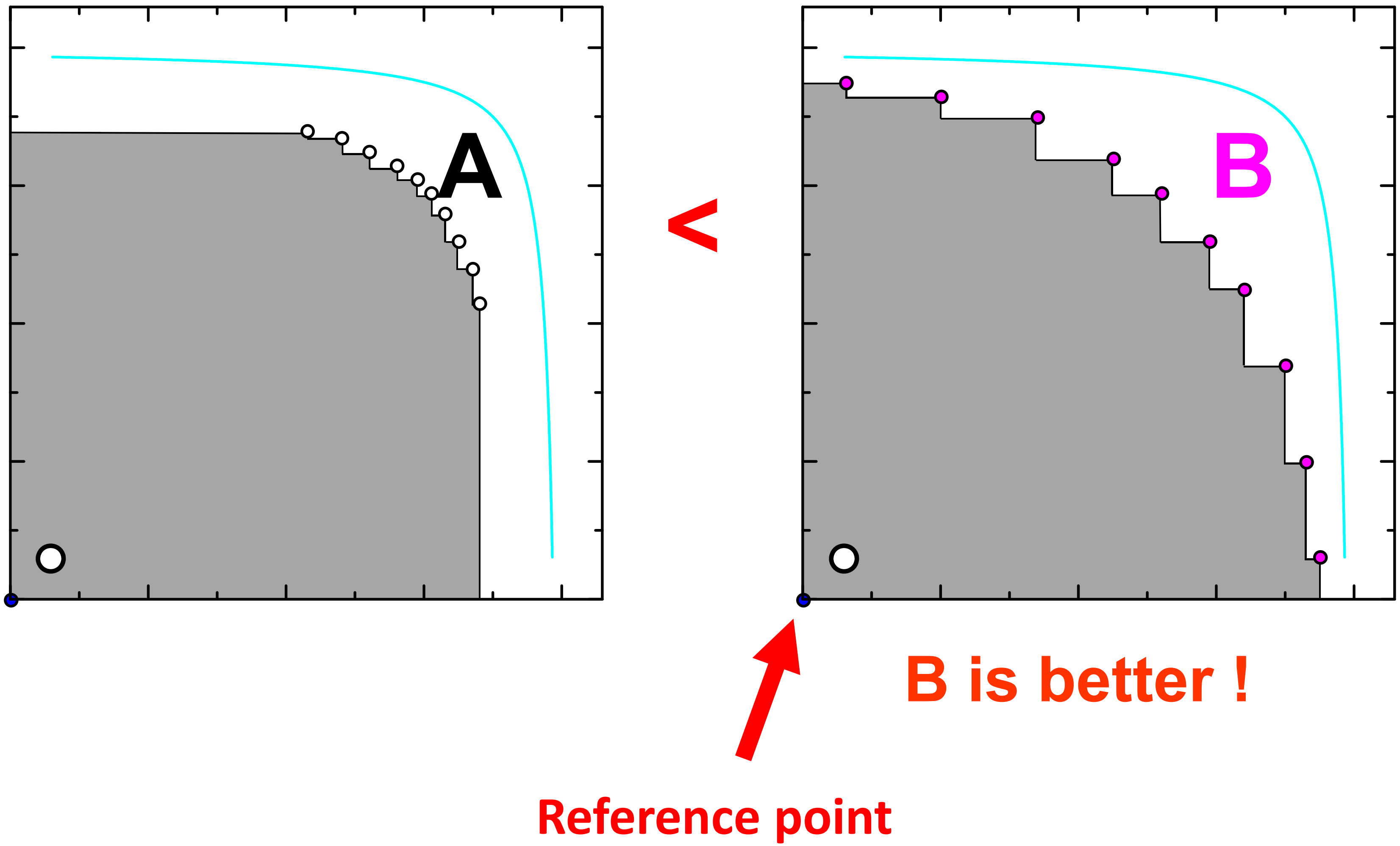
When the reference point is close to the Pareto front:



Difficulties in the use of the HV indicator

HV-based comparison results depend on the reference point.

When the reference point is far from the Pareto front:



Depending on the specification of the reference point, each solution set can be evaluated as the best solution set by HV:

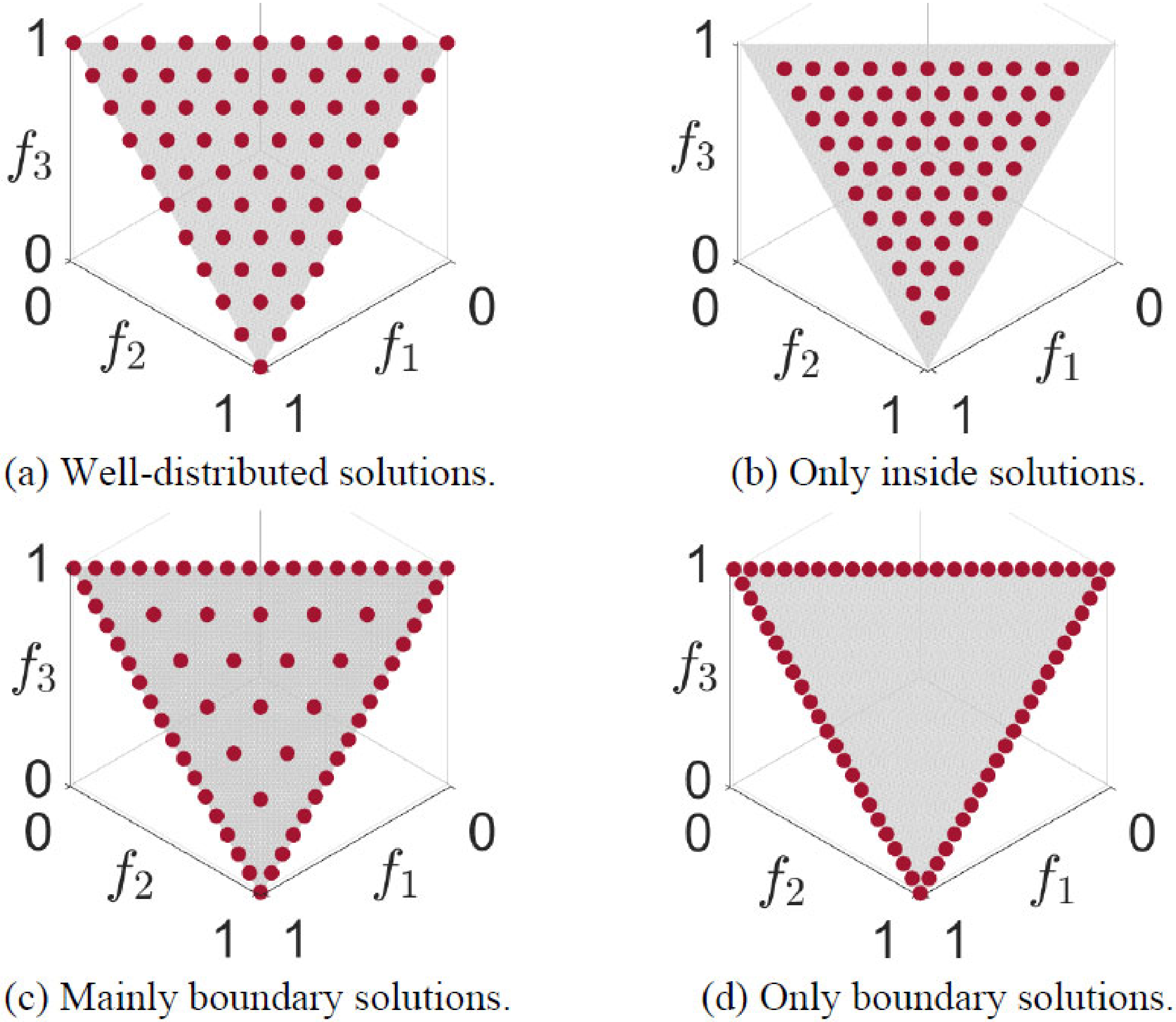
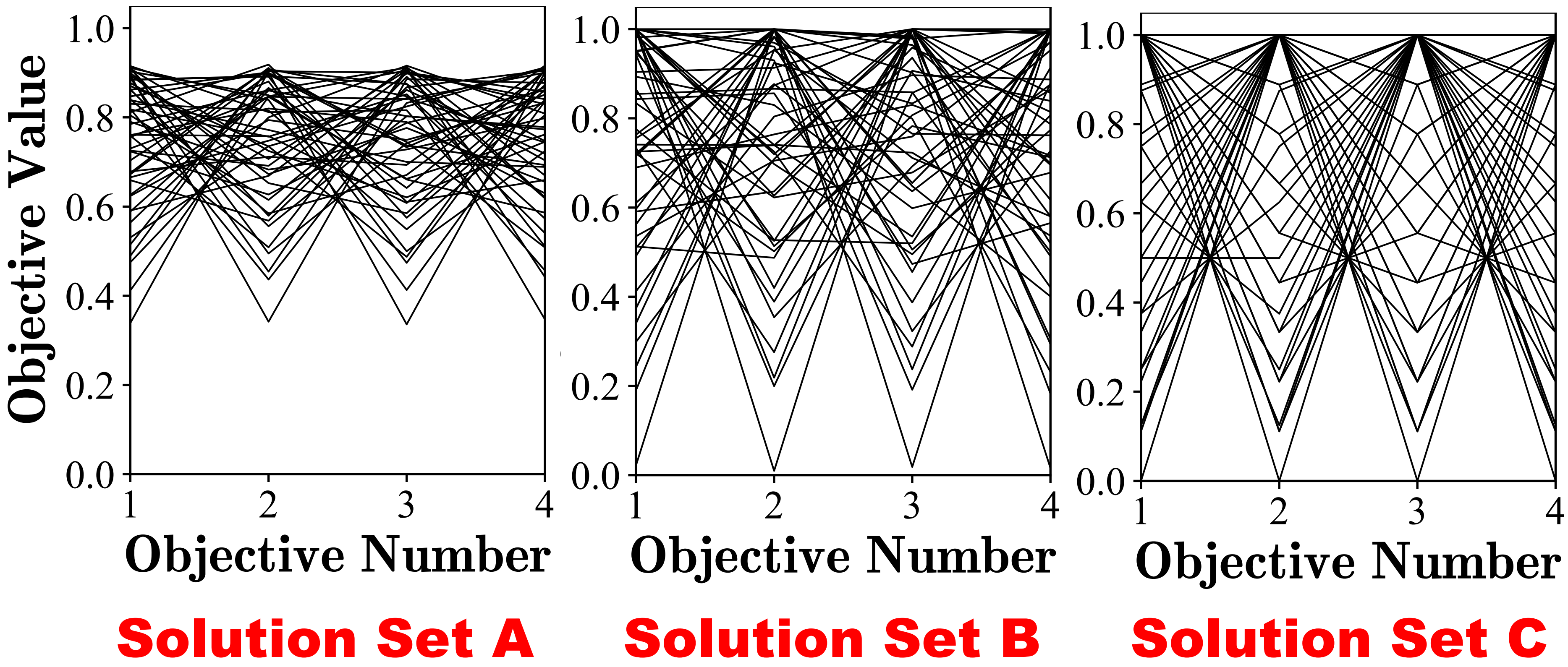


Fig. 20. Four solution sets with 66 solutions on an inverted triangular linear Pareto front.

[Ishibuchi et al. IEEE CIM 2022]

Depending on the specification of the reference point, each solution set can be evaluated as the best solution set by HV:



H. Ishibuchi et al., "Optimal distributions of solutions for hypervolume maximization on triangular and inverted triangular Pareto fronts of four-objective Problems," IEEE SSCI 2019, pp. 1857-1864.

Why ?

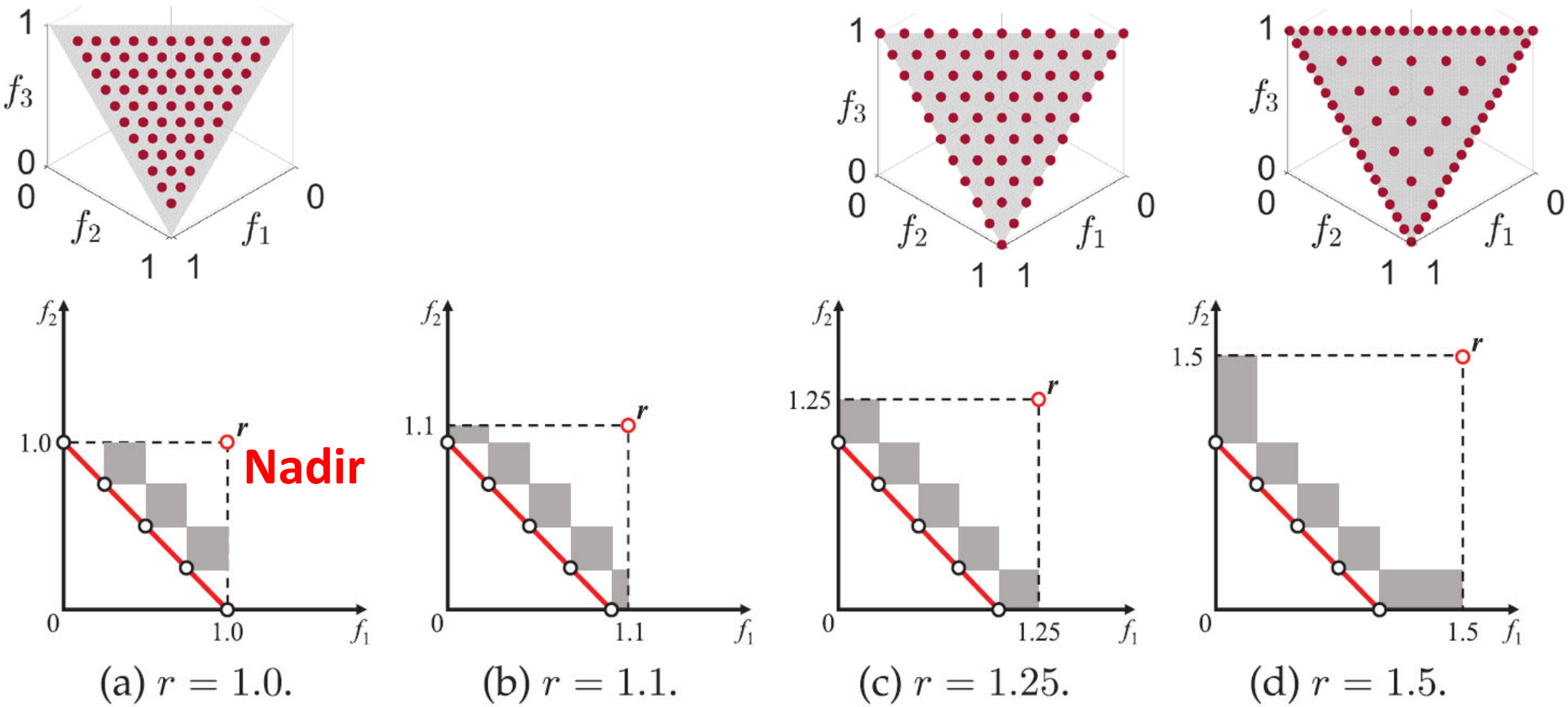


Figure 12: Hypervolume contribution of each of the uniformly distributed solutions on a linear Pareto front. **(Minimization Problem)**

H. Ishibuchi, R. Imada, Y. Setoguchi, and Y. Nojima, "How to specify a reference point in hypervolume calculation for fair performance comparison," *Evolutionary Computation*, 2018

Obtained solution sets by SMS-EMOA

(Near optimal solution distribution for HV maximization. Ishibuchi 2018)

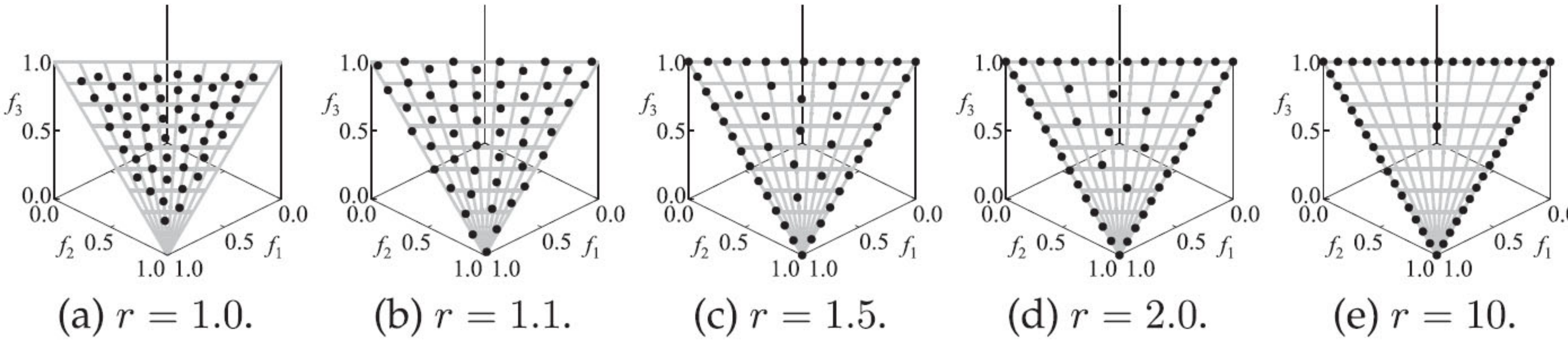


Figure 2: Obtained solution sets for the three-objective normalized Minus-DTLZ1.

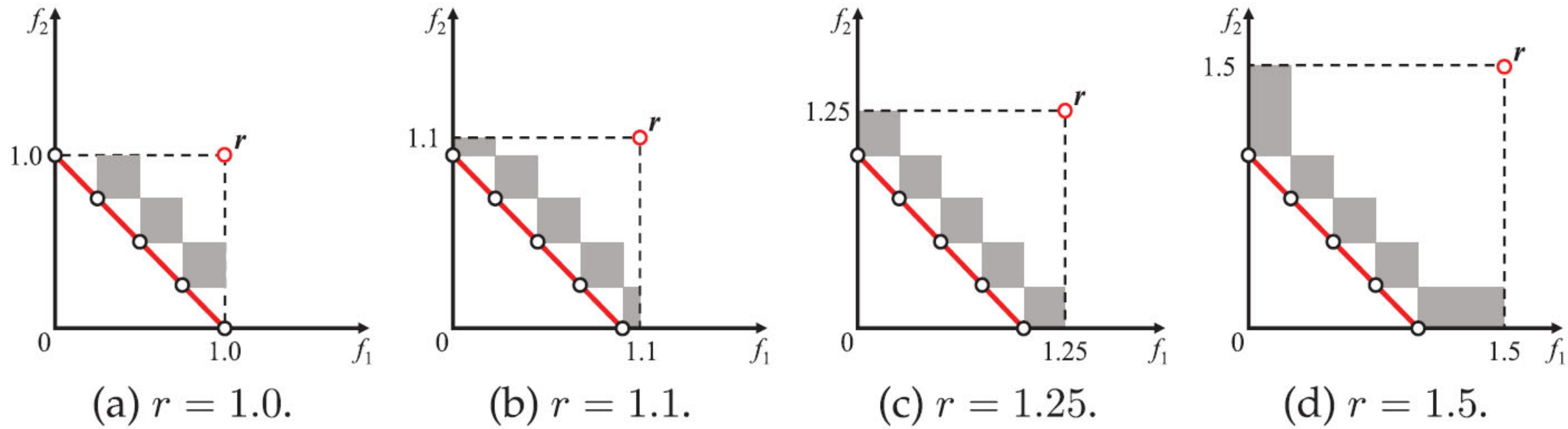
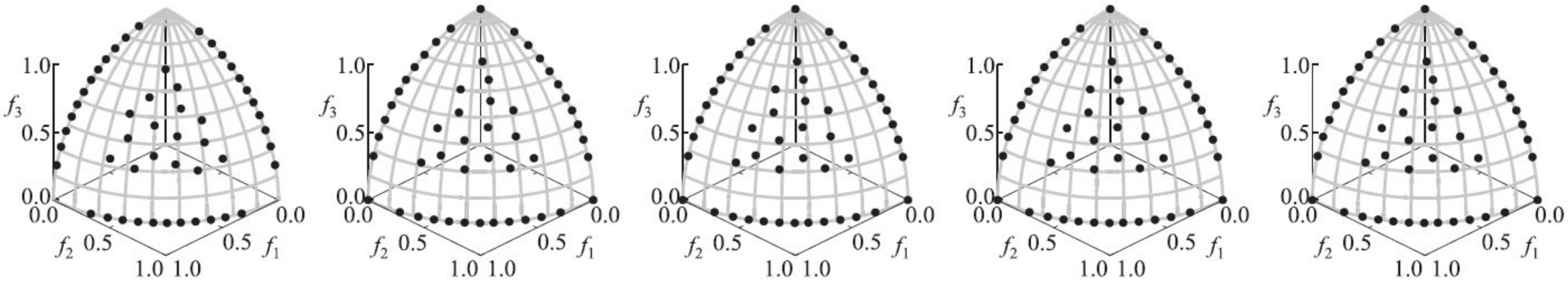


Figure 12: Hypervolume contribution of each of the uniformly distributed solutions on a linear Pareto front.

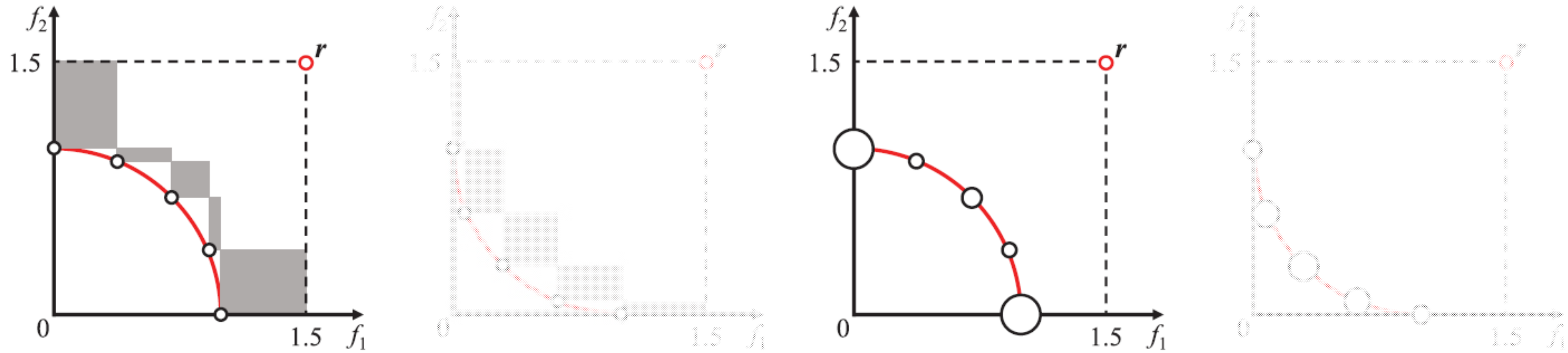
Obtained solution sets by SMS-EMOA

(Near optimal solution distribution for HV maximization. Ishibuchi 2018)



(a) $r = 1.0$. (b) $r = 1.1$. (c) $r = 1.5$. (d) $r = 2.0$. (e) $r = 10$.

