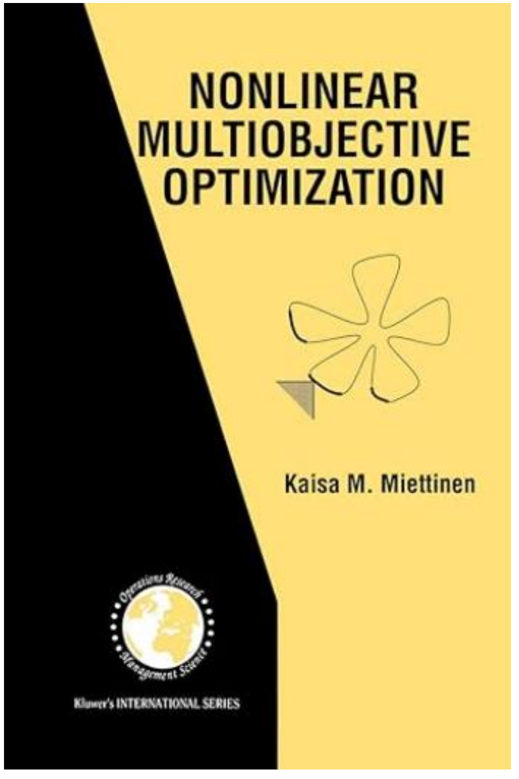
3.1. Weighting Method

$$
\begin{aligned}
&\text{minimize} \quad && \sum_{i=1}^{k} w_i f_i(\mathbf{x}) \\
&\text{subject to} \quad && \mathbf{x} \in S,
\end{aligned}
\tag{3.1.1}
$$

where $w_i \geq 0$ for all $i = 1, \ldots, k$ and $\sum_{i=1}^{k} w_i = 1$.



3.2. ε-Constraint Method

In the ε-constraint method, introduced in Haimes et al. (1971), one of the objective functions is selected to be optimized and all the other objective functions are converted into constraints by setting an upper bound to each of them. The problem to be solved is now of the form

$$\begin{aligned} &\text{minimize} && f_\ell(\mathbf{x}) \\ &\text{subject to} && f_j(\mathbf{x}) \le \varepsilon_j \quad \text{for all} \quad j = 1, \dots, k, \quad j \neq \ell, \\ &&& \mathbf{x} \in S, \end{aligned} \tag{3.2.1}$$

where $\ell \in \{1, \dots, k\}$. Problem (3.2.1) is called an *ε-constraint problem*.

3.3. Hybrid Method

At this point it is worth mentioning a method combining the weighting method and the ε-constraint method. This method is described in Corley (1980) and Wendell and Lee (1977) in slightly different forms. The name hybrid method is introduced in Chankong and Haimes (1983a, b).

The *hybrid problem* to be solved is

$$
\begin{aligned}
&\text{minimize} \quad && \sum_{i=1}^{k} w_i f_i(\mathbf{x}) \\
&\text{subject to} \quad && f_j(\mathbf{x}) \le \varepsilon_j \ \text{ for all } \ j = 1, \ldots, k, \\
& && \mathbf{x} \in S,
\end{aligned}
\tag{3.3.1}
$$

where $w_i > 0$ for all $i = 1, \ldots, k$.

3.4. Method of Weighted Metrics

Weighted L_p and Weighted Tchebycheff

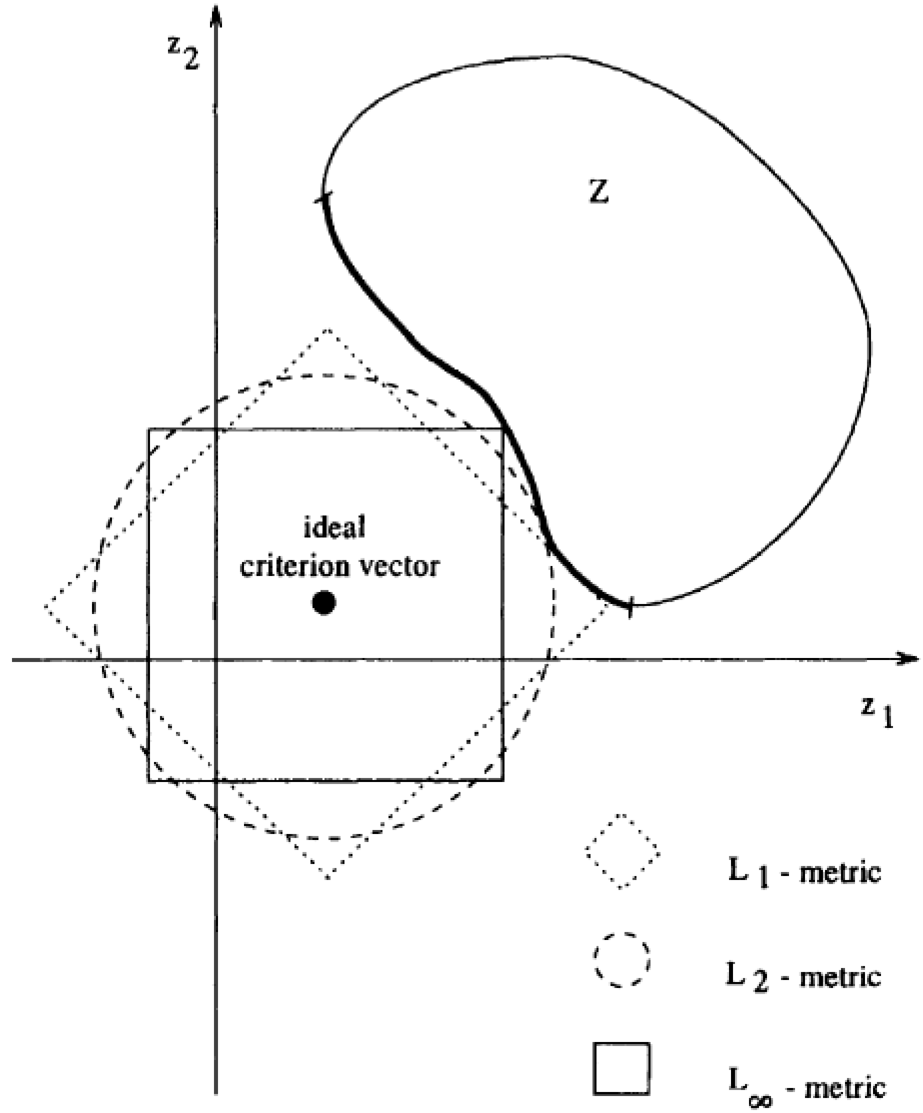
We assume that $w_i \geq 0$ for all $i = 1, \ldots, k$ and $\sum_{i=1}^{k} w_i = 1$. We obtain different solutions by altering the weighting coefficients w_i in the weighted L_p- and Tchebycheff metrics. The *weighted L_p-problem* for minimizing distances is now of the form

$$\text{(3.4.1)} \qquad \begin{aligned} &\text{minimize} \quad \left(\sum_{i=1}^{k} w_i |f_i(\mathbf{x}) - z_i^\star|^p \right)^{1/p} \\ &\text{subject to} \quad \mathbf{x} \in S \end{aligned}$$

for $1 \leq p < \infty$. The *weighted Tchebycheff problem* is of the form

$$\text{(3.4.2)} \qquad \begin{aligned} &\text{minimize} \quad \max_{i=1,\ldots,k} \left[w_i |f_i(\mathbf{x}) - z_i^\star| \right] \\ &\text{subject to} \quad \mathbf{x} \in S. \end{aligned}$$

$$\boxed{\left(\sum_{i=1}^{k}\left|f_i(\mathbf{x})-z_i^\star\right|^p\right)^{1/p}}$$



It is suggested in Steuer (1986) and Steuer and Choo (1983) that the weighted Tchebycheff problem be varied by an augmentation term. In this case, the distance between the utopian objective vector and the feasible objective region is measured by an *augmented weighted Tchebycheff metric*. The *augmented weighted Tchebycheff problem* is of the form

$$
\begin{aligned}
(3.4.5) \qquad & \text{minimize} \quad \max_{i=1,\dots,k} \left[w_i |f_i(\mathbf{x}) - z_i^{\star\star}| \right] + \rho \sum_{i=1}^{k} |f_i(\mathbf{x}) - z_i^{\star\star}| \\
& \text{subject to} \quad \mathbf{x} \in S,
\end{aligned}
$$

where ρ is a sufficiently small positive scalar.

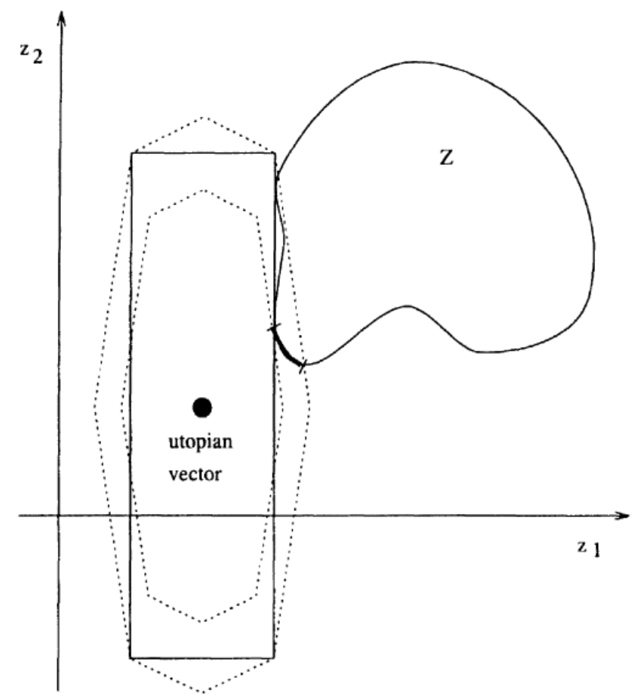


Figure 3.4.2. Augmented weighted Tchebycheff problem.

3.5. Achievement Scalarizing Function Approach

For example, weakly Pareto optimal solutions can be generated with any reference point $\bar{\mathbf{z}} \in \mathbf{R}^k$ by solving the problem

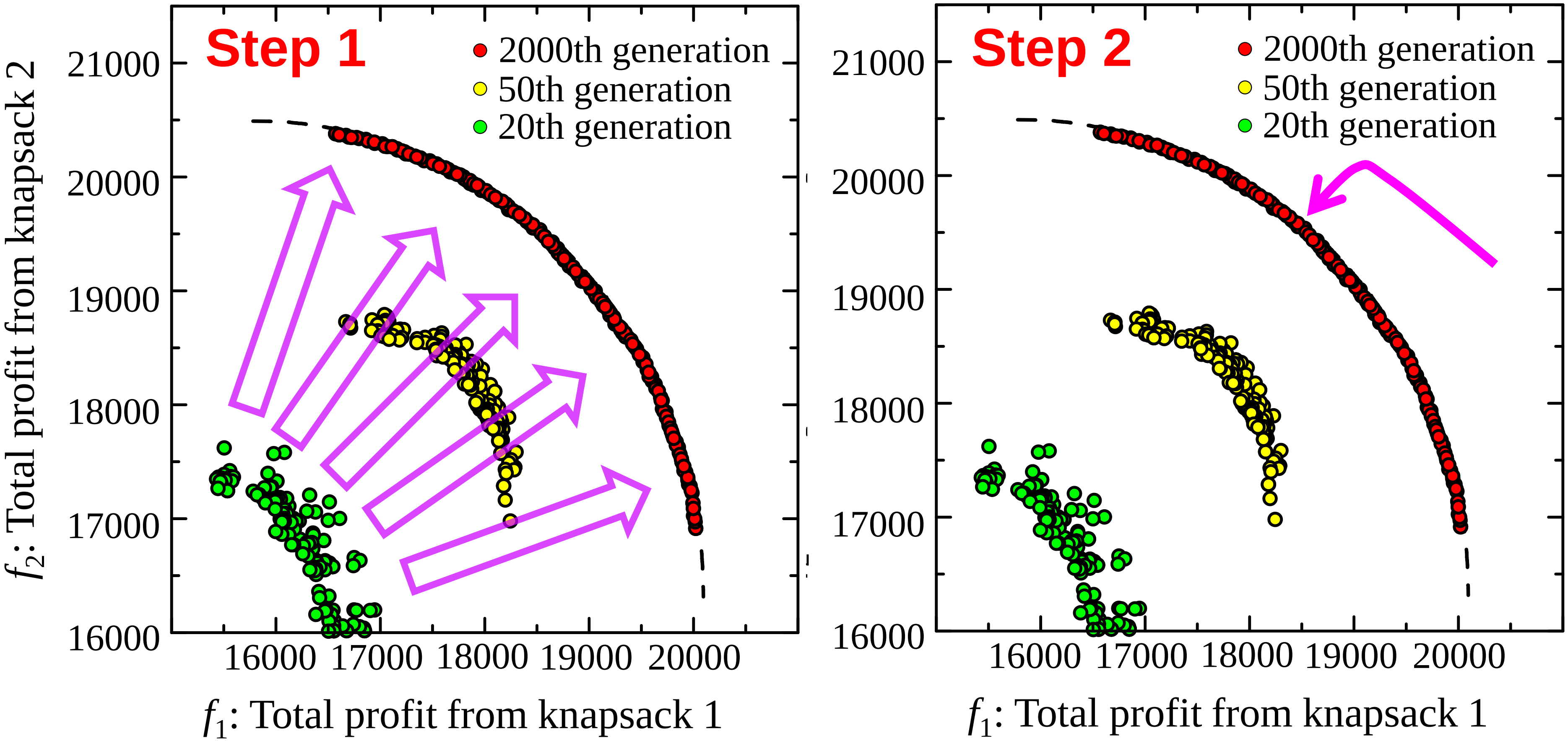
$$
\begin{aligned}
&\text{minimize} \quad \max_{i=1,\dots,k} \left[w_i(f_i(\mathbf{x}) - \bar{z}_i) \right] \\
&\text{subject to} \quad \mathbf{x} \in S.
\end{aligned} \tag{3.5.1}
$$

It differs from weighted Tchebycheff problem (3.4.2) only in that the absolute value signs are missing. This change ensures that weakly Pareto optimal solutions are produced independently of the feasibility or infeasibility of the reference point.

Optimization Methods

1. Introduction.
2. Greedy algorithms for combinatorial optimization.
3. LS and neighborhood structures for combinatorial optimization.
4. Variable neighborhood search, neighborhood descent, SA, TS, EC.
5. Branch and bound algorithms, and subset selection algorithms.
6. Linear programming problem formulations and applications.
7. Linear programming algorithms.
8. Integer linear programming algorithms.
9. Unconstrained nonlinear optimization and gradient descent.
10. Newton's methods and Levenberg-Marquardt modification.
11. Quasi-Newton methods and conjugate direction methods.
12. Nonlinear optimization with equality constraints.
13. Nonlinear optimization with inequality constraints.
14. Problem formulation and concepts in multi-objective optimization.
15. Search for single final solution in multi-objective optimization.
16: Search for multiple solutions in multi-objective optimization.

Basic Idea of Decision Making in EMO
(Evolutionary Multiobjective Optimization)

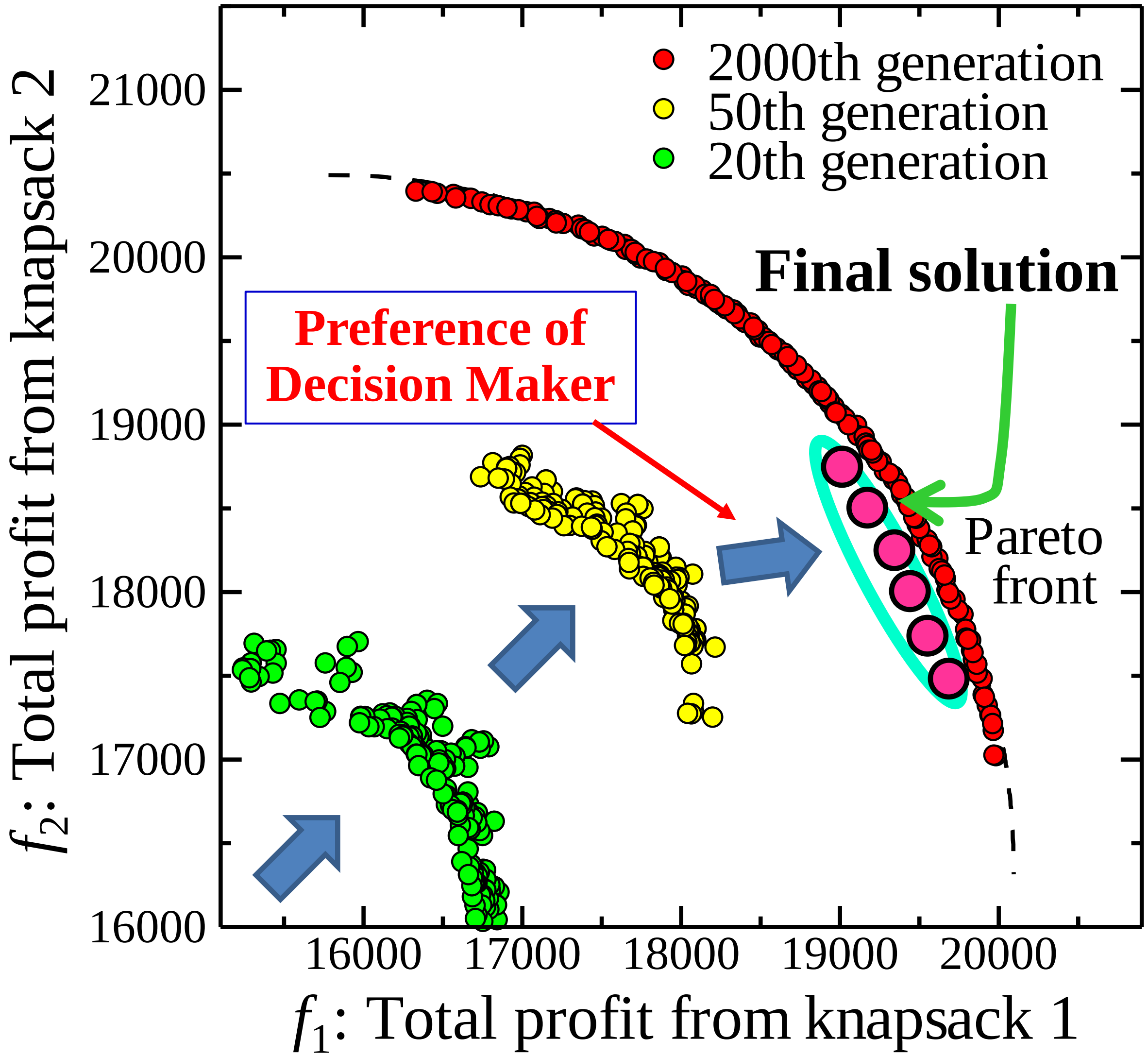


Step 1: Search for non-dominated solutions along the Pareto front.

Step 2: Selection of a single solution from the obtained solutions by the decision maker.

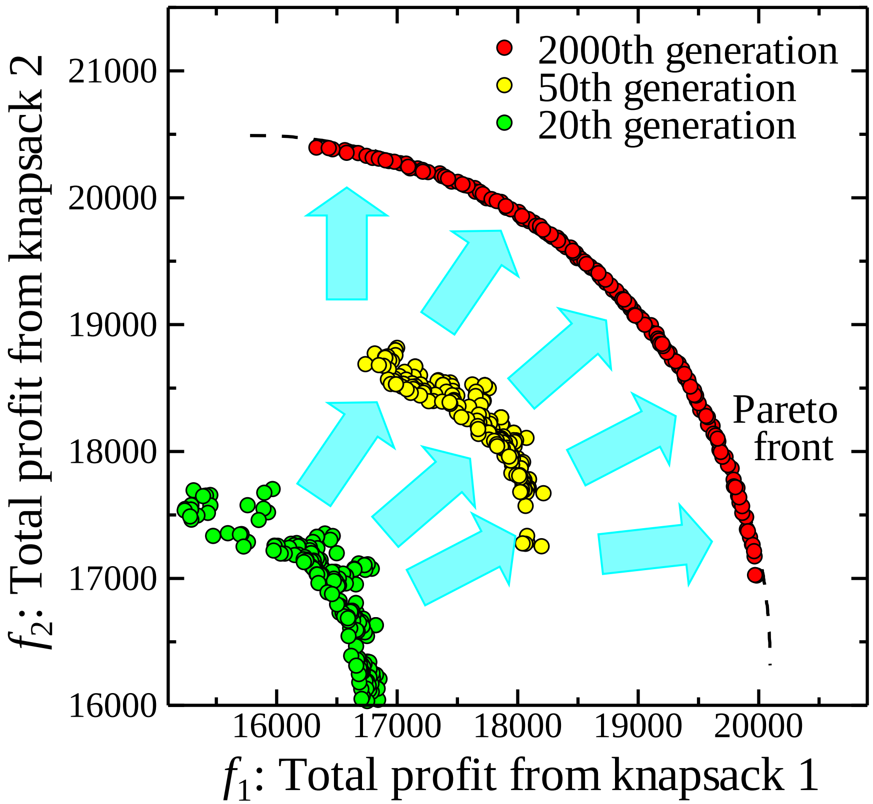
Combination of Two Approach. Interactive EMO Algorithms

If the decision maker’s preference is available in the middle of evolution, it is a good idea to focus on the preferred region.



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

Important Issue in EMO-Approach: How to search for a well-distributed solution set along the Pareto front.



Desired search behavior of EMO algorithms

Many EMO algorithms and test problems are available through the Internet:

jMetal (for Java users)

J.J. Durillo, and A. J. Nebro, “jMetal: A Java framework for multi-objective optimization,” Advances in Engineering Software (2011).

PlatEMO (for MATLAB users)

Y. Tian, R. Cheng, X. Zhang, and Y. Jin, “PlatEMO: A MATLAB platform for evolutionary multi-objective optimization,” *IEEE Computational Intelligence Magazine* (2017)

Pymoo (for Python users)

J.Blank and K. Deb, “Pymoo: Multi-objective optimization in Python,” *IEEE Access* (2020).

How to evaluate the obtained solution set ?

Performance Evaluation and Performance Indicators

E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. G. da Fonseca, "Performance assessment of multiobjective optimizers: An analysis and review," *IEEE Trans. on Evolutionary Computation* 7 (2), 117-132, 2003.

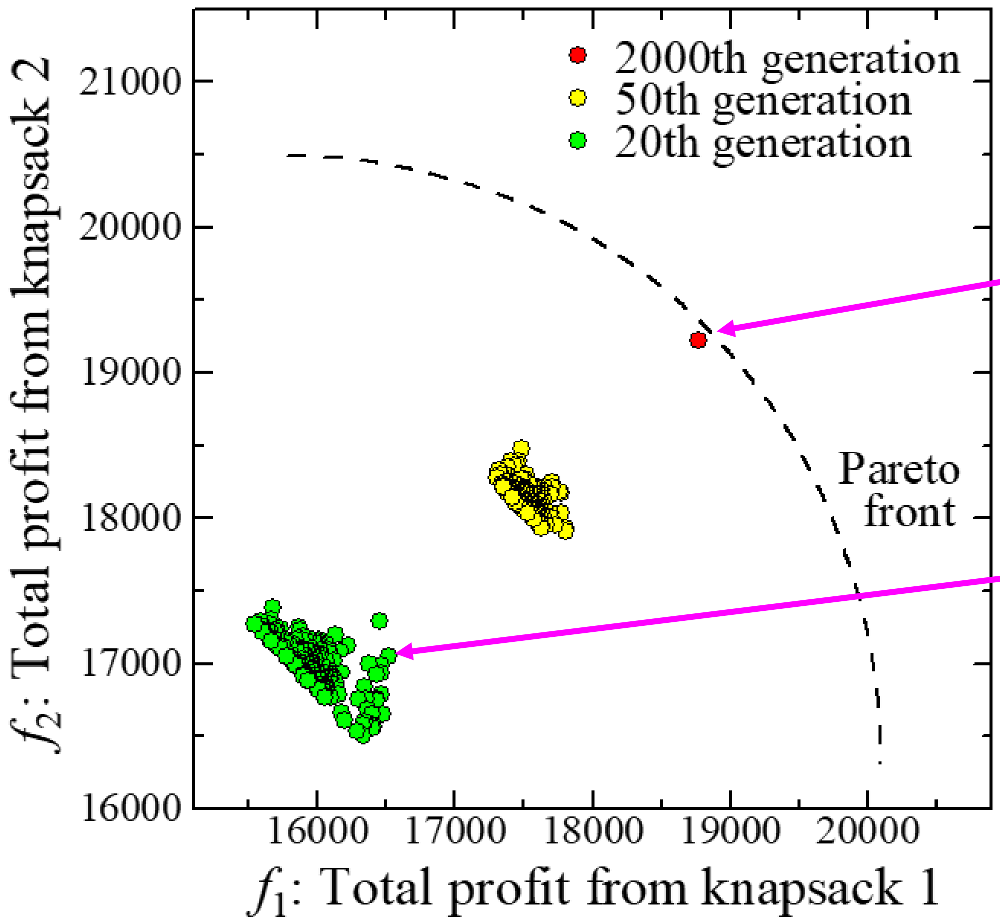
M. Li and X. Yao, Quality evaluation of solution sets in multiobjective optimisation: A survey, ACM Computing Surveys, 52(2), 2019.

H. Ishibuchi, L. M. Pang, and K. Shang, Difficulties in fair performance comparison of multi-objective evolutionary algorithms, IEEE Computational Intelligence Magazine, 17 (1), 86-101, 2022.

In standard single-objective optimization

The performance of an algorithm
= The performance of the finally obtained solution

Algorithm performance = Solution performance



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

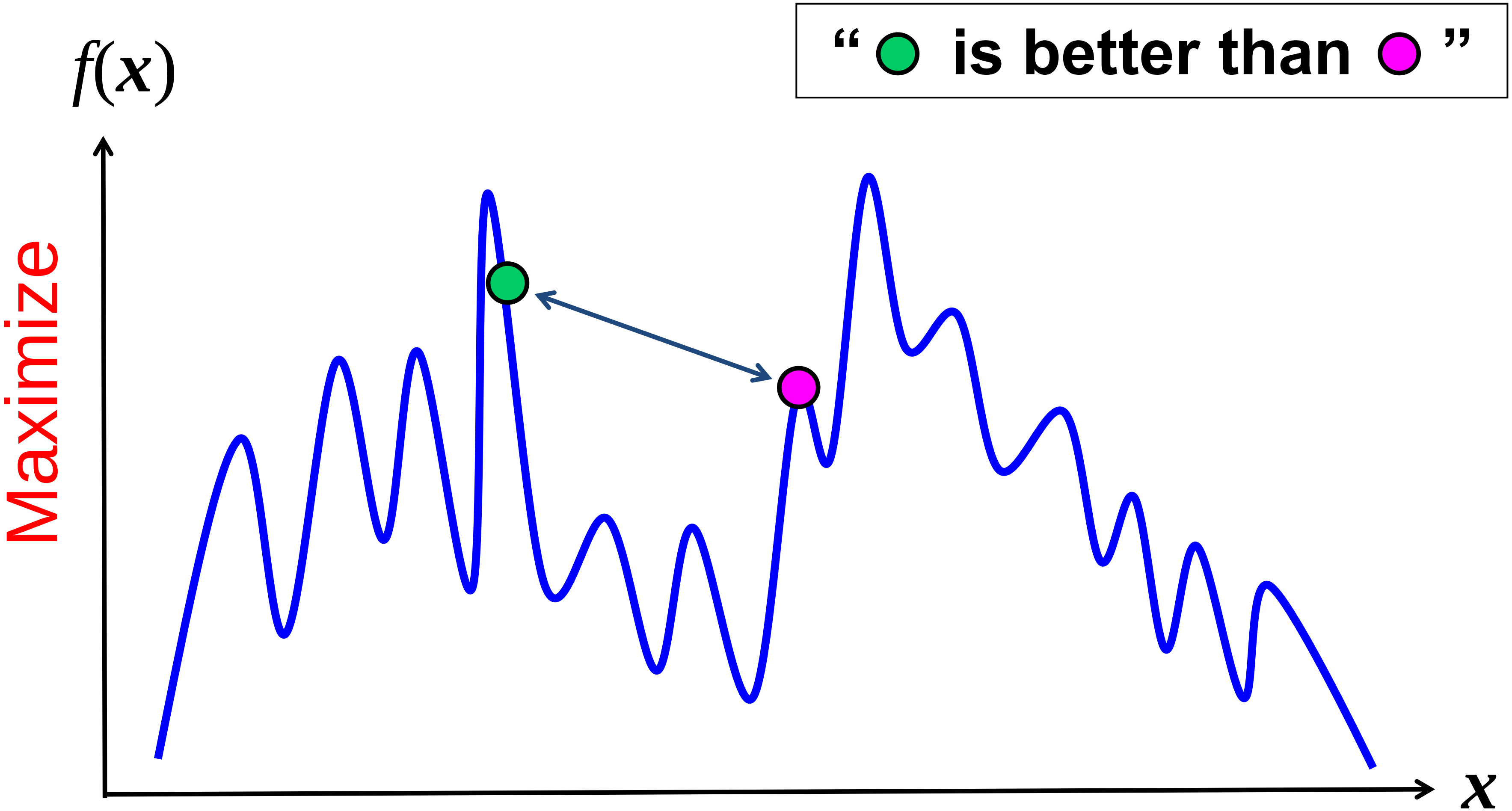
Algorithm performance =
This solution's performance

Each solution can be evaluated in the same (or similar) manner (i.e., based on the value of $f_1(\boldsymbol{x}) + f_2(\boldsymbol{x})$).

In standard single-objective optimization

The final result of optimization is a single solution.

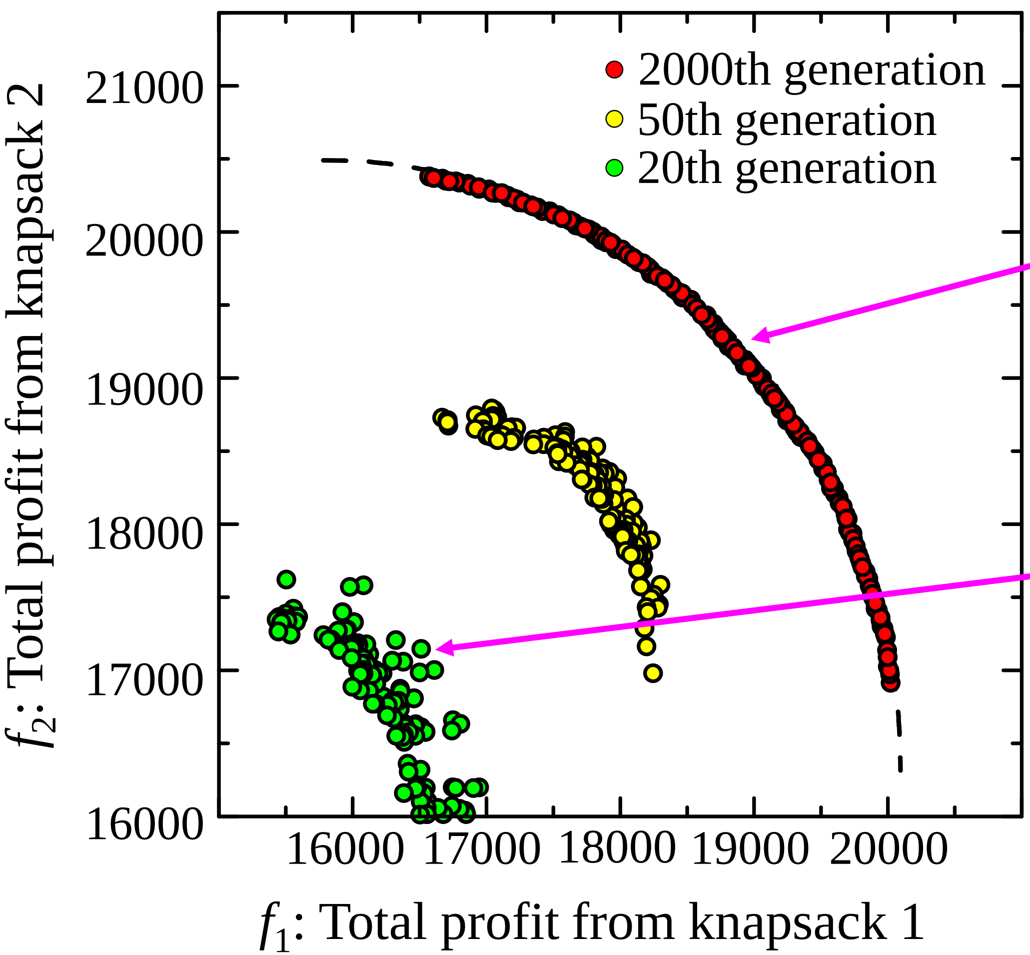
Comparison of solutions is easy.



In evolutionary multi-objective optimization (EMO)

The performance of an algorithm
= The performance of the finally obtained solution set

Algorithm performance = Population performance



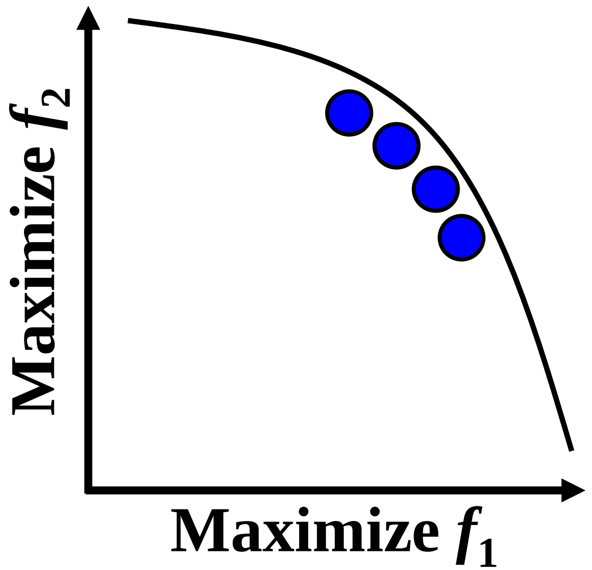
Algorithm performance =
This population's performance

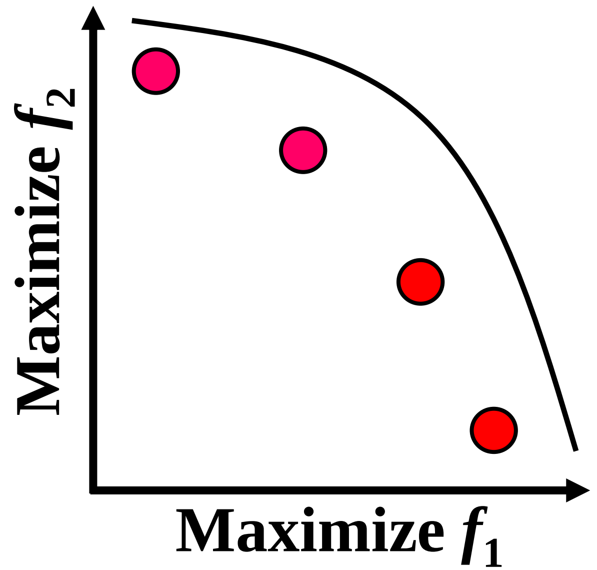
Each solution is evaluated in a totally different manner.

In evolutionary multi-objective optimization (EMO)

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

Which is a better solution set?





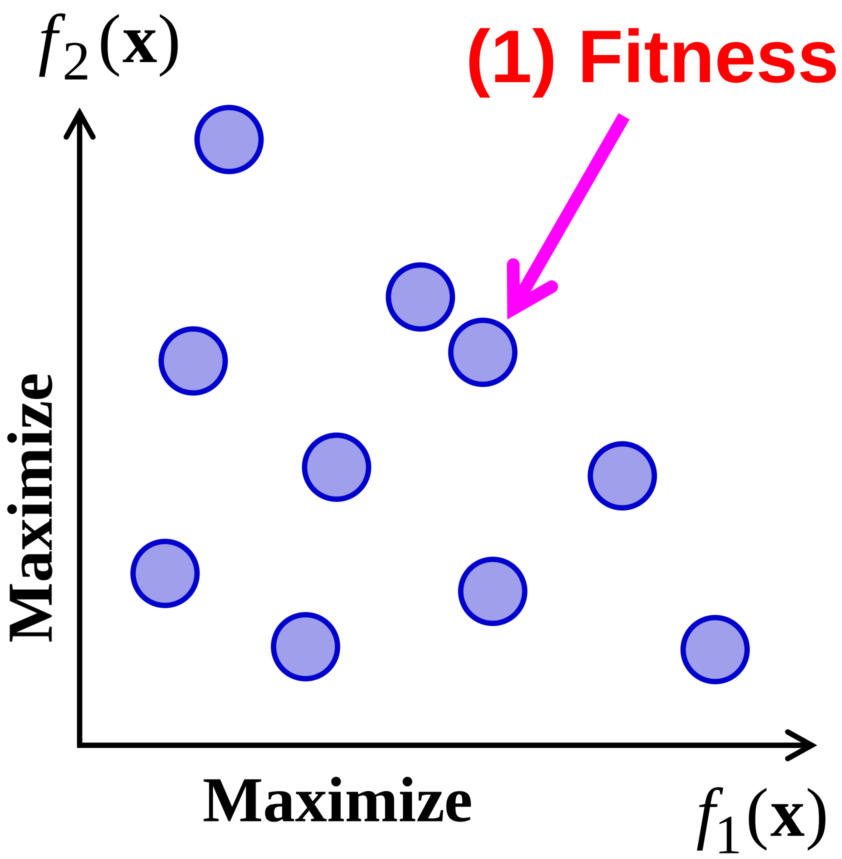
Two Levels of Evaluation

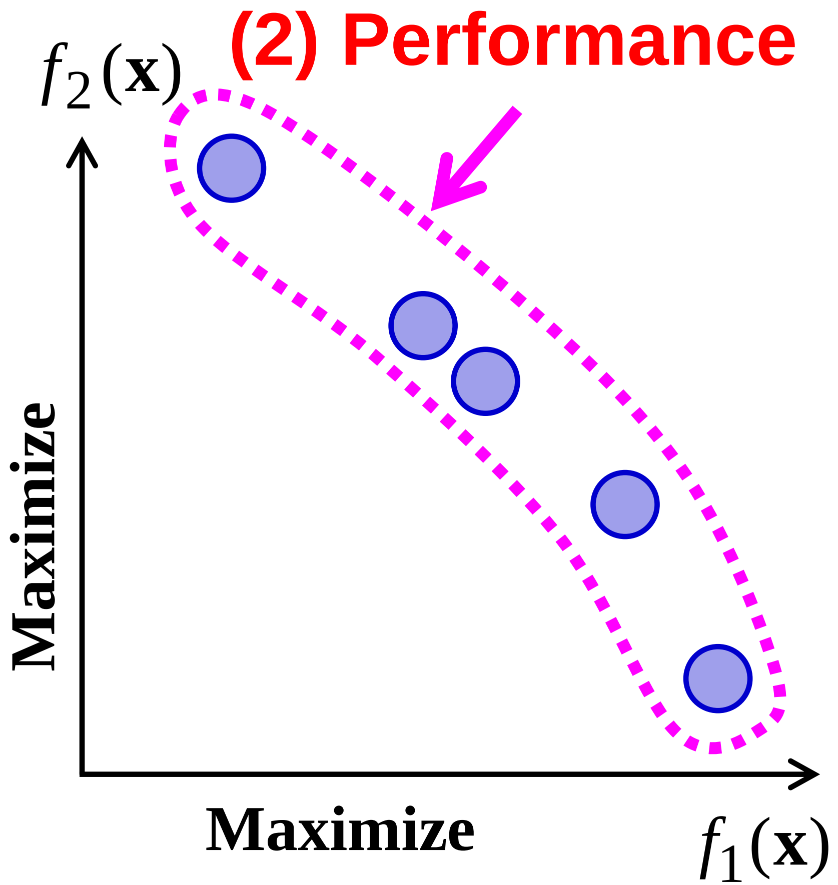
(1) Solution level:

for solution evaluation within an algorithm

(2) Solution set level:

for algorithm evaluation (solution set evaluation)

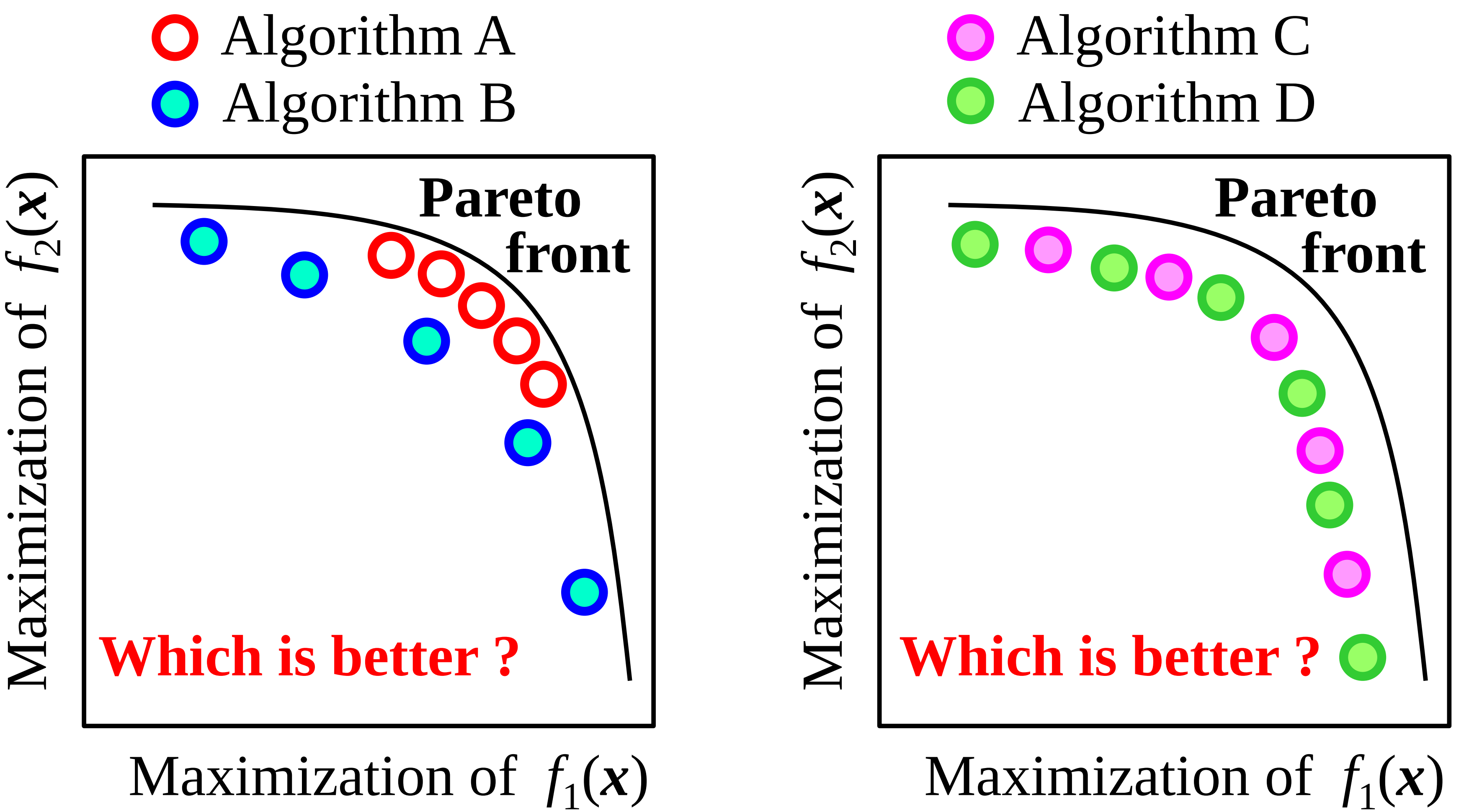




In evolutionary multi-objective optimization (EMO)

Algorithm Evaluation (Solution Set Level):

Performance comparison is not very clear.

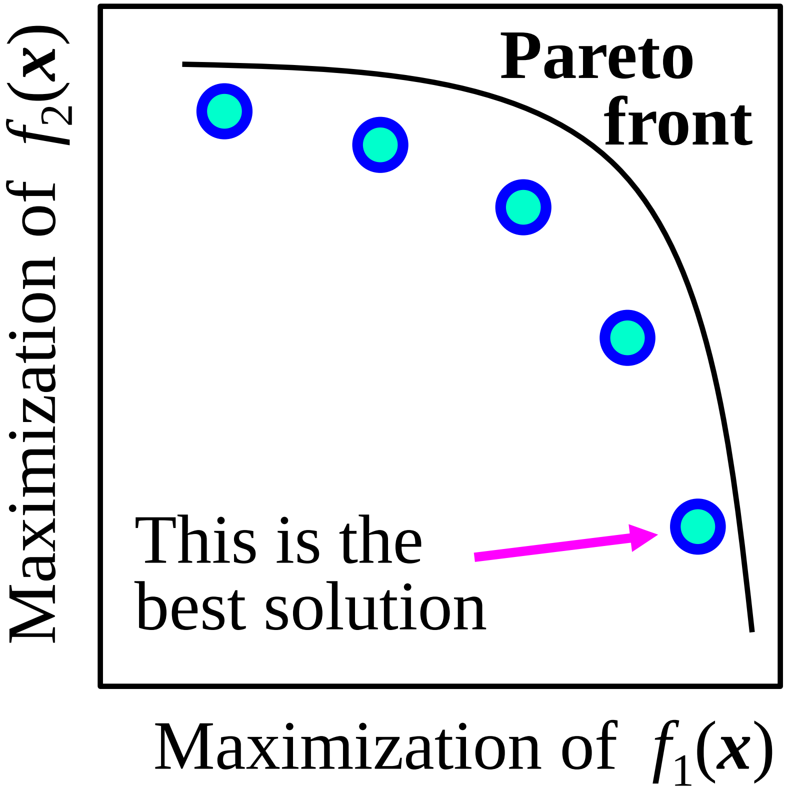


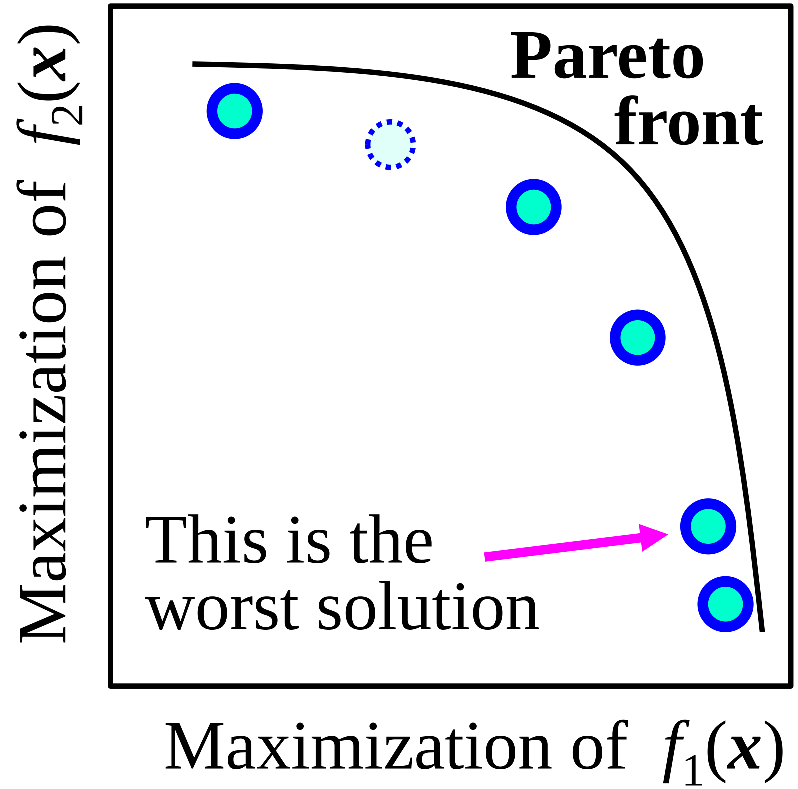
In evolutionary multi-objective optimization (EMO)

Evaluation of each solution:

based on the relation with the other solutions.

==> The same solution can have different evaluation results.





Similar Case: Genetics-based machine learning

Classifier
If ***x*** is ... then class 1
If ***x*** is ... then class 1
If ***x*** is ... then class 2
. . .   . . .
If ***x*** is ... then class 3

Michigan Approach (Similar to EMO):

Solution (Individual) = A single rule
Fitness = Performance of a single rule
Population = A set of rules = A single classifier
Algorithm evaluation = Performance of the final population
(Performance of a set of individuals).

Pittsburgh Approach (Similar to Standard Optimization):

Solution (Individual) = A set of rules = A single classifier
Fitness = Performance of a classifier
Population = A set of classifiers
Algorithm evaluation = Performance of the best classifier
(Performance of the best individual)

Population

Classifier
If ***x*** is ... then class 1
If ***x*** is ... then class 1
If ***x*** is ... then class 2
. . .   . . .
If ***x*** is ... then class 3

Classifier
If ***x*** is ... then class 1
If ***x*** is ... then class 1
If ***x*** is ... then class 2
. . .   . . .
If ***x*** is ... then class 3

. . .

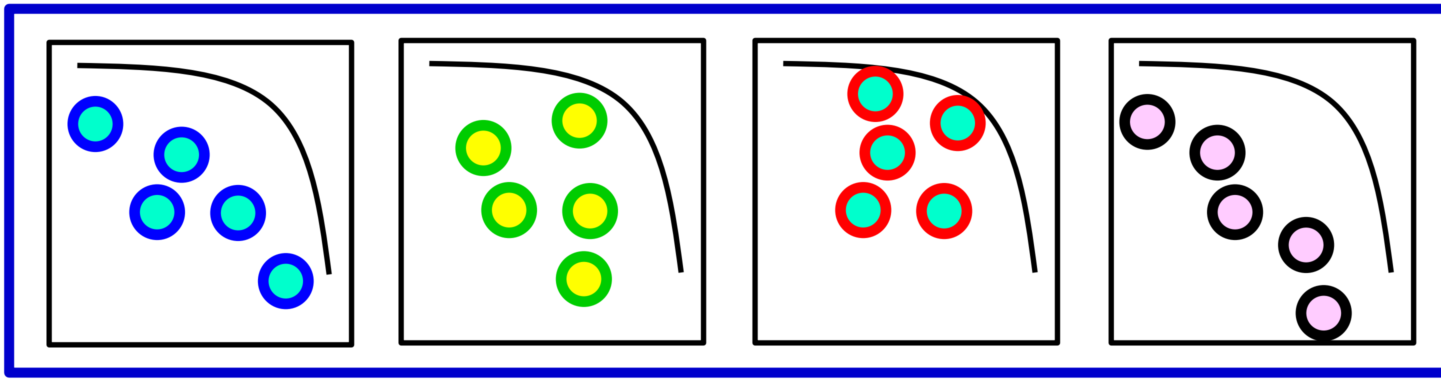
Classifier
If ***x*** is ... then class 1
If ***x*** is ... then class 1
If ***x*** is ... then class 2
. . .   . . .
If ***x*** is ... then class 3

Set-based EMO Framework (Similar to Pittsburgh Approach

Individual = A set of solutions (i.e., solution set)

Fitness evaluation = Performance of a solution set

Population



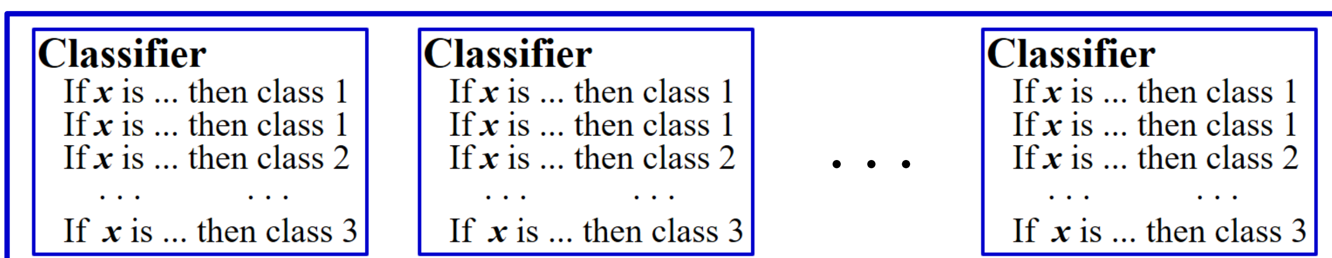
Pittsburgh Approach (Similar to Standard Optimization):

Solution (Individual) = A set of rules = A single classifier

Fitness = Performance of a classifier

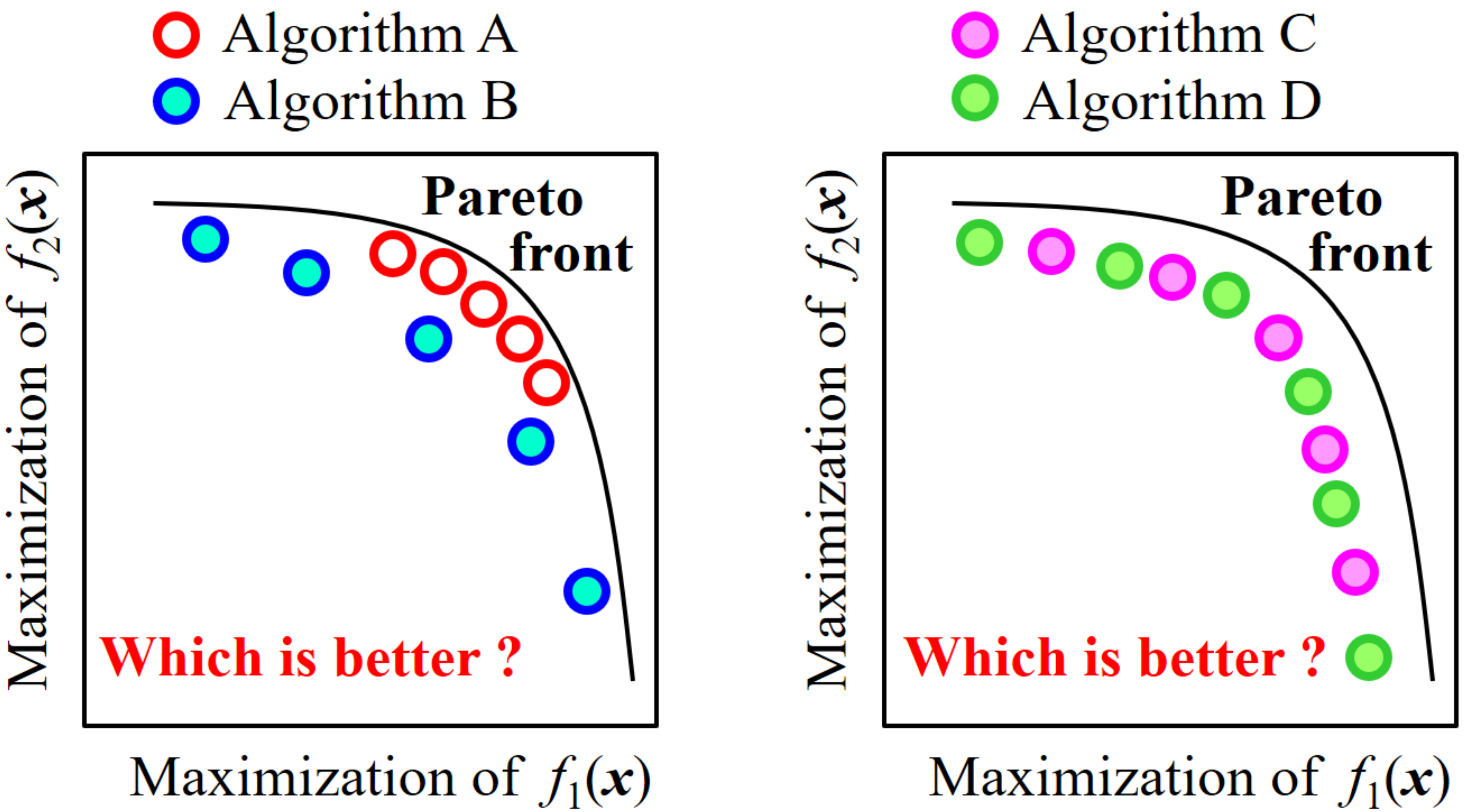
Population = A set of classifiers

Population



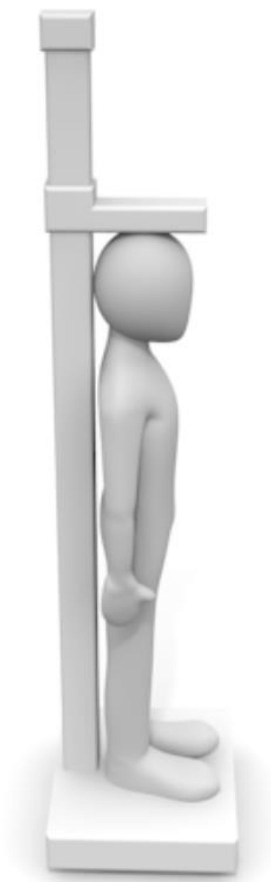
Indicators: Performance evaluation of a solution set

- **Convergence to the Pareto front.**
- **Spread of a solution set (Broadness): Diversity of solutions.**
- **Uniformity of solutions: Diversity of solutions.**
- **Overall performance.**



Indicators: Performance evaluation of a solution set

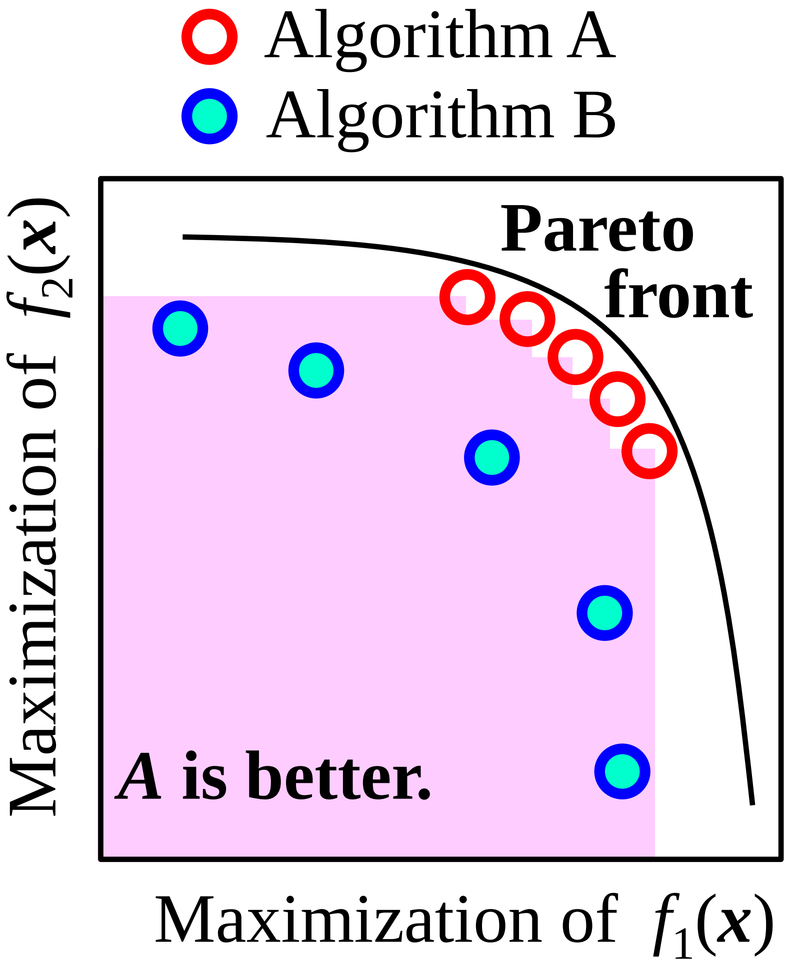
- **Convergence to the Pareto front.**
- **Spread of a solution set (Broadness): Diversity of solutions.**
- **Uniformity of solutions: Diversity of solutions.**
- **Overall performance.**

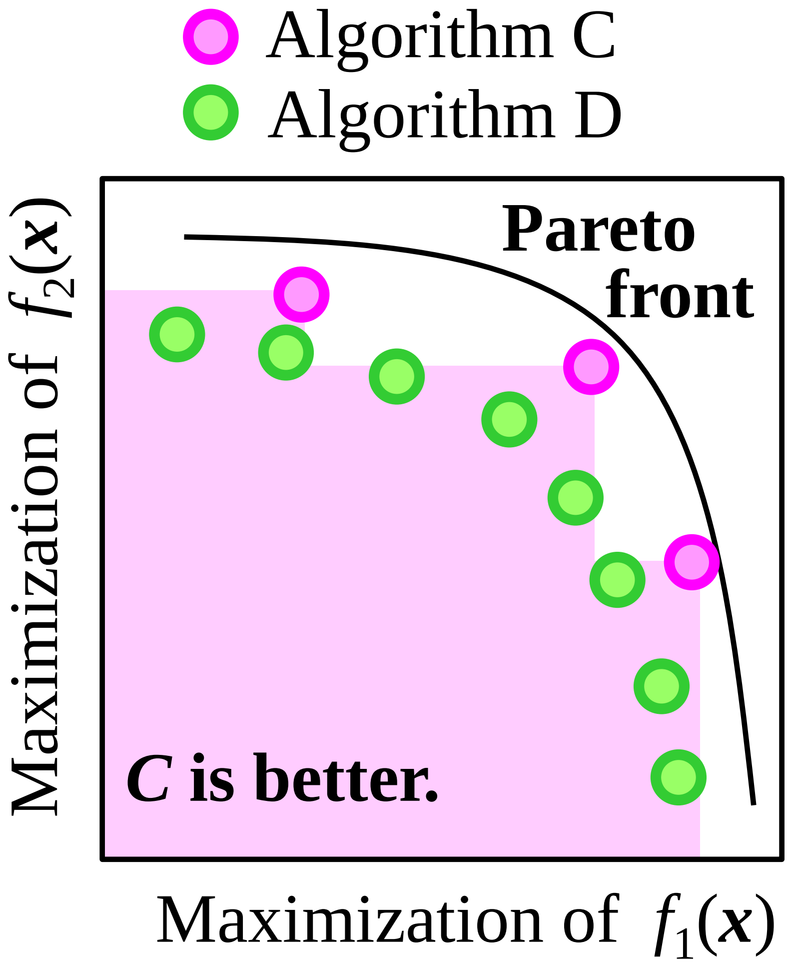


One Important Feature of Indicators

Pareto Compliance

If a solution set *A* is better than another solution set *B*,
A is always evaluated as a better solution set than *B* by an indicator.