Figure 8: Obtained solution sets for the three-objective DTLZ2 problem.

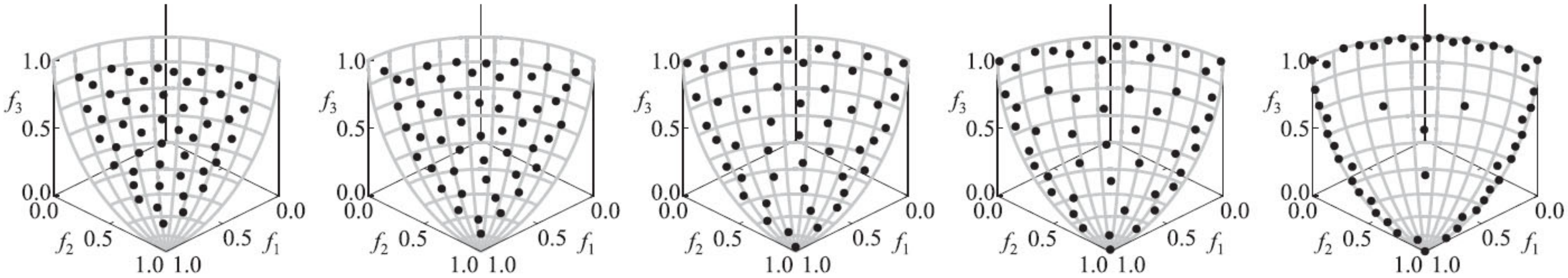


(a) Two-objective DTLZ2. (b) Two-objective Minus-DTLZ2. (c) Two-objective DTLZ2. (d) Two-objective Minus-DTLZ2.

Figure 10: Hypervolume contribution of each of the five uniformly distributed solutions.

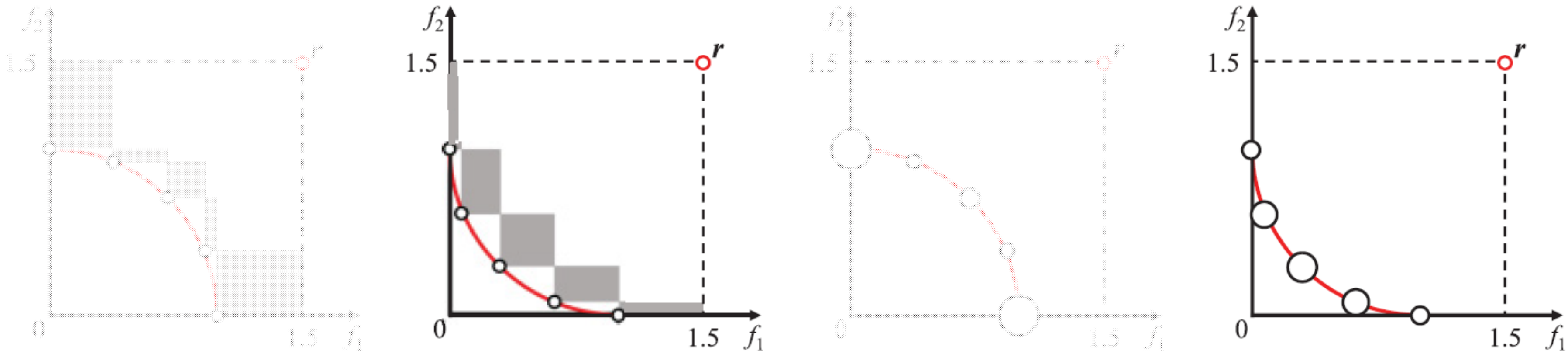
Obtained solution sets by SMS-EMOA

(Near optimal solution distribution for HV maximization. Ishibuchi 2018)



(a) $r = 1.0$. (b) $r = 1.1$. (c) $r = 1.5$. (d) $r = 2.0$. (e) $r = 10$.

Figure 9: Obtained solution sets for the three-objective Minus-DTLZ2 problem.



(a) Two-objective DTLZ2. (b) Two-objective Minus-DTLZ2. (c) Two-objective DTLZ2. (d) Two-objective Minus-DTLZ2.

Figure 10: Hypervolume contribution of each of the five uniformly distributed solutions.

How to specify the reference point ?

One Idea:

Each solution in the uniformly distributed solution set has the same HV contribution.

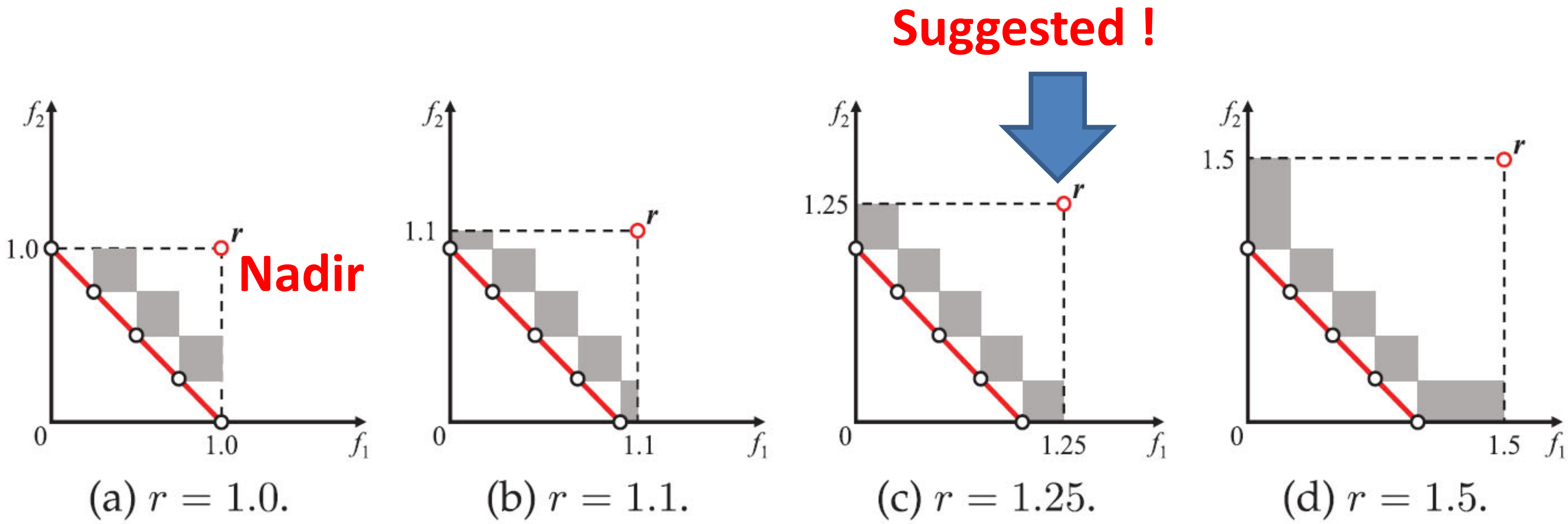


Figure 12: Hypervolume contribution of each of the uniformly distributed solutions on a linear Pareto front. **(Minimization Problem)**

H. Ishibuchi, R. Imada, Y. Setoguchi, and Y. Nojima, "How to specify a reference point in hypervolume calculation for fair performance comparison," *Evolutionary Computation*, 2018

Reference Point Specification for Fair Comparison

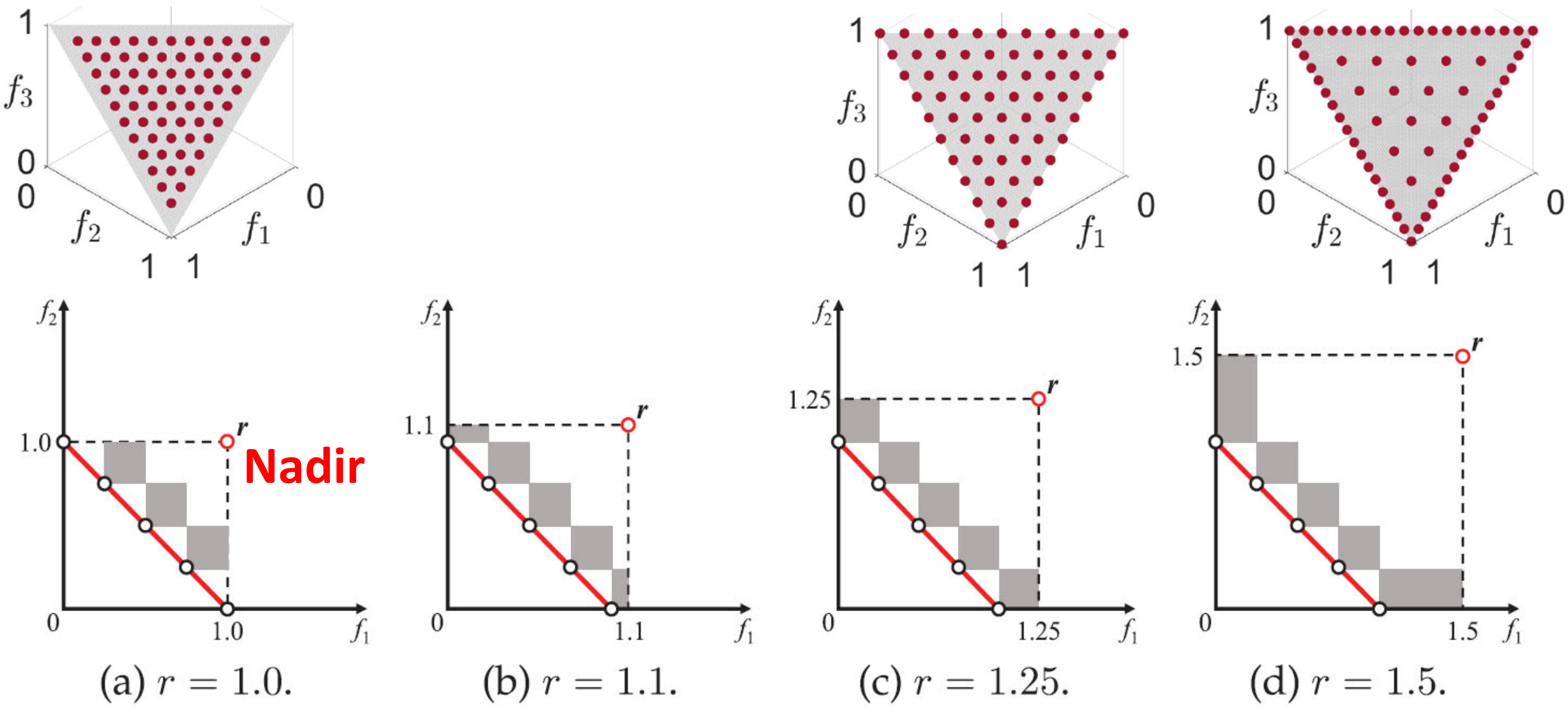
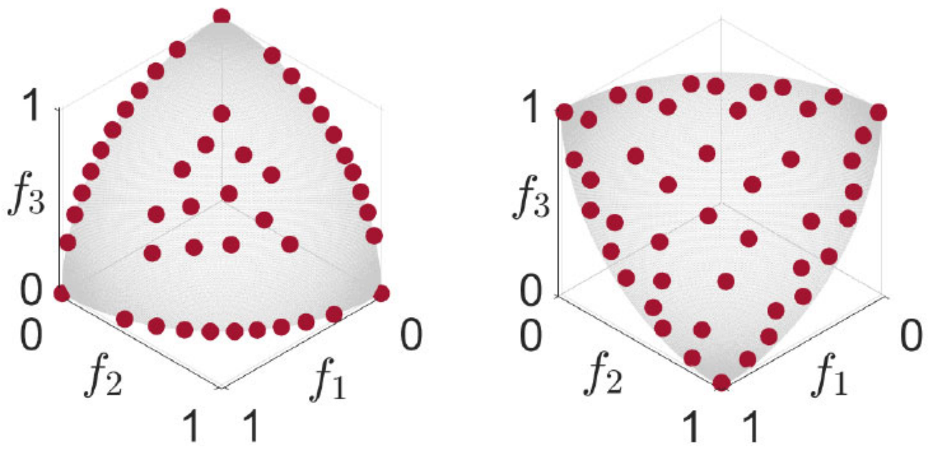


Figure 12: Hypervolume contribution of each of the uniformly distributed solutions on a linear Pareto front.

This suggestion is useful for linear Pareto fronts.

Difficulties in the use of the HV indicator

Uniformly distributed solution sets are not the best solution set for HV when the Pareto front is non-linear.



(a) DTLZ2. (b) Minus-DTLZ2.

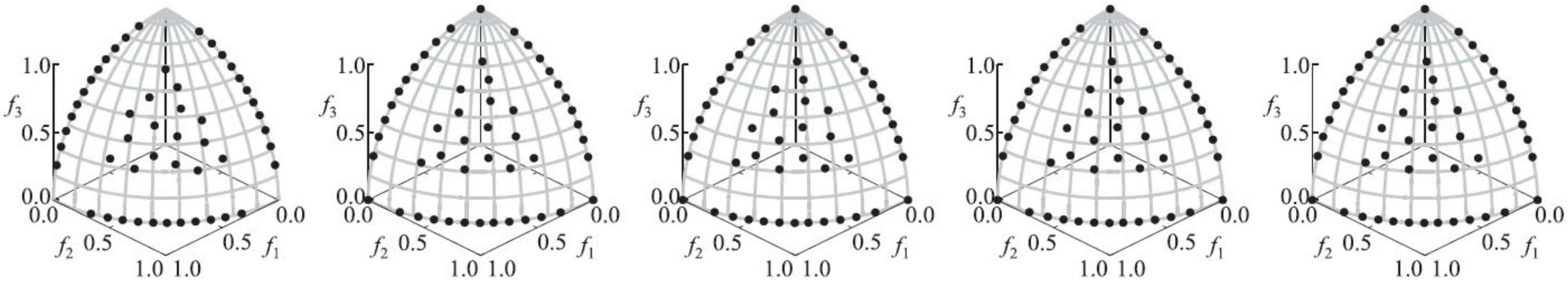
Fig. 18. Hypervolume optimal distributions of 45 solutions for $\boldsymbol{r}$ = (2, 2, 2).

HV optimal solution sets

[Ishibuchi et al. IEEE CIM 2022]

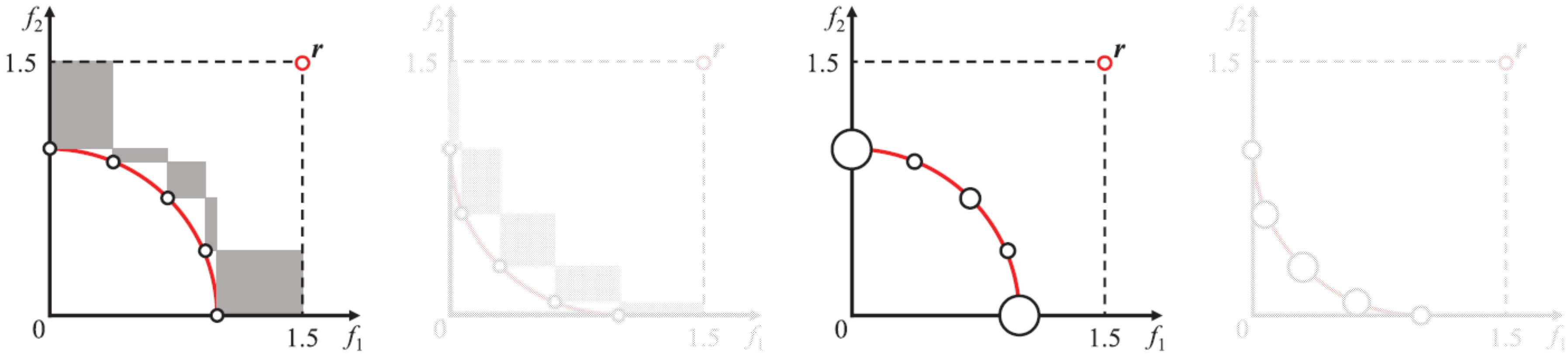
Obtained solution sets by SMS-EMOA

(Near optimal solution distribution for HV maximization. Ishibuchi 2018)



(a) $r = 1.0$. (b) $r = 1.1$. (c) $r = 1.5$. (d) $r = 2.0$. (e) $r = 10$.

Figure 8: Obtained solution sets for the three-objective DTLZ2 problem.

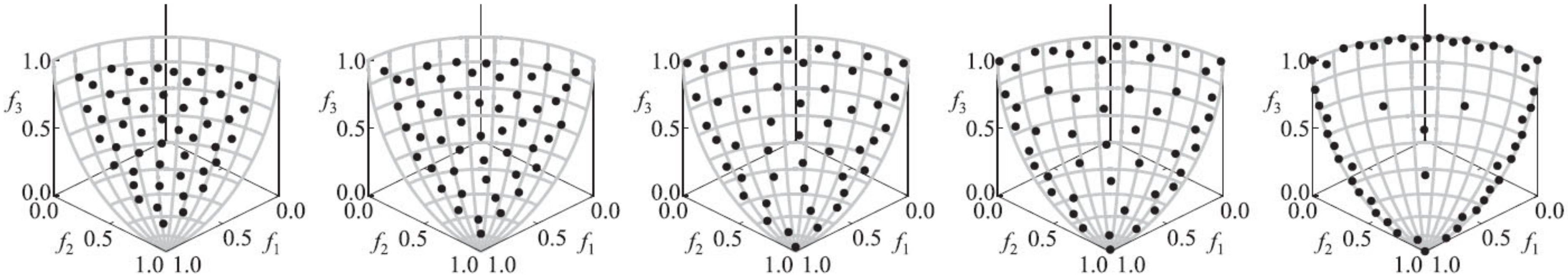


(a) Two-objective DTLZ2. (b) Two-objective Minus-DTLZ2. (c) Two-objective DTLZ2. (d) Two-objective Minus-DTLZ2.

Figure 10: Hypervolume contribution of each of the five uniformly distributed solutions.

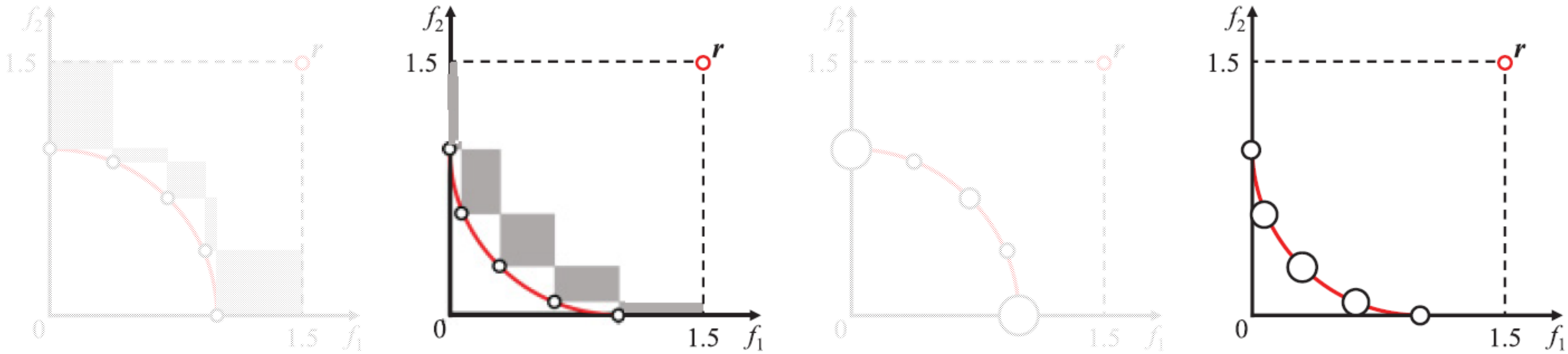
Obtained solution sets by SMS-EMOA

(Near optimal solution distribution for HV maximization. Ishibuchi 2018)



(a) $r = 1.0$. (b) $r = 1.1$. (c) $r = 1.5$. (d) $r = 2.0$. (e) $r = 10$.

Figure 9: Obtained solution sets for the three-objective Minus-DTLZ2 problem.

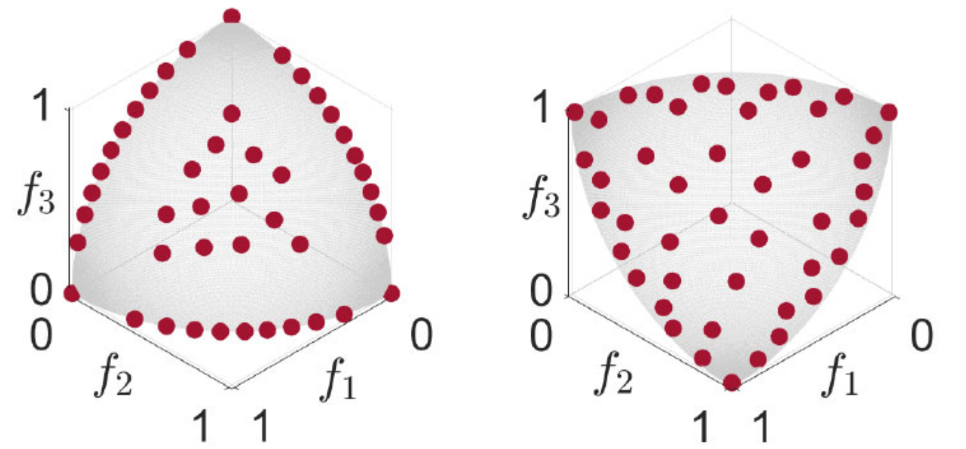


(a) Two-objective DTLZ2. (b) Two-objective Minus-DTLZ2. (c) Two-objective DTLZ2. (d) Two-objective Minus-DTLZ2.

Figure 10: Hypervolume contribution of each of the five uniformly distributed solutions.

Difficulties in the use of the HV indicator

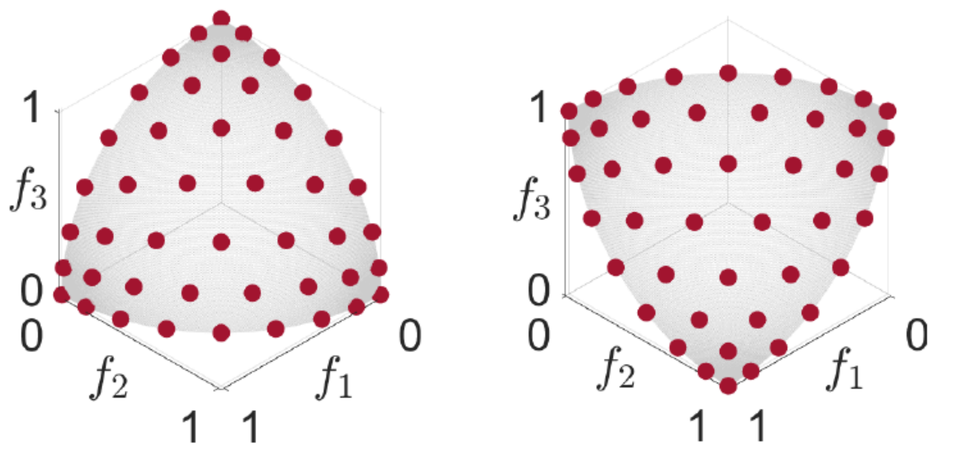
Uniformly distributed solution sets are not the best solution set for HV when the Pareto front is non-linear.