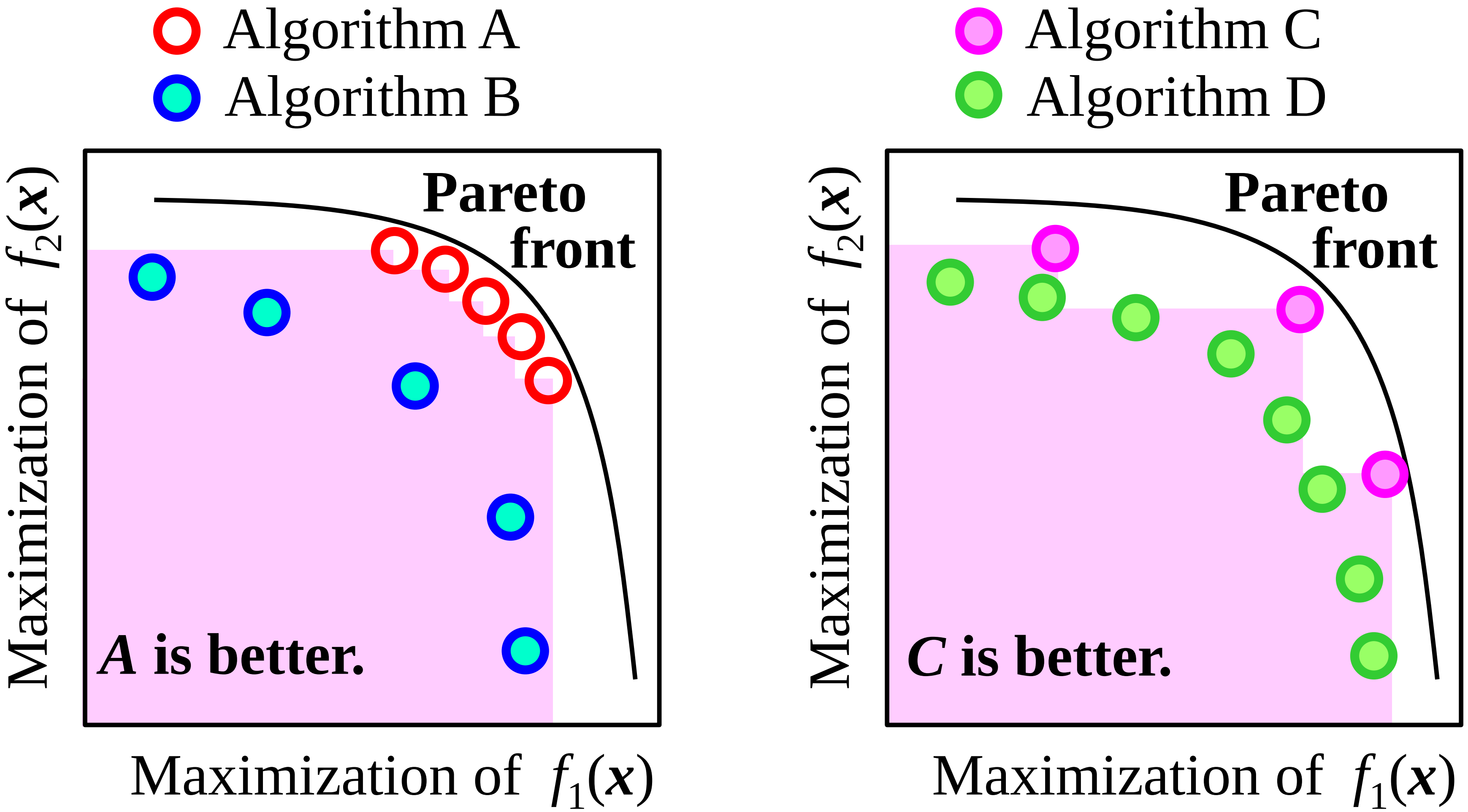


One Important Feature of Indicators

Pareto Compliance

How to define?

If **a solution set *A* is better than another solution set *B***,
A is always evaluated as a better solution set than *B* by an indicator.



Dominance Relation between Solutions *x* and *y*

Maximize $\boldsymbol{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}),\ f_2(\boldsymbol{x}),\ \ldots,\ f_m(\boldsymbol{x}))$

Pareto Dominance: $\boldsymbol{x}$ is dominated by $\boldsymbol{y}$ ($\boldsymbol{y}$ is better than $\boldsymbol{x}$)

$$\boldsymbol{f}(\boldsymbol{x}) \prec (\boldsymbol{y}) \iff \forall i,\ f_i(\boldsymbol{x}) \le f_i(\boldsymbol{y}) \text{ and } \exists j,\ f_j(\boldsymbol{x}) < f_j(\boldsymbol{y}).$$

Weak Pareto Dominance: $\boldsymbol{x}$ is weakly dominated by $\boldsymbol{y}$

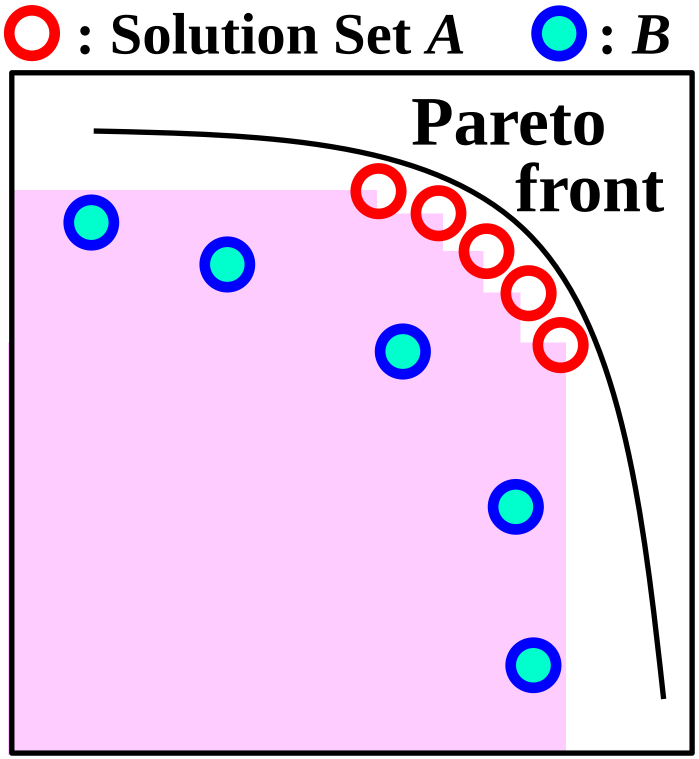
($\boldsymbol{y}$ is not worse than $\boldsymbol{x}$)

$$\boldsymbol{f}(\boldsymbol{x}) \preceq \boldsymbol{f}(\boldsymbol{y}) \iff \forall i,\ f_i(\boldsymbol{x}) \le f_i(\boldsymbol{y}).$$ This relation includes $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{y})$

Another explanation of Pareto dominance (the same meaning):

$$\boldsymbol{f}(\boldsymbol{x}) \prec \boldsymbol{f}(\boldsymbol{y}) \iff \boldsymbol{f}(\boldsymbol{x}) \preceq \boldsymbol{f}(\boldsymbol{y}) \text{ and } \boldsymbol{f}(\boldsymbol{x}) \ne \boldsymbol{f}(\boldsymbol{y}).$$

Relation between Non-Dominated Solution Sets *A* and *B*

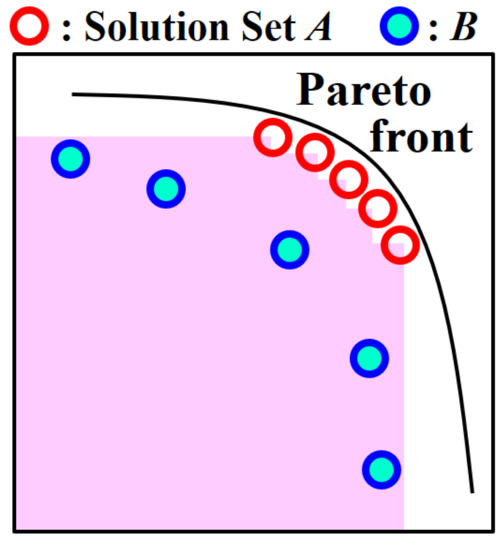


Important:
A and *B* are non-dominated solution sets.
That is, all solutions are non-dominated in each solution set (as shown in this figure).

Weak Pareto Dominance: $\boldsymbol{f}(\boldsymbol{x}) \preceq \boldsymbol{f}(\boldsymbol{y}) \Leftrightarrow \forall i,\ f_i(\boldsymbol{x}) \leq f_i(\boldsymbol{y})$.

Pareto Dominance: $\boldsymbol{f}(\boldsymbol{x}) \prec \boldsymbol{f}(\boldsymbol{y}) \Leftrightarrow \boldsymbol{f}(\boldsymbol{x}) \preceq \boldsymbol{f}(\boldsymbol{y})$ and $\boldsymbol{f}(\boldsymbol{x}) \neq \boldsymbol{f}(\boldsymbol{y})$.

Relation between Non-Dominated Solution Sets *A* and *B*



Weak Pareto Dominance (B is weakly dominated by A).

$$B \preceq A \Leftrightarrow \forall \boldsymbol{b} \in B, \exists \boldsymbol{a} \in A, \boldsymbol{b} \preceq \boldsymbol{a}.$$

This relation includes $A = B$.

Better Relation (A is better than B).

$$B \lhd A \Leftrightarrow B \preceq A \text{ and } A \neq B.$$

This relation excludes $A = B$.

Weak Pareto Dominance: $\boldsymbol{f}(\boldsymbol{x}) \preceq \boldsymbol{f}(\boldsymbol{y}) \Leftrightarrow \forall i,\ f_i(\boldsymbol{x}) \leq f_i(\boldsymbol{y})$.

Pareto Dominance: $\boldsymbol{f}(\boldsymbol{x}) \prec \boldsymbol{f}(\boldsymbol{y}) \Leftrightarrow \boldsymbol{f}(\boldsymbol{x}) \preceq \boldsymbol{f}(\boldsymbol{y})$ and $\boldsymbol{f}(\boldsymbol{x}) \neq \boldsymbol{f}(\boldsymbol{y})$.

Important Note: *A* and *B* are non-dominated solution sets

Weak Pareto Dominance (*B* is weakly dominated by *A*).

$$B \preceq A \Leftrightarrow \forall \boldsymbol{b} \in B, \exists \boldsymbol{a} \in A, \boldsymbol{b} \preceq \boldsymbol{a}.$$

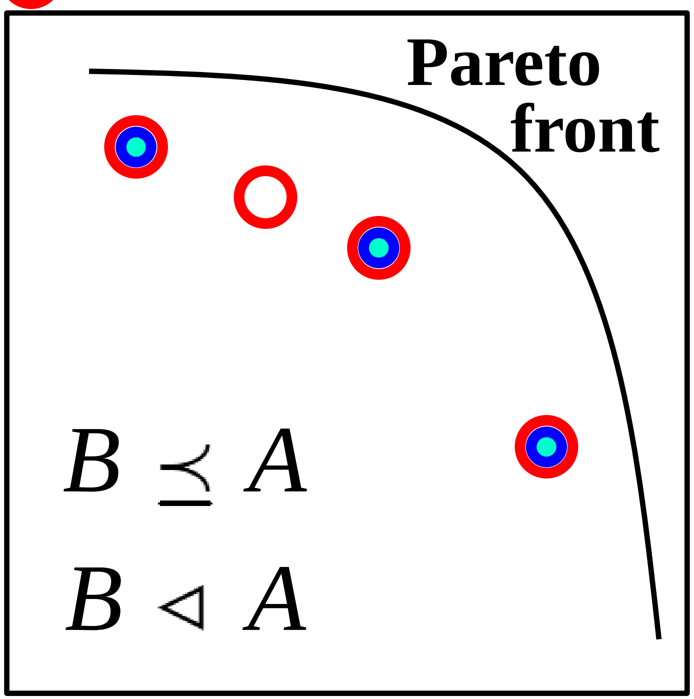
This relation includes $A = B$.

Better Relation (*A* is better than *B*).

$$B \lhd A \Leftrightarrow B \preceq A \ \ and \ \ A \neq B.$$

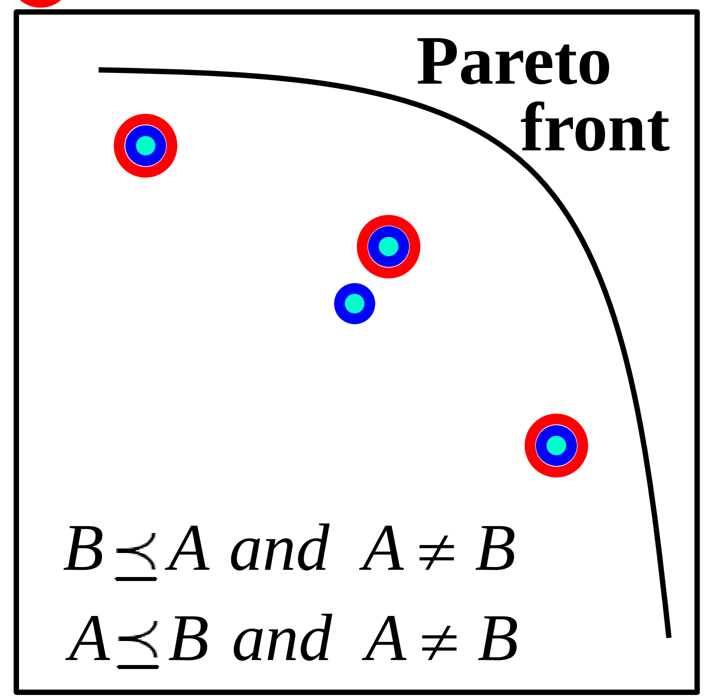
B is not a non-dominated solution set.

○: Solution Set *A* ●: *B*



A is better than *B*.

○: Solution Set *A* ●: *B*



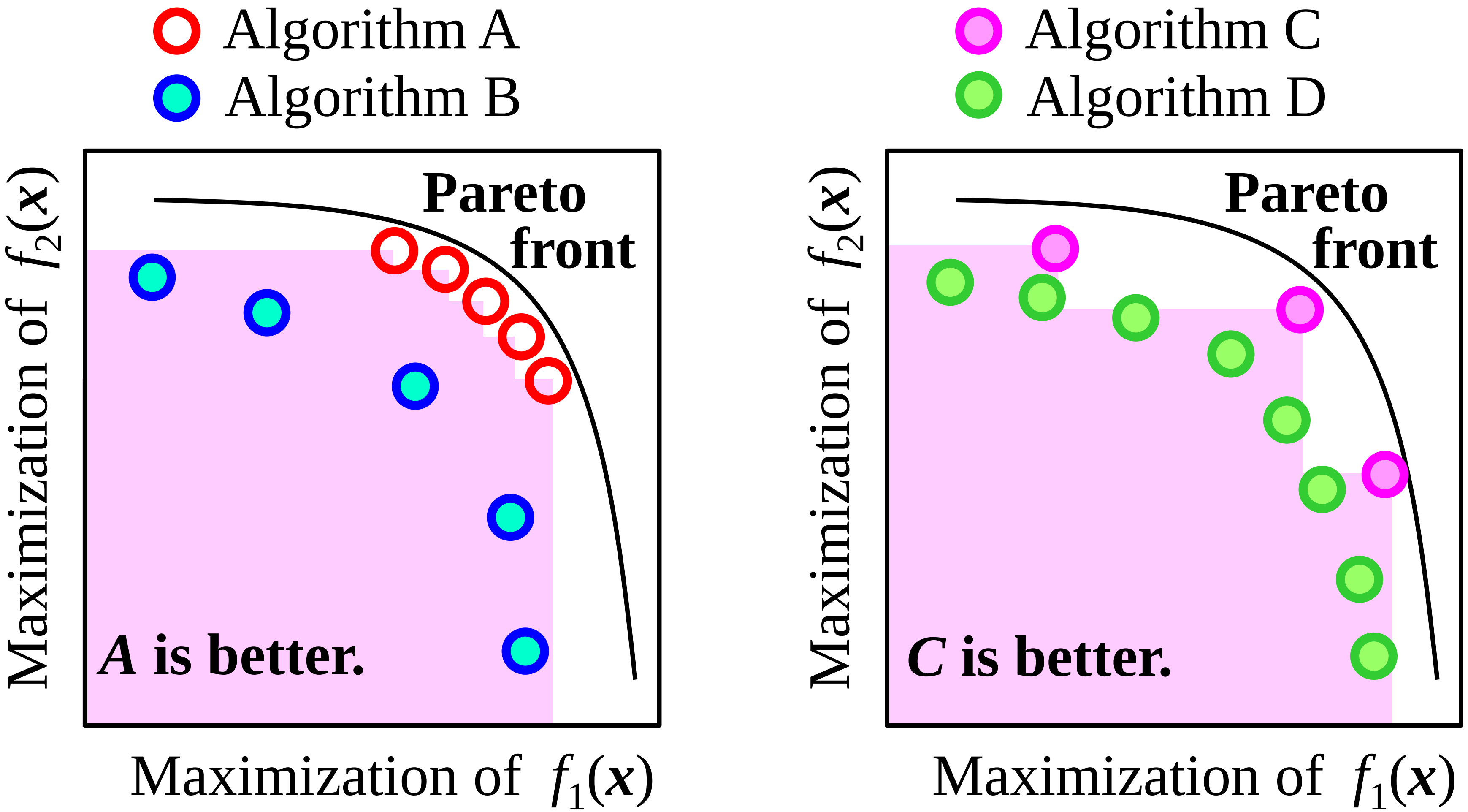
A is not better than *B*.

One Important Feature of Indicators

Pareto Compliance (Pareto Compliant Indicator)

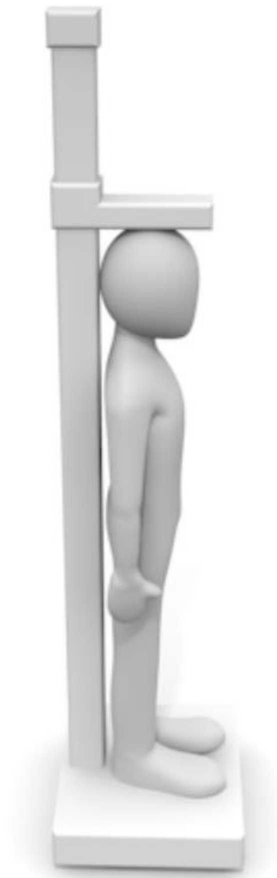
If a solution set *A* is better than another solution set *B*,

A is always evaluated as a better solution set than *B* by an indicator.



Indicators: Performance evaluation of a solution set

- Convergence to the Pareto front.
- Spread of a solution set (Broadness): Diversity of solutions.
- Uniformity of solutions: Diversity of solutions.
- Overall performance.

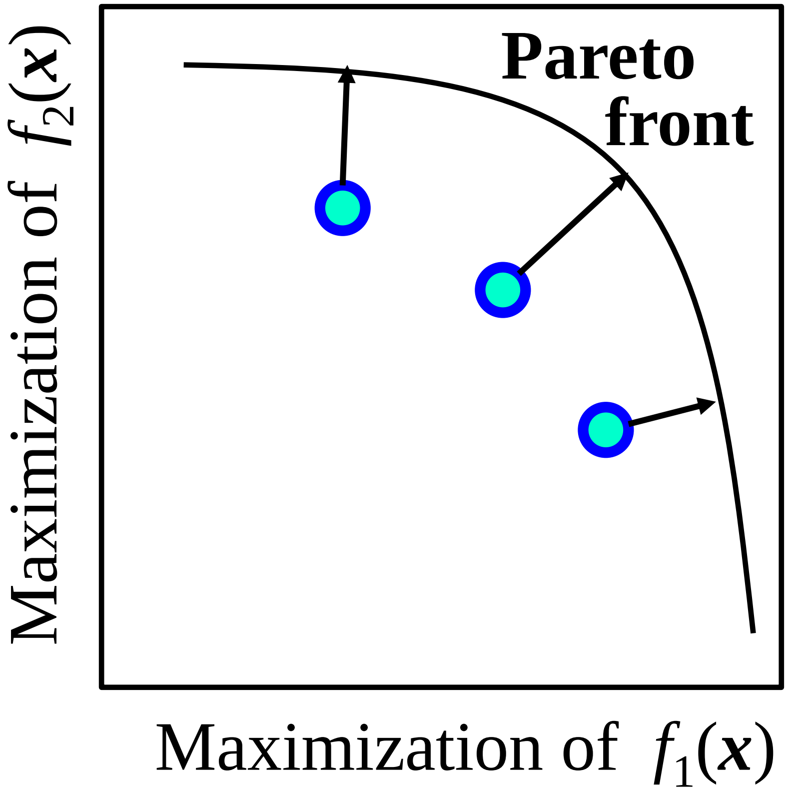



Performance Indicator for Convergence

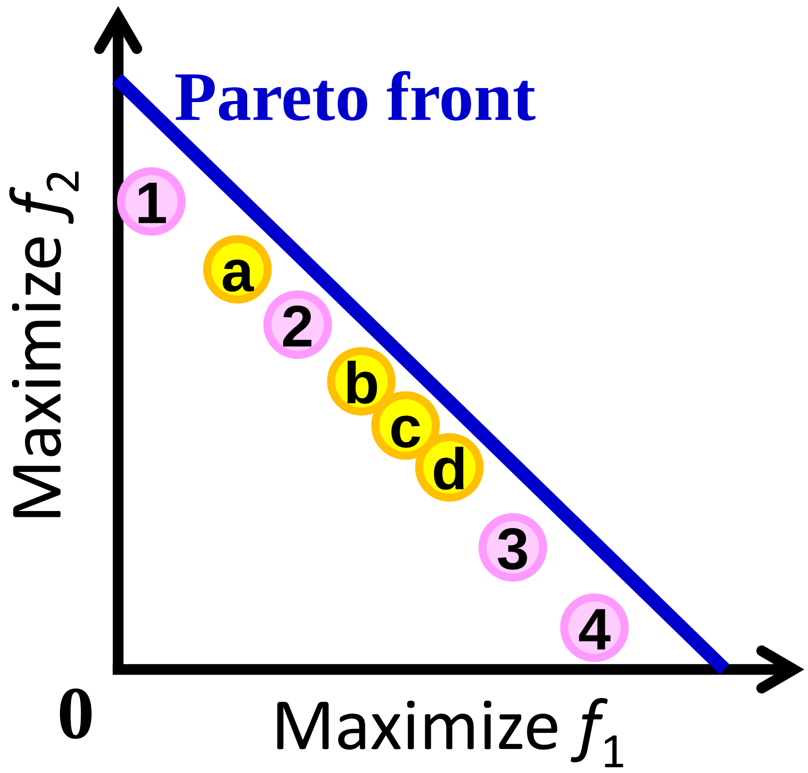
GD (Generational Distance): The smaller, the better

Basic Idea:

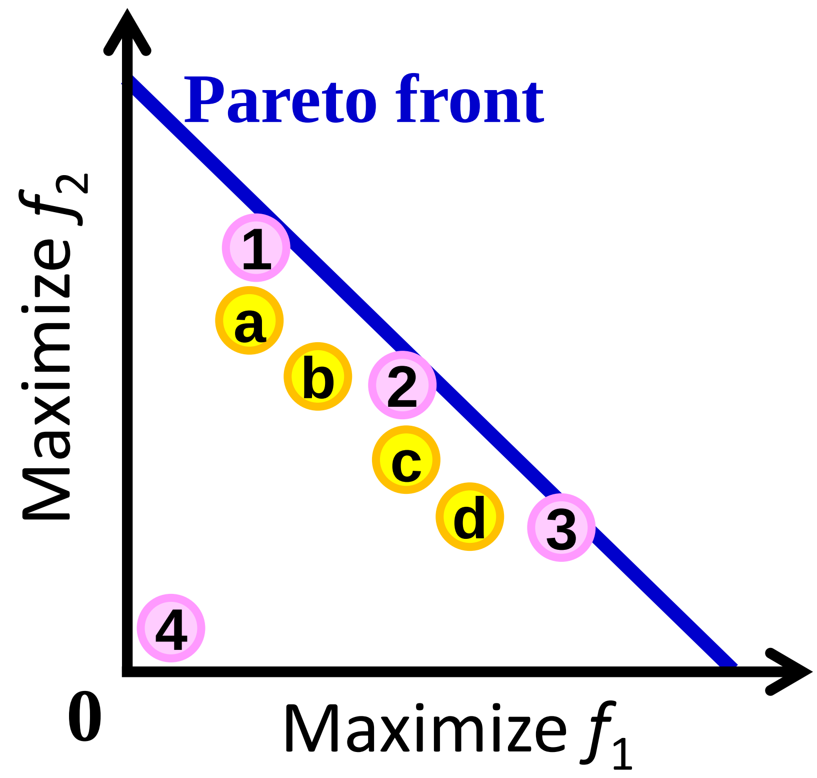
The average distance from each solution to the Pareto front.



Which is better based on GD: Pink or Yellow ?



Example 1



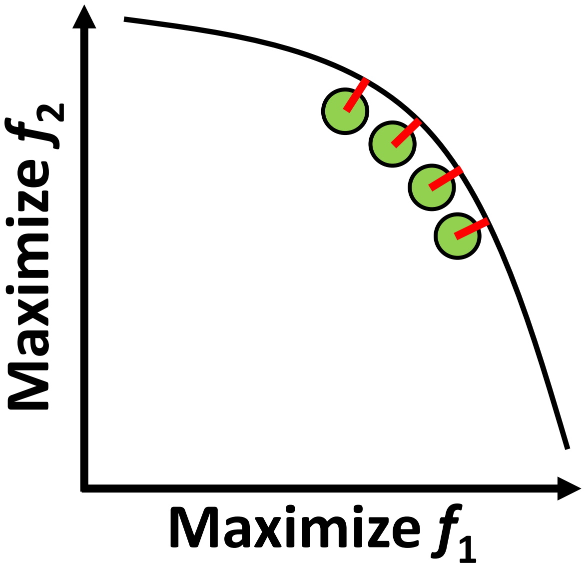
Example 2

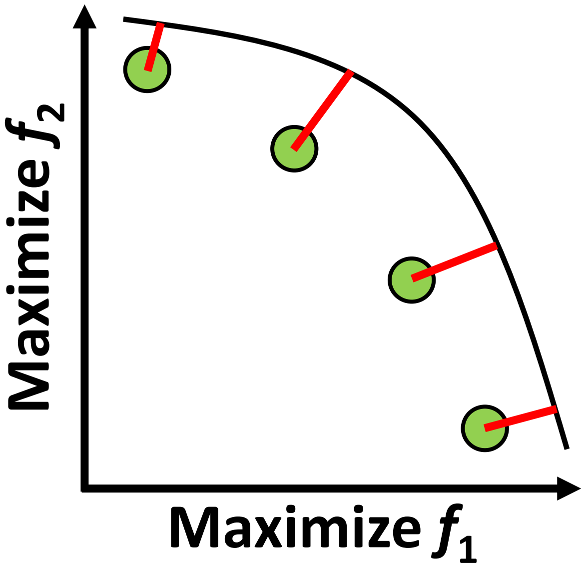
Performance Indicator for Convergence

GD (Generational Distance)

Basic Idea:

The average distance from each solution to the Pareto front.





Small GD ==>
Good Convergence

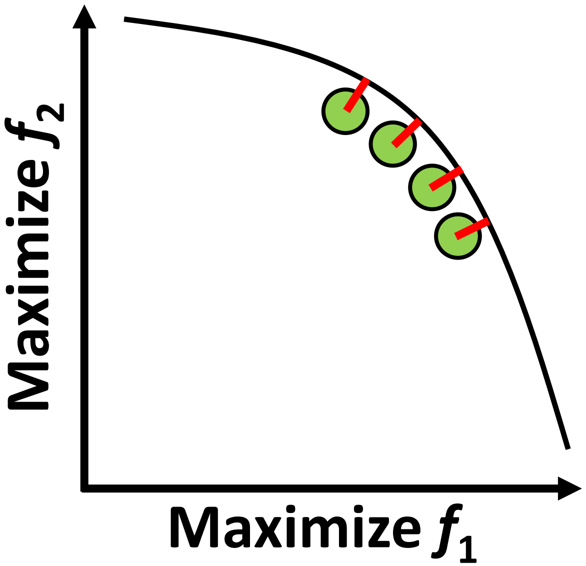
Large GD ==>
Poor Convergence

Performance Indicator for Convergence

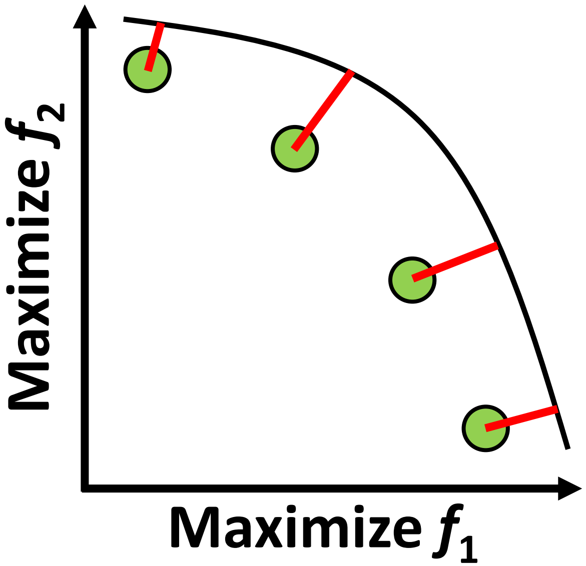
GD (Generational Distance)

Question:

Is GD a Pareto compliant indicator? Your Answer: __________.



Small GD ==>
Good Convergence



Large GD ==>
Poor Convergence

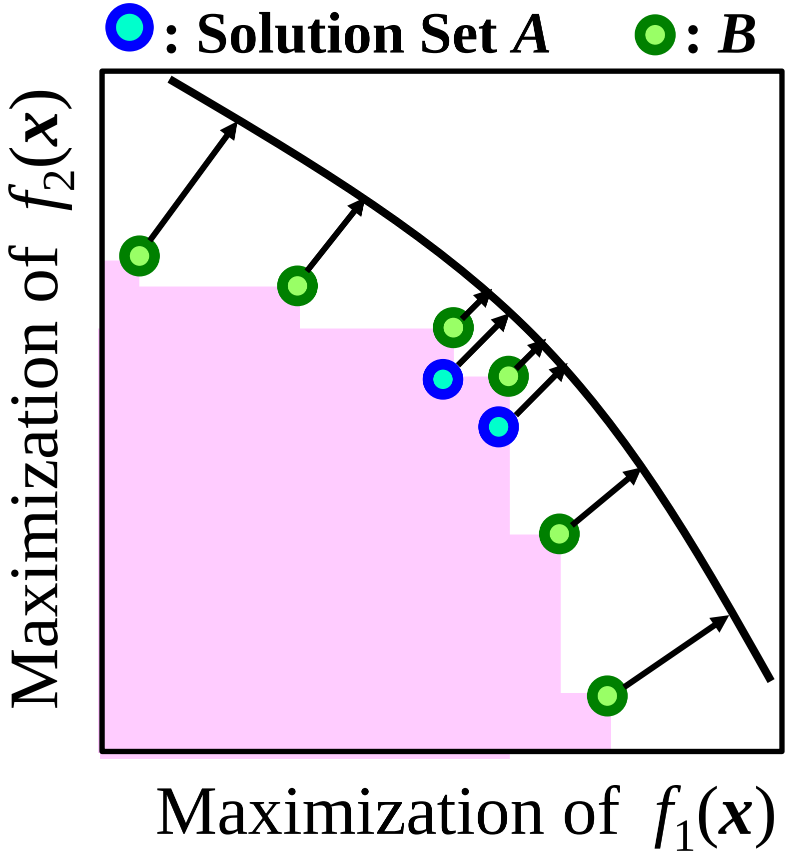
Performance Indicator for Convergence

GD (Generational Distance)

The main difficulty of GD:

GD is not Pareto compliant.

Each solution in A is dominated by at least one solution in B and $A \neq B$ (i.e., B is better than A). However, A has a better (smaller) GD value.



Performance Indicator for Convergence

GD (Generational Distance)

Related difficulty of GD:

GD favors a smaller solution set.

Smaller solution sets often have better GD values (i.e., GD values are improved by removing solutions that are not close to the Pareto front).

In the figure:

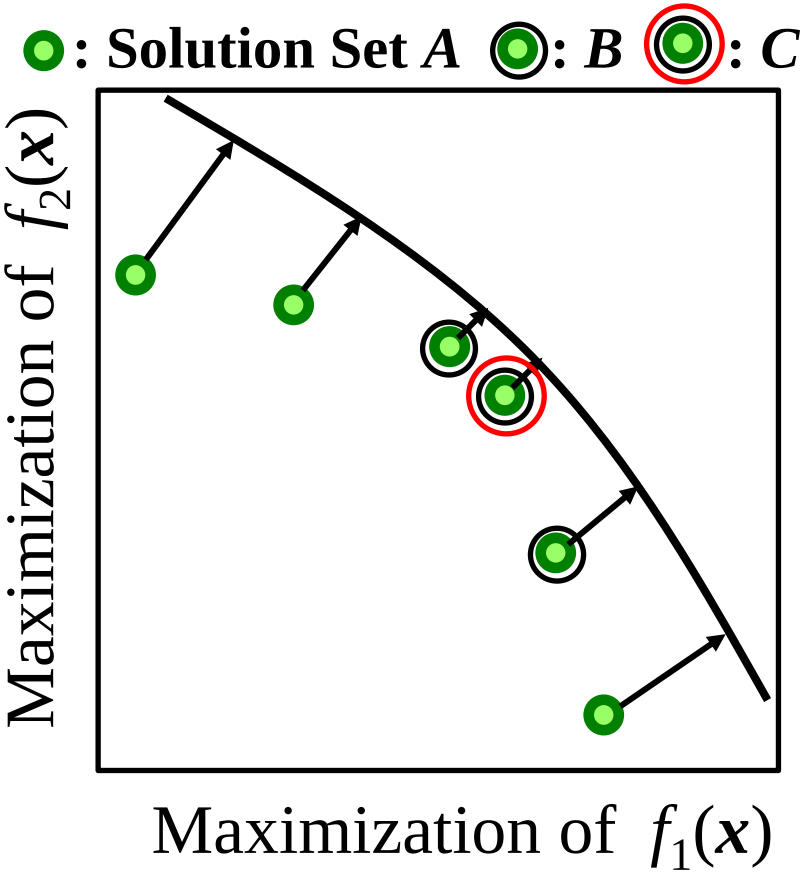
$$A \supset B \supset C$$

$$GD(A) > GD(B) > GD(C)$$

Pareto dominance-based evaluation:

A is better than *B* and *C*,

and *B* is better than *C*.

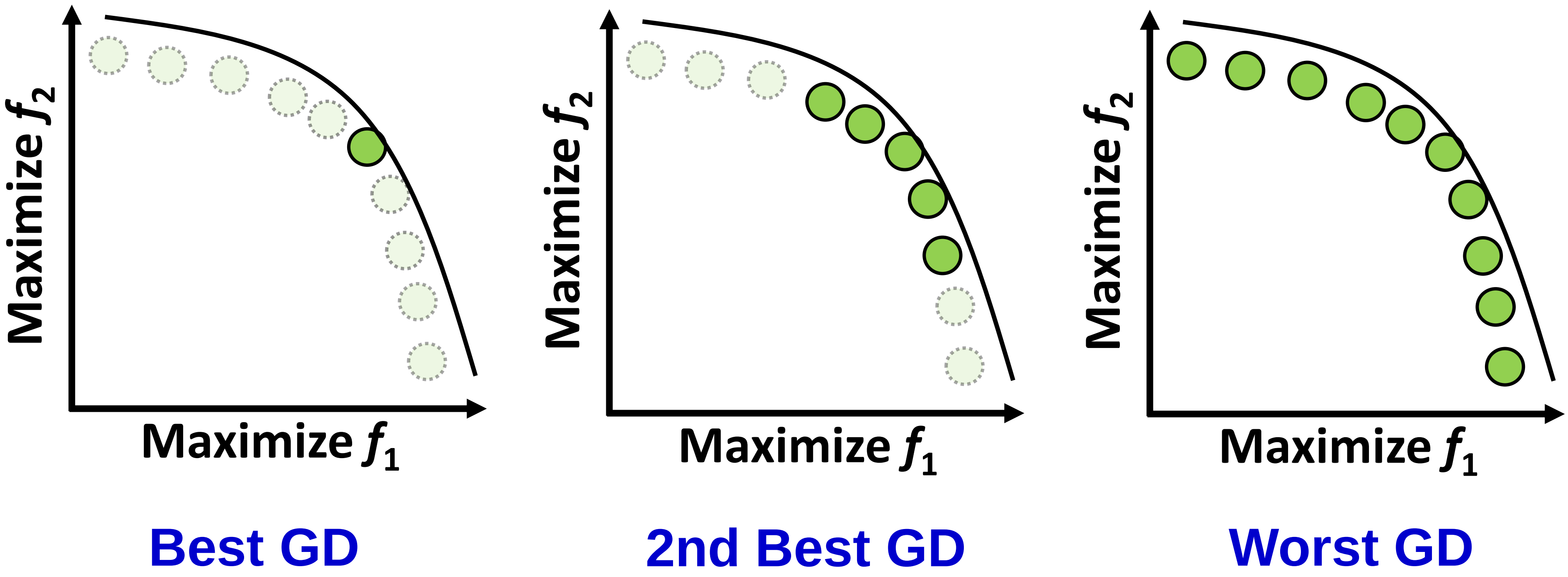


Performance Indicator for Convergence

GD (Generational Distance)

Related difficulty of GD:

GD favors a smaller solution set.



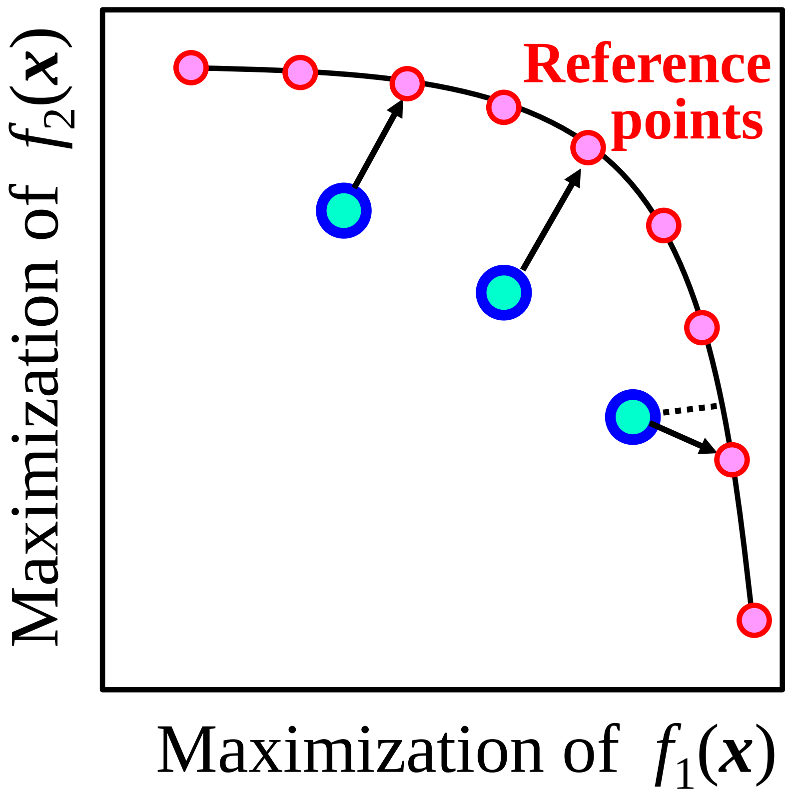
Performance Indicator for Convergence

GD (Generational Distance)

When the distance calculation to the Pareto front is difficult:

Step 1: Generate a number of reference points on the Pareto front.

Step 2: Calculate the average distance from each solution to the nearest reference point.



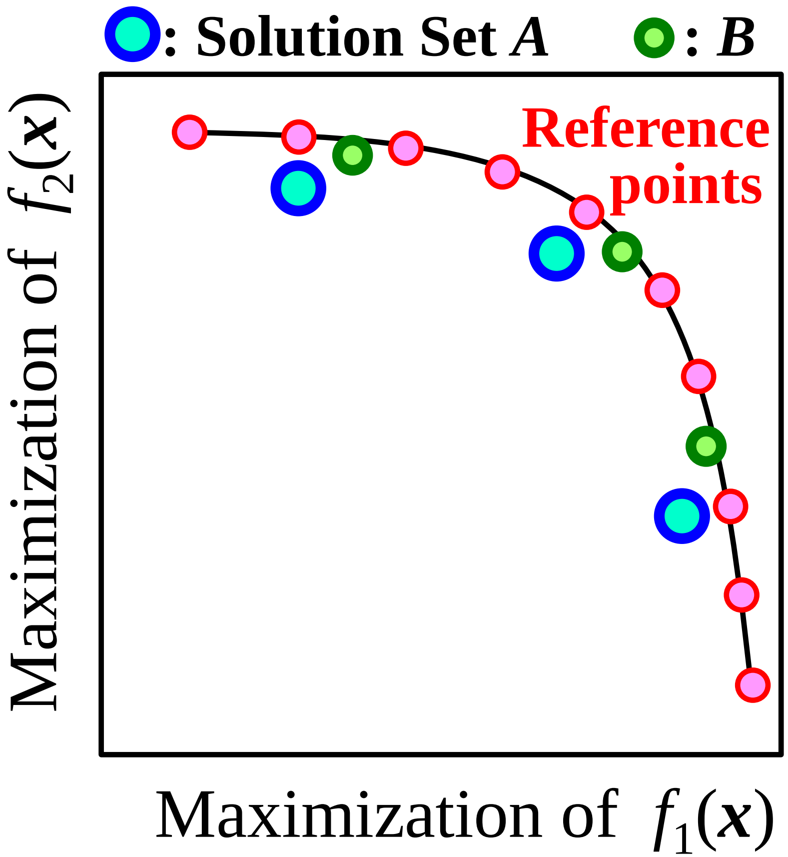
Performance Indicator for Convergence

GD (Generational Distance)

One potential difficulty when we use the reference points:

Dependency on the reference points

While solution set *B* is clearly closer to the Pareto front than solution set *A* (i.e., *B* has a better convergence), solution set *A* has a better GD value.



Performance Indicator for Convergence

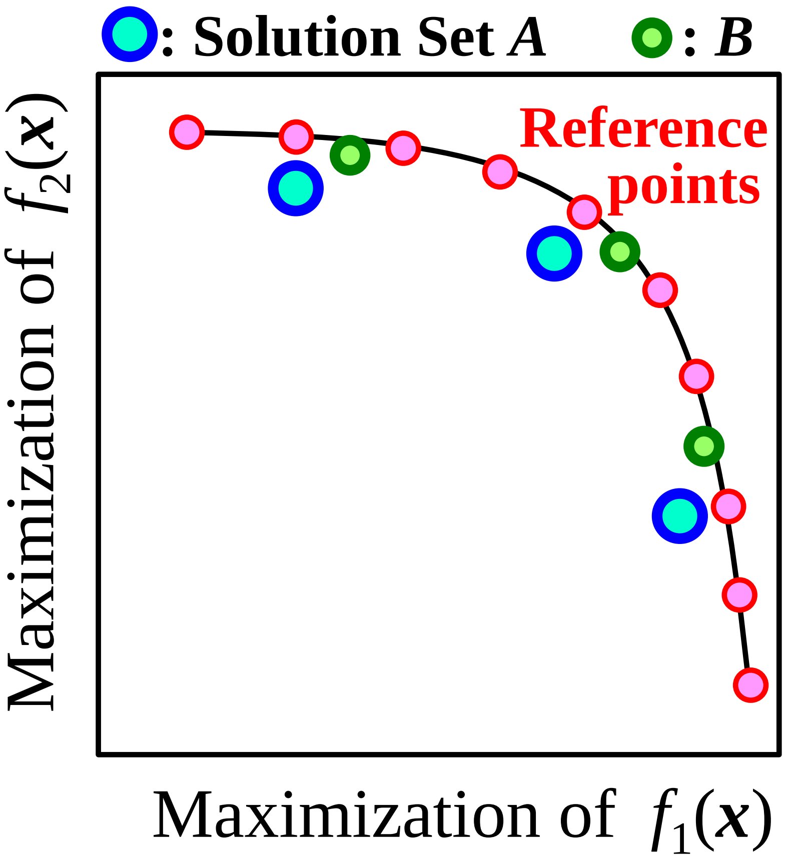
GD (Generational Distance)

One potential difficulty when we use the reference points:

Dependency on the reference points

While solution set B is clearly closer to the Pareto front than solution set A (i.e., B has a better convergence), solution set A has a better GD value.

It is advisable to calculate the distance to the Pareto front (without using reference points).

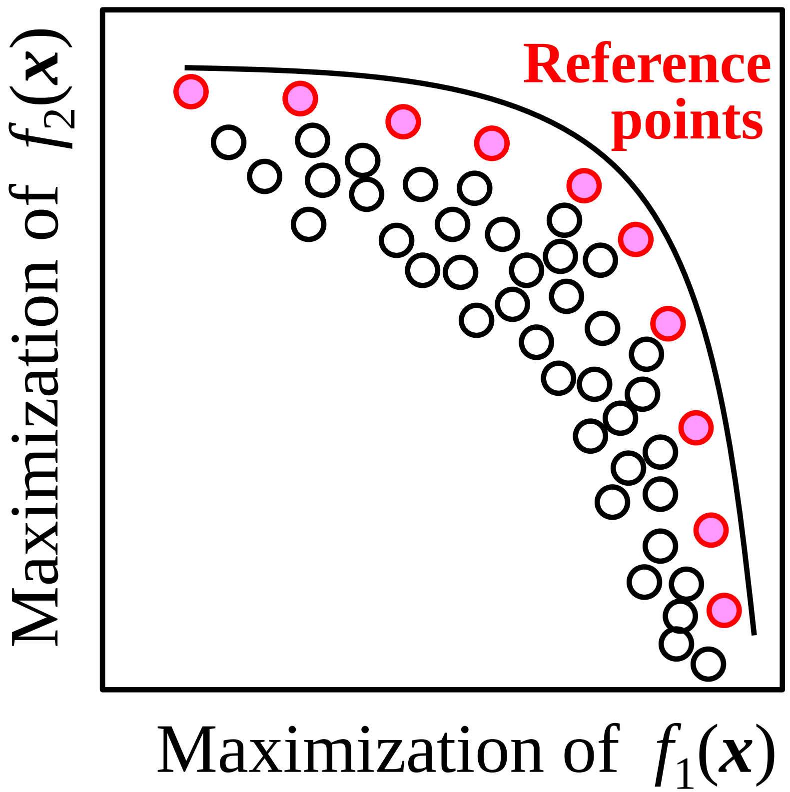


Performance Indicator for Convergence

GD (Generational Distance)

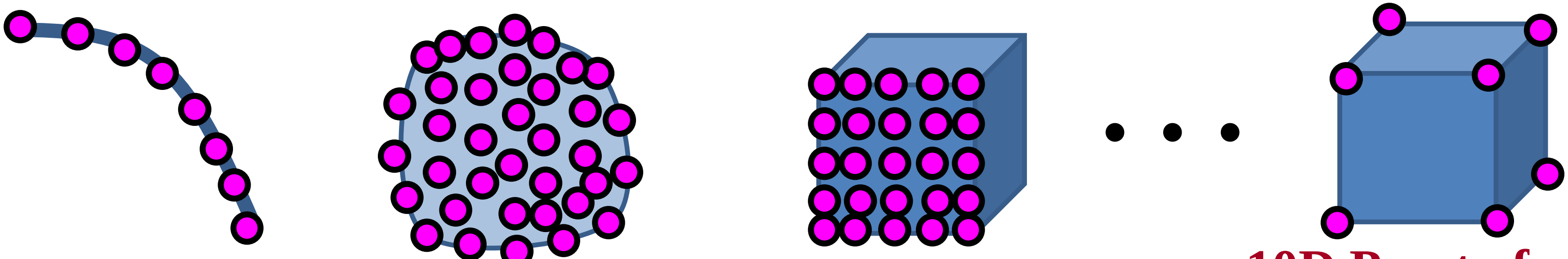
When the Pareto front is unknown,
how can we generate reference points for GD calculation?

1. Merge the obtained solutions by all algorithms into a single set.
2. Choose only non-dominated solutions from the merged solution set.



Difficulty: It is very difficult to create enough reference points for many-objective problems

2-Objective problem | **3-Objective problem** | **4-Objective problem** | . . . | **11-Objective problem**



1D Pareto front
(100 solutions)

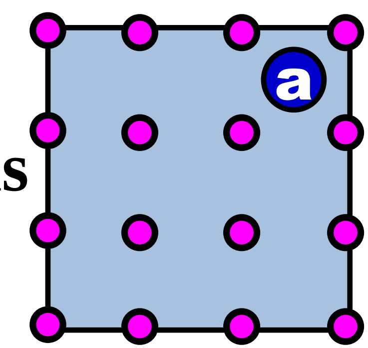
2D Pareto front
(100 x 100)

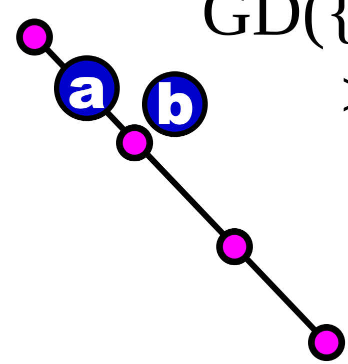
3D Pareto front
(100 x 100 x 100)

10D Pareto front
(1024 Corners)

11-Objective problem

4^{10} **= 1,048,576 solutions**





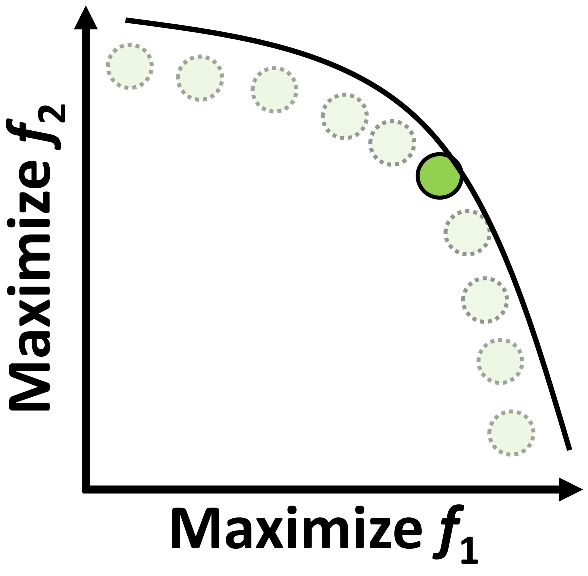
GD({a})
> GD({b})

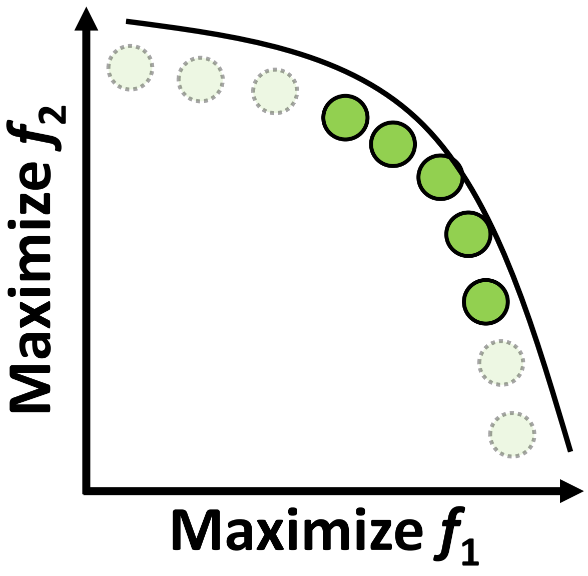
Performance Indicator for Convergence

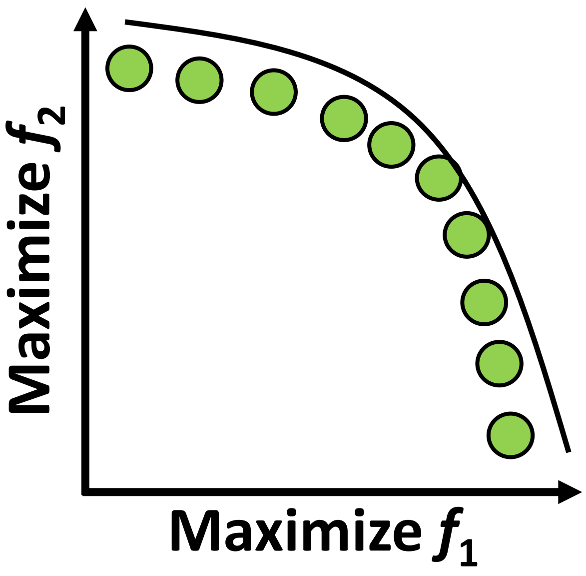
GD (Generational Distance)

Important Note:

GD can evaluate only the convergence (not the overall performance).







Other Ideas to Evaluate the Convergence

For Discrete (Combinatorial) Optimization Problem

When all the Pareto optimal solutions are known,

(1) Percentage of the obtained Pareto optimal solutions.

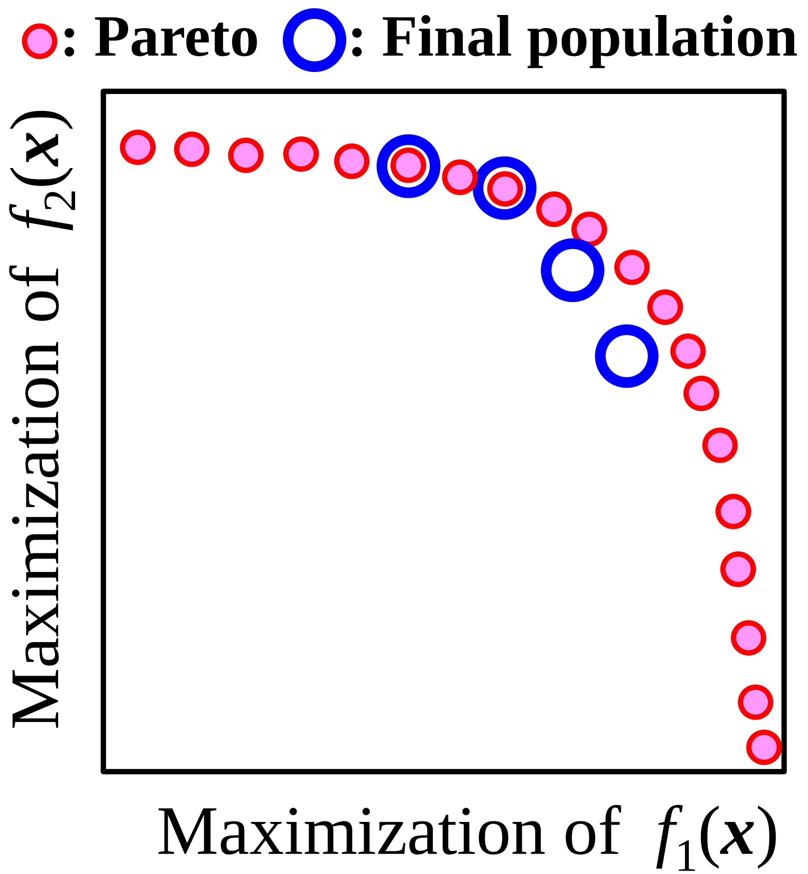
(2) Percentage of the Pareto optimal solutions in the final population.

(1) 10% (2 our of 20) of the Pareto optimal solutions are obtained by a single run of an EMO algorithm.

This is an overall performance indicator.

(2) **50% (2 out of 4)** of the final population of a single run of an EMO algorithm are Pareto optimal solutions.

This is a pure convergence indicator.



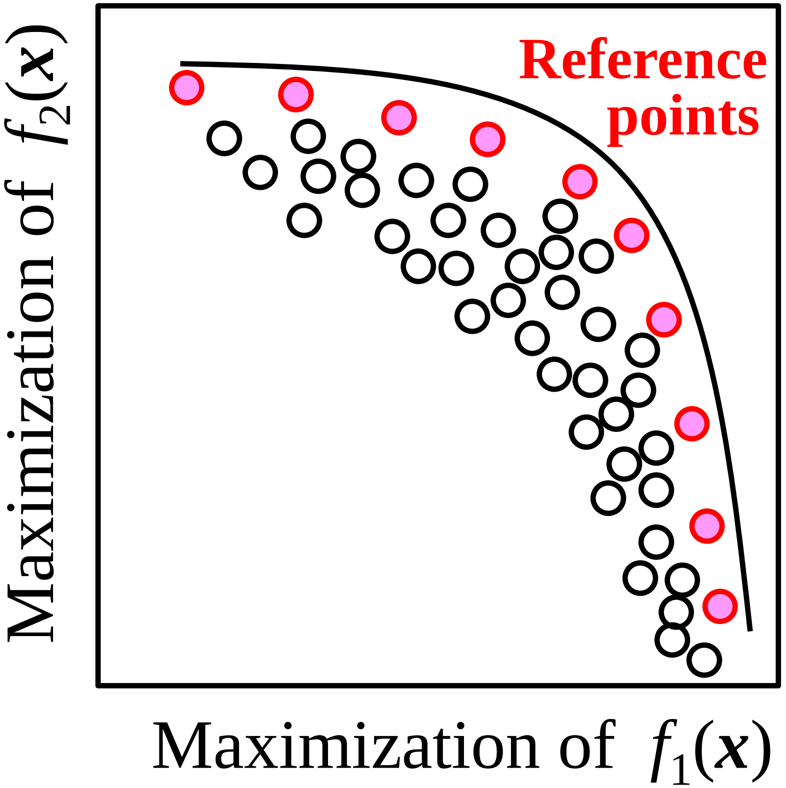
The larger, the better

Other Ideas to Evaluate the Convergence

For Discrete (Combinatorial) Optimization Problem

When Pareto optimal solutions are unknown:

Using all the obtained solutions by all algorithm and choosing only non-dominated solutions, we can construct an approximate Pareto optimal solution set (in the same manner as the construction of the reference point set for GD calculation). Then we can calculate the percentage.



Performance Indicator for Diversity (Spread)

Maximum Spread: The larger, the better

Basic Idea:

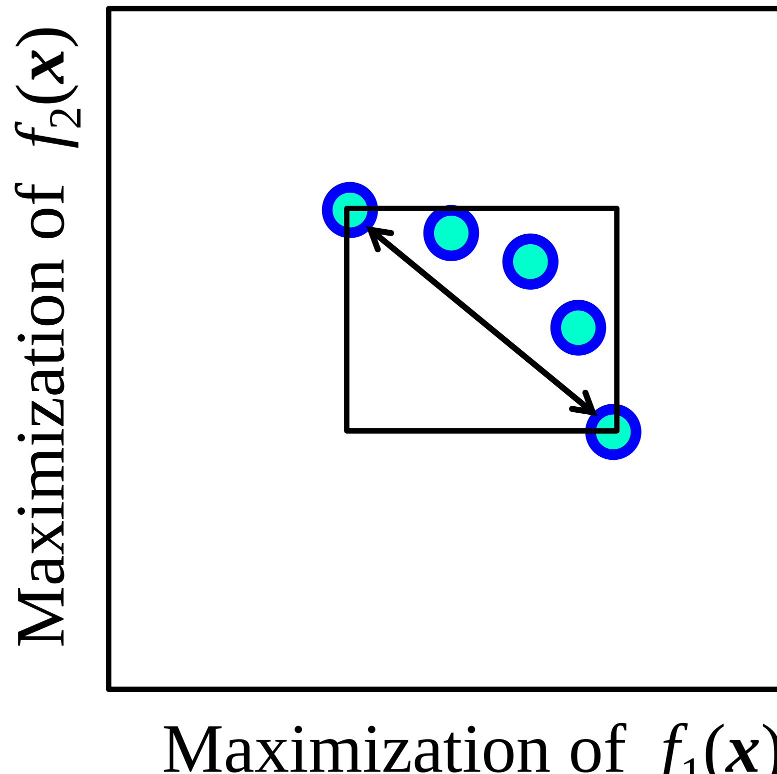
The length of the diagonal of a hyperbox including all solutions.