(a) DTLZ2. (b) Minus-DTLZ2.

Fig. 18. Hypervolume optimal distributions of 45 solutions for r = (2, 2, 2).

HV optimal solution sets



(a) DTLZ2. (b) Minus-DTLZ2.

Fig. 19. Solution sets generated by the 45 uniform weight vectors.

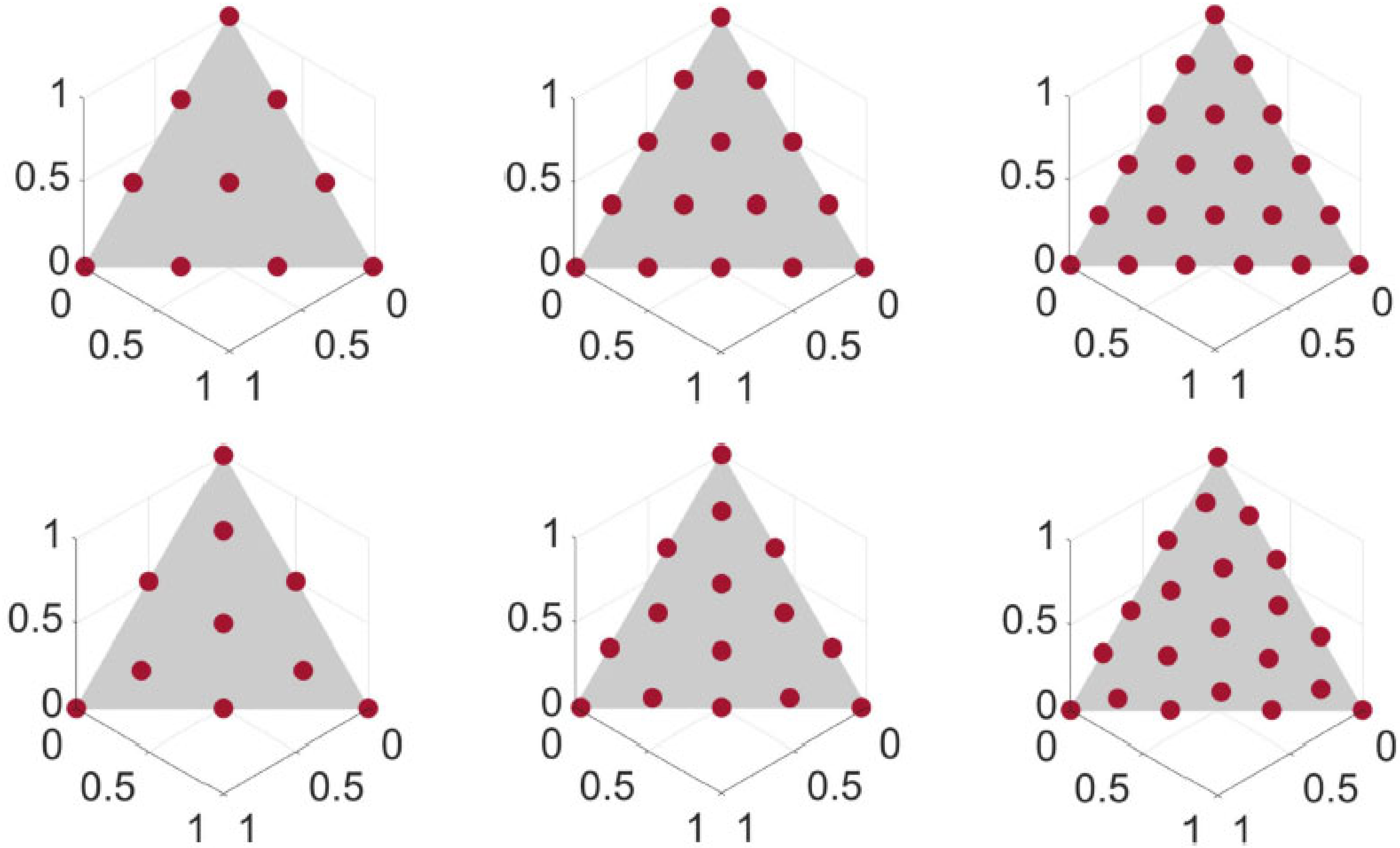
More uniform solution sets

[Ishibuchi et al. IEEE CIM 2022]

Recent Results: Even for linear triangular Pareto fronts, uniformly distributed solutions are not always optimal for HV.

Uniform distribution (smaller HV)

Better distribution (larger HV)



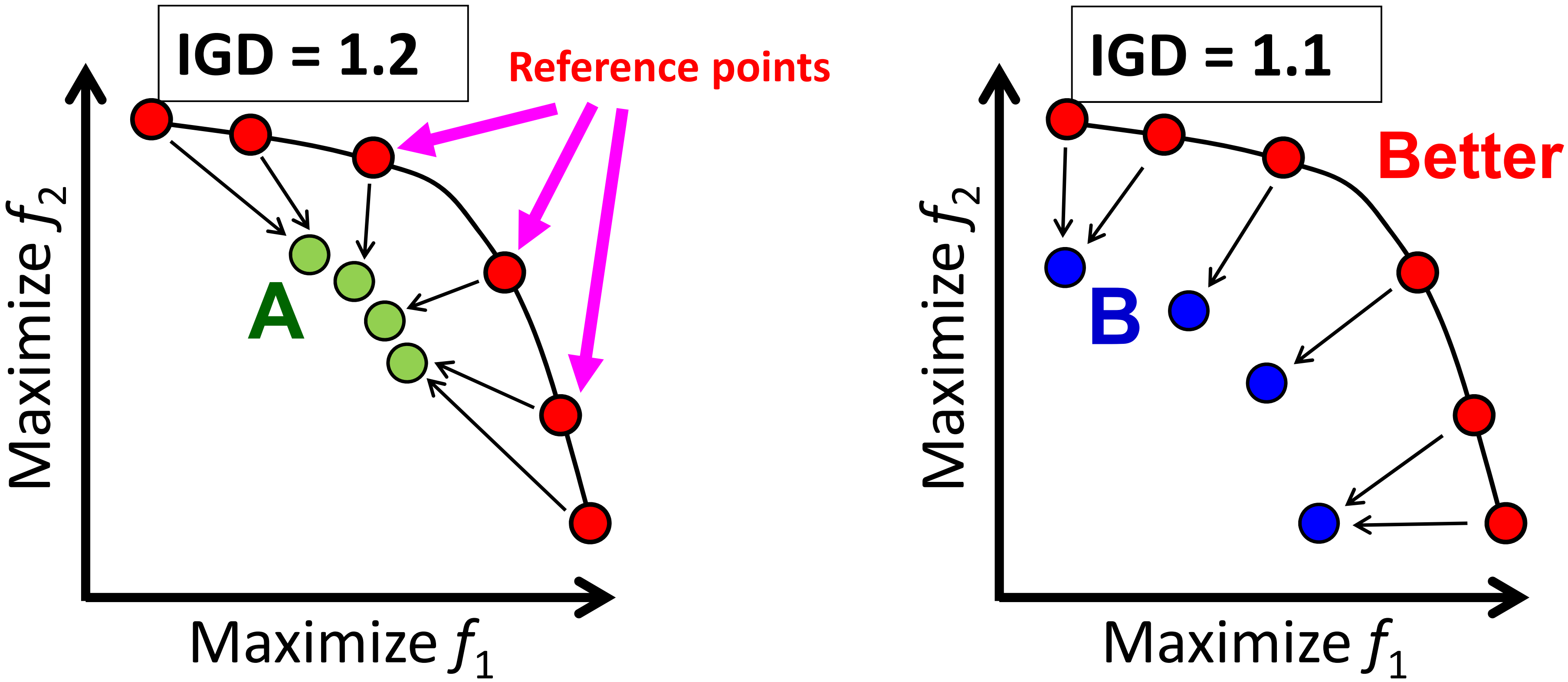
K. Shang, H. Ishibuchi, W. Chen, Y. Nan, and W. Liao, "Hypervolume-optimal μ-distributions on line/plane-based Pareto fronts in three dimensions," IEEE Trans. on Evolutionary Computation, vol. 26, no. 2, pp. 349-363, April 2022.

Difficulties in the use of the IGD indicator

IGD-based comparison results depend on the reference point set.

(**IGD:** The average distance from each reference point to the nearest solution)

Q. Which is the better solution set between A and B (Two-Objective Maximization Problem)

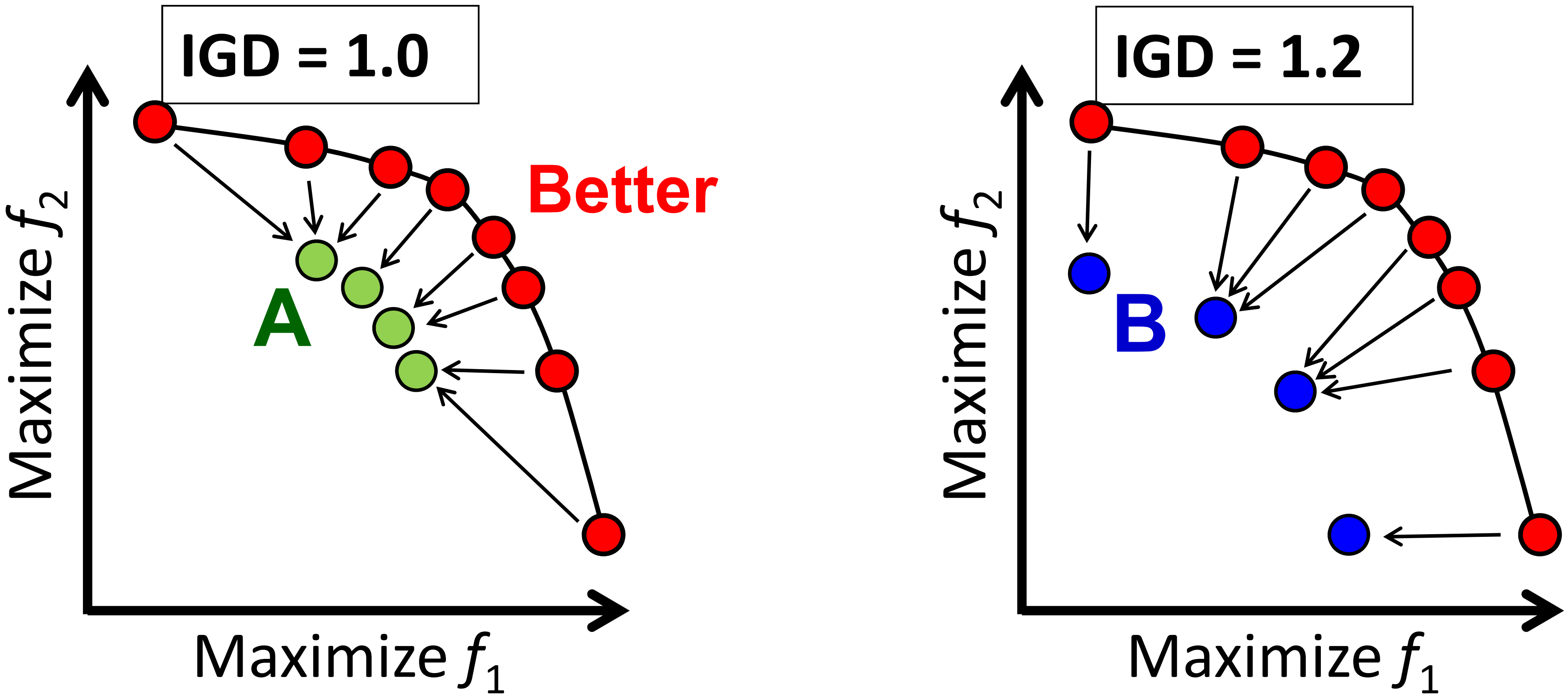


Difficulties in the use of the IGD indicator

IGD-based comparison results depend on the reference point set.

(**IGD:** The average distance from each reference point to the nearest solution)

Q. Which is the better solution set between A and B
(Two-Objective Maximization Problem)

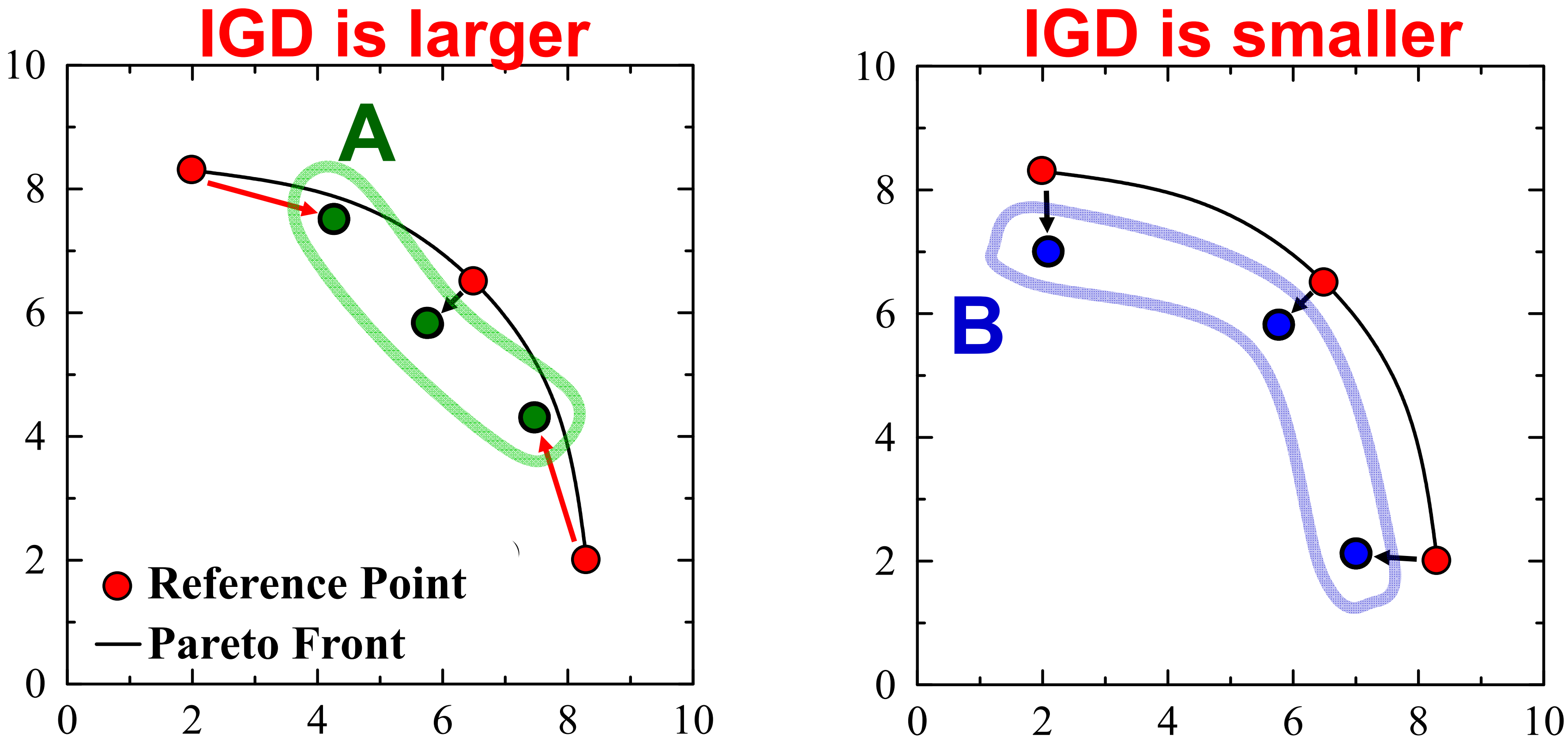


Difficulties in the use of the IGD indicator

IGD is not Pareto compliant

Q. Which is the better solution set between A and B (Two-Objective Maximization Problem)

IGD-based comparison result: B is better.

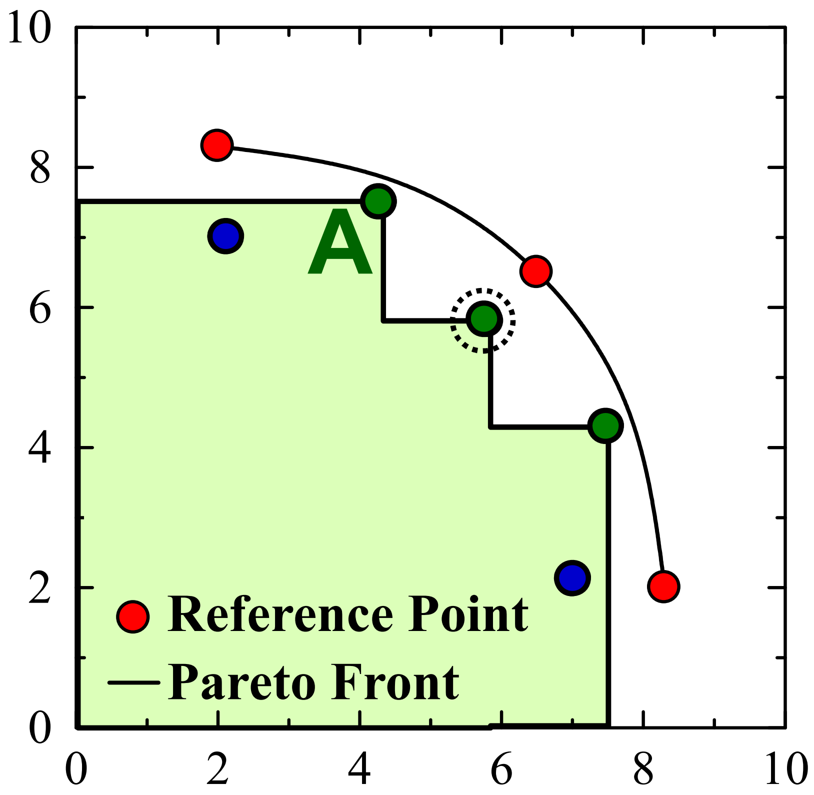


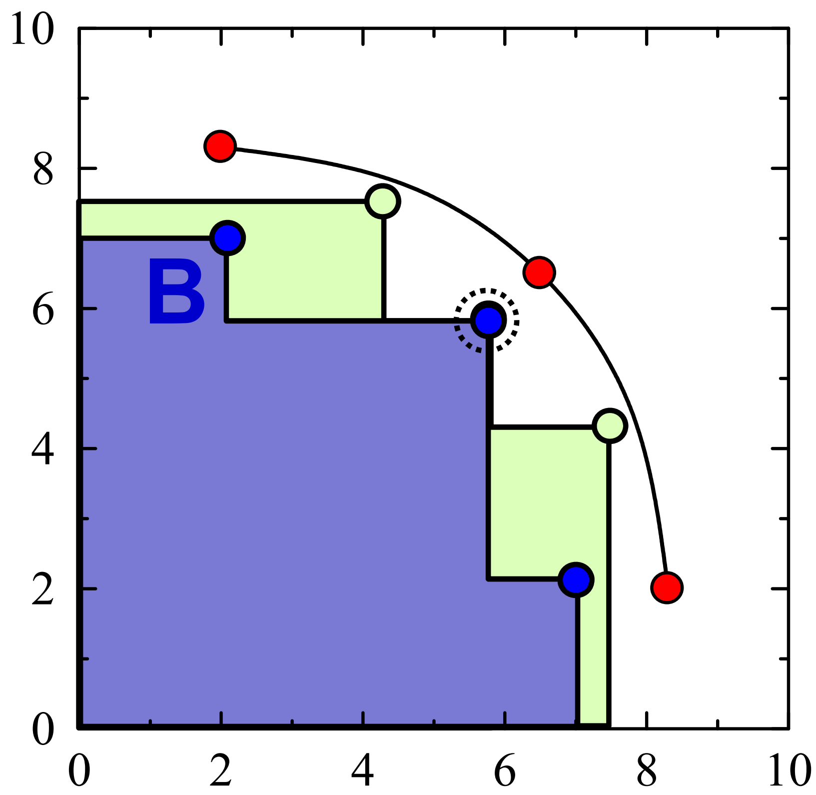
Difficulties in the use of the IGD indicator

IGD is not Pareto compliant

Q. Which is the better solution set between A and B (Two-Objective Maximization Problem)

A. A is better than B based on Pareto dominance.



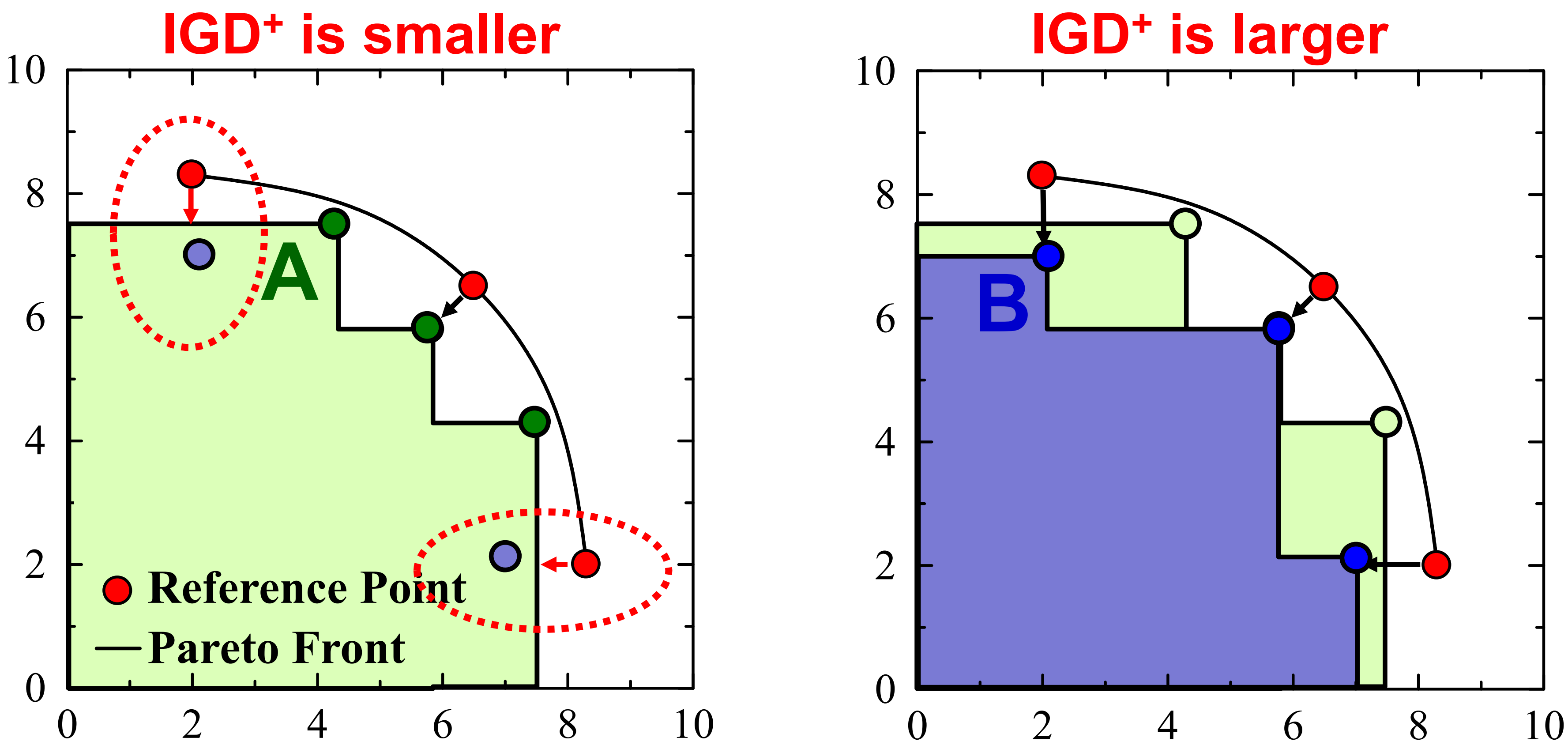


Difficulties in the use of the IGD indicator

IGD$^+$ is weakly Pareto compliant

Q. **Which is the better solution set between A and B (Two-Objective Maximization Problem)**

IGD$^+$-based comparison result: B is better.