Solution Set: $A = \{\boldsymbol{a}_1, \boldsymbol{a}_2, ..., \boldsymbol{a}_N\}$ where $\boldsymbol{a}_i = (a_{i1}, a_{i2}, ..., a_{im})$

Maximum Spread:

$$D = \sqrt{\sum_{j=1}^{m} \left(\max_{\boldsymbol{a}_i \in A}\{a_{ij}\} - \min_{\boldsymbol{a}_i \in A}\{a_{ij}\} \right)^2}$$

$$\boldsymbol{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}),\ f_2(\boldsymbol{x}),\ ...,\ f_m(\boldsymbol{x}))$$



Performance Indicator for Diversity (Spread)

Maximum Spread: **The larger, the better**

Question:

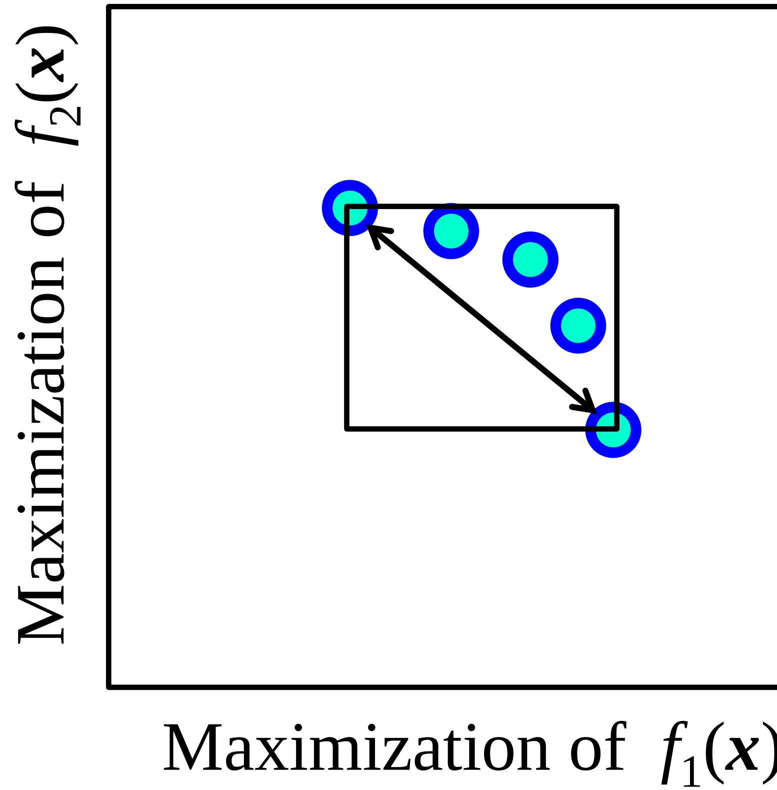
Is Spread a Pareto compliant indicator ? Your answer: ________ .

Solution Set: $A = \{\boldsymbol{a}_1, \boldsymbol{a}_2, ..., \boldsymbol{a}_N\}$ where $\boldsymbol{a}_i = (a_{i1}, a_{i2}, ..., a_{im})$

Maximum Spread:

$$D = \sqrt{\sum_{j=1}^{m} \left(\max_{\boldsymbol{a}_i \in A} \{a_{ij}\} - \min_{\boldsymbol{a}_i \in A} \{a_{ij}\} \right)^2}$$

$$\boldsymbol{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}),\ f_2(\boldsymbol{x}),\ ...,\ f_m(\boldsymbol{x}))$$



Performance Indicator for Diversity (Spread)

Maximum Spread

The maximum spread indicator is not Pareto compliant.

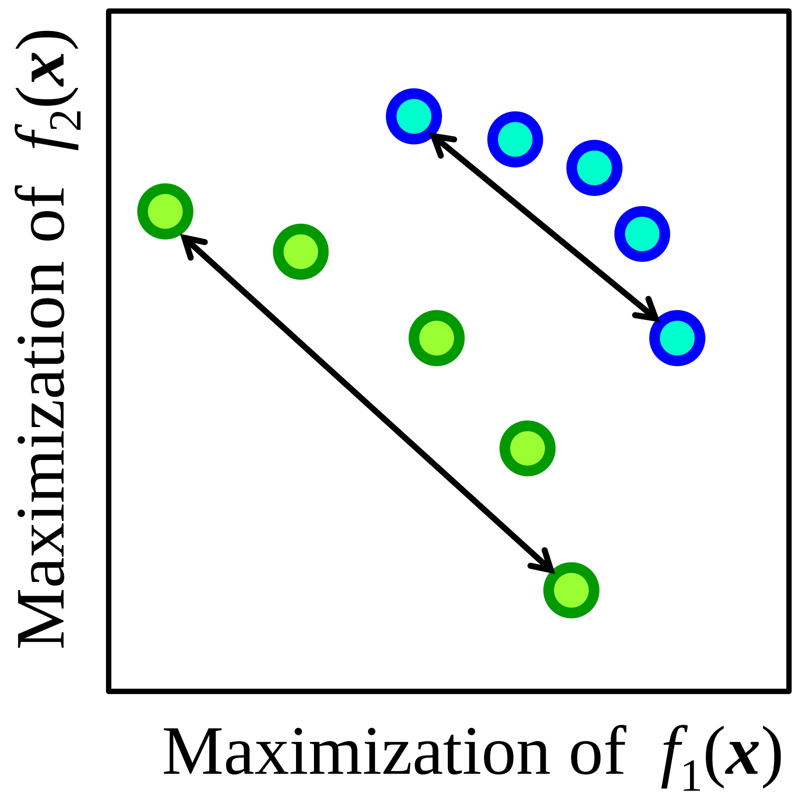
It can evaluate only the spread (not the overall performance).

Solution Set: $A = \{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_N\}$ where $\boldsymbol{a}_i = (a_{i1}, a_{i2}, \ldots, a_{im})$

Maximum Spread:

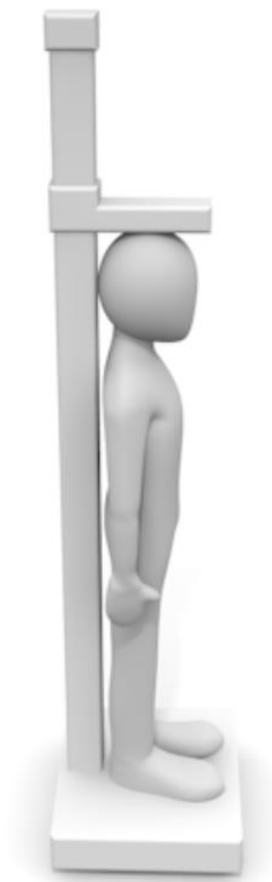
$$D = \sqrt{\sum_{j=1}^{m} \left(\max_{\boldsymbol{a}_i \in A}\{a_{ij}\} - \min_{\boldsymbol{a}_i \in A}\{a_{ij}\} \right)^2}$$

$$\boldsymbol{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}),\ f_2(\boldsymbol{x}),\ \ldots,\ f_m(\boldsymbol{x}))$$



Indicators: Performance evaluation of a solution set

- **Convergence to the Pareto front: GD**
- **Spread of a solution set (Broadness): Maximum Spread**
- **Uniformity of solutions**
- **Overall performance.**



Performance Indicator for Diversity (Uniformity)

Spacing: The smaller, the better

Basic Idea:

Standard deviation of the distance from each solution to the nearest neighbor.

Solution Set: $A = \{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_N\}$ where $\boldsymbol{a}_i = (a_{i1}, a_{i2}, \ldots, a_{im})$

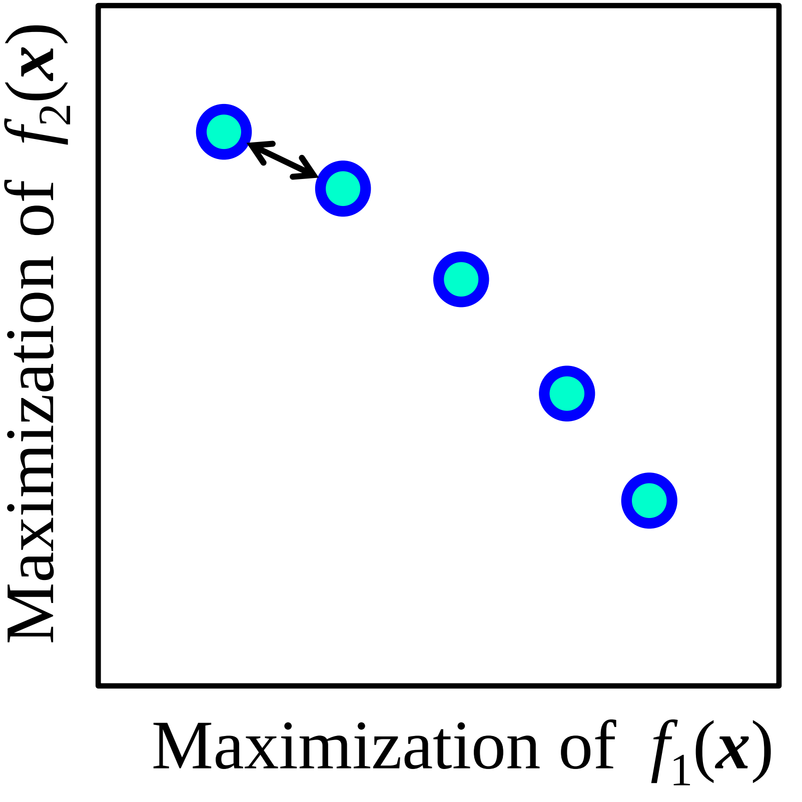
Spacing $$S = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(d_i - \bar{d})^2}$$

where

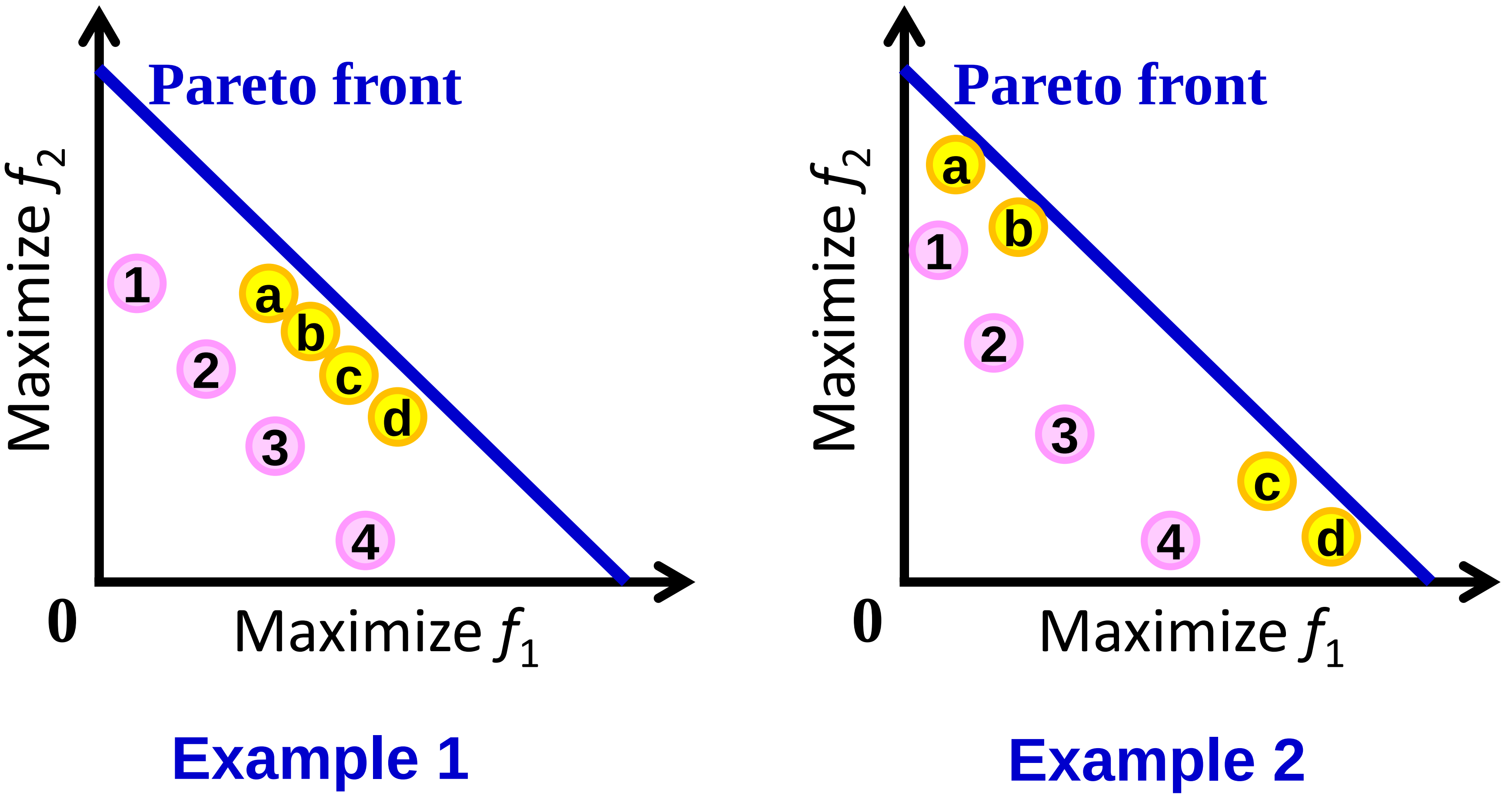
$$d_i = \min_{\substack{\boldsymbol{a}_k \in A \\ k \neq i}} distance(\boldsymbol{a}_i, \boldsymbol{a}_k)$$

$$\bar{d} = \frac{1}{N}\sum_{i=1}^{N} d_i$$

$$\boldsymbol{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}),\ f_2(\boldsymbol{x}),\ \ldots,\ f_m(\boldsymbol{x}))$$



Which is better based on Spacing: Pink or Yellow ?

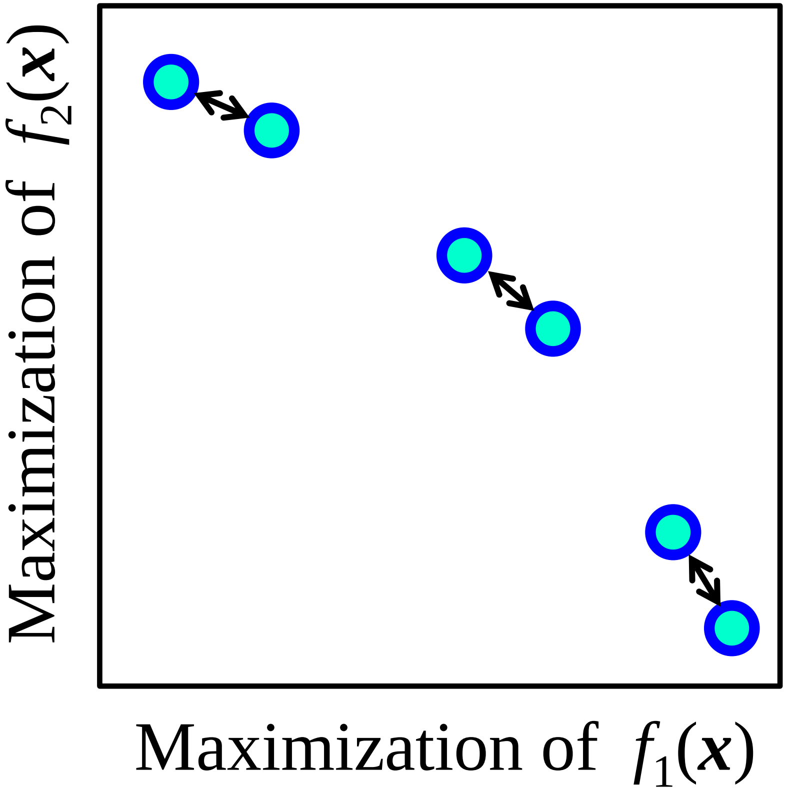


Performance Indicator for Diversity (Uniformity)

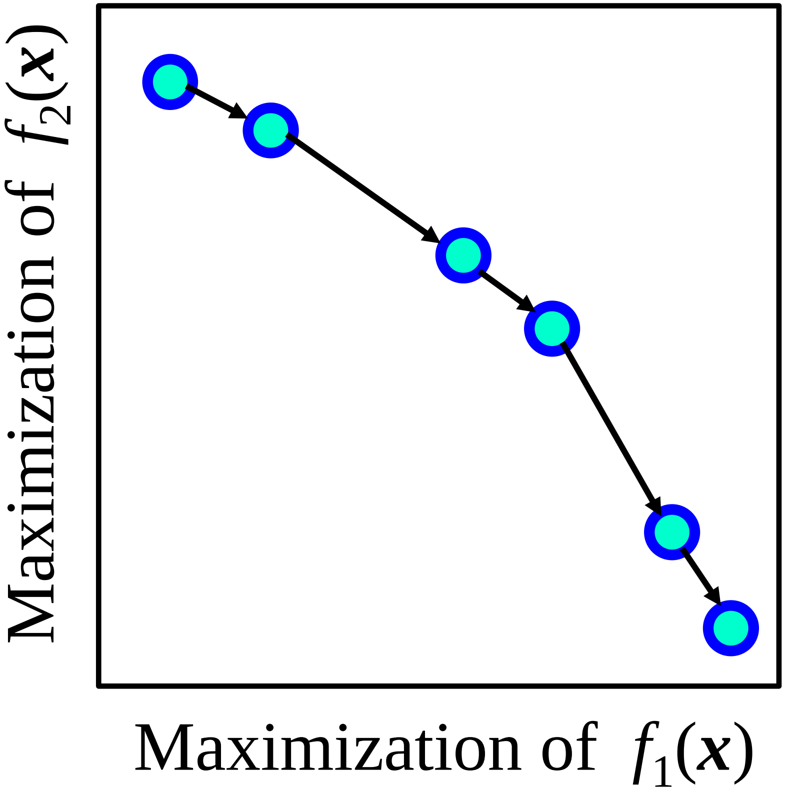
Spacing

This definition does not work as intended.

Nearest Neighbor Distance (based on the definition)



Intention (?)
The distance to the next solution



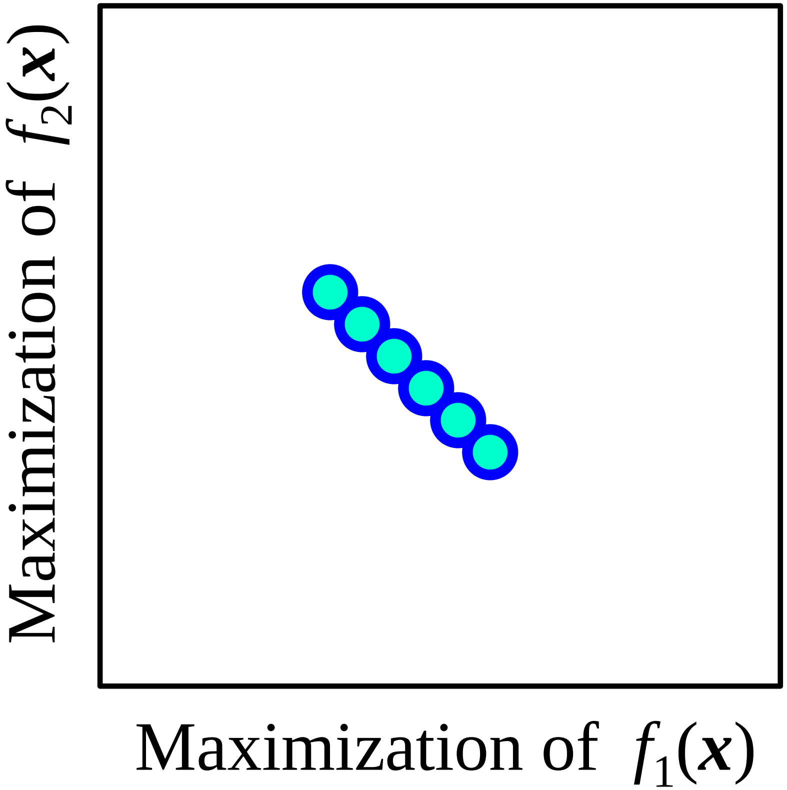
Performance Indicator for Diversity (Uniformity)

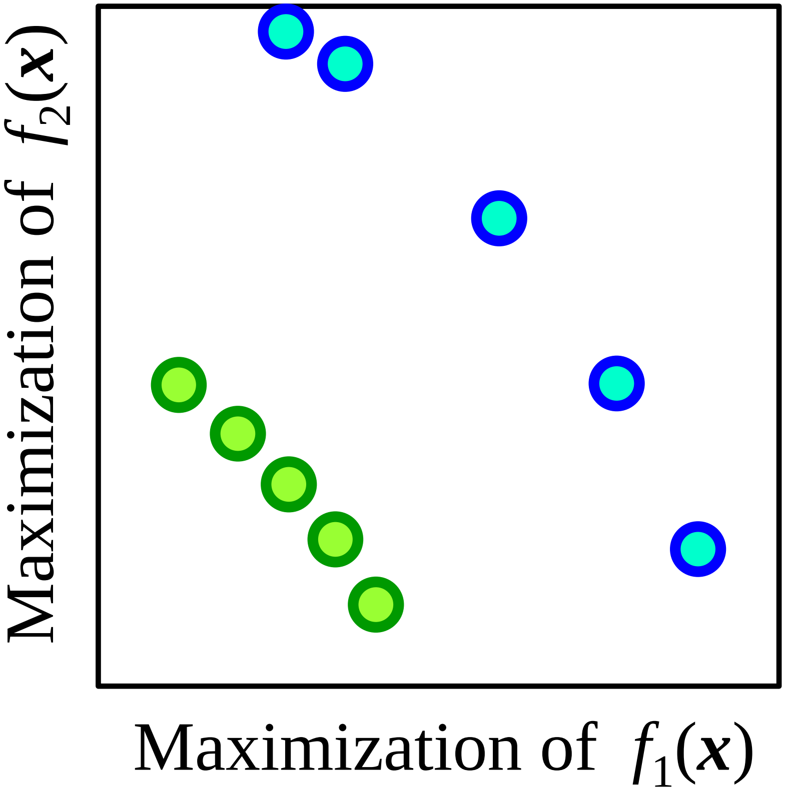
Spacing

Other difficulties:

(1) A concentrated solution set has a good value.

(2) The spacing indicator is not Pareto compliant.





Performance Indicator for Diversity (Uniformity)

Spread: The smaller, the better

$$\Delta = \frac{d_1^e + d_2^e + \sum_{i=1}^{N-1} | d_i - \bar{d} |}{d_1^e + d_2^e + (N-1)\,\bar{d}}$$

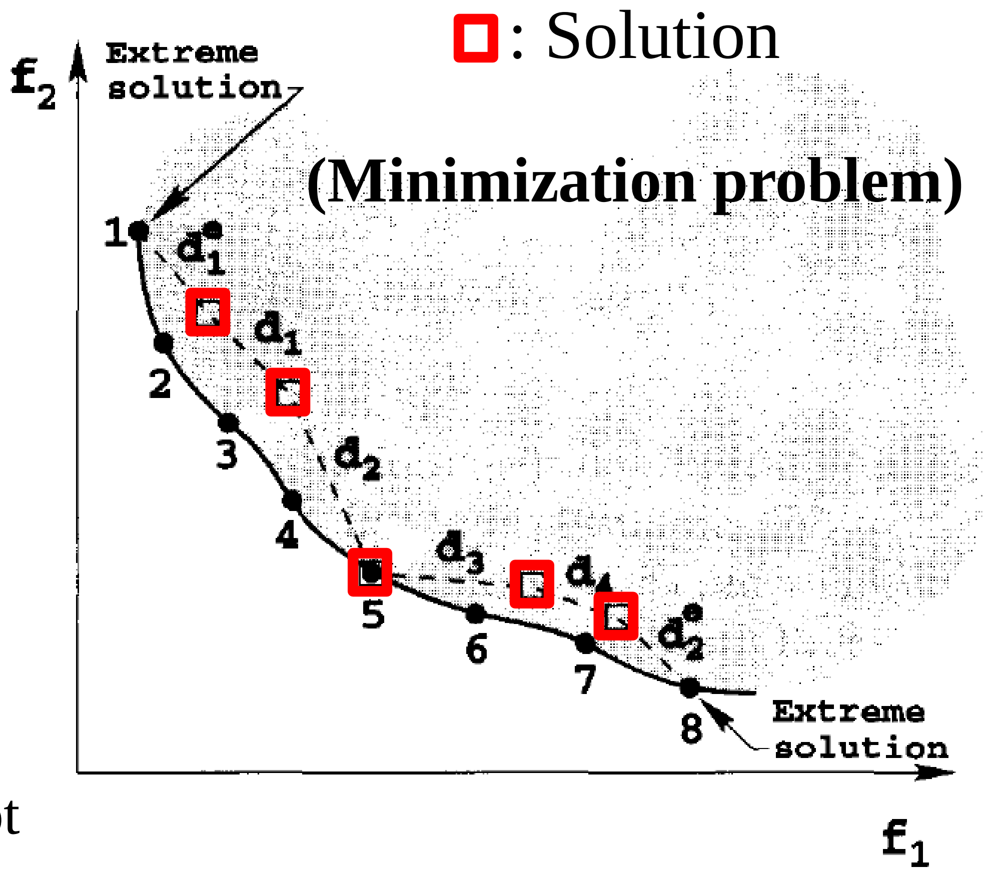


Figure 188 Distances from the extreme solutions.

K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*, John Wiley & Sons, Chichester, 2001.

Positive comments:

- d_i is the distance to the next solution (**Uniformity**).
- The distance from each extreme point of the Pareto front is also used (**Spread:** Overall diversity)

Negative comments:

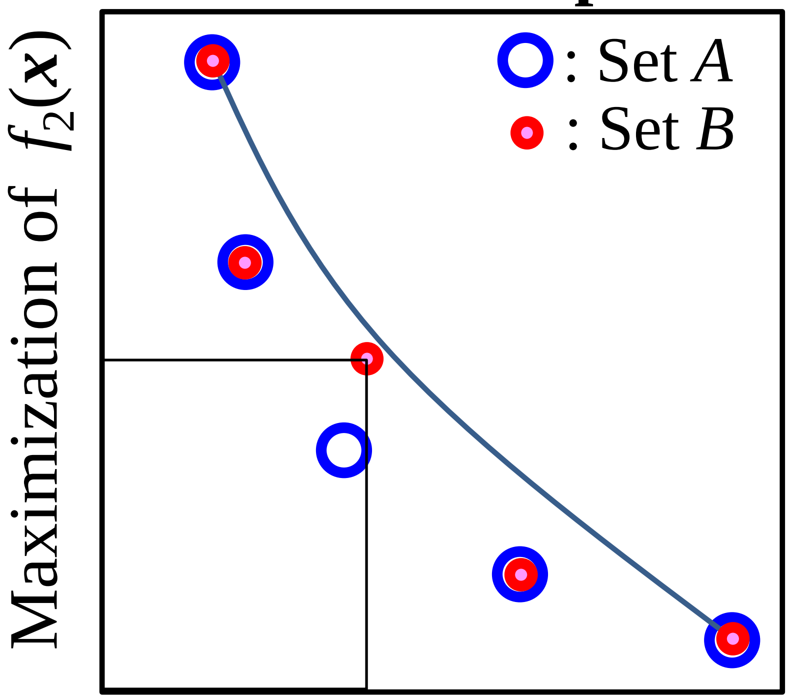
- The distance to the next point is not easy to define for the case of three or more objectives.
- Not Pareto compliant.

Performance Indicator for Diversity (Uniformity)

Spread

$$\Delta = \frac{d_1^e + d_2^e + \sum_{i=1}^{N-1} | d_i - \bar{d} |}{d_1^e + d_2^e + (N-1)\,\bar{d}}$$

Not Pareto compliant



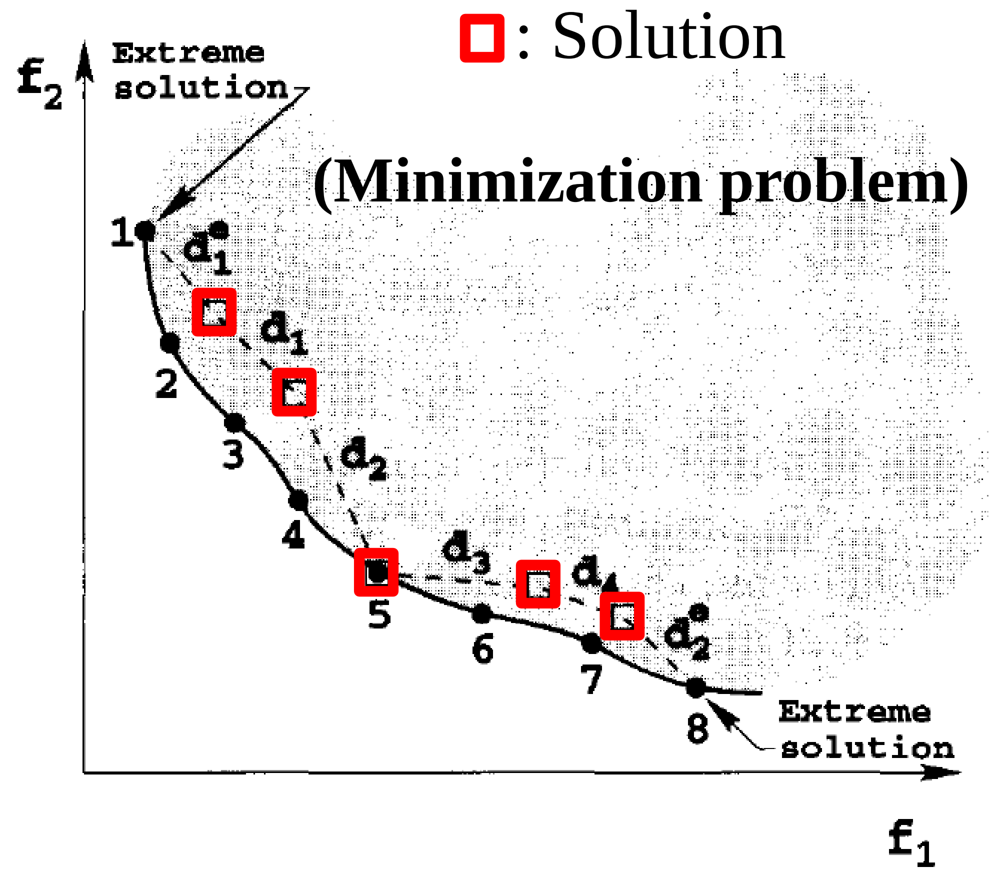


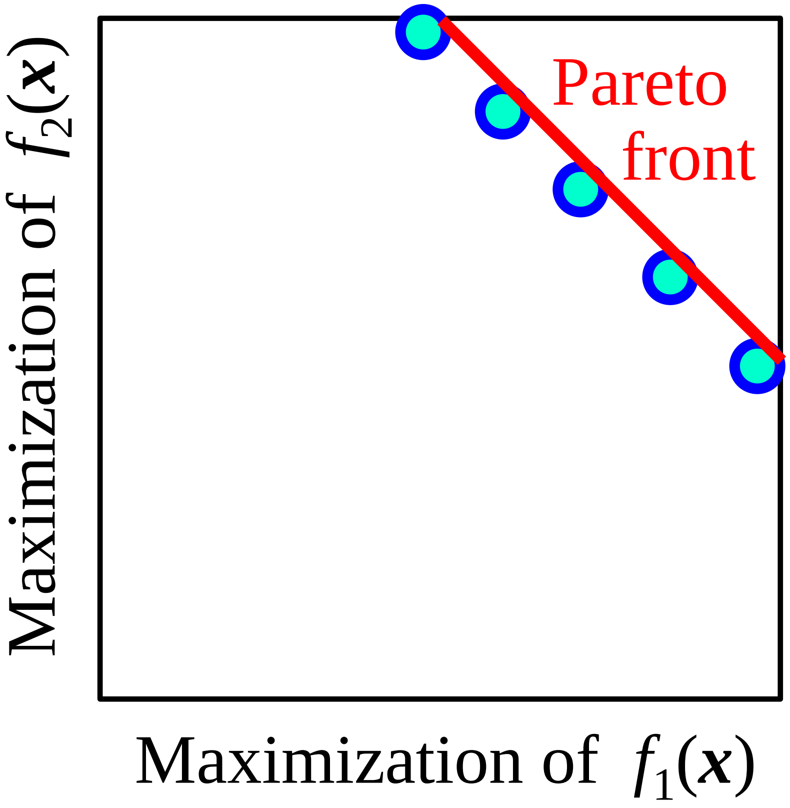
Figure 188 Distances from the extreme solutions.

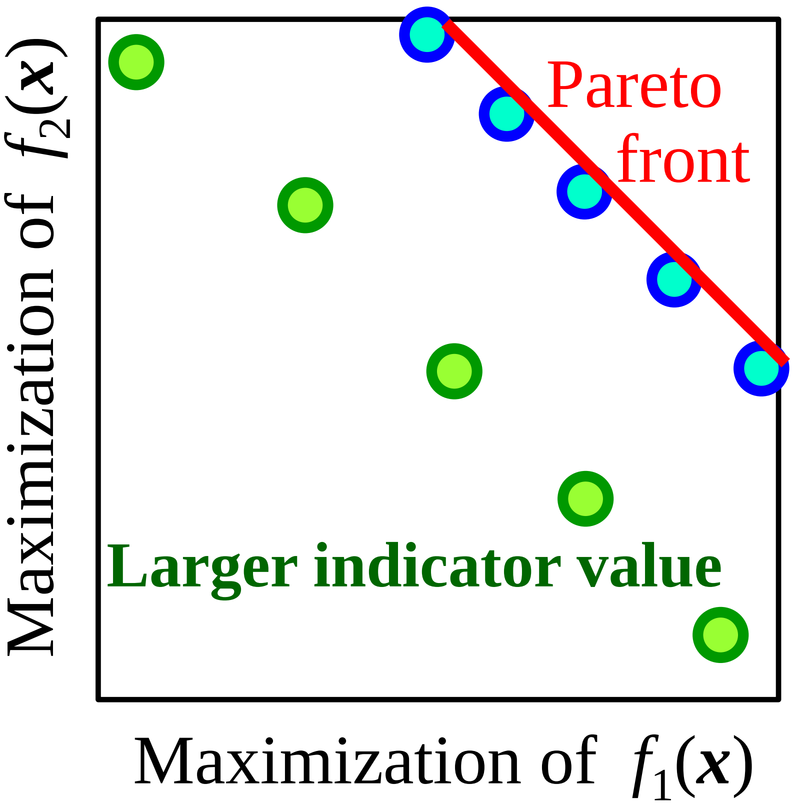
Performance Indicator for Diversity (Uniformity): Uniformity

Minimum distance between solutions: The larger, the better

(1) If all solutions are on the Pareto front, this is a good indicator.
The maximization of this indicator means the best uniformity.

(2) If solutions are not converged, this is not a good indicator.

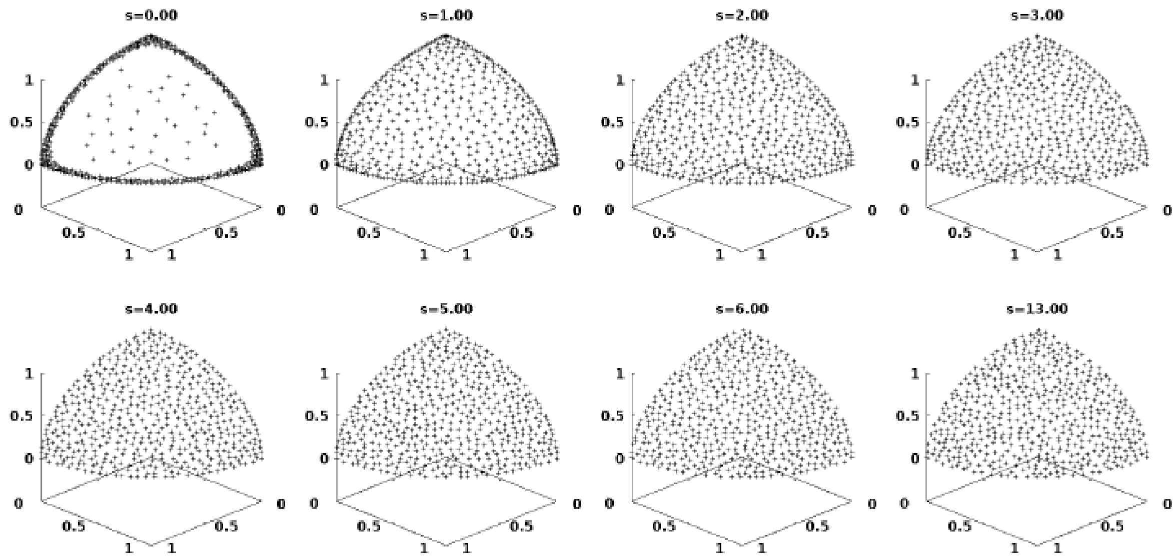




Performance Indicator for Diversity (Spread & Uniformity)

Riesz *s*-Energy

$$E_s(A) = \sum_{\boldsymbol{a} \in A} \sum_{\substack{\boldsymbol{b} \in A \\ \boldsymbol{b} \neq \boldsymbol{a}}} \frac{1}{\| \boldsymbol{a} - \boldsymbol{b} \|^s} \qquad (s = m \quad 1 \text{ (suggested value)})$$



(c) DTLZ2 3D

Fig. 3: Approximated Riesz s-energy optimal distributions, varying the value of the parameter s.

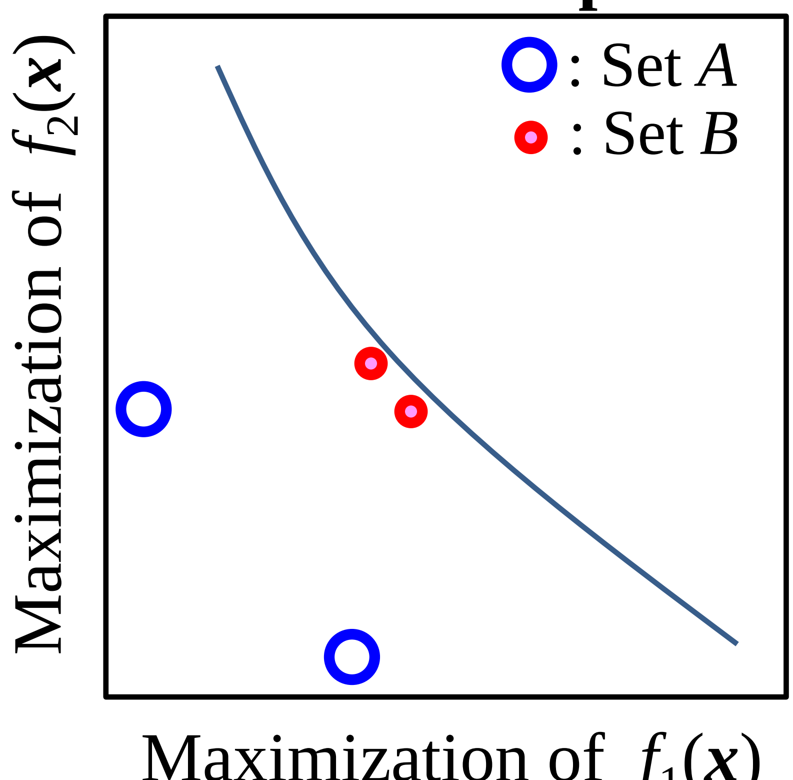
J. G. Falcon-Cardona, H. Ishibuchi, and C. A. C. Coello, "Riesz *s*-energy-based reference sets for multi-objective optimization," CEC 2020.

Performance Indicator for Diversity (Spread & Uniformity)

Riesz *s*-Energy

$$E_s(A) = \sum_{\boldsymbol{a} \in A} \sum_{\substack{\boldsymbol{b} \in A \\ \boldsymbol{b} \neq \boldsymbol{a}}} \frac{1}{\| \boldsymbol{a} - \boldsymbol{b} \|^s} \qquad (s = m \quad 1 \text{ (suggested value)})$$

Not Pareto compliant



Performance Indicators

Convergence: GD (OK)

Spread: Maximum Spread (OK)

Uniformity: Spacing (Questionable)

Spread & Uniformity: Deb's spread for two-objective problems

A current active research topic for three and more objectives.

In Proceedings of EMO 2019

Kalyanmoy Deb, Sunith Bandaru, and Haitham Seada, **Generating Uniformly Distributed Points on a Unit Simplex for Evolutionary Many-Objective Optimization.**

Alberto Rodríguez Sánchez, Antonin Ponsich, Antonio López Jaimes, and Saúl Zapotecas Martínez: **A Parallel Tabu Search Heuristic to Approximate Uniform Designs for Reference Set Based MOEAs**

In Proceedings of IEEE CEC 2020

J. G. Falcon-Cardona, H. Ishibuchi, and C. A. C. Coello, **Riesz *s*-energy-based Reference Sets for Multi-objective Optimization**.

Overall Performance Indicators

IGD (inverted generational distance) indicator

IGD$^+$ (inverted generational distance plus) indicator

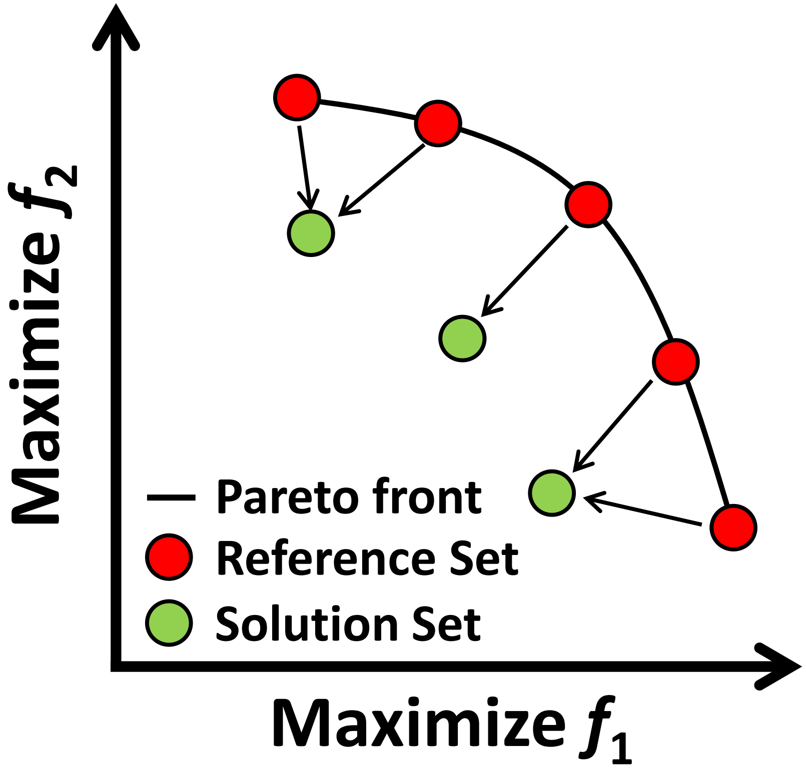
-indicator

HV (hypervolume) indicator

Performance Indicator: IGD

IGD: Inverted Generational Distance

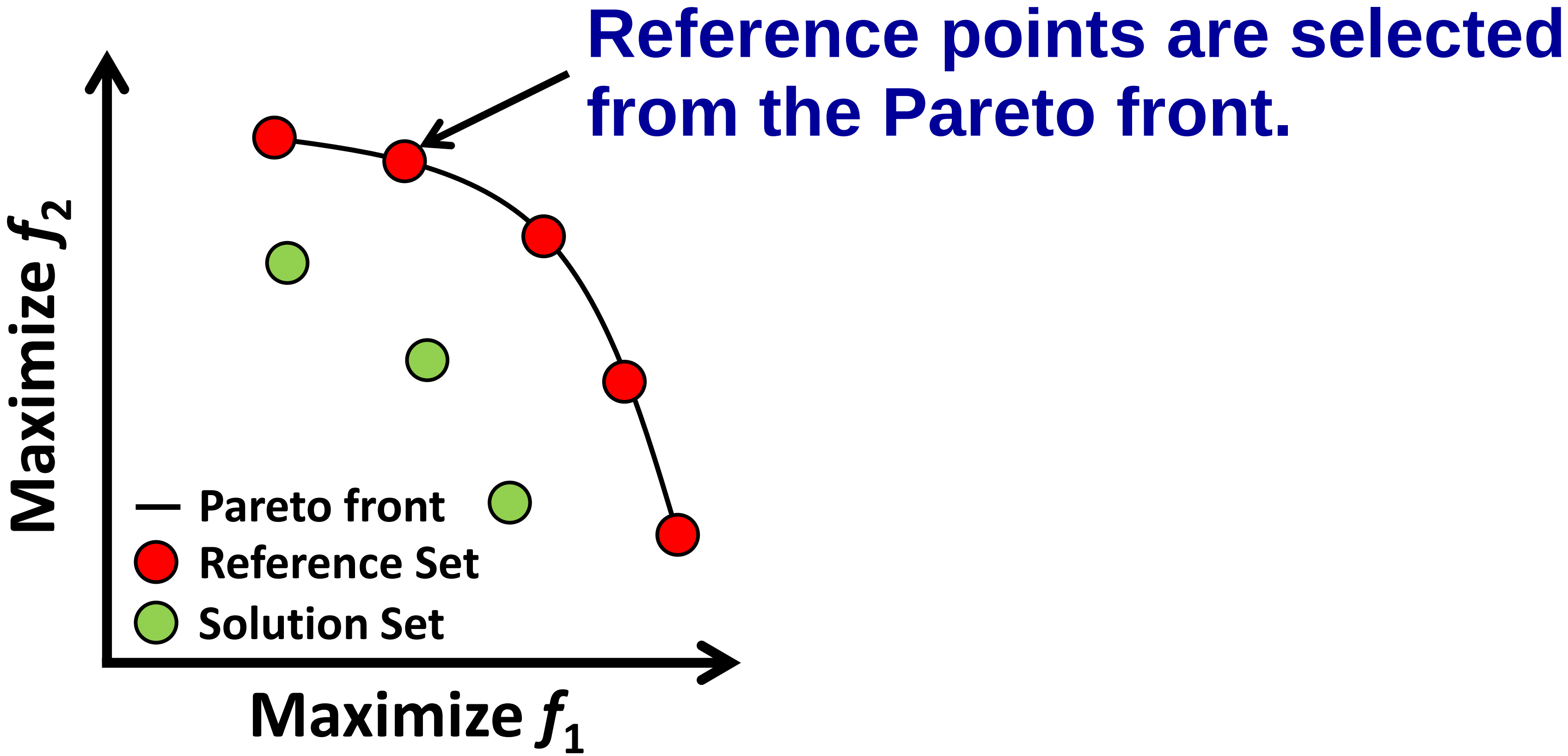
Average distance from each reference point on the Pareto front to the nearest solution.



Performance Indicator: IGD

IGD: Inverted Generational Distance

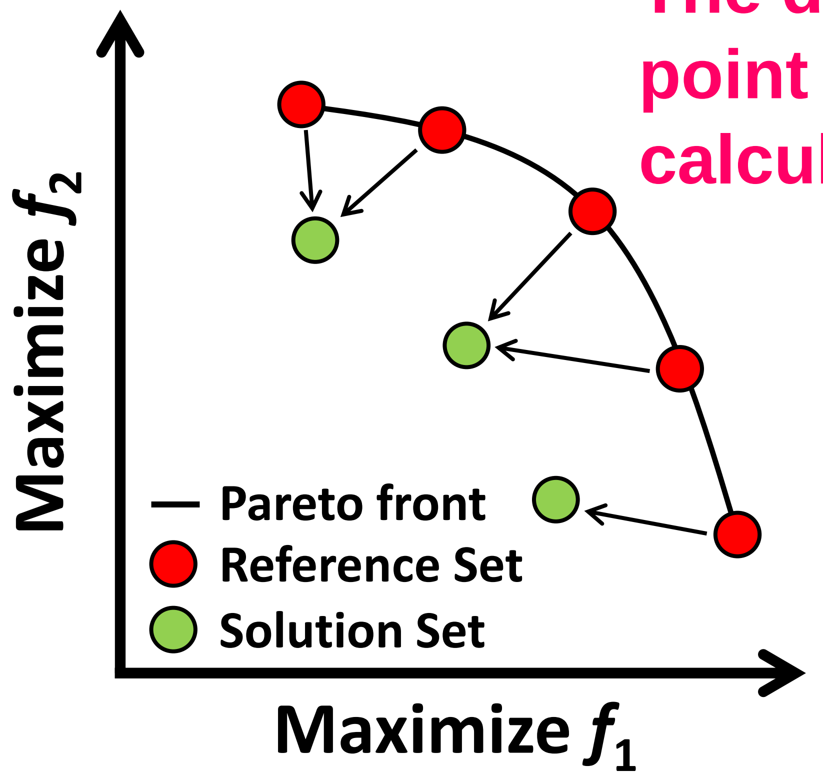
Average distance from each reference point on the Pareto front to the nearest solution.



Performance Indicator: IGD

IGD: Inverted Generational Distance

Average distance from each reference point on the Pareto front to the nearest solution.

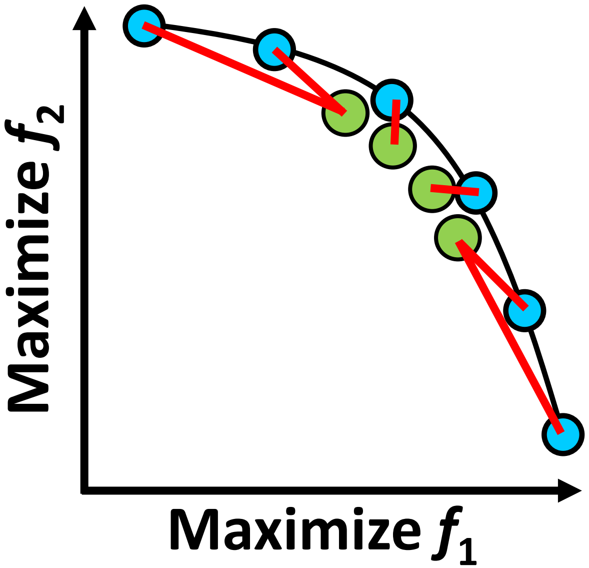


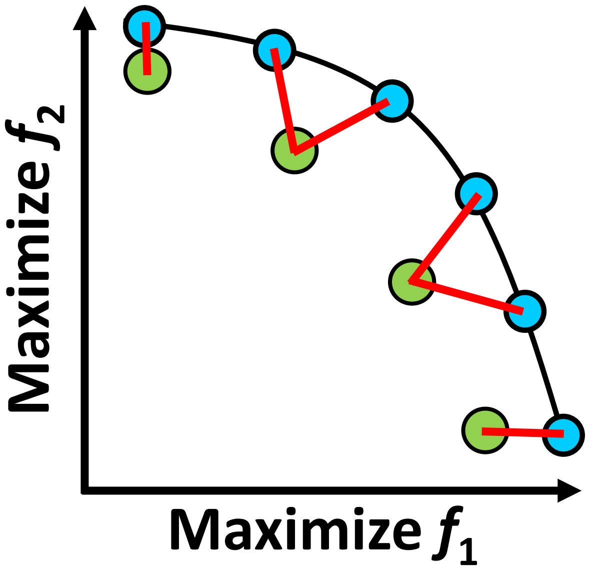
The distance from each reference point to its nearest solution is calculated.

Performance Indicator: IGD

IGD: Inverted Generational Distance

Average distance from each reference point on the Pareto front to the nearest solution.





Due to poor diversity,
IGD is not very good.

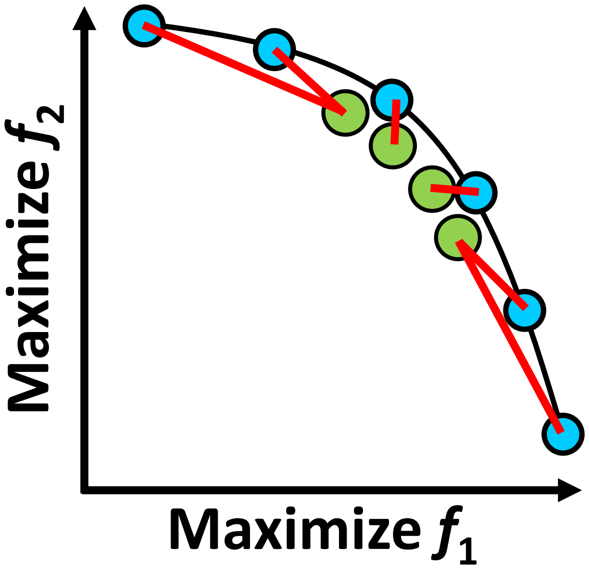
Due to poor convergence
IGD is not very good.

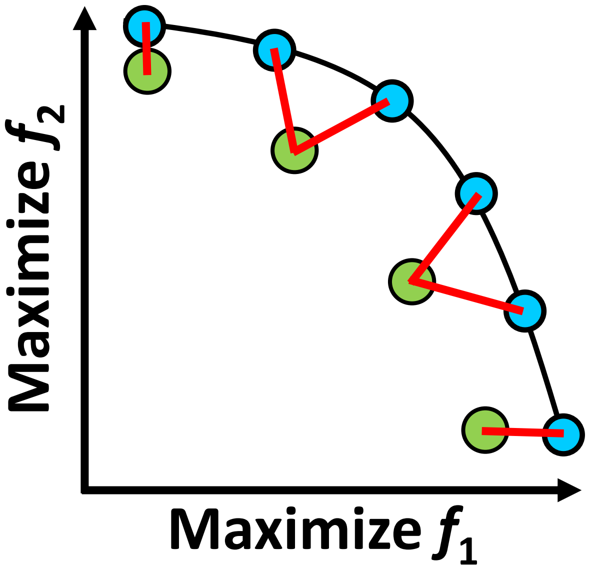
Performance Indicator: IGD

IGD: Inverted Generational Distance

Good IGD needs good convergence and good diversity.

==> The IGD can evaluate the overall performance.





Due to poor diversity,
IGD is not very good.

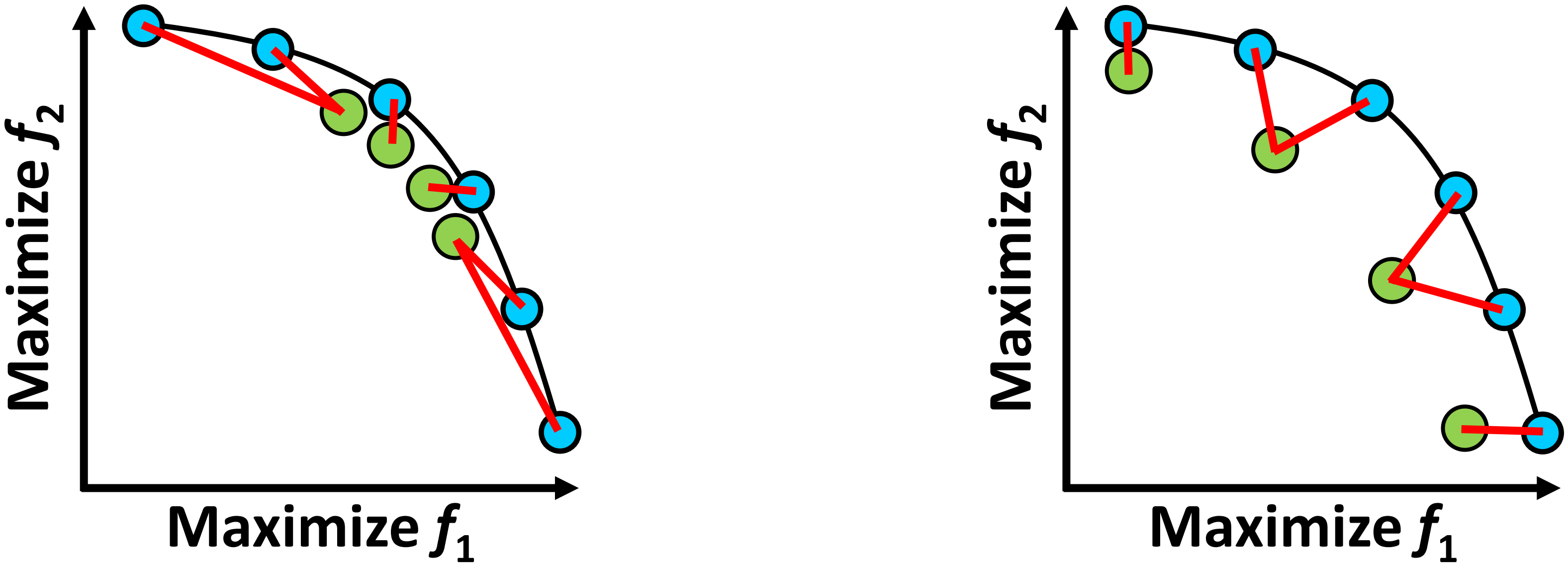
Due to poor convergence
IGD is not very good.

Performance Indicator: IGD

IGD: Inverted Generational Distance

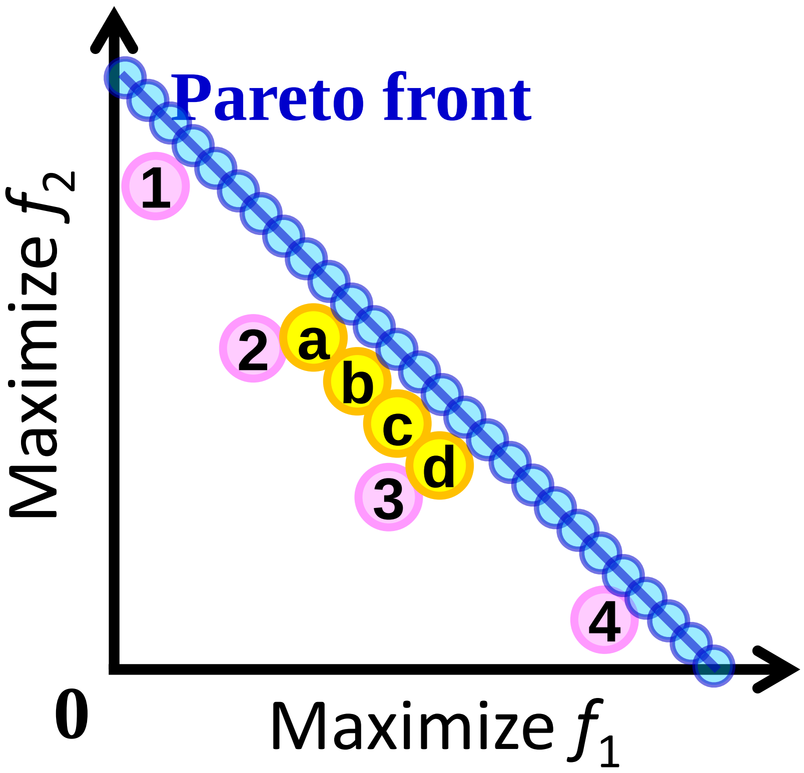
Good IGD needs good convergence and good diversity.

==> The IGD can evaluate the overall performance.

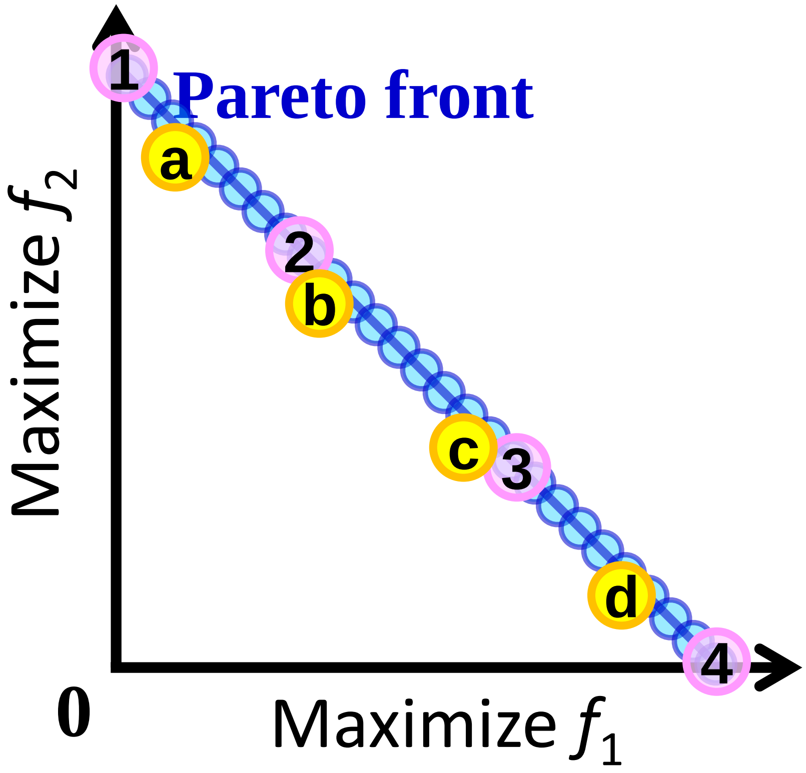


Question: Is IGD a Pareto compliant indicator ?
Your answer:________.

Which is better based on IGD: Pink or Yellow ?