Suggestions about Indicators

(i) To use multiple indicators (e.g., HV, IGD, IGD$^+$). Reviewers often suggest the use of multiple indicators.

(ii) To use multiple reference points in the HV indicator (i.e., multiple comparison results). This can be viewed as using multiple indicators.

(iii) To use a large number of uniformly distributed reference points for IGD and IGD$^+$ (e.g., 100,000 reference points).

Suggestions about Indicators

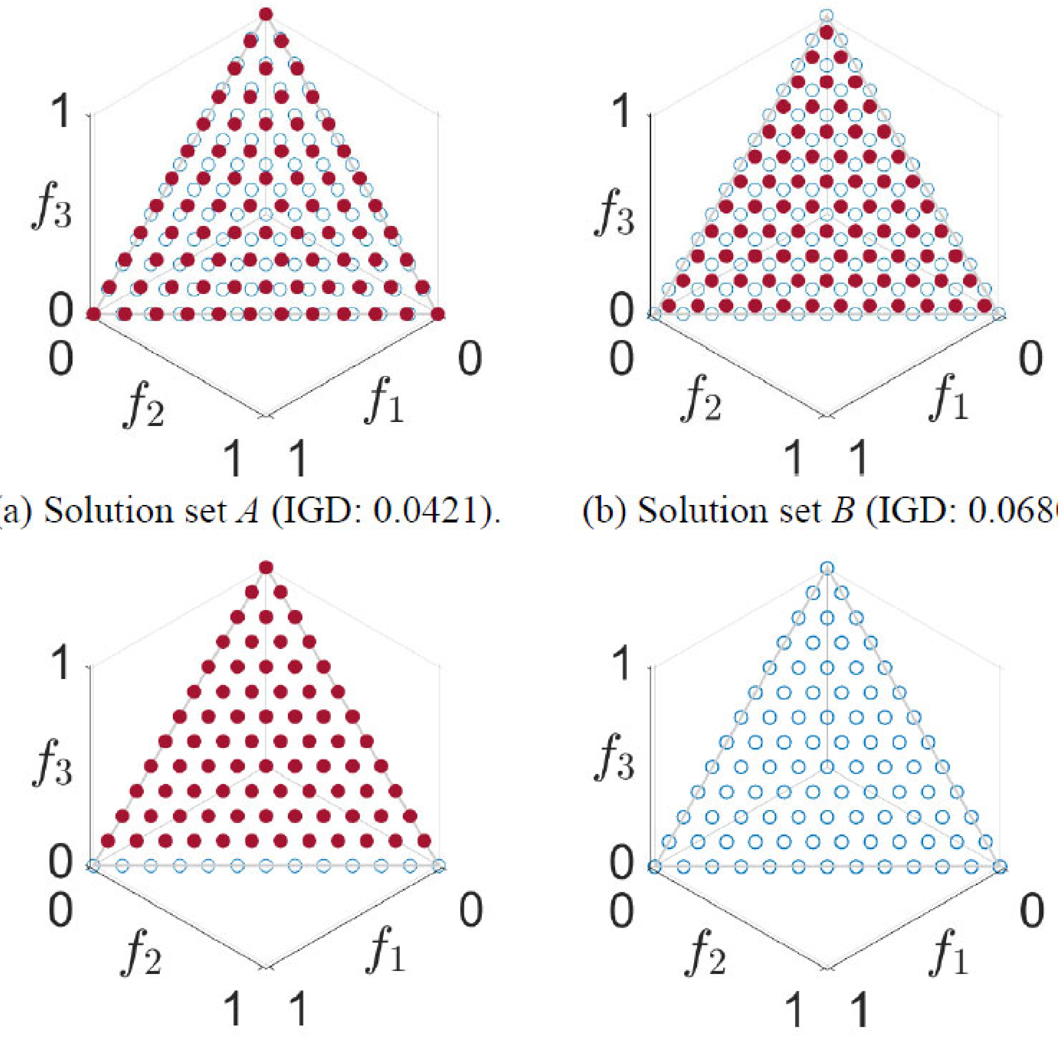
(i) To use multiple indicators (e.g., HV, IGD, IGD^+). Reviewers often suggest the use of multiple indicators.

(ii) To use multiple reference points in the HV indicator (i.e., multiple comparison results). This can be viewed as using multiple indicators.

(iii) To use a large number of uniformly distributed reference points for IGD and IGD^+ (e.g., 100,000 reference points).

(Uniform reference points are not always good)

Depending on the reference point set, each of the following solution sets can be evaluated as the best solution set.



(a) Solution set A (IGD: 0.0421). (b) Solution set B (IGD: 0.0680).

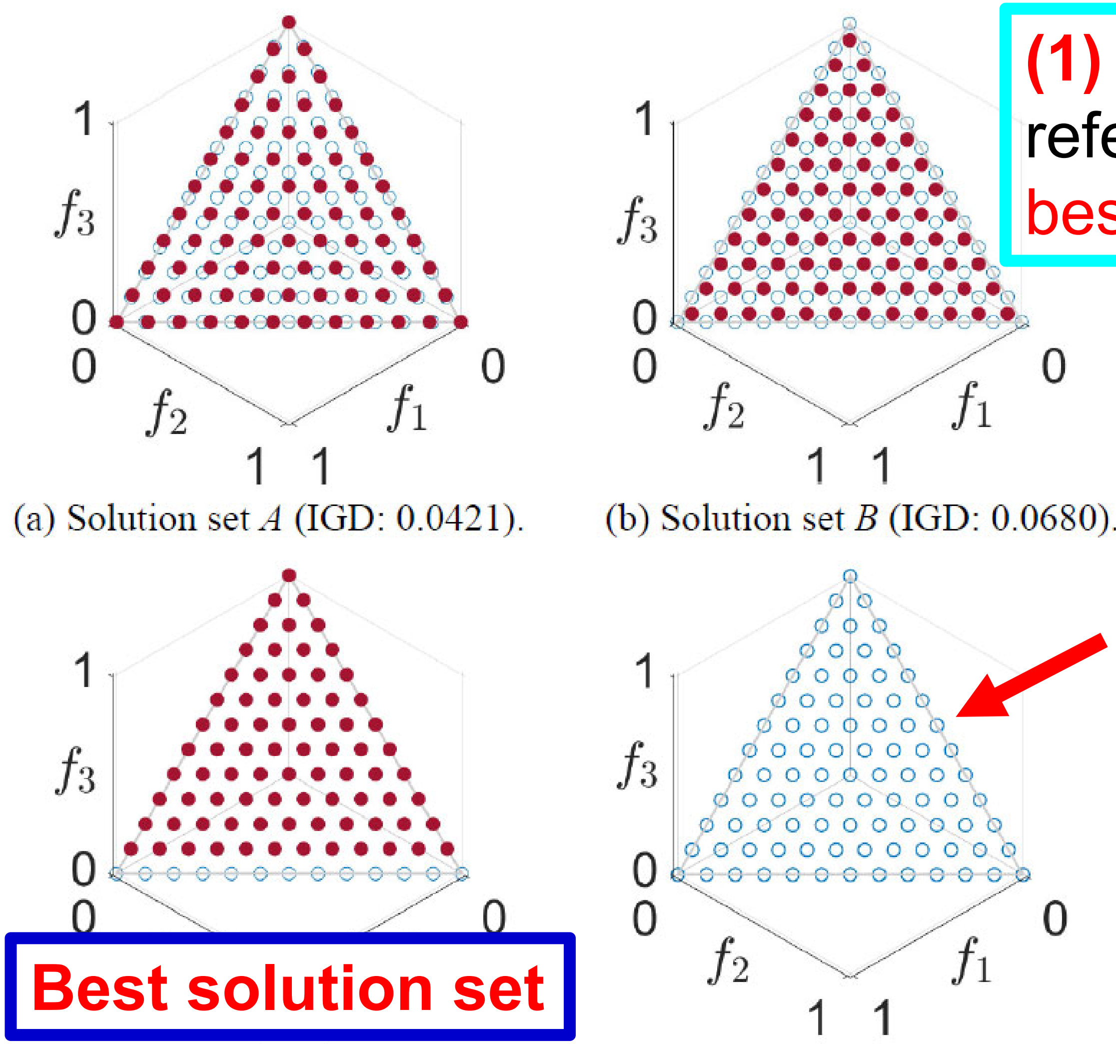
(c) Solution set C (IGD: 0.0168). (d) Reference point set ($H = 12$).

Fig. 22. Three solution sets with 78 solutions and 91 reference points. The IGD value of each solution set is shown in parentheses.

[CIM 2022]

Depending on the reference point set, each of the following solution sets can be evaluated as the best solution set.

(1) When (d) is used as the reference point set, (c) is the best solution set for IGD.



Reference point set

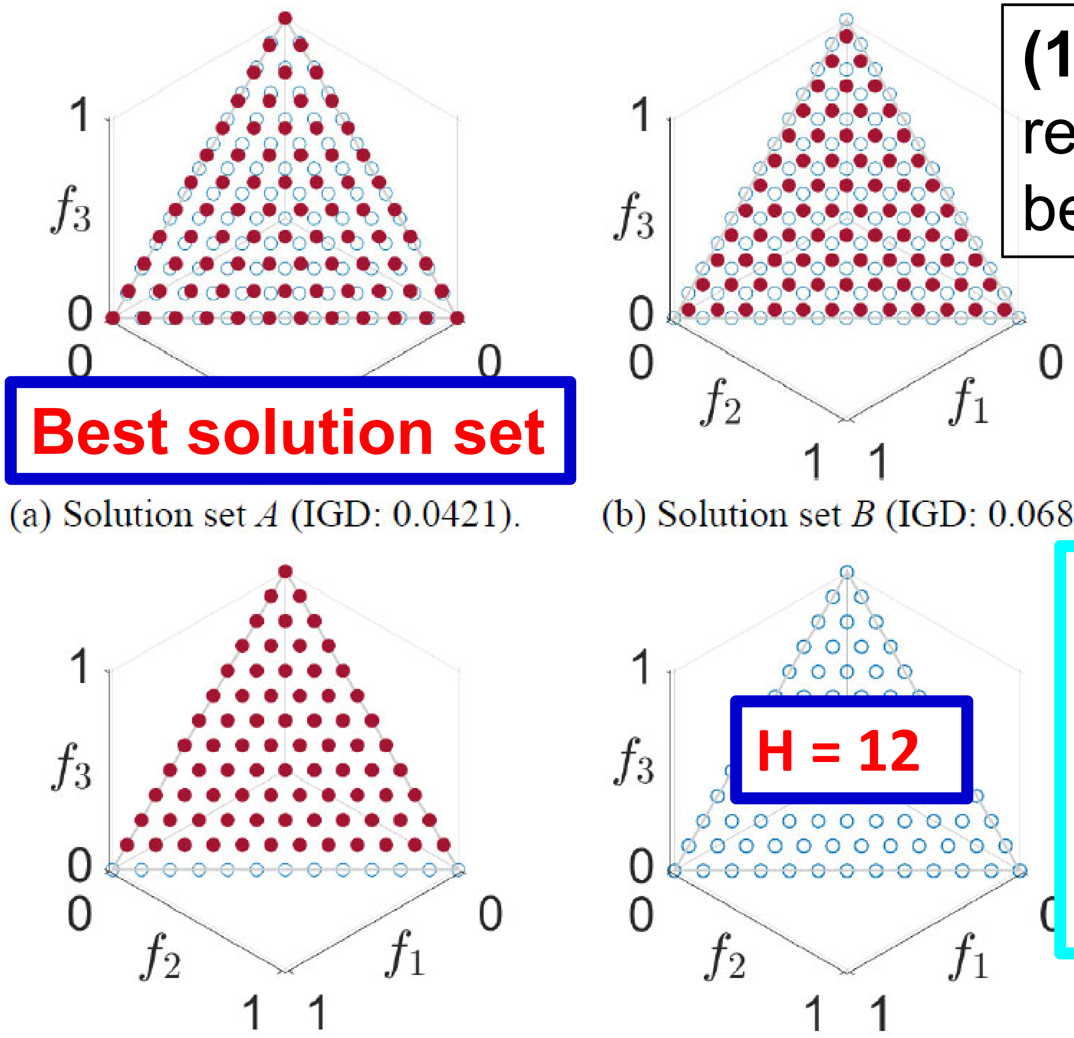
(a) Solution set *A* (IGD: 0.0421).
(b) Solution set *B* (IGD: 0.0680).
(c) Solution set *C* (IGD: 0.0168).
(d) Reference point set (H = 12).

Fig. 22. Three solution sets with 78 solutions and 91 reference points. The IGD value of each solution set is shown in parentheses.

[CIM 2022]

Depending on the reference point set, each of the following solution sets can be evaluated as the best solution set.

(1) When (d) is used as the reference point set, (c) is the best solution set for IGD.



(a) Solution set A (IGD: 0.0421). (b) Solution set B (IGD: 0.0680).

(c) Solution set C (IGD: 0.0168). (d) Reference point set (H = 12).

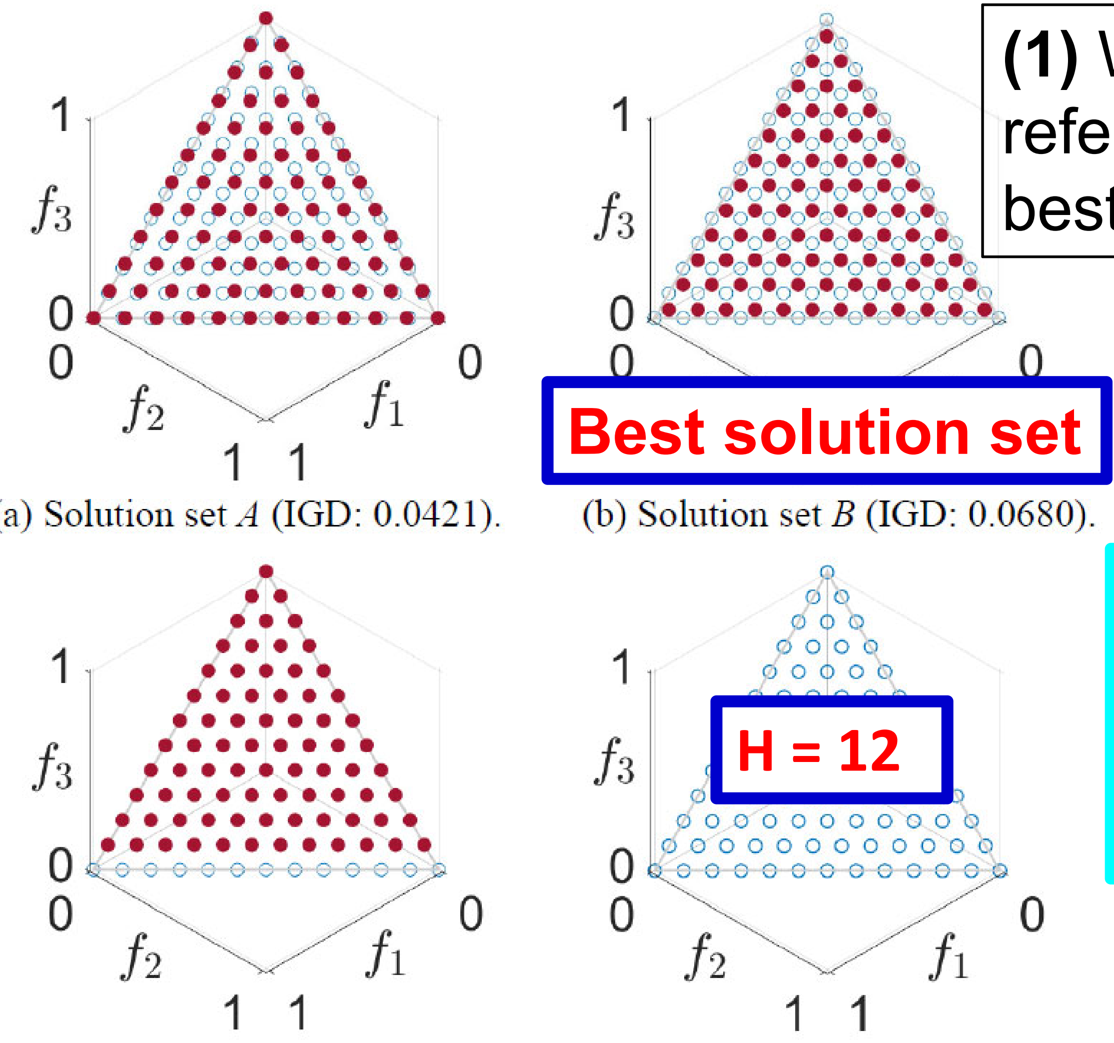
Fig. 22. Three solution sets with 78 solutions and 91 reference points. The IGD value of each solution set is shown in parentheses.

(2) When more reference points (H = 13, ..., 18) are used as a reference point set, (a) is the best solution set for IGD. **(Good !)**

[CIM 2022]

Depending on the reference point set, each of the following solution sets can be evaluated as the best solution set.

(1) When (d) is used as the reference point set, (c) is the best solution set for IGD.



(a) Solution set A (IGD: 0.0421). (b) Solution set B (IGD: 0.0680).

(c) Solution set C (IGD: 0.0168). (d) Reference point set ($H = 12$).

Fig. 22. Three solution sets with 78 solutions and 91 reference points. The IGD value of each solution set is shown in parentheses.

(3) For more reference points (e.g., 20,100 points), (b) is the best solution set for IGD.

[CIM 2022]

How to specify a set of reference points

Current Standard:

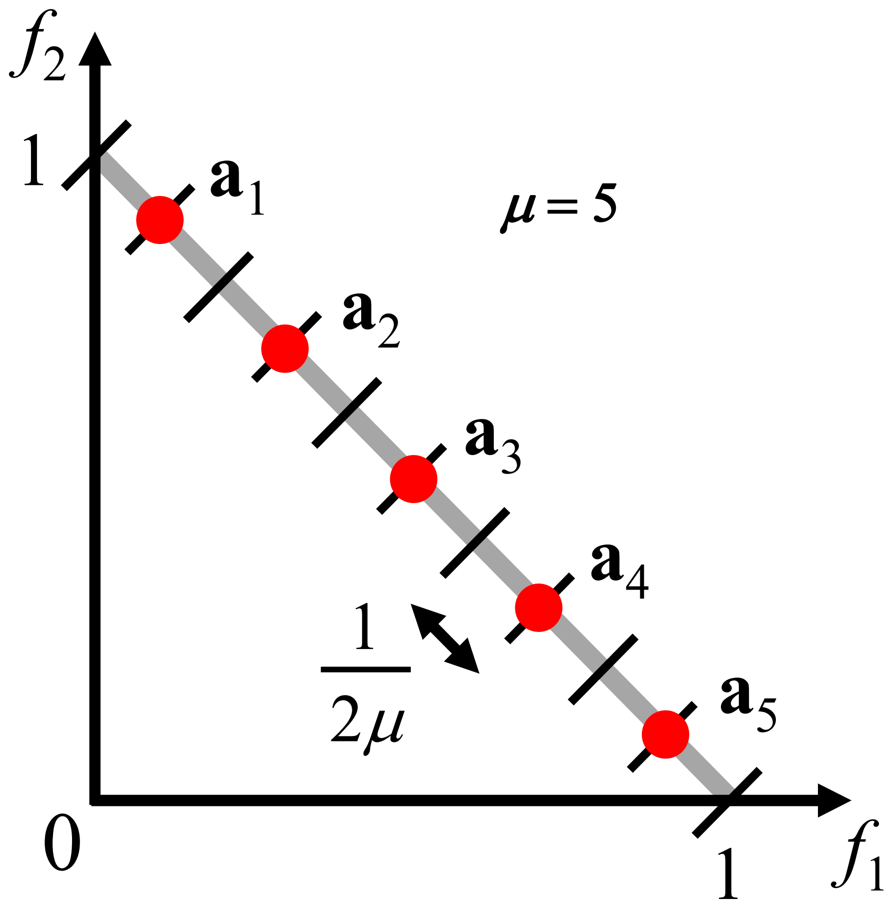
Use of a large number of uniformly distributed solutions as reference points for IGD calculation.

This is not always a good method.

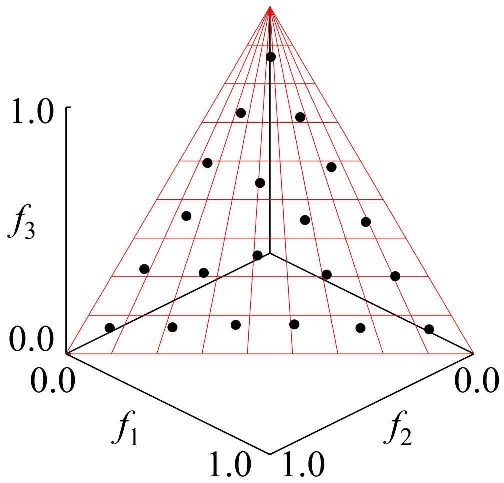
H. Ishibuchi, R. Imada, Y. Setoguchi, and Y. Nojima, "Reference point specification in inverted generational distance for triangular linear Pareto front," IEEE Trans. on Evolutionary Computation, 2018.

Optimal Distribution of Solutions for IGD

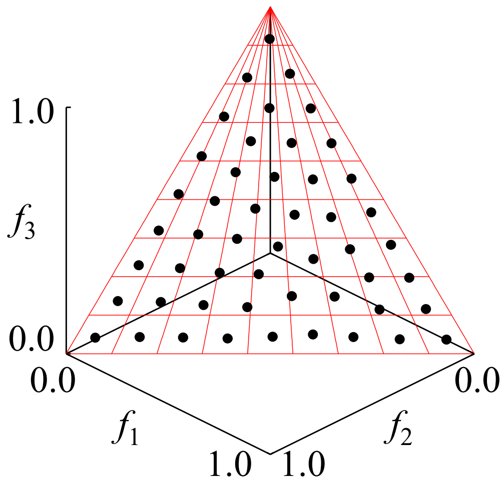
When an infinitely large number of uniformly distributed reference points on the Pareto front are used, the best distribution of solutions is as follows (*m* : population size)



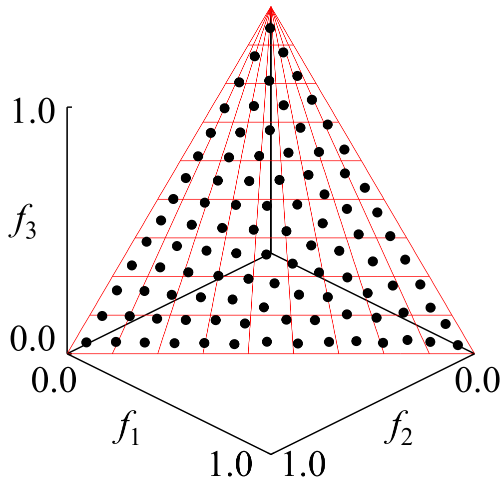
Optimal Distributions for IGD are not always intuitive (when we use a large number of uniform reference points)



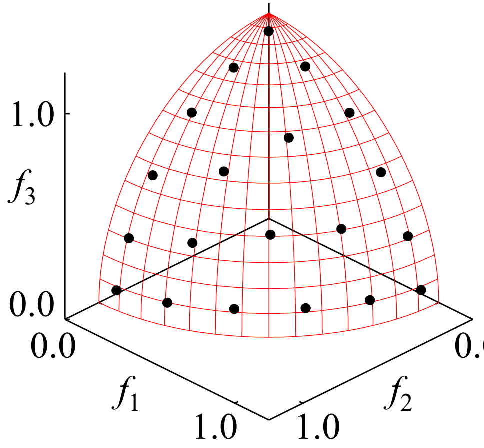
Population Size 20



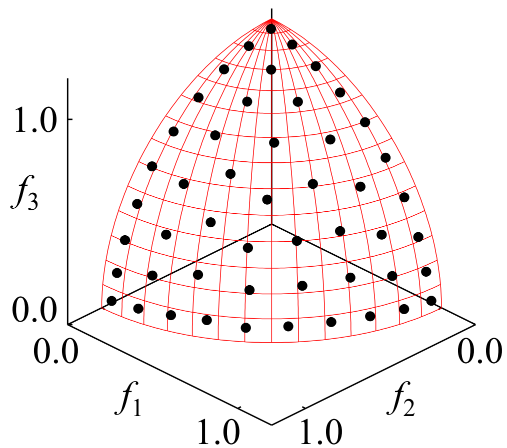
Population size 50



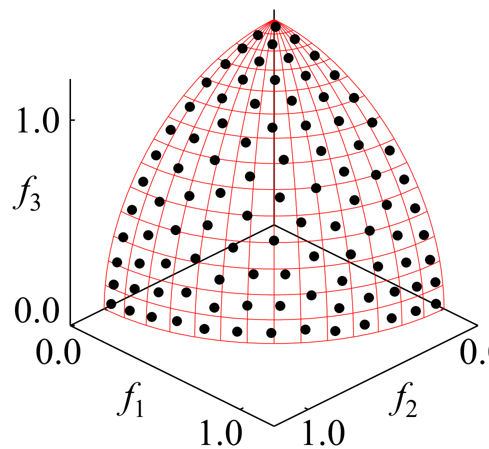
Population size 100



Population Size 20



Population size 50



Population size 100

Depending on the reference point set, each of the following solution sets can be evaluated as the best solution set.

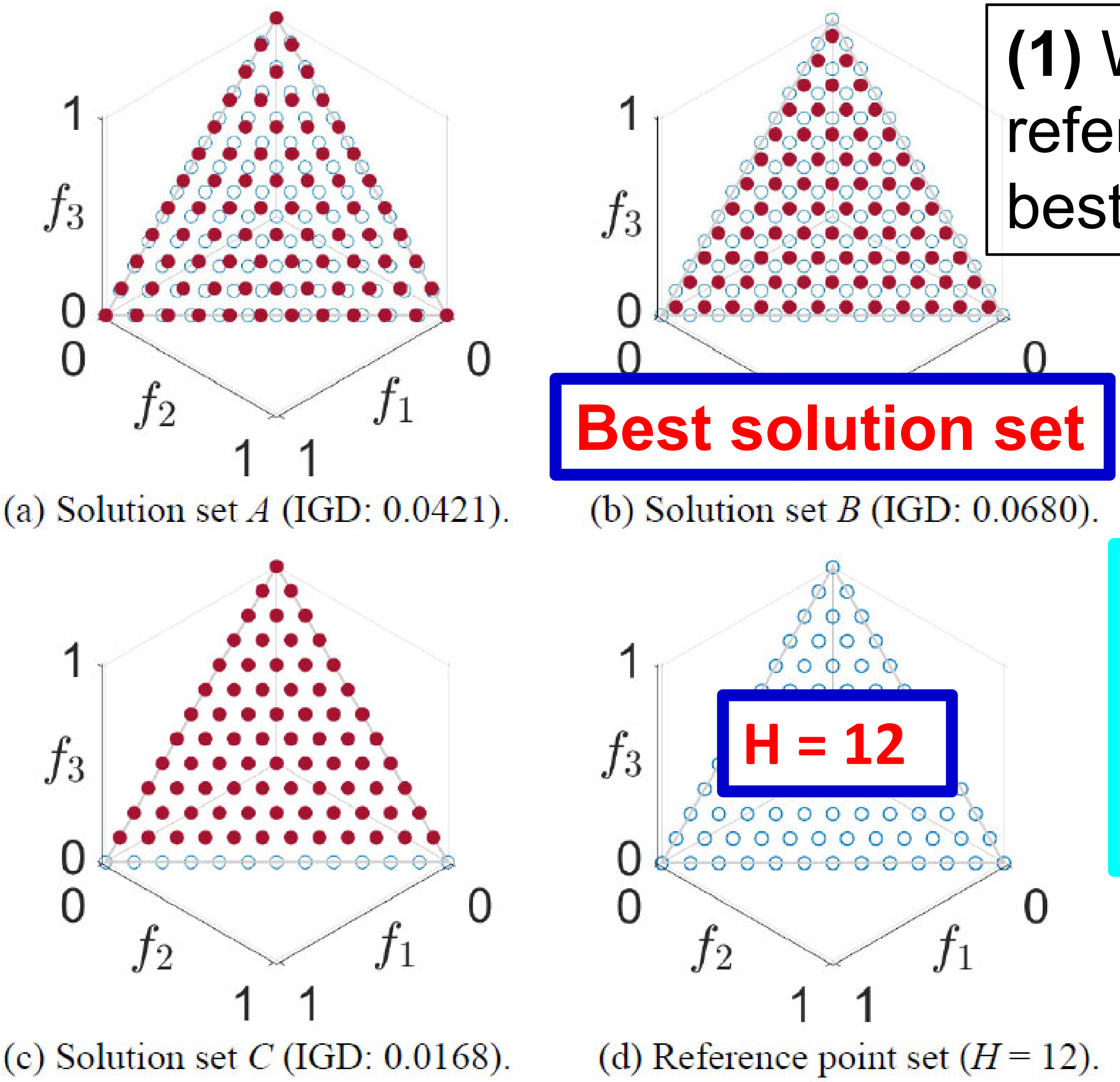


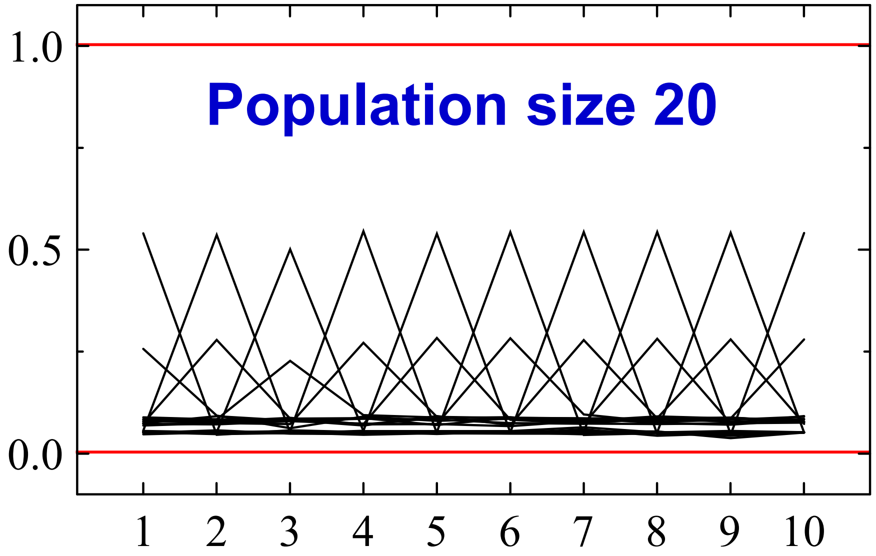
Fig. 22. Three solution sets with 78 solutions and 91 reference points. The IGD value of each solution set is shown in parentheses.

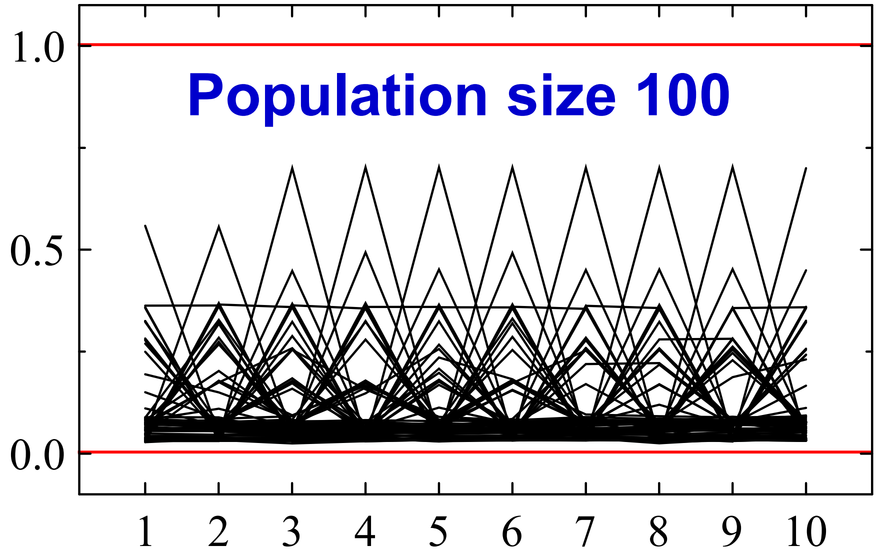
(1) When (d) is used as the reference point set, (c) is the best solution set for IGD.

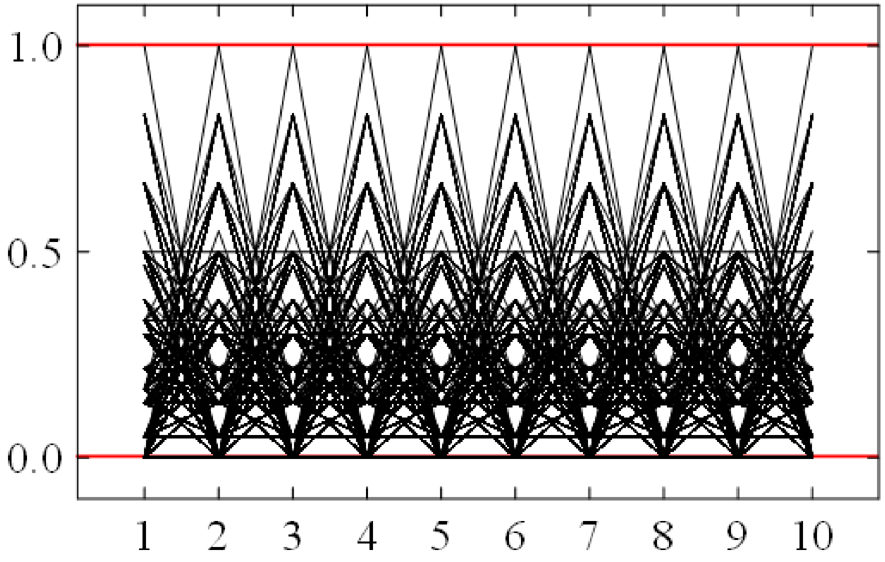
(3) For more reference points (e.g., 20,100 points), (b) is the best solution set for IGD.

[CIM 2022]

Optimal Distributions for IGD are not always intuitive (when we use a large number of uniform reference points)



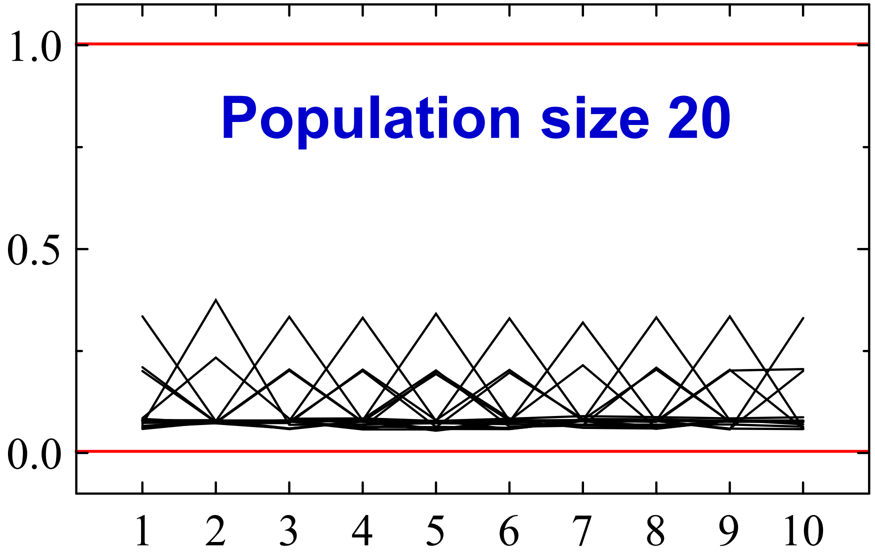


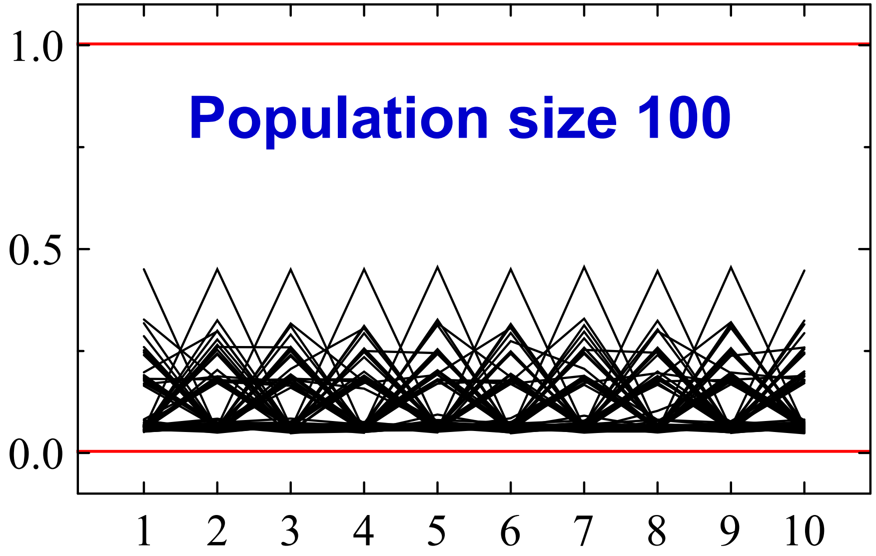


Reference points: Systematically Generated 10,010 points

Optimal Distributions for IGD are not always intuitive (when we use a large number of uniform reference points)

When randomly generated 100,000 reference points are used, the optimal distributions of solutions are as follows:

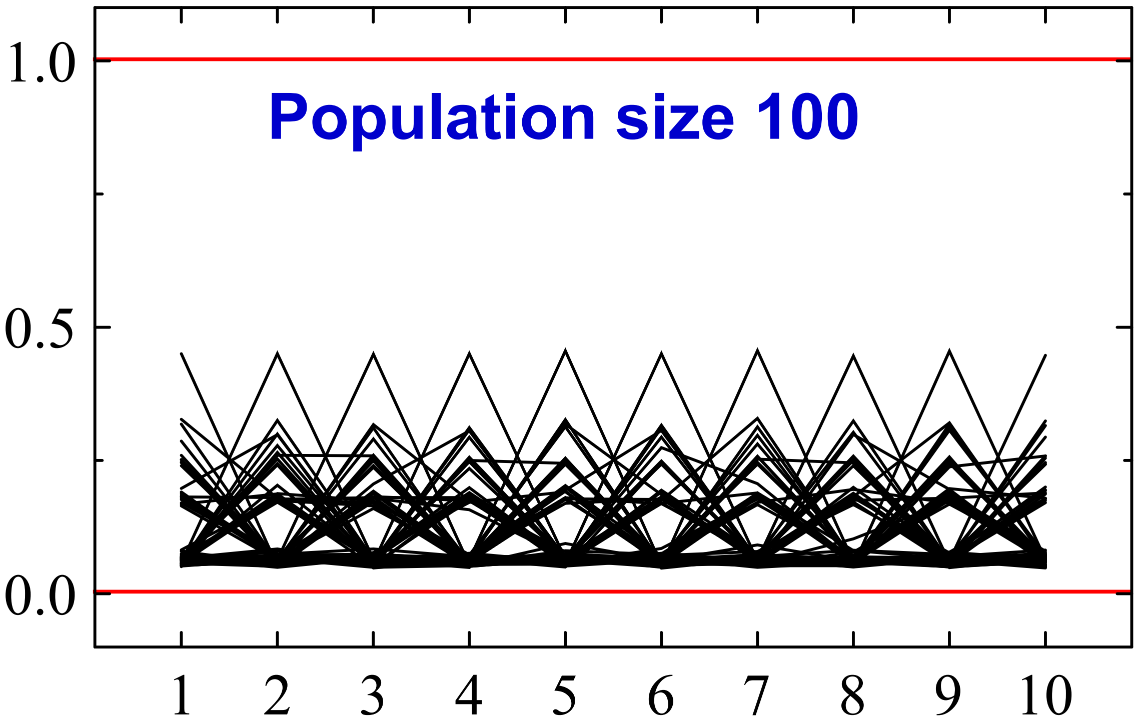


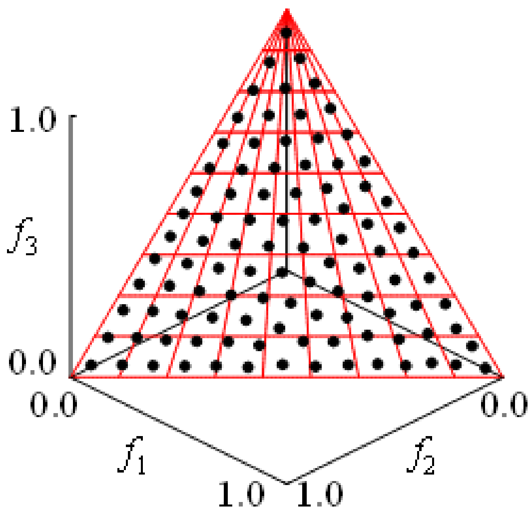


H. Ishibuchi, R. Imada, Y. Setoguchi, and Y. Nojima, "Reference point specification in inverted generational distance for triangular linear Pareto front," IEEE Trans. on Evolutionary Computation, 2018.