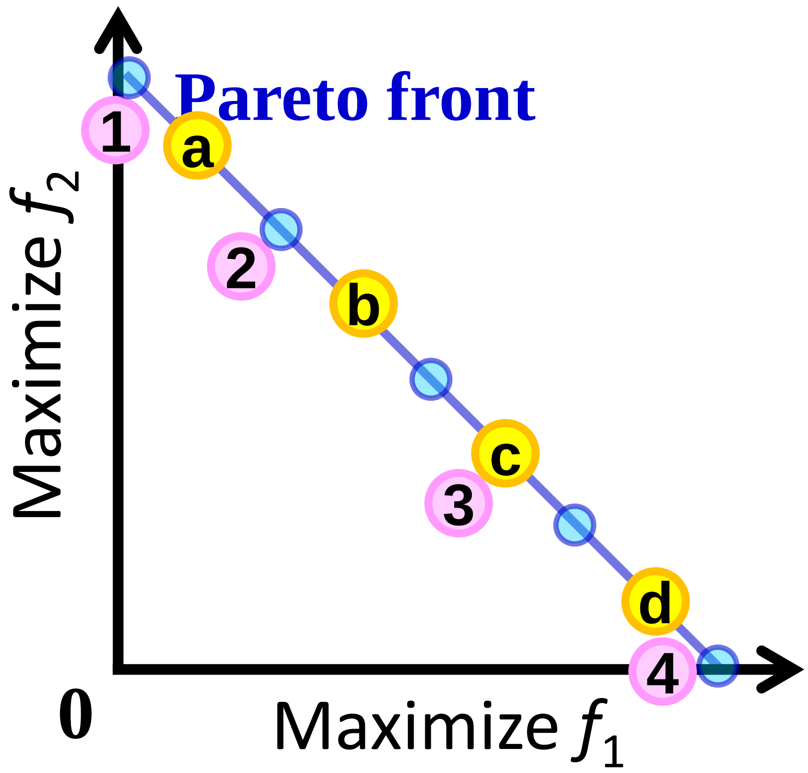


Example 1

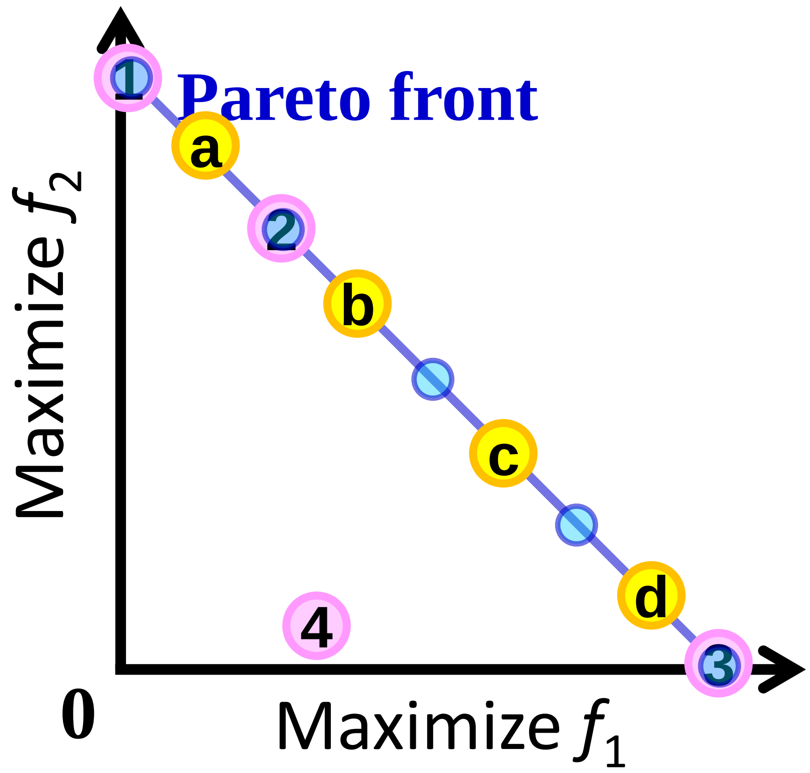


Example 2

Which is better based on IGD: Pink or Yellow ?



Example 1



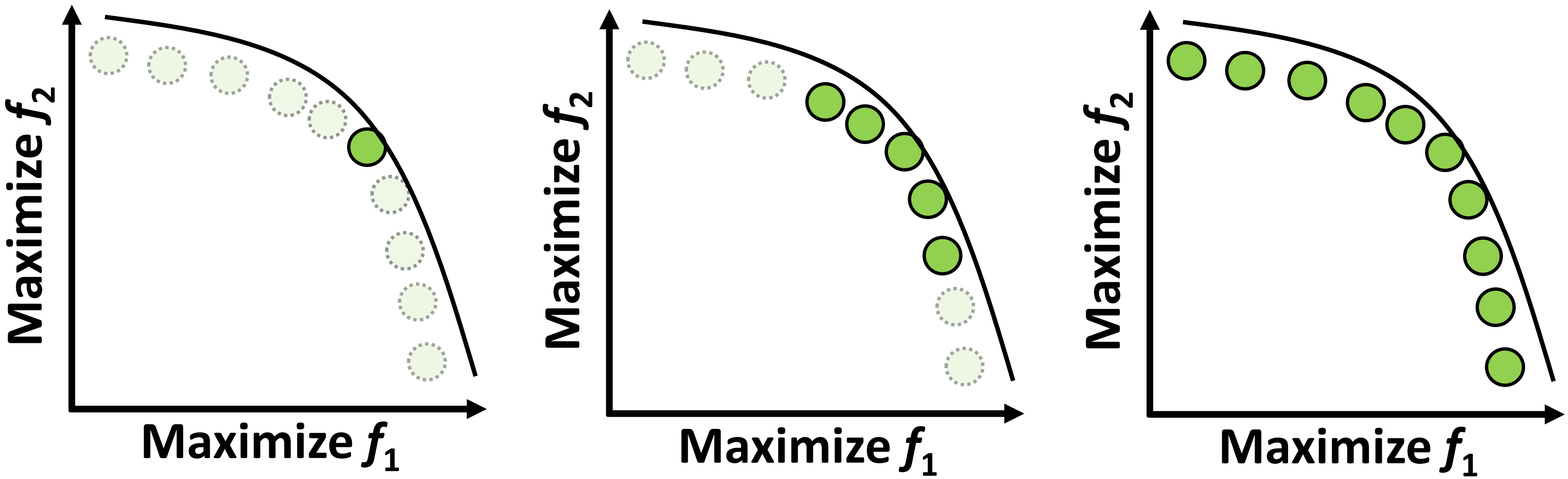
Example 2

Performance Indicator: IGD

IGD: Inverted Generational Distance

Good IGD needs good convergence and good diversity.

==> The IGD can evaluate the overall performance.

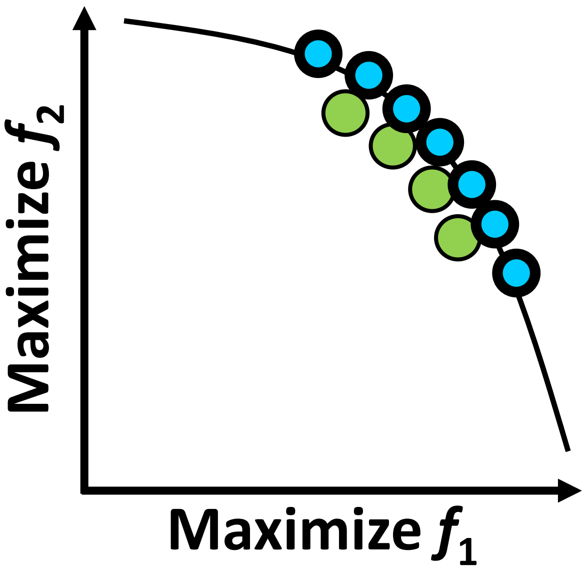


Worst IGD | **2nd Best IGD** | **Best IGD**

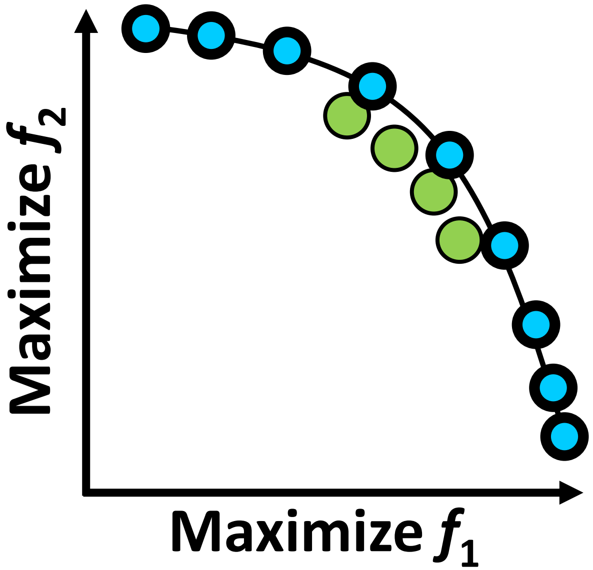
Performance Indicator: IGD

IGD: Inverted Generational Distance

Evaluation results strongly depend on reference points.



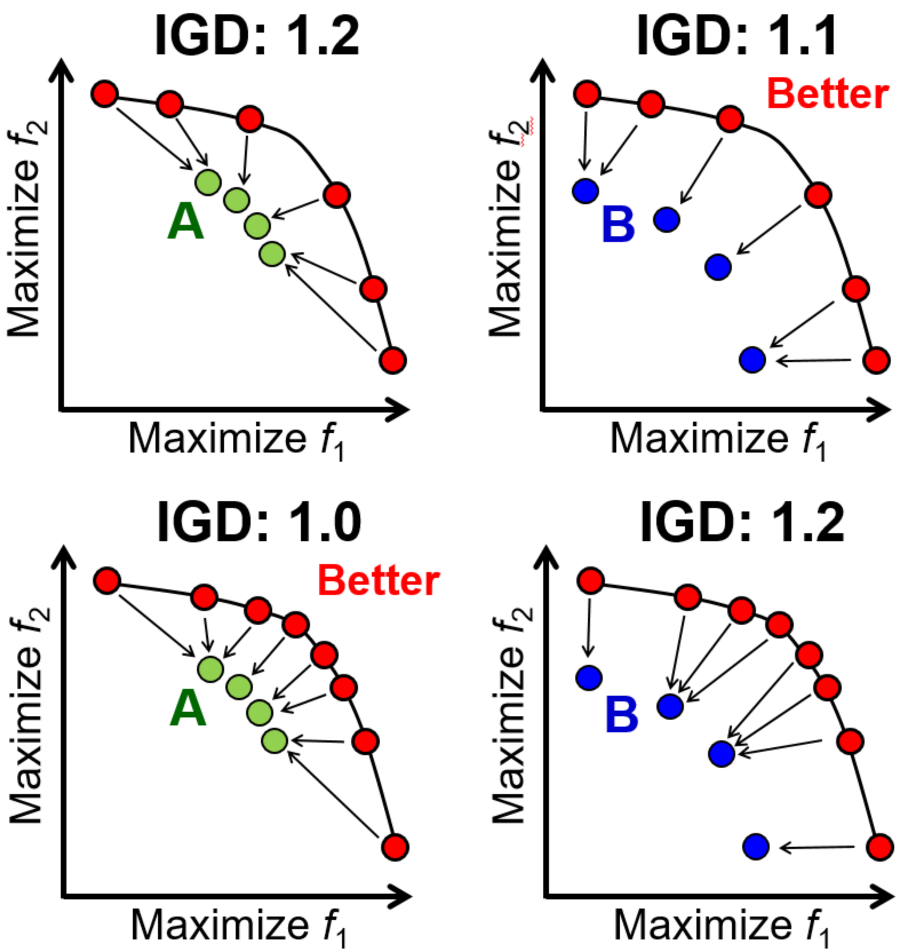
Good evaluation result



Poor evaluation result

IGD: Inverted Generational Distance

Evaluation results strongly depend on reference points.

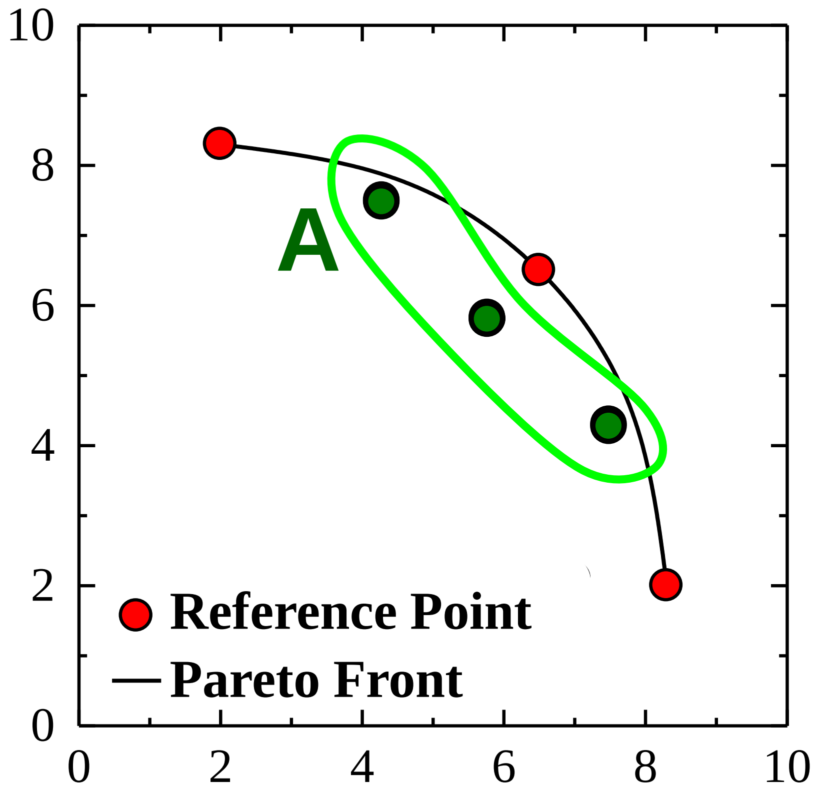


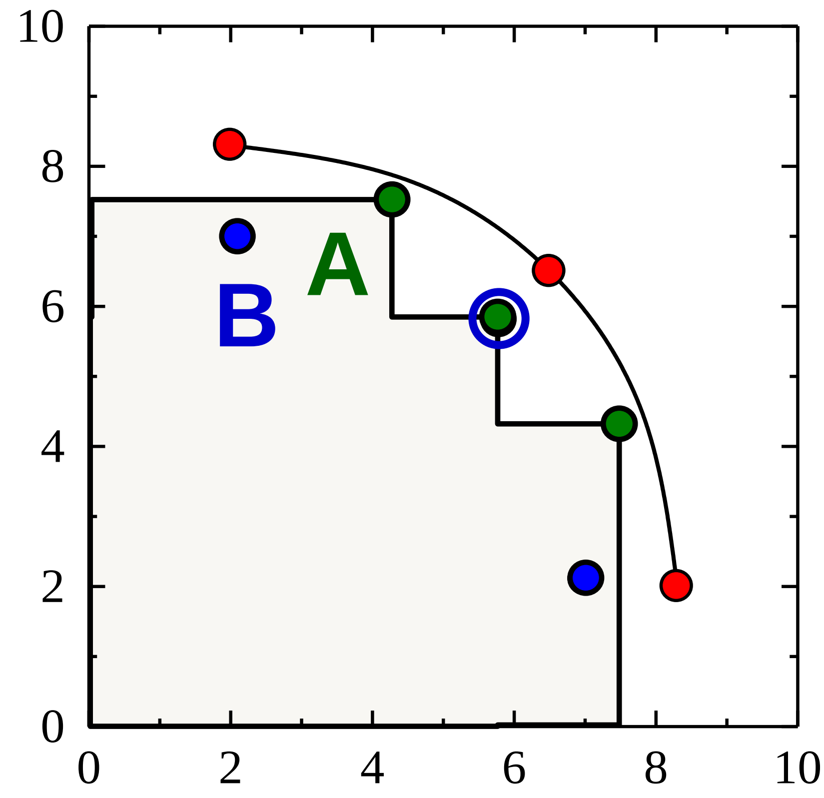
For many-objective problems, it is not easy to appropriately specify the reference point set for IGD calculation.

H. Ishibuchi et al., Reference point specification in inverted generational distance for triangular linear Pareto front, *IEEE TEVC* (2018).

IGD is not Pareto Compliant

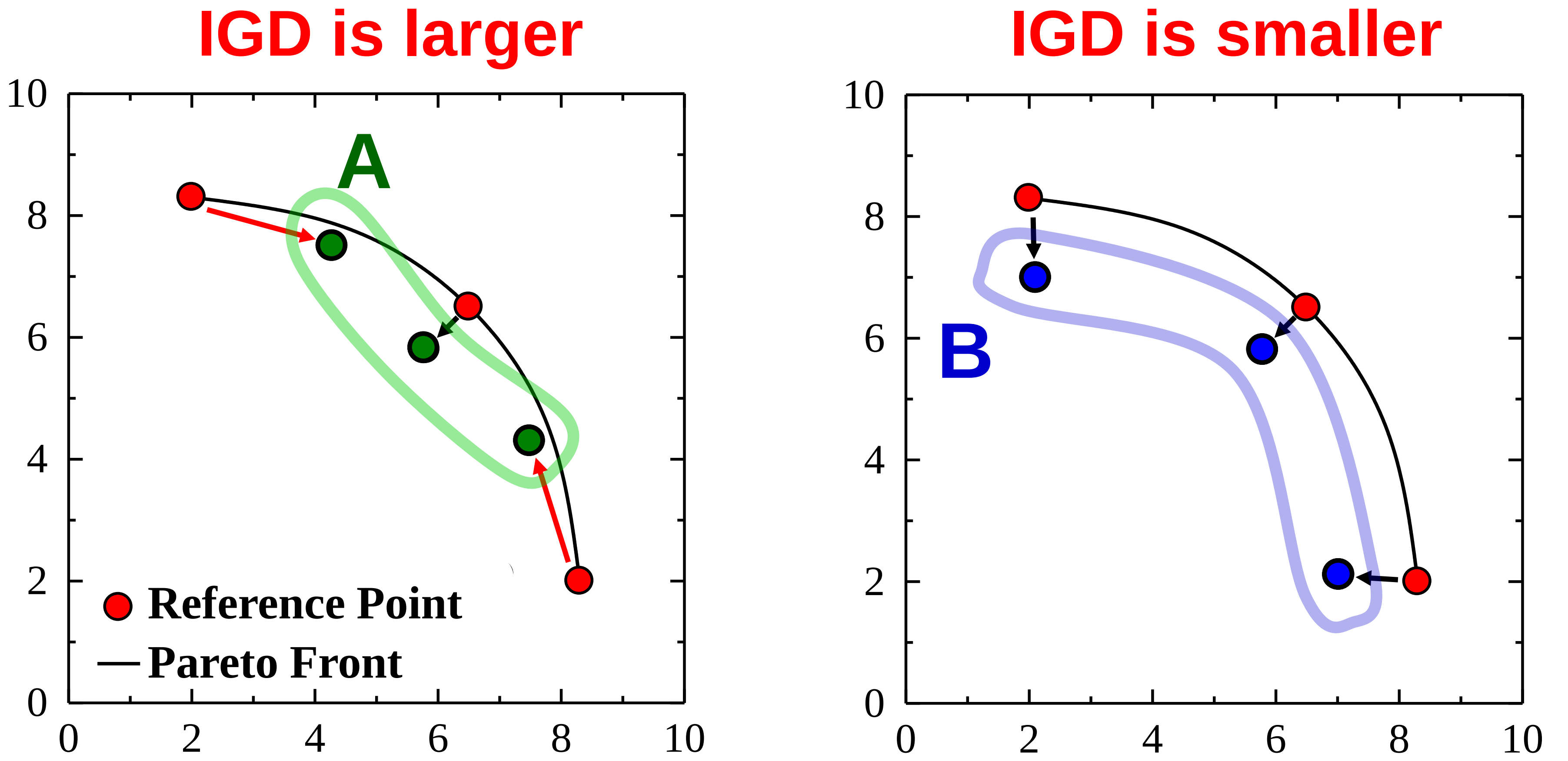
Solution set A dominates B





IGD is not Pareto Compliant

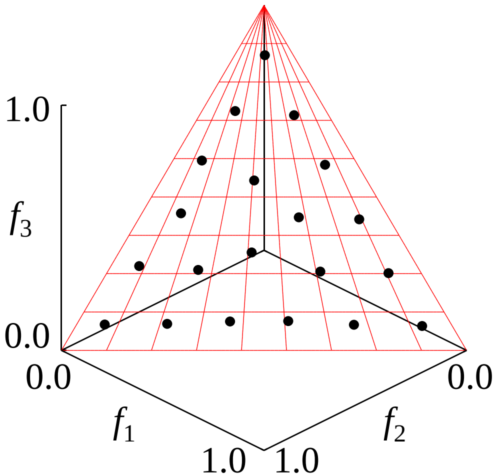
Solution set B is evaluated as being better than A



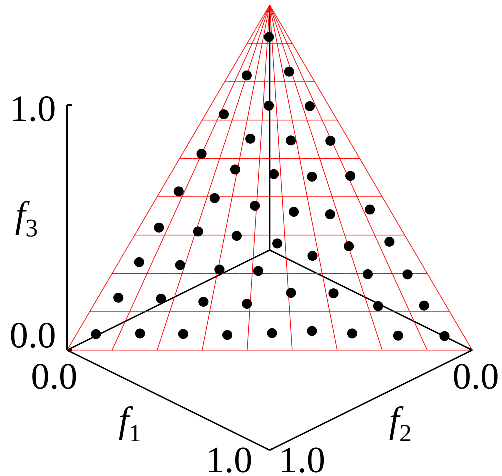
A dominates B (A is better than B).

However, B is evaluated as being better than A.

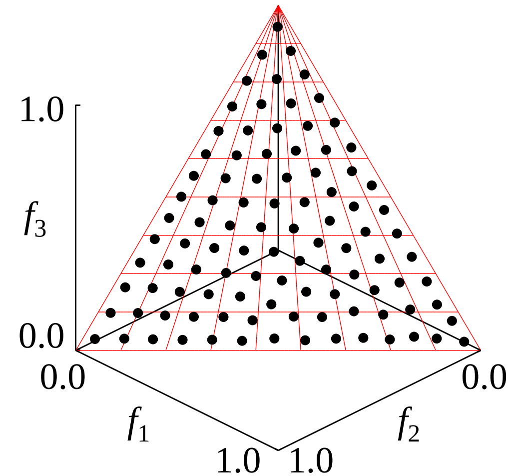
Optimal Distributions for IGD are not always intuitive



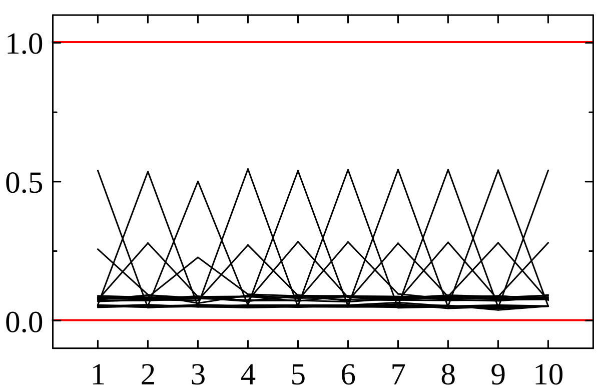
Population size 20



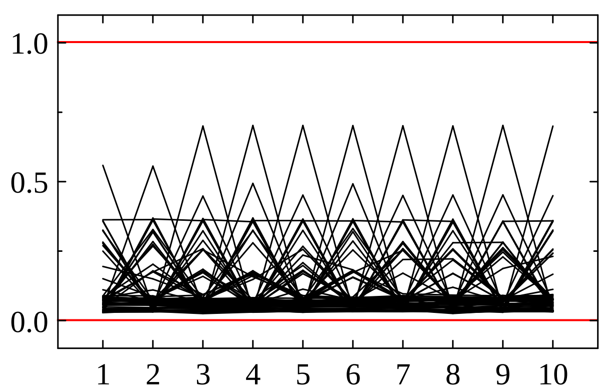
Population size 50



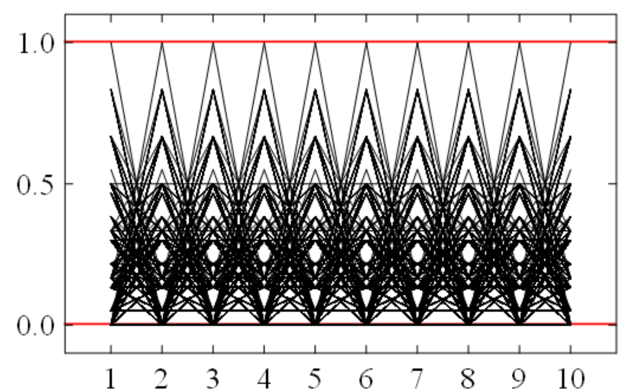
Population size 100



Population size 20



Population size 100



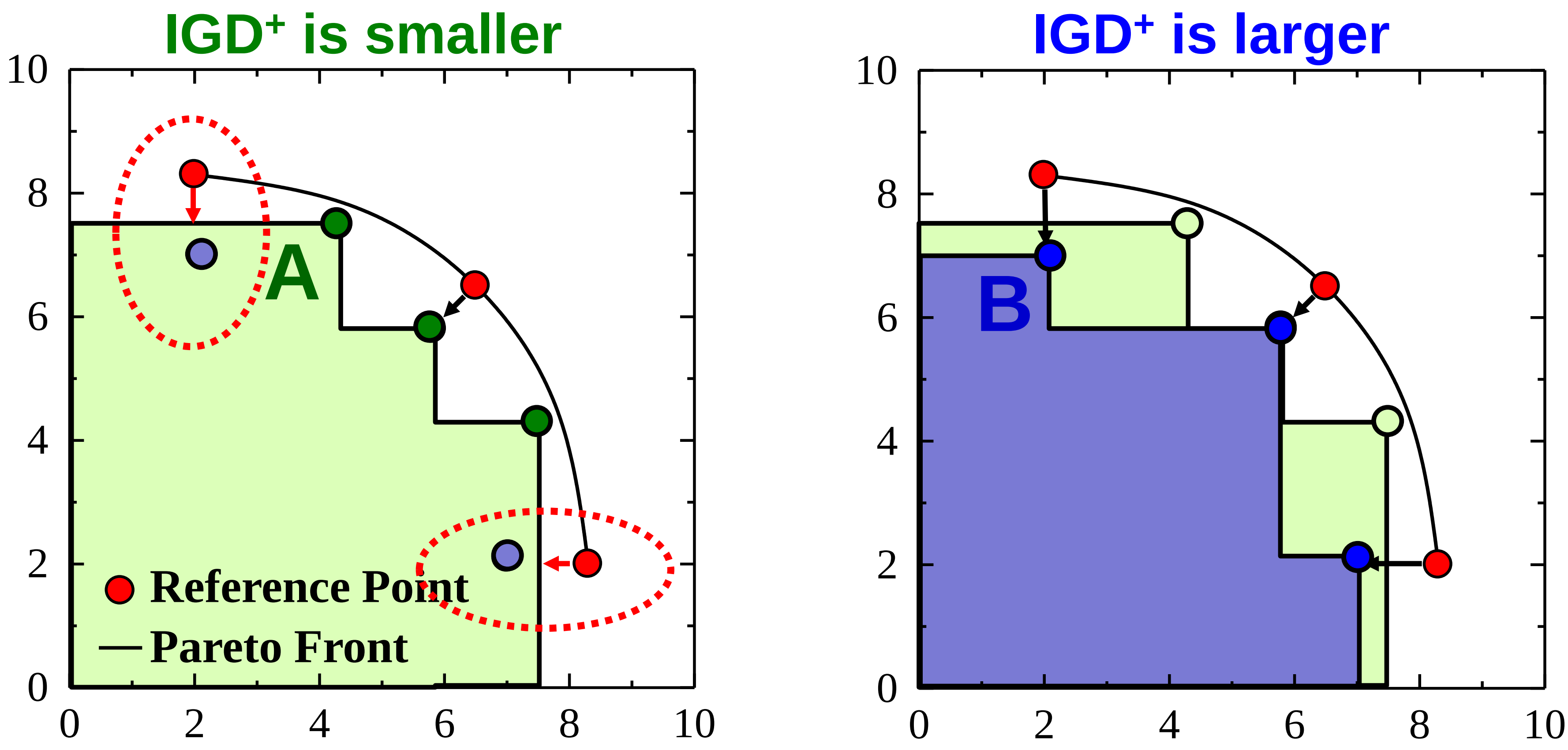
Reference point set

H. Ishibuchi et al., Reference point specification in inverted generational distance for triangular linear Pareto front, *IEEE TEVC* (2018).

IGD$^+$ (Inverted Generational Distance Plus)

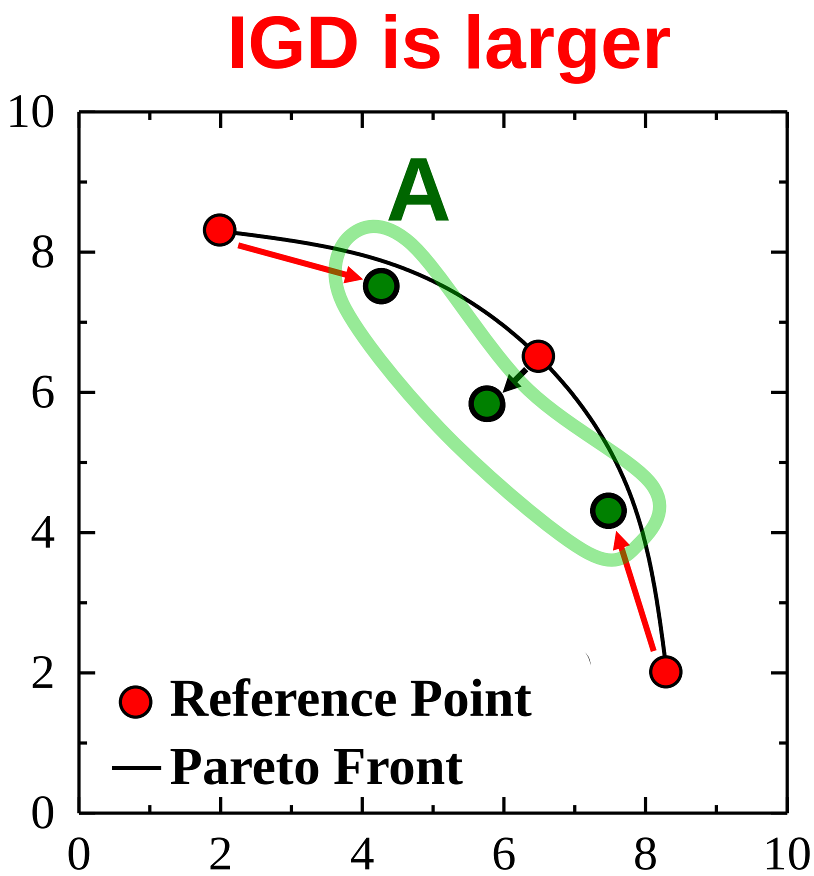
(Ishibuchi et al., EMO 2015, GECCO 2015)

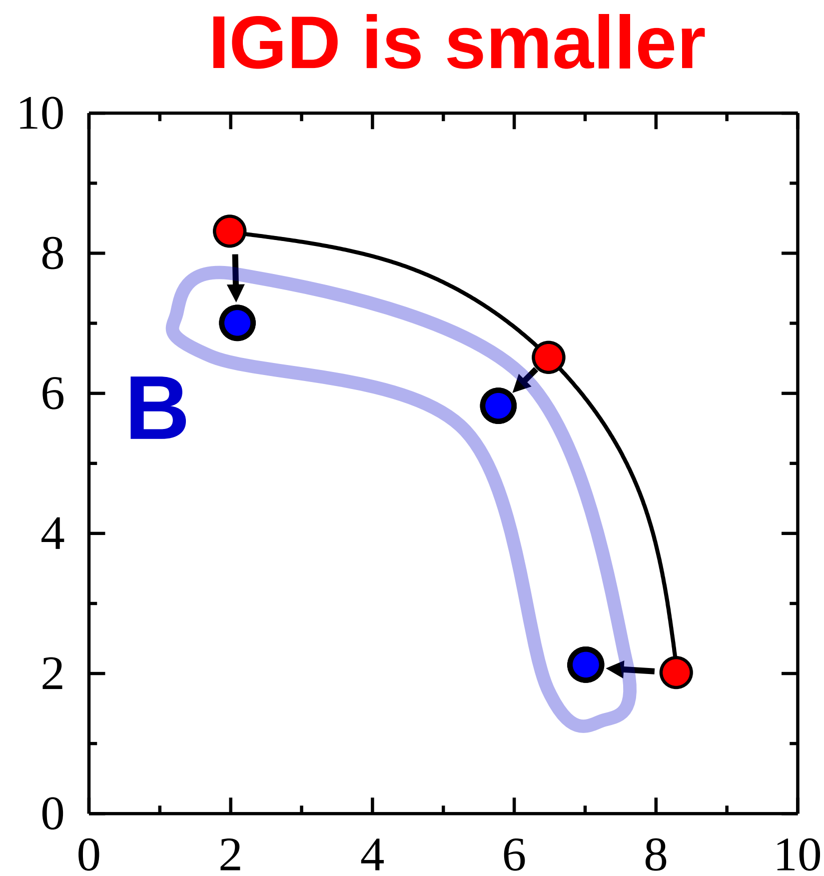
The calculation is from each reference point to the dominated region by the obtained solution set.