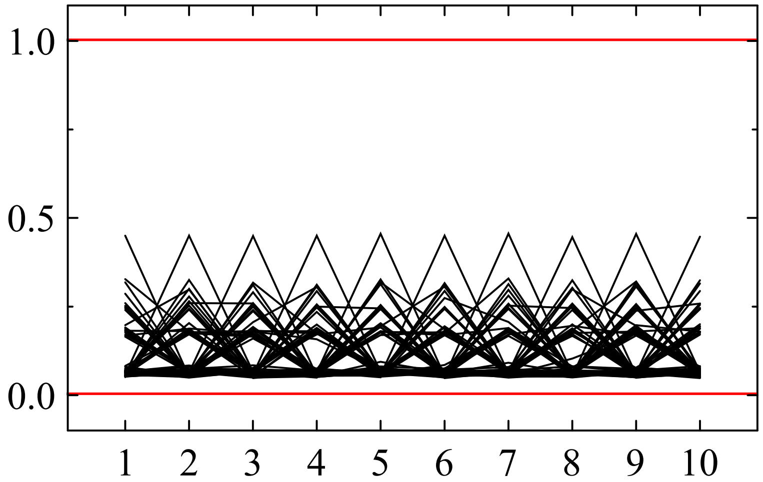
Additional Explanations

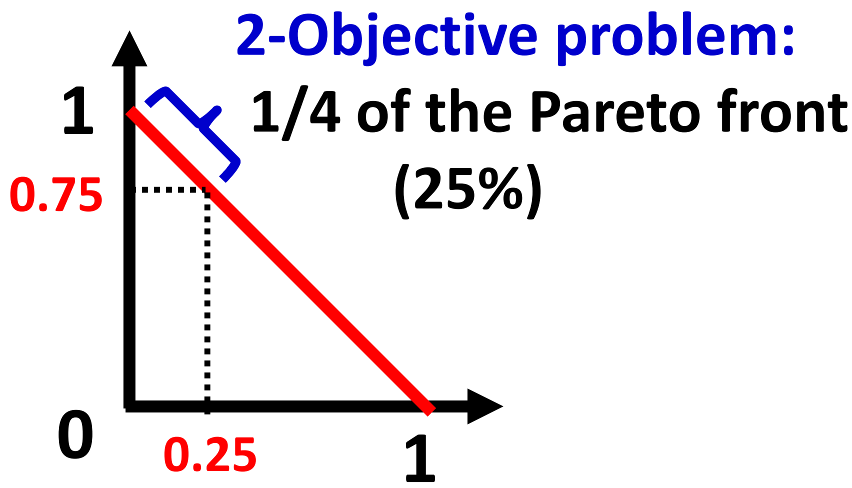
Near IGD-Optimal Distribution of Solutions for Random 100,000 Reference Points

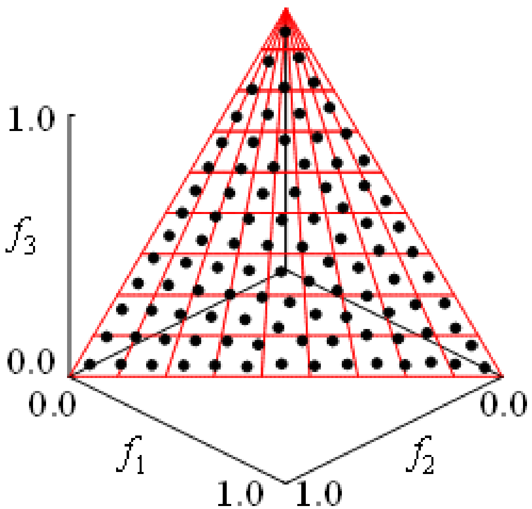




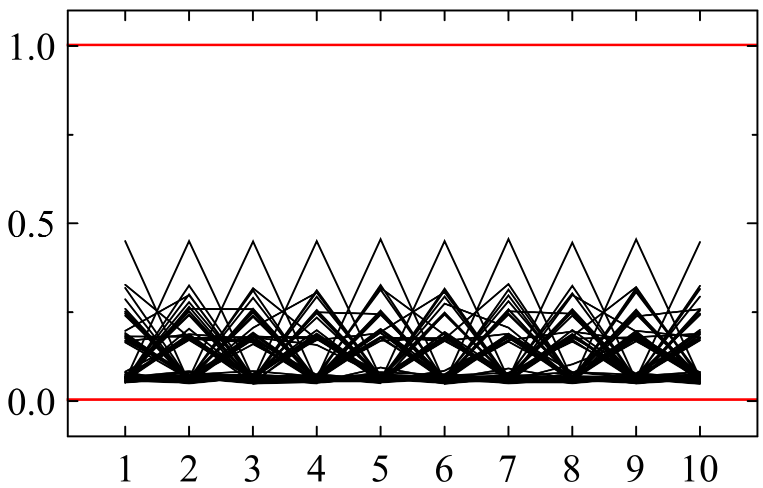
Additional Explanations



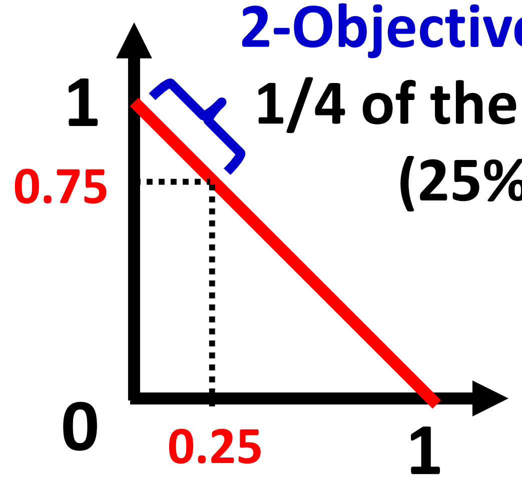


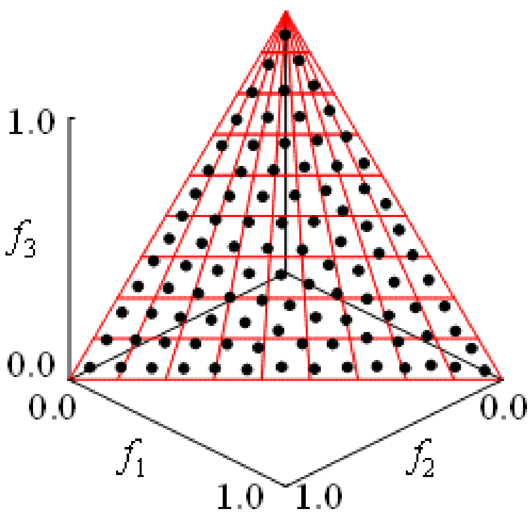


Additional Explanations

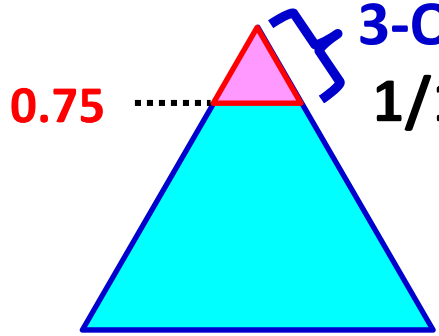


2-Objective problem:
1/4 of the Pareto front (25%)

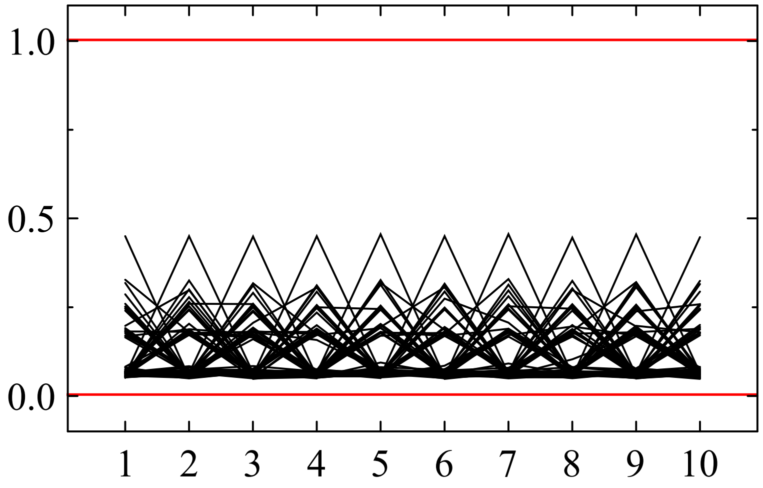


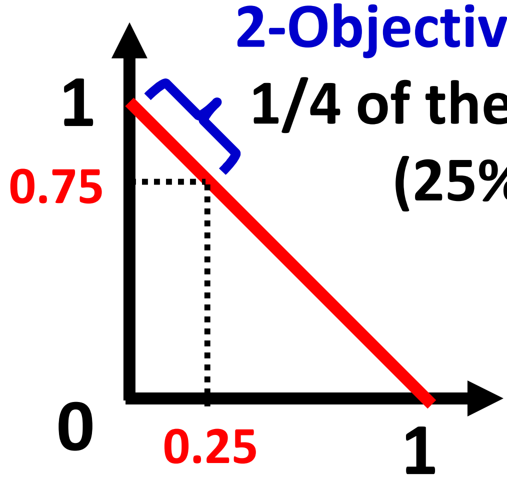


3-Objective problem
1/16 of the Pareto front (12.5%)

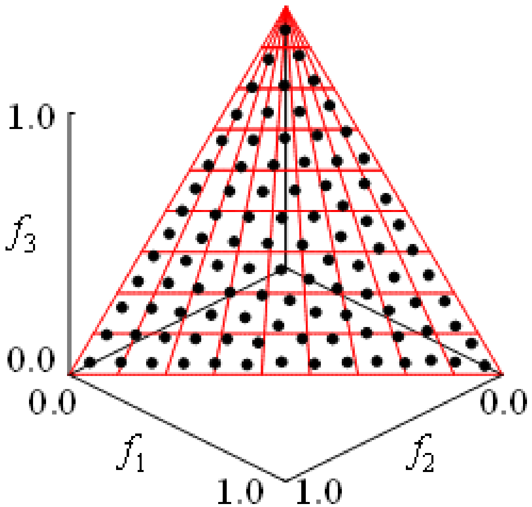


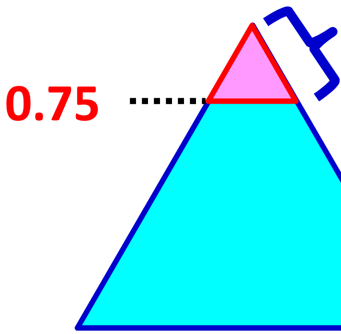
Additional Explanations





2-Objective problem:
1/4 of the Pareto front
(25%)



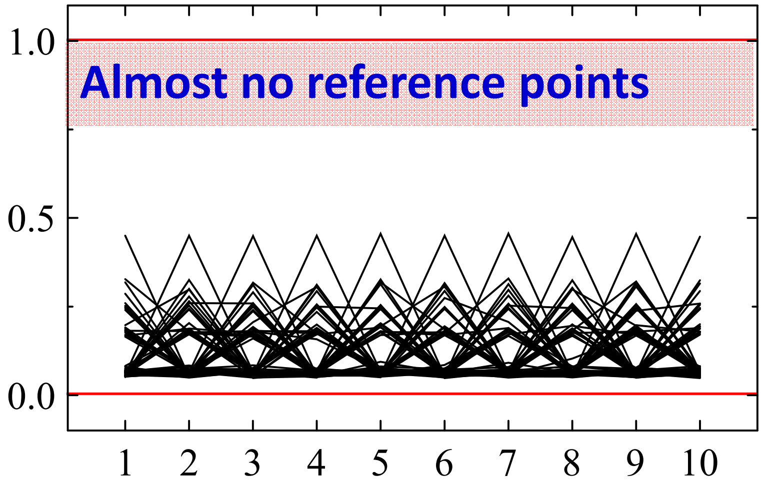


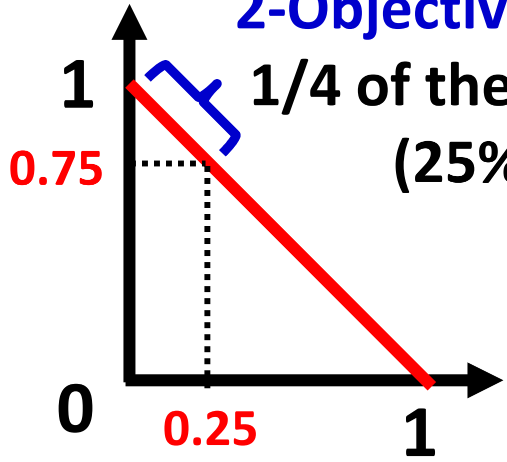
3-Objective problem
1/16 of the Pareto front
(12.5%)

10-objective problem:

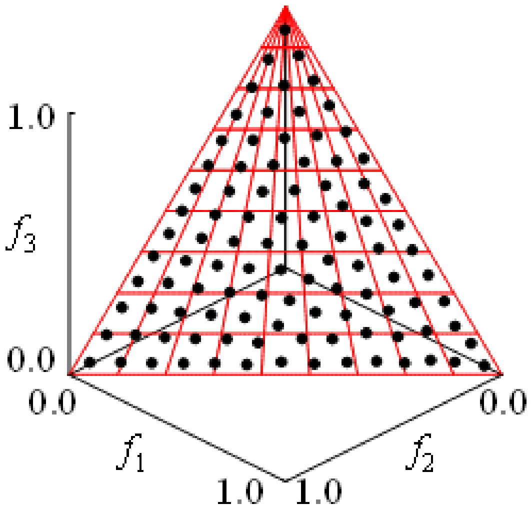
$(1/4)^9$ = 1/262,144 of the Pareto front
(0.00038%)

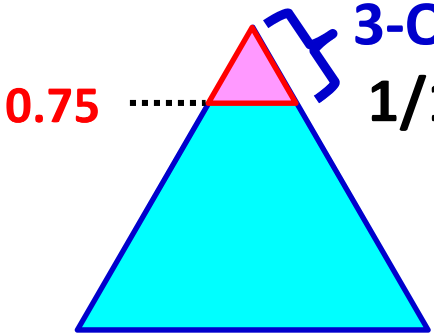
Additional Explanations





2-Objective problem:
1/4 of the Pareto front (25%)





3-Objective problem
1/16 of the Pareto front (12.5%)

10-objective problem:
$(1/4)^9$ = 1/262,144 of the Pareto front (0.00038%)

Difficulties in Fair Performance Comparison of Evolutionary Multi-Objective Optimization Algorithms

(0) Visual Comparison

(1) Specification of Termination Condition

(2) Specification of Population Size

(3) Choice of Performance Indicators (e.g., HV, IGD)

(4) Setting in Performance Indicators (e.g., reference point)

(5) Choice of Test Problems

This Talk is mainly based on my recent paper:

Hisao Ishibuchi, Lie Meng Pang, and Ke Shang, "**Difficulties in Fair Performance Comparison of Multi-Objective Evolutionary Algorithms**" ***IEEE Computational Intelligence Magazine*** **(February 2022)**

Literature Review about the Performance of NSGA-II

Very poor performance of NSGA-II on many-objective test problems has been reported in the literature.

Some Examples:

(1) The average HV value by NSGA-II on the 5-objective DTLZ1 test problem is zero [1].

[1] T. Wagner, N. Beume, and B. Naujoks: "Pareto-, aggregation-, and indicator-based methods in many-objective optimization," **EMO 2007**.

(2) NSGA-II is outperformed by random search on the 10-objective DTLZ2 test problem [2].

[2] S. Mostaghim and H. Schmeck: "Distance based ranking in many-objective particle swarm optimization," **PPSN 2008**.

Literature Review about the Performance of NSGA-II

Very poor performance of NSGA-II on many-objective test problems has been reported in the literature.

Reported results in [3]:

(3) NSGA-II (2002) is clearly outperformed by MOEA/D (2007), NSGA-III (2014), MOEA/DD (2015) and θ-DEA (2016) on many-objective DTLZ and WFG test problems.

[3] H. Ishibuchi, Y. Setoguchi, H. Masuda, and Y. Nojima: "Performance of decomposition-based many-objective algorithms strongly depends on Pareto front shapes," ***IEEE Trans. on Evolutionary Computation* (2017)**.

Literature Review about the Performance of NSGA-II

Very poor performance of NSGA-II on many-objective test problems has been reported in the literature.

Reported results in [3]:

(3) NSGA-II (2002) is clearly outperformed by MOEA/D (2007), NSGA-III (2014), MOEA/DD (2015) and θ-DEA (2016) on many-objective DTLZ and WFG test problems.

[3] H. Ishibuchi, Y. Setoguchi, H. Masuda, and Y. Nojima: "Performance of decomposition-based many-objective algorithms strongly depends on Pareto front shapes," ***IEEE Trans. on Evolutionary Computation*** **(2017)**.

We have also reported very good performance of NSGA-II in the same paper [3].

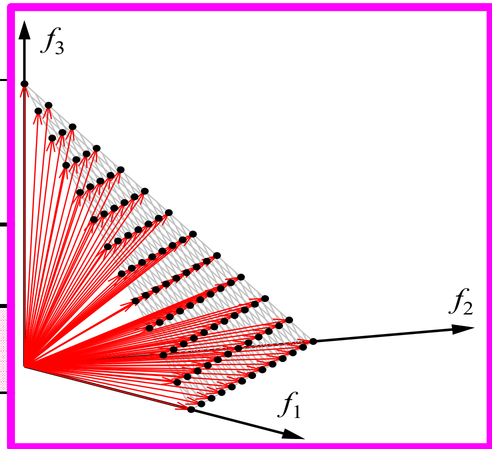
Best algorithm percentage for each test problems with 3, 5, 8 and 10 objectives: IGD-based comparison results with randomly generated 100,000 reference points.

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**

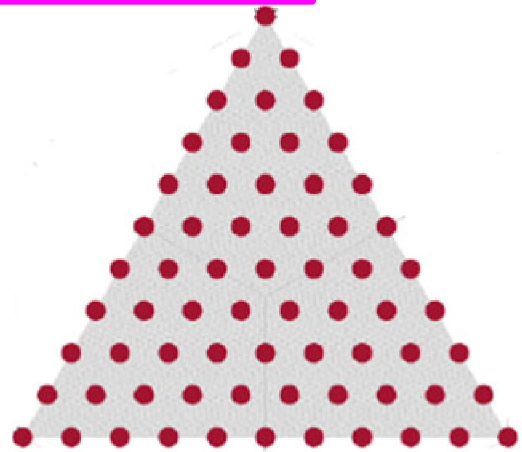
[Ishibuchi et al. IEEE CIM 2022]

NSGA-II is outperformed by many-objective algorithms on many-objective DTLZ problems

Best algorithm percentage for each test problems with 3, 5, 8 and 10 objectives: IGD-based comparison results with randomly generated 100,000 reference points.



Algorithm	DTLZ1-4	WFG1-9	[illegible]	…inus-…FG1-9
NSGA-III	6.25	16.67	[illegible]	5.56
θ-DEA	18.75	**44.44**	[illegible]	0.00
MOEA/DD	**62.50**	11.11	[illegible]	0.00
MOEA/D-PBI	12.50	5.56	0…	[illegible]
MOEA/D-Tch	0.00	5.56	0…	[illegible]
MOEA/D-WS	0.00	0.00	6…	[illegible]
MOEA/D-IPBI	0.00	0.00	12…	[illegible]
NSGA-II	0.00	16.67	**68…**	[illegible]



[Ishibuchi et al. IEEE CIM 2022]

NSGA-II is outperformed by many-objective algorithms on many-objective DTLZ problems

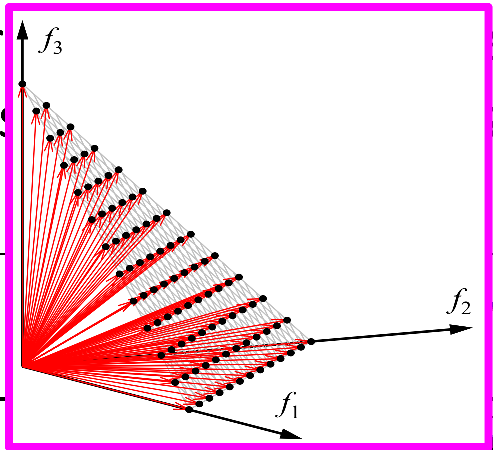
Best algorithm percentage for each test problems with 3, 5, 8 and 10 objectives: IGD-based comparison results with randomly generated 100,000 reference points.

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**

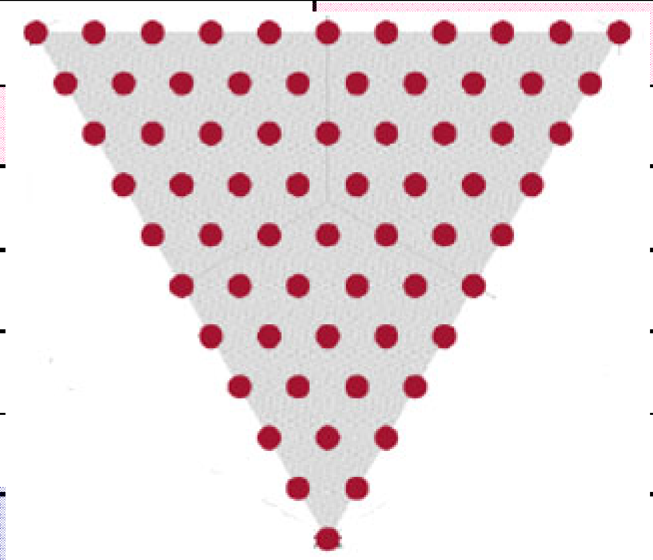
[Ishibuchi et al. IEEE CIM 2022]

NSGA-II outperforms many-objective algorithms on many-objective Minus-DTLZ problems

Best algorithm per[...]est problems with 3, 5, 8 and 10 objectives: IGD-bas[...]sults with randomly generated 100,000 reference [...]



Algorithm		G1-9	Minus-DTLZ1-4	Minus-WFG1-9
NSGA-III		.67	12.50	5.56
θ-DEA			0.00	0.00
MOEA/DD			0.00	0.00
MOEA/D-PBI			0.00	0.00
MOEA/D-Tch			0.00	8.33
MOEA/D-WS			6.25	0.00
MOEA/D-IPBI			12.50	16.67
NSGA-II			**68.75**	**69.44**



[Ishibuchi et al. IEEE CIM 2022]

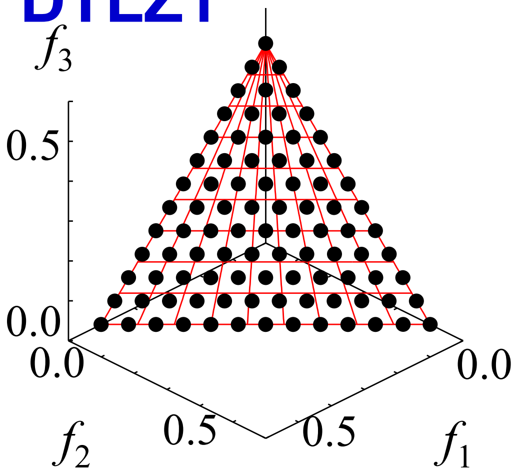
DTLZ: Minimize $f_i(\boldsymbol{x})$

Minus-DTLZ: Minimize $-f_i(\boldsymbol{x})$

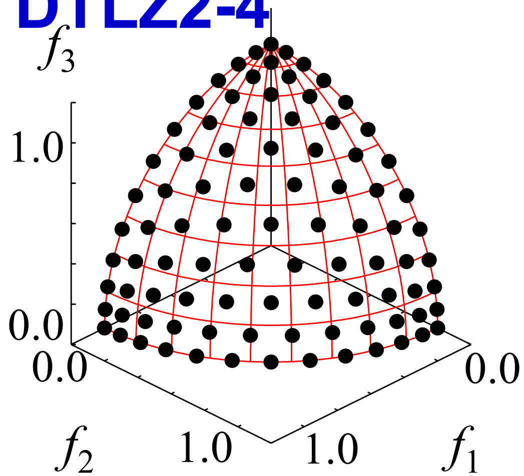
DTLZ: Minimize $f_i(\boldsymbol{x})$

Minus-DTLZ: Minimize $-f_i(\boldsymbol{x})$

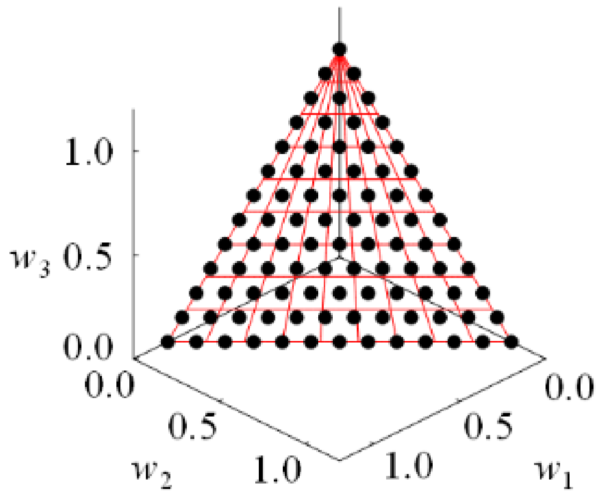
DTLZ1



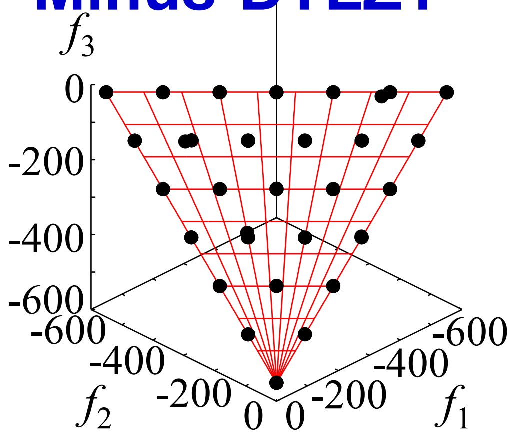
DTLZ2-4



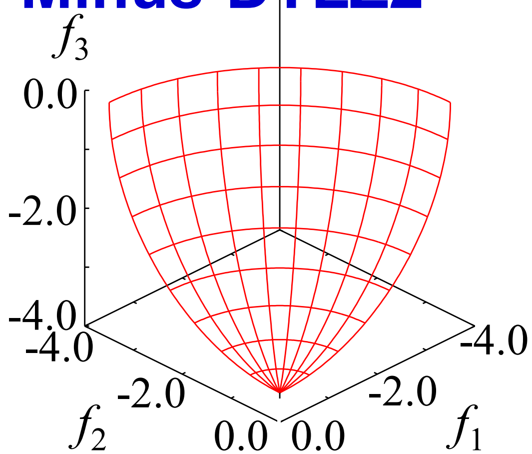
Weight Vectors



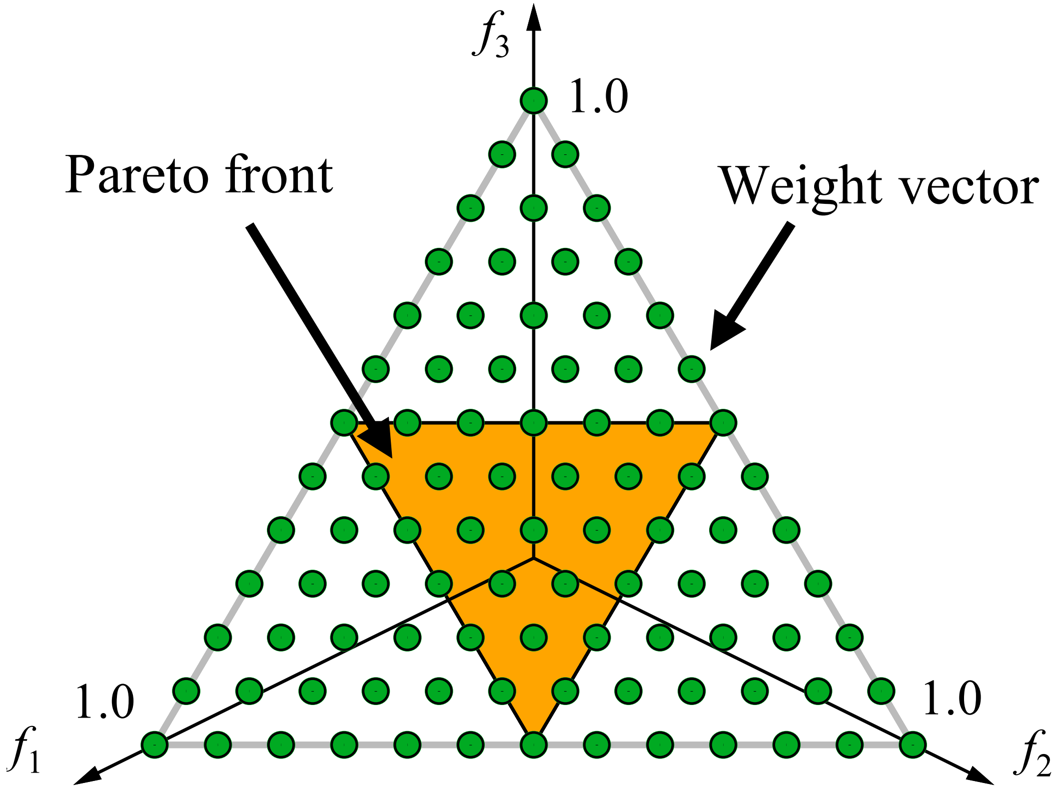
Minus-DTLZ1

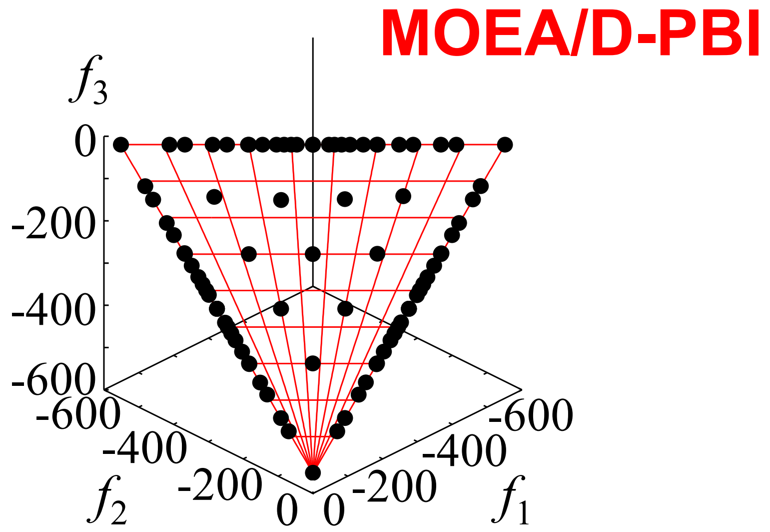


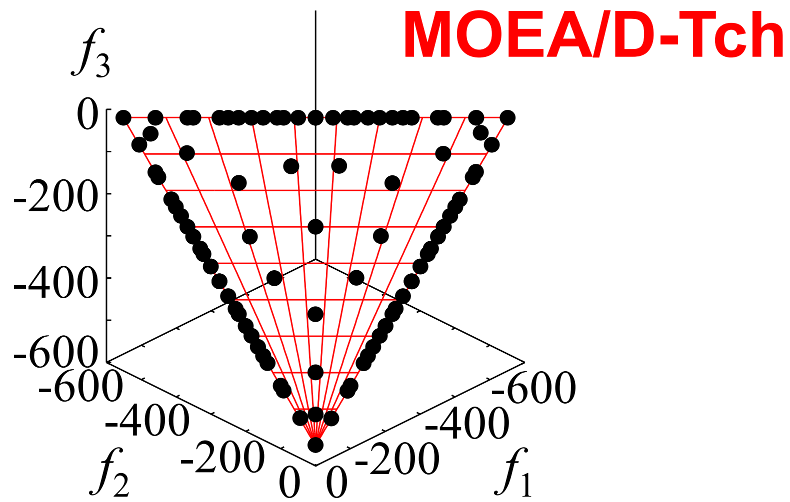
Minus-DTLZ2



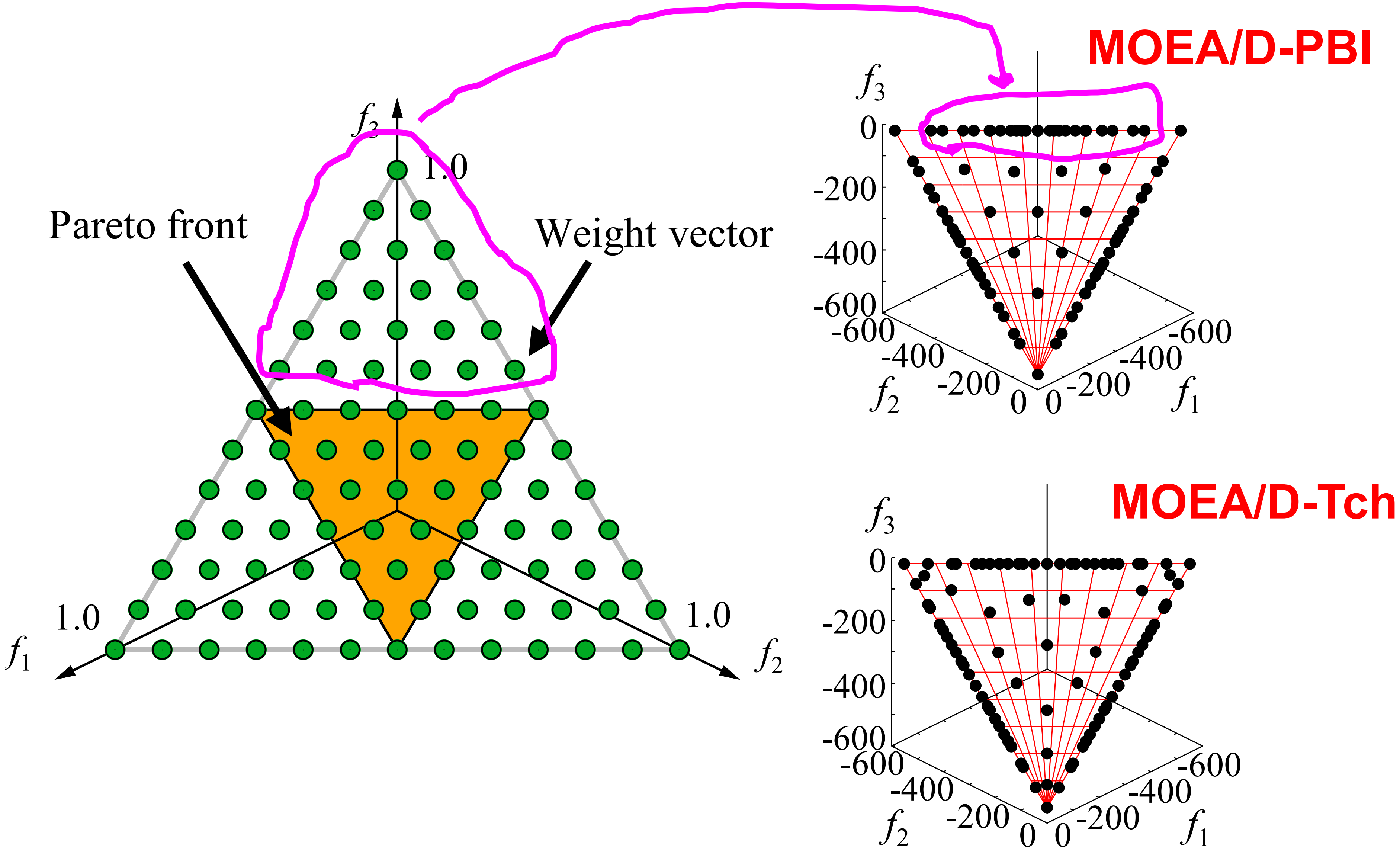
H. Ishibuchi, Y. Setoguchi, H. Masuda, and Y. Nojima, "Performance of decomposition-based many-objective algorithms strongly depends on Pareto front shapes," IEEE TEVC 2017.







H. Ishibuchi, Y. Setoguchi, H. Masuda, and Y. Nojima, “Performance of decomposition-based many-objective algorithms strongly depends on Pareto front shapes,” IEEE TEVC 2017.



Best algorithm percentage for each test problems with 3, 5, 8 and 10 objectives: IGD-based comparison results with randomly generated 100,000 reference points.

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**

[Ishibuchi et al. IEEE CIM 2022]

NSGA-II outperforms many-objective algorithms on many-objective Minus-DTLZ problems

Difficulties in Fair Performance Comparison of Evolutionary Multi-Objective Optimization Algorithms

(1) Specification of Termination Condition

(2) Specification of Population Size

(3) Choice of Performance Indicators (e.g., HV, IGD)

(4) Setting in Performance Indicators (e.g., reference point)

(5) Choice of Test Problems

Performance comparison results depend on the choice of test problems. It may be needed to use a wide variety of test problems including realistic test problems.

R. Tanabe and H. Ishibuchi, “An easy-to-use real-world multi-objective optimization problem suite,” Applied Soft Computing, vol. 89, April 2020.

Recent Development: Proposal of Real-World Problem Sets

[1] Tanabe, R., Ishibuchi, H.: **An easy-to-use real-world multi-objective optimization problem suite.** Applied Soft Computing, Article 106078 (2020).

[2] He, C., Tian, Y., Wang, H., Jin, Y.: **A repository of real-world datasets for data-driven evolutionary multiobjective optimization.** Complex & Intelligent Systems 6, 189-197 (2020).

[3] Kumar, A., Wu, G., Ali, M. Z., Luo, Q., Mallipeddi, R., Suganthan, P. N., Das, S.: **A benchmark-suite of real-world constrained multi-objective optimization problems and some baseline results.** Swarm and Evolutionary Computation 67, Article 100961 (2021).

Some examples of Irregular Pareto Fronts from [2].

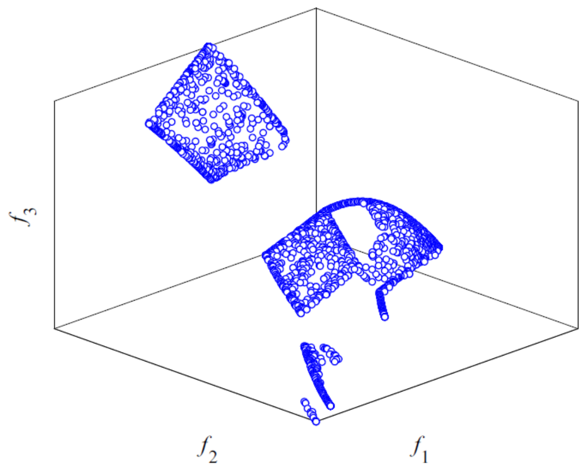


Fig. 3 The approximate POF of DDMOP2

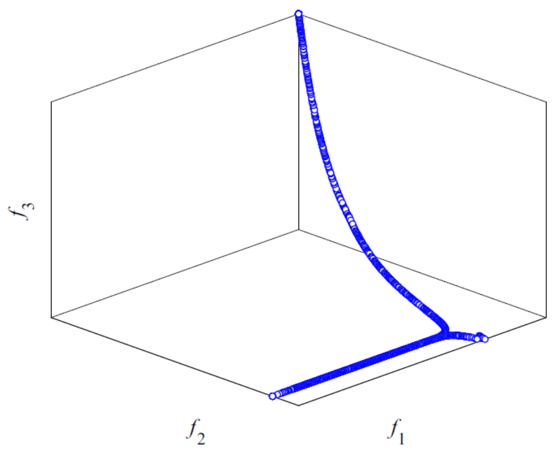


Fig. 4 The approximate POF of DDMOP3

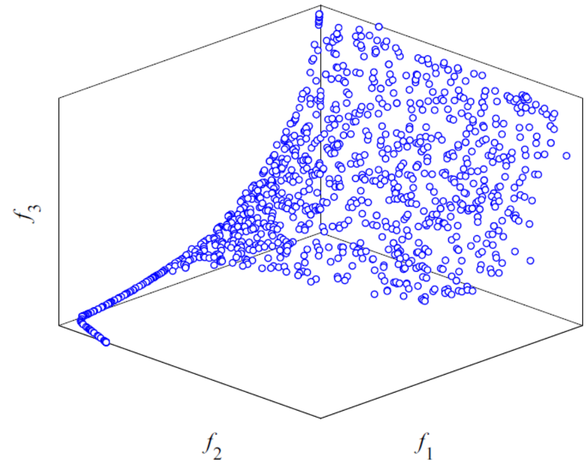


Fig. 5 The approximate POF of DDMOP5

Interesting Research Topic

New EMO Algorithm Design for Real-World Problems

Some examples of Pareto Fronts from [2].

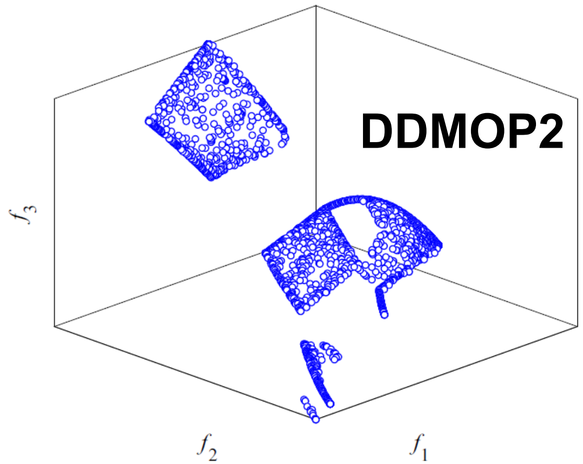


Fig. 3 The approximate POF of DDMOP2

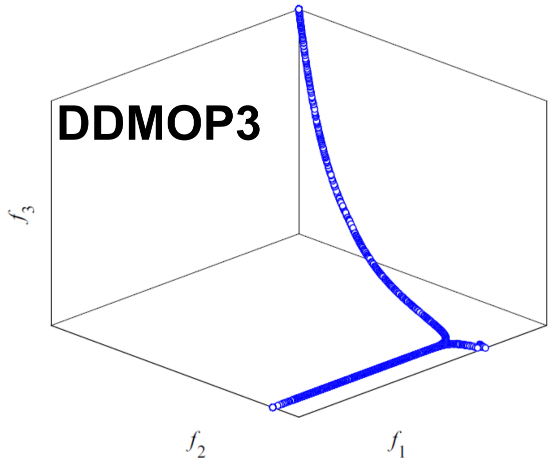


Fig. 4 The approximate POF of DDMOP3

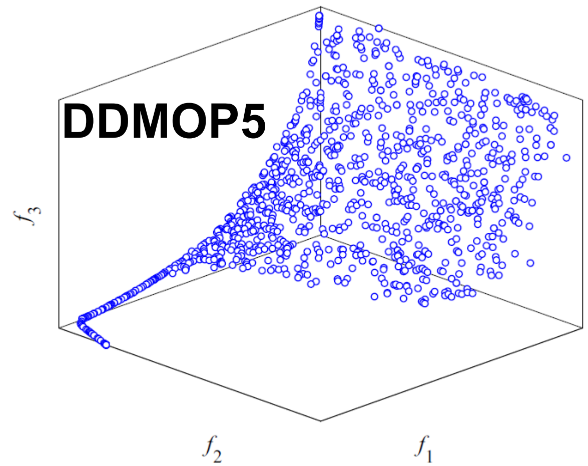


Fig. 5 The approximate POF of DDMOP5

[1] **16 RE problems**: Tanabe, R., Ishibuchi, H.: **An easy-to-use real-world multi-objective optimization problem suite.** *Applied Soft Computing*, 106078 (2020).

[2] **7 DDMOP problems**: He, C., Tian, Y., Wang, H., Jin, Y.: **A repository of real-world datasets for data-driven evolutionary multiobjective optimization.** *Complex & Intelligent Systems* 6, 189-197 (2020).

[3] **50 RCM problems**: Kumar, A., Wu, G., Ali, M. Z., Luo, Q., Mallipeddi, R., Suganthan, P. N., Das, S.: **A benchmark-suite of real-world constrained multi-objective optimization problems and some baseline results.** *Swarm and Evolutionary Computation* 67, 100961 (2021).

Interesting Research Topic

New EMO Algorithm Design for Real-World Problems

Some examples of Pareto Fronts from [2].

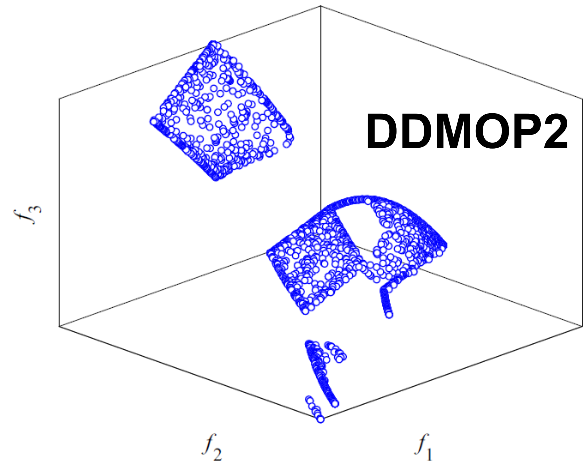


Fig. 3 The approximate POF of DDMOP2

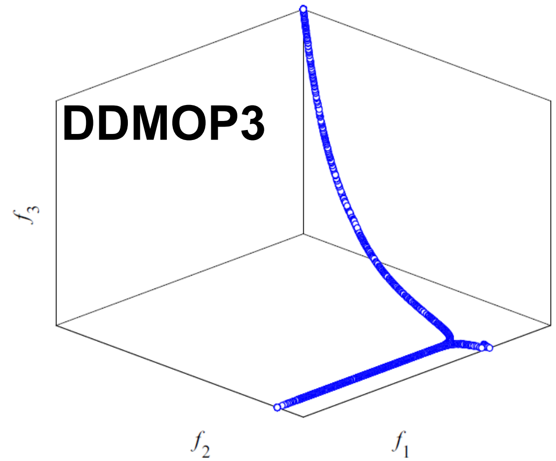
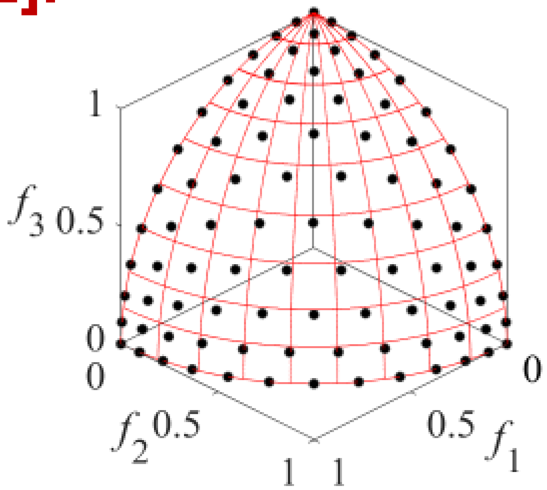


Fig. 4 The approximate POF of DDMOP3



MOEA/D on DTLZ2

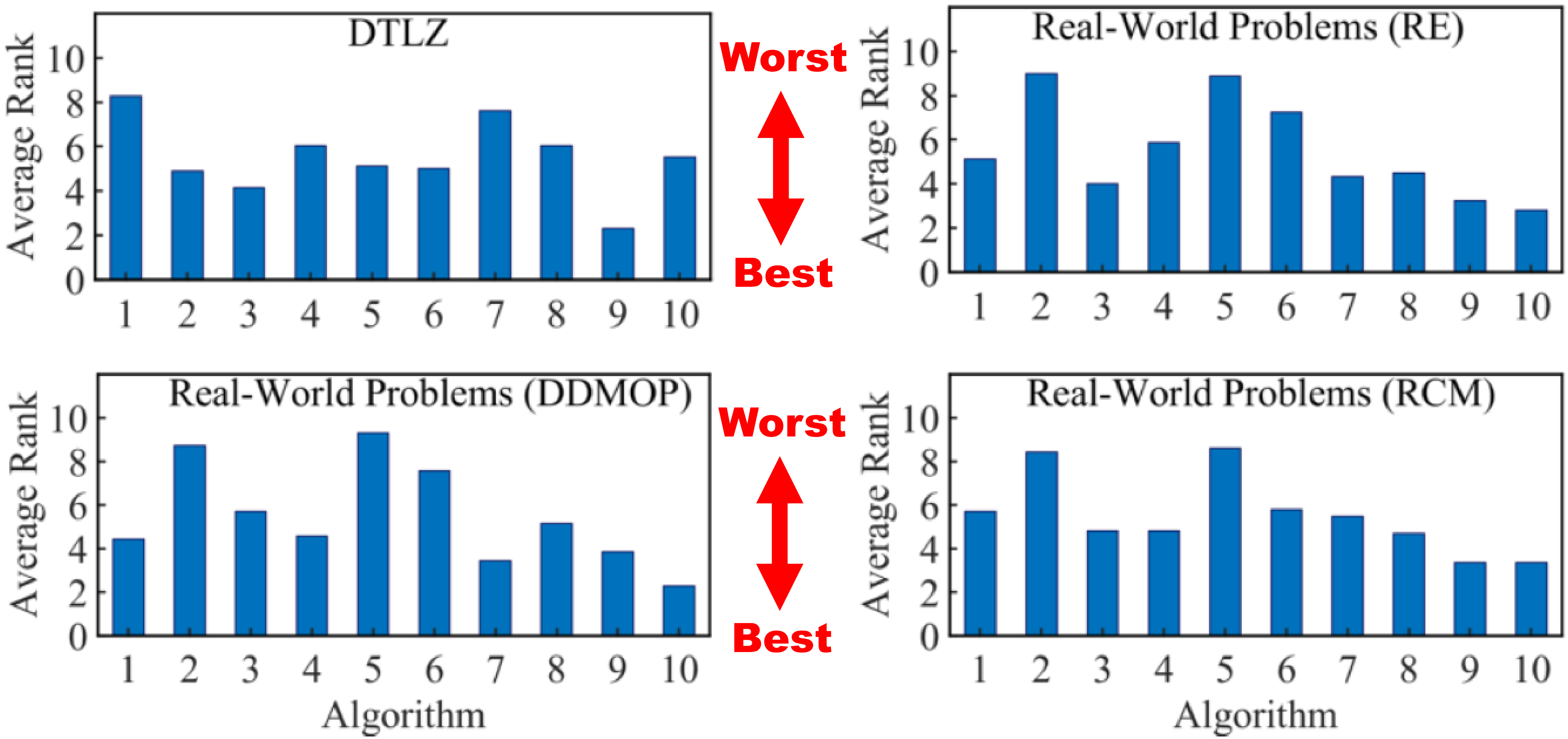
[1] **16 RE problems**: Tanabe, R., Ishibuchi, H.: **An easy-to-use real-world multi-objective optimization problem suite.** *Applied Soft Computing*, 106078 (2020).

[2] **7 DDMOP problems**: He, C., Tian, Y., Wang, H., Jin, Y.: **A repository of real-world datasets for data-driven evolutionary multiobjective optimization.** *Complex & Intelligent Systems* 6, 189-197 (2020).

[3] **50 RCM problems**: Kumar, A., Wu, G., Ali, M. Z., Luo, Q., Mallipeddi, R., Suganthan, P. N., Das, S.: **A benchmark-suite of real-world constrained multi-objective optimization problems and some baseline results.** *Swarm and Evolutionary Computation* 67, 100961 (2021).

Comparison Results on DTLZ and Real-world Problems

1: NSGA-II 2: MOEA/D-PBI 3: SMS-EMOA & HypE 4: NSGA-III 5: MOEA/DD 6: RVEA 7: SparseEA 8: DEA-GNG 9: R2HCA-EMOA 10: PREA



Totally different comparison results of ten EMO algorithms between the test problem DTLZ and the real-world problems

H. Ishibuchi, Y. Nan, and L. M. Pang, "Performance evaluation of multi-objective evolutionary algorithms using artificial and real-world problems," *Proc. EMO 2023*.

Interesting Research Topic

New EMO Algorithm Design for Real-World Problems

Some examples of Pareto Fronts from [2].

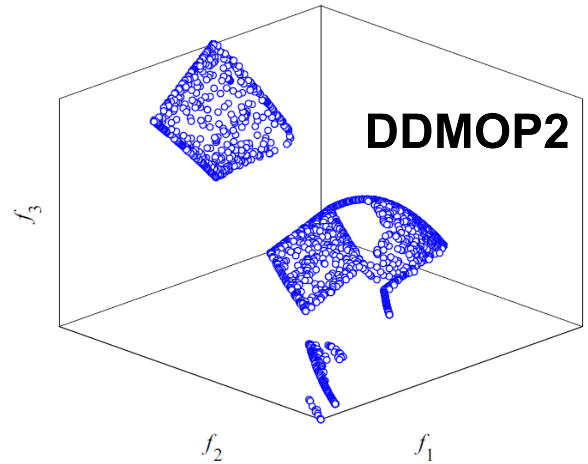


Fig. 3 The approximate POF of DDMOP2

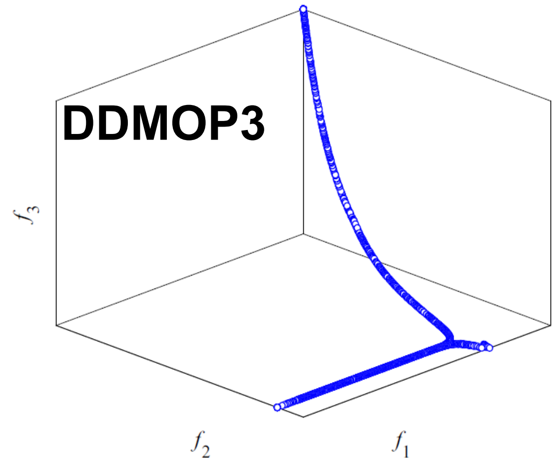
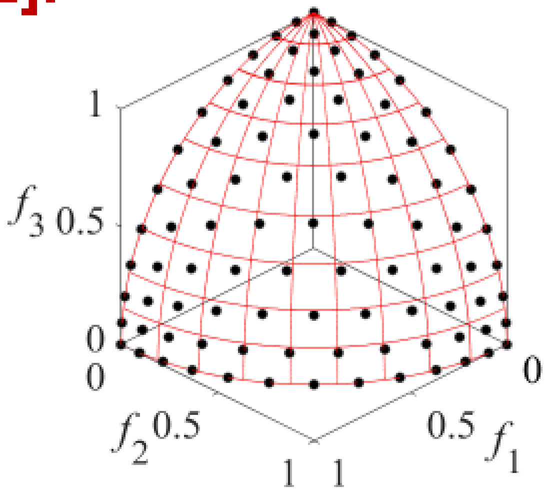


Fig. 4 The approximate POF of DDMOP3



MOEA/D on DTLZ2

[1] **16 RE problems**: Tanabe, R., Ishibuchi, H.: **An easy-to-use real-world multi-objective optimization problem suite.** *Applied Soft Computing*, 106078 (2020).

[2] **7 DDMOP problems**: He, C., Tian, Y., Wang, H., Jin, Y.: **A repository of real-world datasets for data-driven evolutionary multiobjective optimization.** *Complex & Intelligent Systems* 6, 189-197 (2020).

[3] **50 RCM problems**: Kumar, A., Wu, G., Ali, M. Z., Luo, Q., Mallipeddi, R., Suganthan, P. N., Das, S.: **A benchmark-suite of real-world constrained multi-objective optimization problems and some baseline results.** *Swarm and Evolutionary Computation* 67, 100961 (2021).

Suggestions:

(i) To focus on a class of problems (not all problems), and to demonstrate the usefulness of the proposed algorithm on that class of problems.

(ii) To design a general robust and flexible framework which can be easily modified for different problems, and demonstrate its usefulness on various problems.

Conclusion:

Fair performance comparison of EMO algorithms is difficult.
We can easily obtain clearly different comparison results.

Today's Contents

(0) Visual Comparison
(1) Specification of Termination Condition
(2) Specification of Population Size
(3) Choice of Performance Indicators (e.g., HV, IGD)
(4) Setting in Performance Indicators (e.g., reference point)
(5) Choice of Test Problems

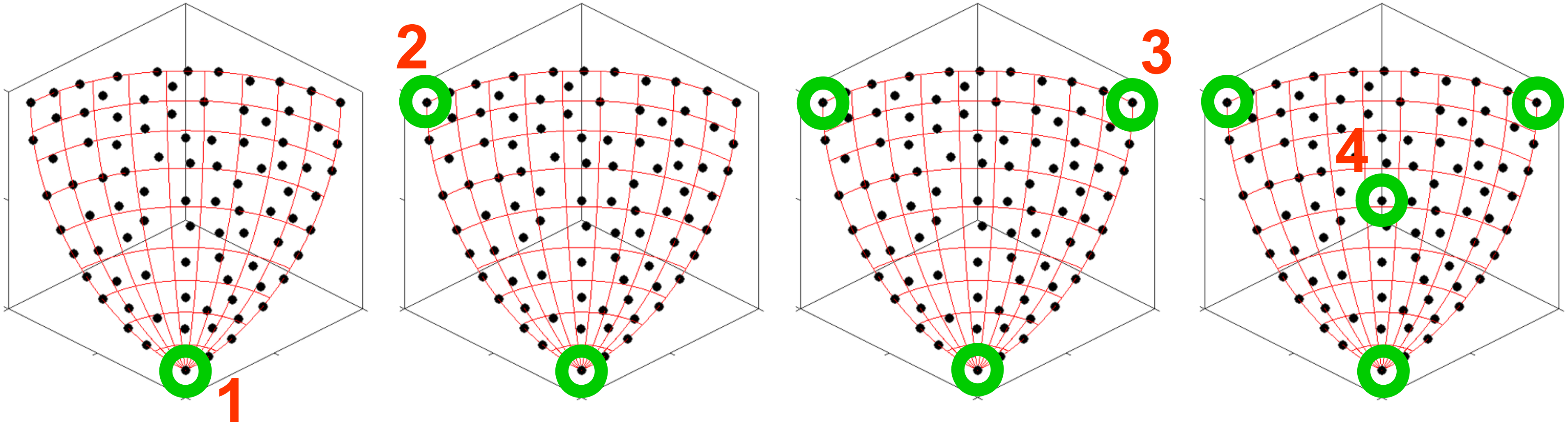
This Talk is mainly based on my recent paper:

Hisao Ishibuchi, Lie Meng Pang, and Ke Shang, "**Difficulties in Fair Performance Comparison of Multi-Objective Evolutionary Algorithms**" ***IEEE Computational Intelligence Magazine*** **(February 2022)**

Solution Selection in Our Computational Experiment

A simple distance-based subset selection in [13].

(i) Selection of the first solution: An extreme solution.
(ii) Selection of the others: The most distant solution from the selected ones.

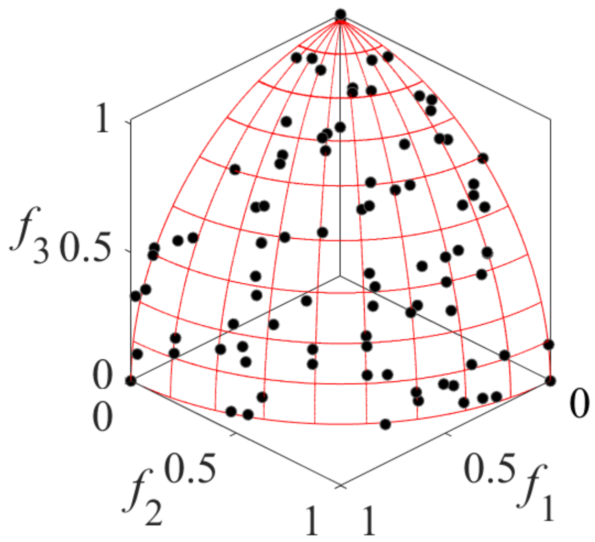


[13] H. K. Singh, K. S. Bhattacharjee, and T. Ray, “Distance-based subset selection for benchmarking in evolutionary multi/many-objective optimization” **IEEE TEVC (2019)**.

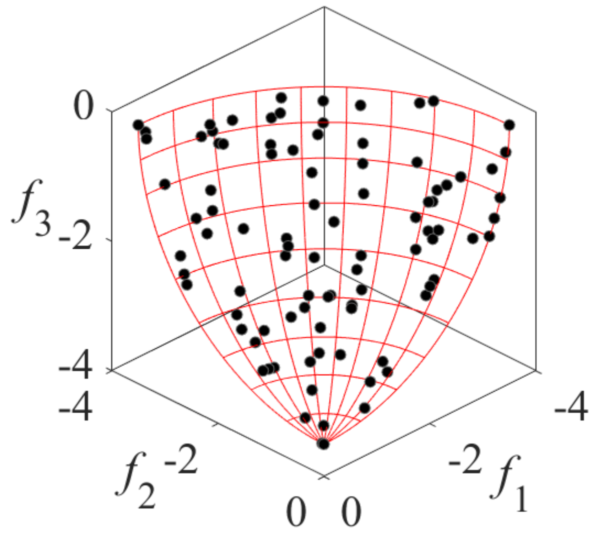
For further discussions on solution selection, see

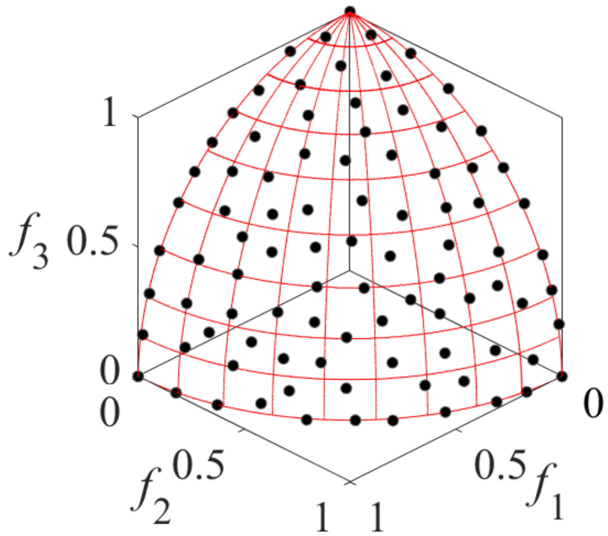
[14] H. Ishibuchi, L. M. Pang, and K. Shang, "Solution subset selection for final decision making in evolutionary multi-objective optimization" **arXiv (2020)**.

Experimental Results (NSGA-II)

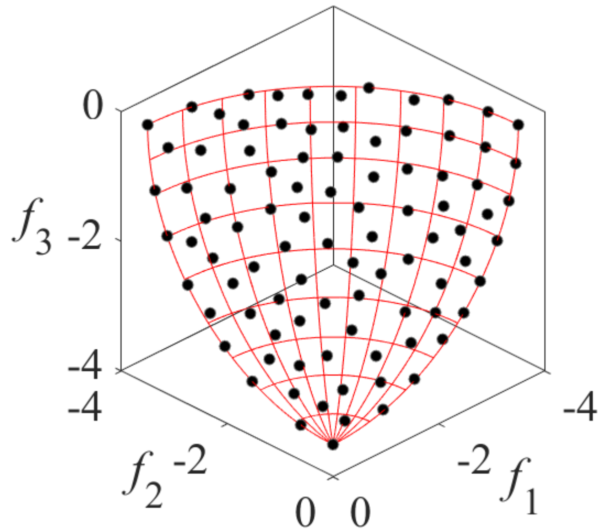


Final Population





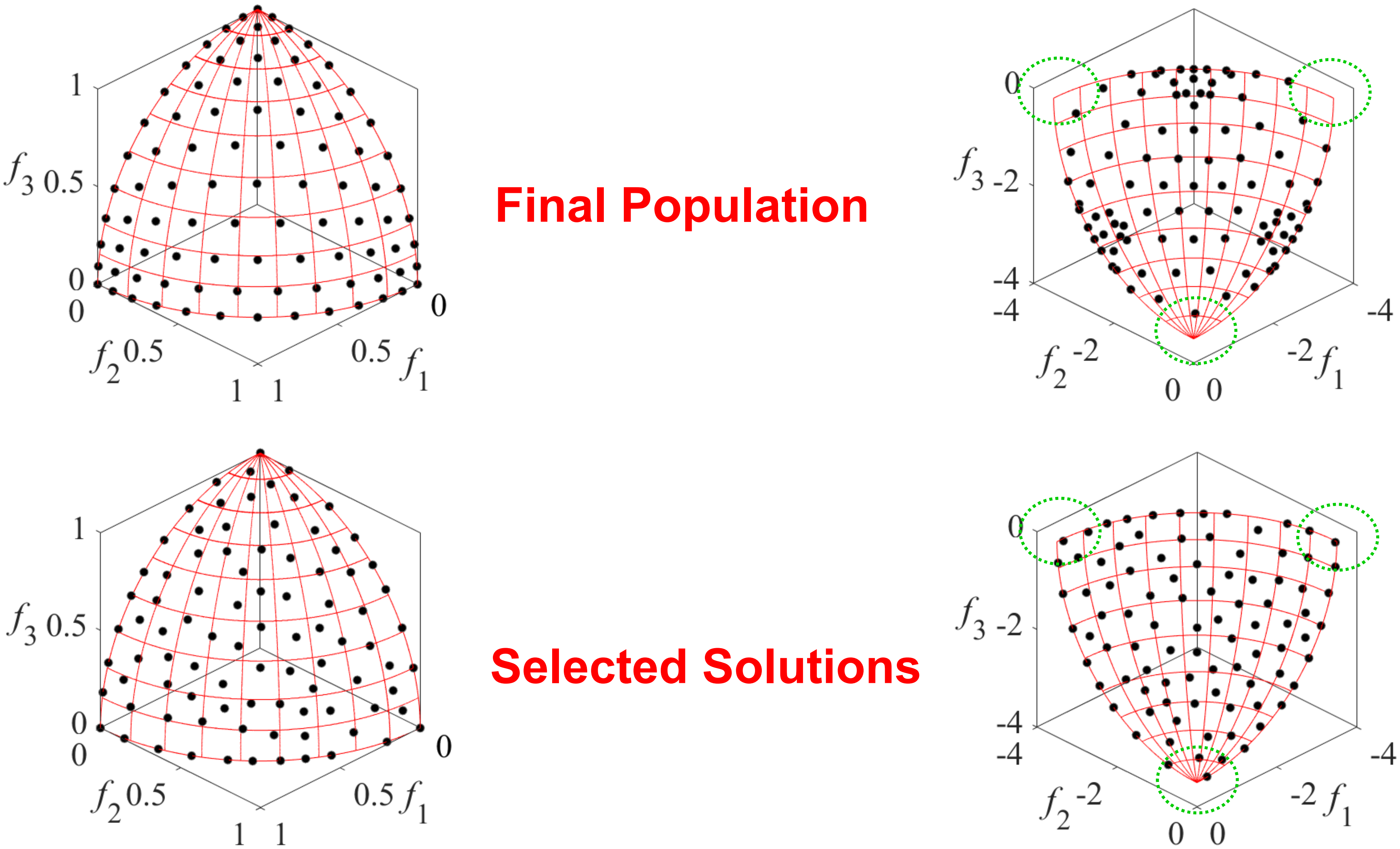
Selected Solutions



(a) NSGA-II on DTLZ2.

(b) NSGA-II on Minus-DTLZ2.

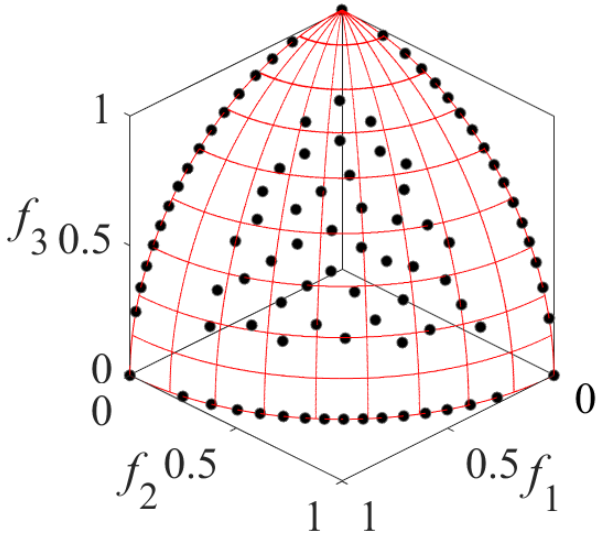
Experimental Results (MOEA/D)

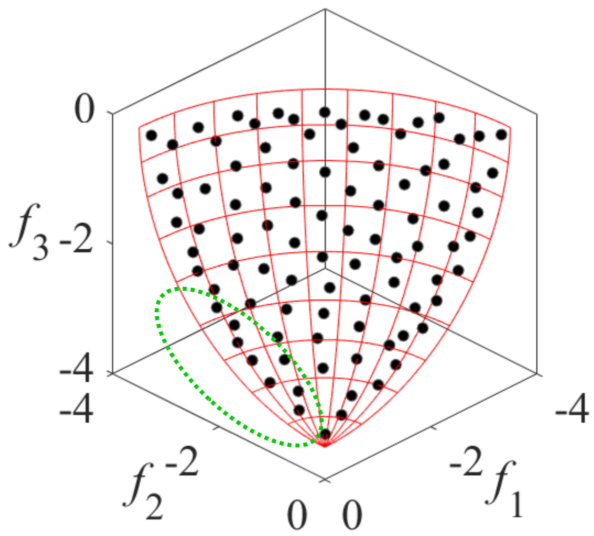


(a) MOEA/D on DTLZ2.

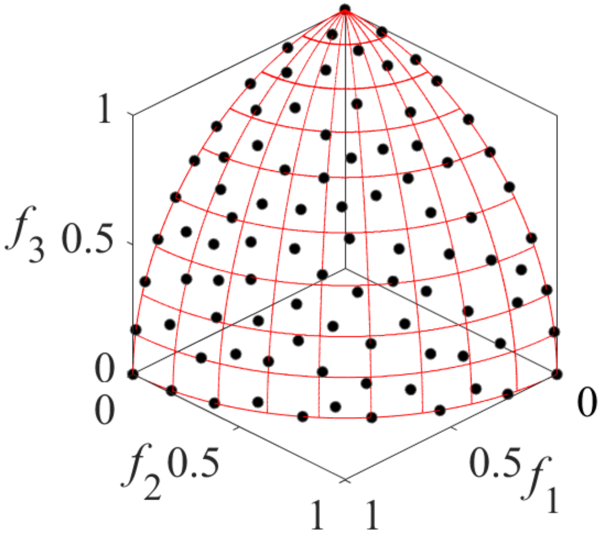
(b) MOEA/D on Minus-DTLZ2.

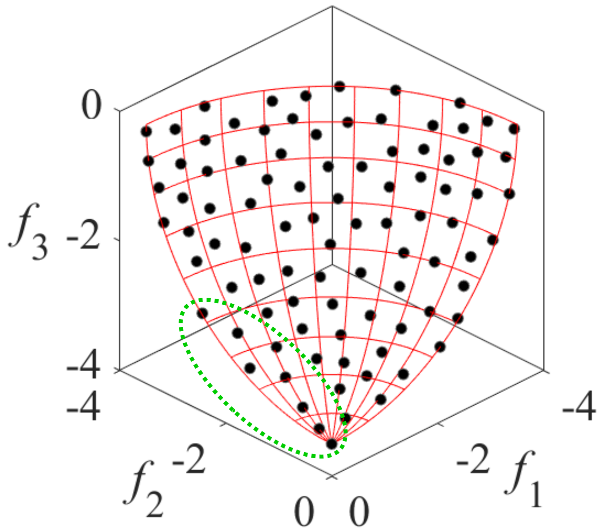
Experimental Results (SMS-EMOA)





Final Population





Selected Solutions

(a) SMS-EMOA on DTLZ2.

(b) SMS-EMOA on Minus-DTLZ2.