IGD is not Pareto Compliant

Solution set B is evaluated as being better than A





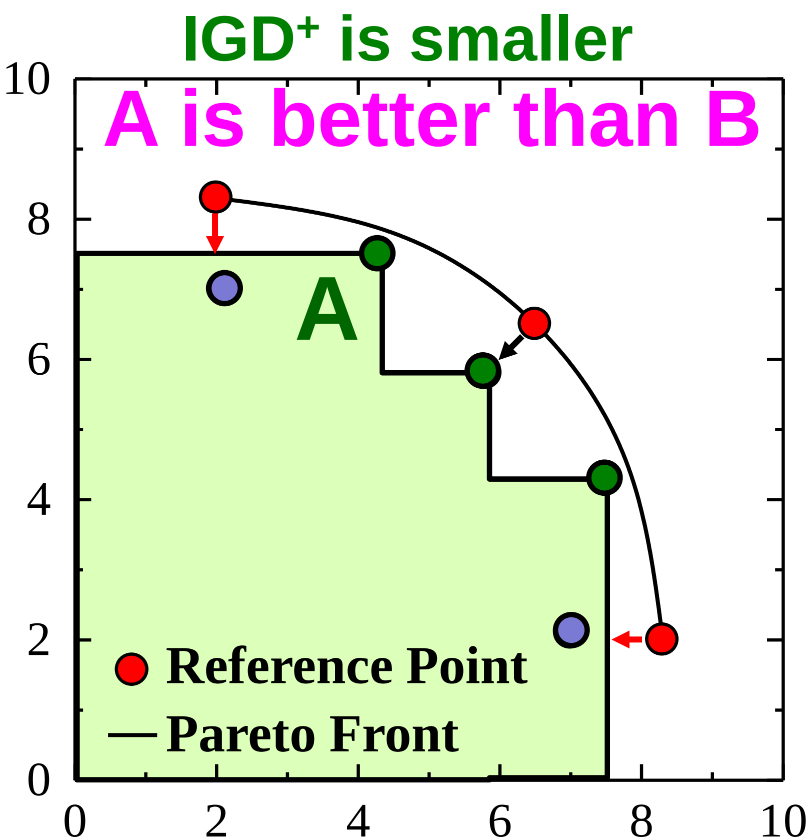
A dominates B (A is better than B).

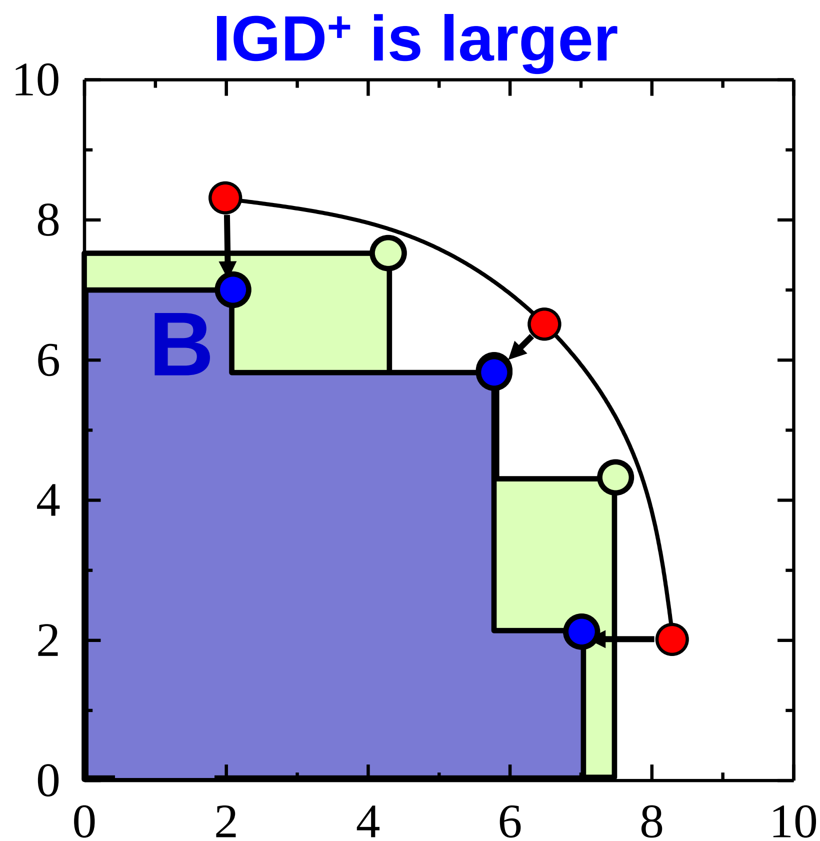
However, B is evaluated as being better than A.

Weak Pareto Compliance

IGD$^+$ is not inconsistent with the Pareto dominance relation between solution sets.

IGD$^+$ is a weakly Pareto compliant indicator.





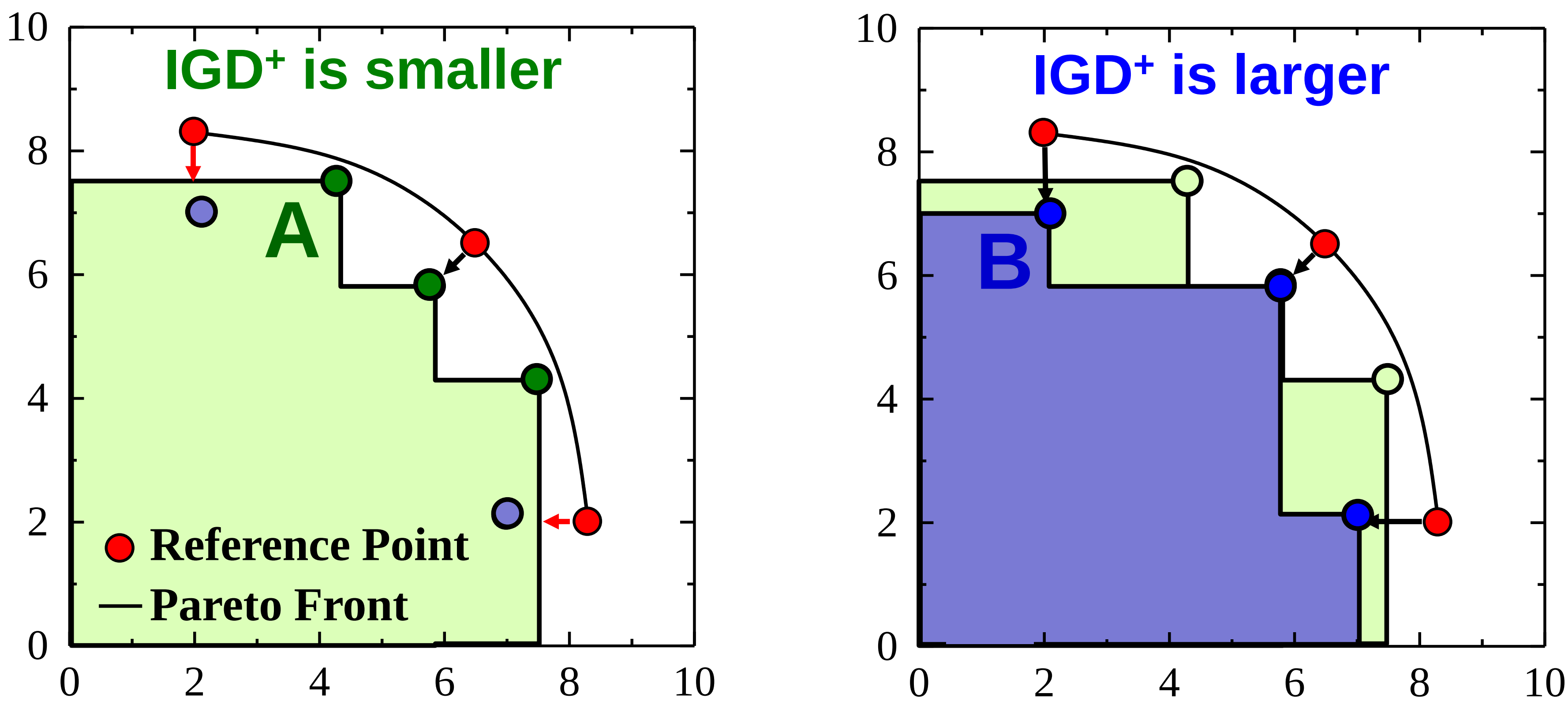
Weak Pareto Compliance

Pareto Compliance

$$B \lhd A \;\Rightarrow\; IGD^{+}(B) > IGD^{+}(A) \quad \text{(This does not hold)}$$

Weak Pareto Compliance.

$$B \lhd A \;\Rightarrow\; IGD^{+}(B) \geq IGD^{+}(A) \quad \text{(This holds)}$$



Weak Pareto Compliance

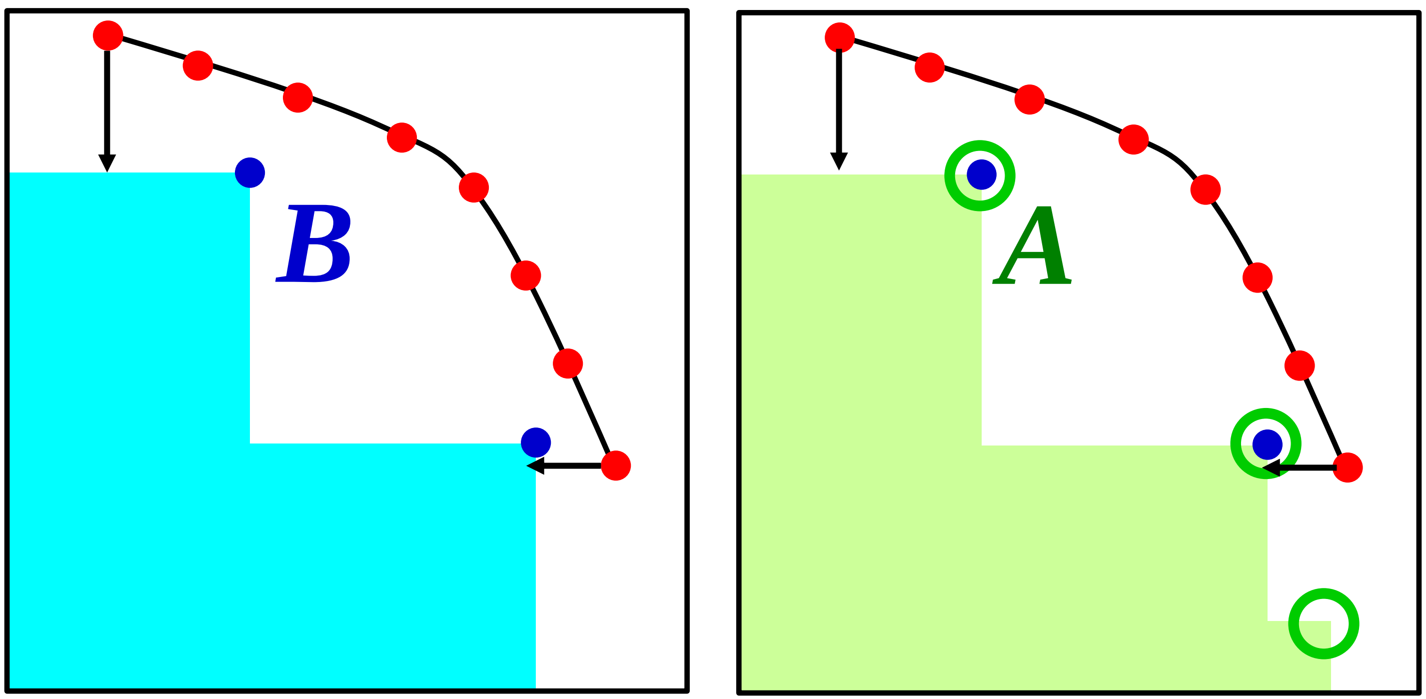
Pareto Compliance

$$B \lhd A \;\Rightarrow\; IGD^{+}(B) > IGD^{+}(A) \quad \text{(This does not hold)}$$

Weak Pareto Compliance.

$$B \lhd A \;\Rightarrow\; IGD^{+}(B) \geq IGD^{+}(A) \quad \text{(This holds)}$$

● Solution set B ○ Solution set A



$IGD^{+}(A)$

$= IGD^{+}(B)$

Performance Indicators

in jMetal 5 Web Site

Welcome to the **jMetal 5** Web Site

jMetal is ...

jMetal stands for **Metaheuristic Algorithms in Java**, and it is an object-oriented Java-based framework for multi-objective optimization with metaheuristics.

Summary of features

- Multi-objective algoritms: NSGA-II, SPEA2, PAES, PESA-II, OMOPSO, MOCell, AbYSS, MOEA/D, GDE3, IBEA, SMPSO, SMPSOhv, SMS-EMOA, MOEA/D-STM, MOCHC, MOMBI, MOMBI-II, NSGA-III, WASF-GA, GWASF-GA

- Quality indicators: hypervolume, spread, generational distance, inverted generational distance, inverted generational distance plus, additive epsilon.

Overall Performance Indicators

IGD (inverted generational distance) indicator

IGD^{+} (inverted generational distance plus) indicator

-indicator (Additive epsilon)

HV (hypervolume) indicator

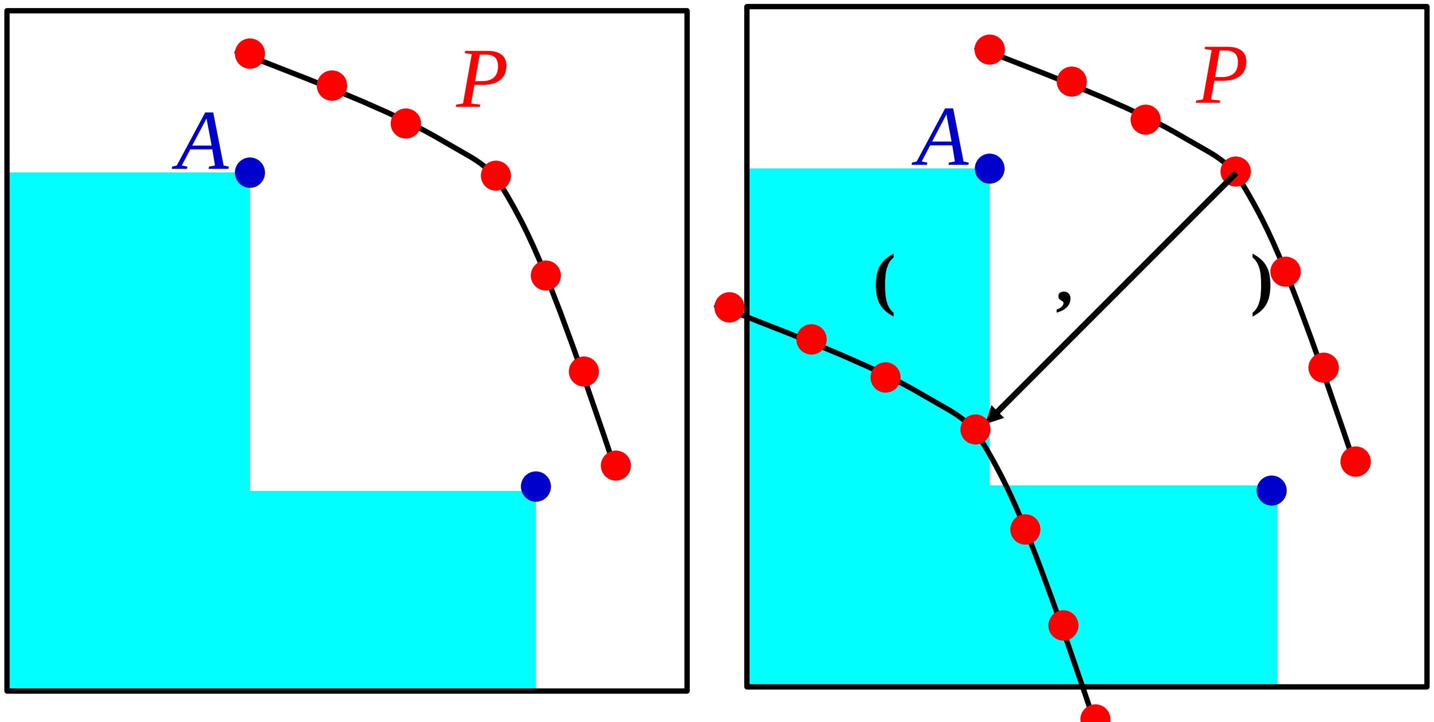
-indicator (Additive epsilon)

Solution Set $A = \{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots\}$, $\boldsymbol{a}_i = (a_{i1}, \ldots, a_{im})$ for $i = 1, 2, \ldots,$

Reference Set $P = \{\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots\}$, $\boldsymbol{p}_i = (p_{i1}, \ldots, p_{im})$ for $i = 1, 2, ..$

$$\boxed{I_{\varepsilon+}(A, P) = \min\{\varepsilon \in \Re \mid P_{\varepsilon-} \preceq A\}}$$

$P \quad = \{\boldsymbol{p}_1 \quad , \boldsymbol{p}_2 \quad , \ldots\}, \quad \boldsymbol{p}_i \quad = (p_{i1} \quad , \ldots, p_{im}$
)



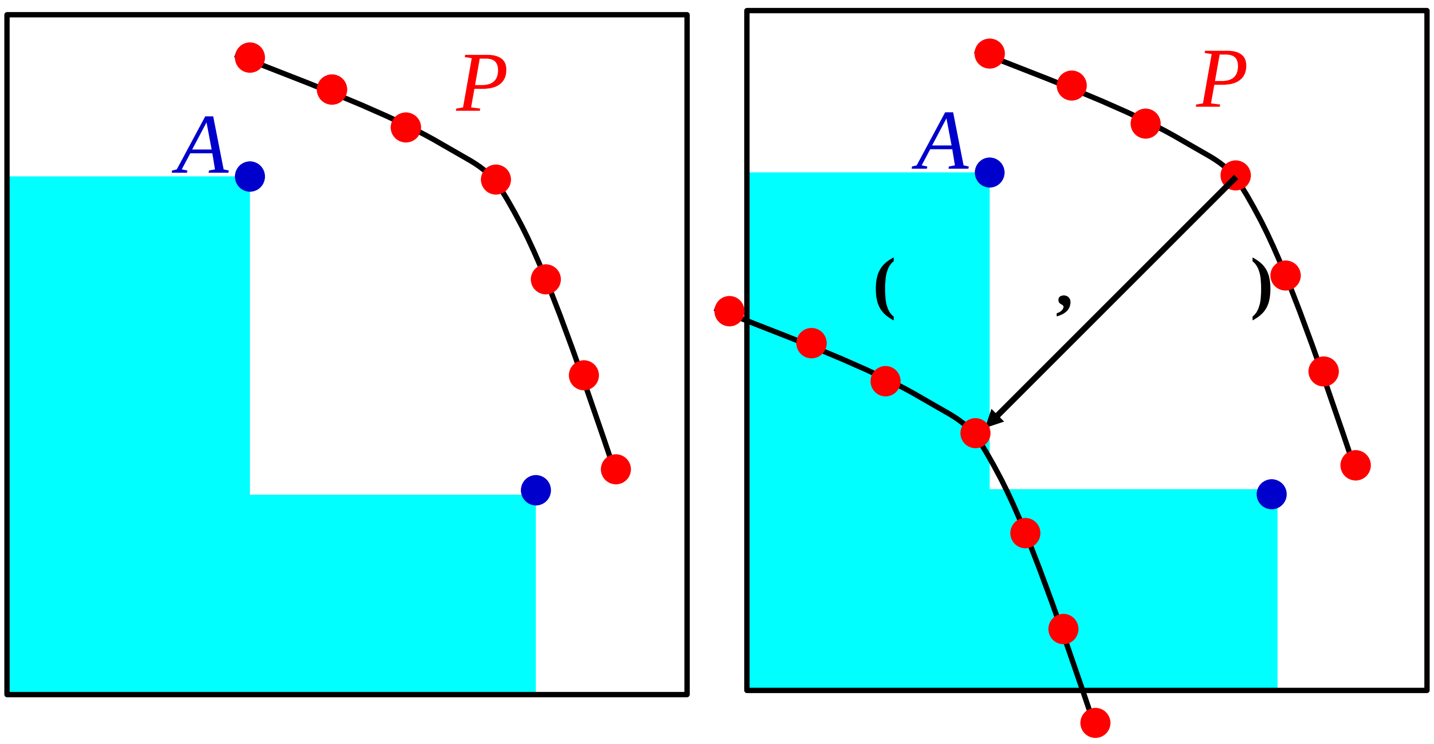
for $i = 1, 2, \ldots,$

-indicator (Additive epsilon)

$$I_{\varepsilon+}(A,P) = \min\{\varepsilon \in \Re \mid P_{\varepsilon-} \preceq A\} \quad P \quad = \{\boldsymbol{p}_1 \quad , \boldsymbol{p}_2 \quad , \ldots\}$$

Basic Idea:

The minimum change (the minimum shift) of *P* so that the solution set *A* weakly dominates the reference set *P*.



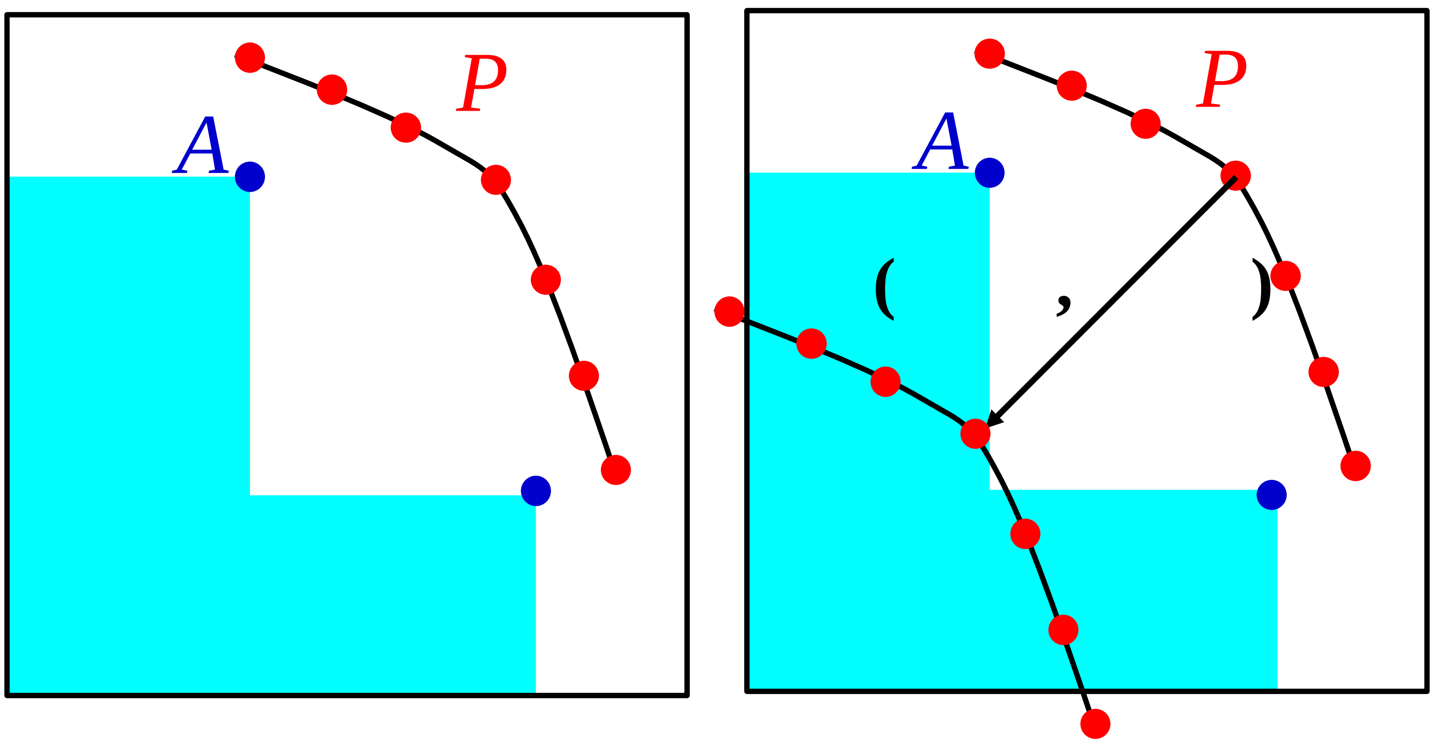
-indicator (Additive epsilon)

$$I_{\varepsilon+}(A,P) = \min\{\varepsilon \in \Re \mid P_{\varepsilon-} \preceq A\} \quad P = \{\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots\}$$

Basic Idea:

The minimum change (the minimum shift) of *P* so that the solution set *A* weakly dominates the reference set *P*.

All reference points in *P* should be moved into the dominated region by the solution set *A*.



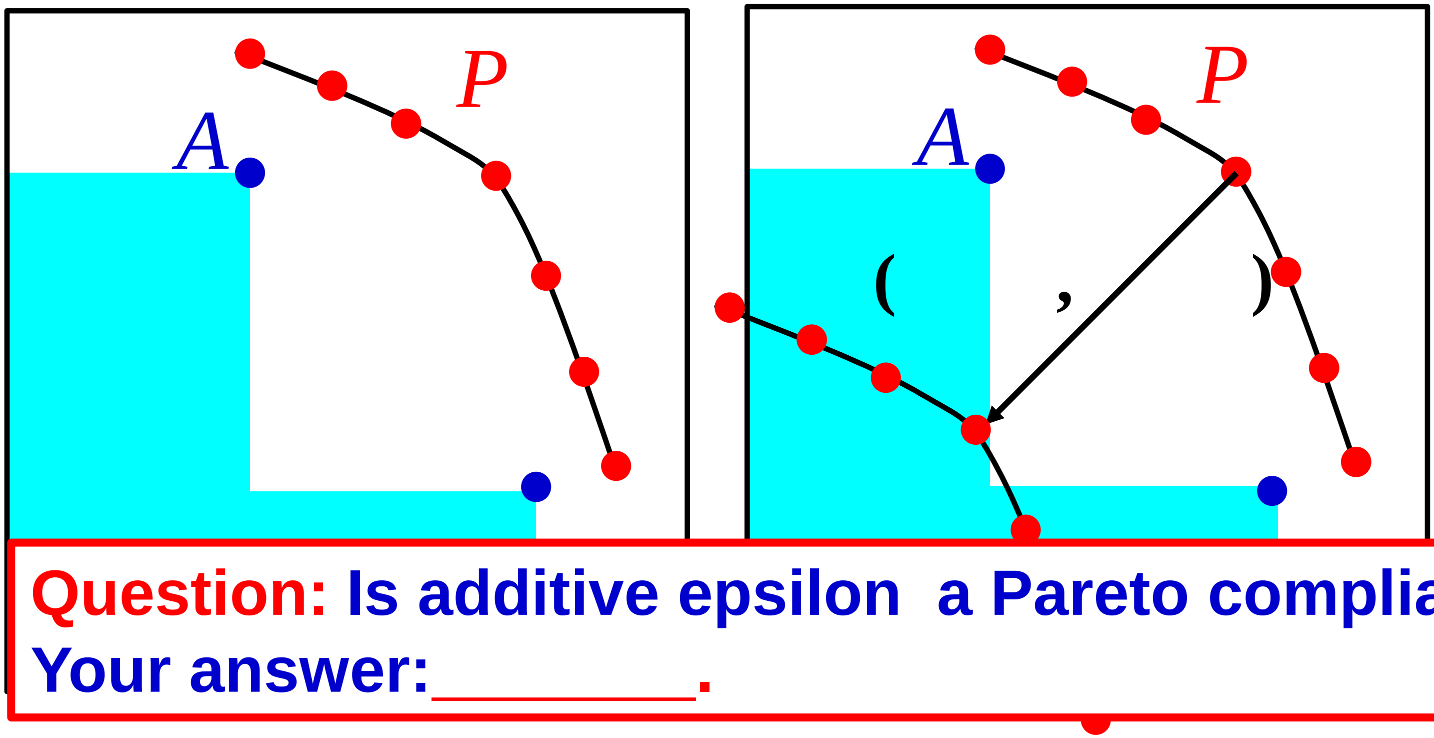
-indicator (Additive epsilon)

$$I_{\varepsilon+}(A,P) = \min\{\varepsilon \in \Re \mid P_{\varepsilon-} \preceq A\} \quad P \quad = \{\boldsymbol{p}_1 \quad , \ \boldsymbol{p}_2 \quad , \ldots\}$$

Basic Idea:

The minimum change (the minimum shift) of *P* so that the solution set *A* weakly dominates the reference set *P*.

All reference points in *P* should be moved into the dominated region by the solution set *A*.



Question: **Is additive epsilon a Pareto compliant indicator ?**

Your answer:________.

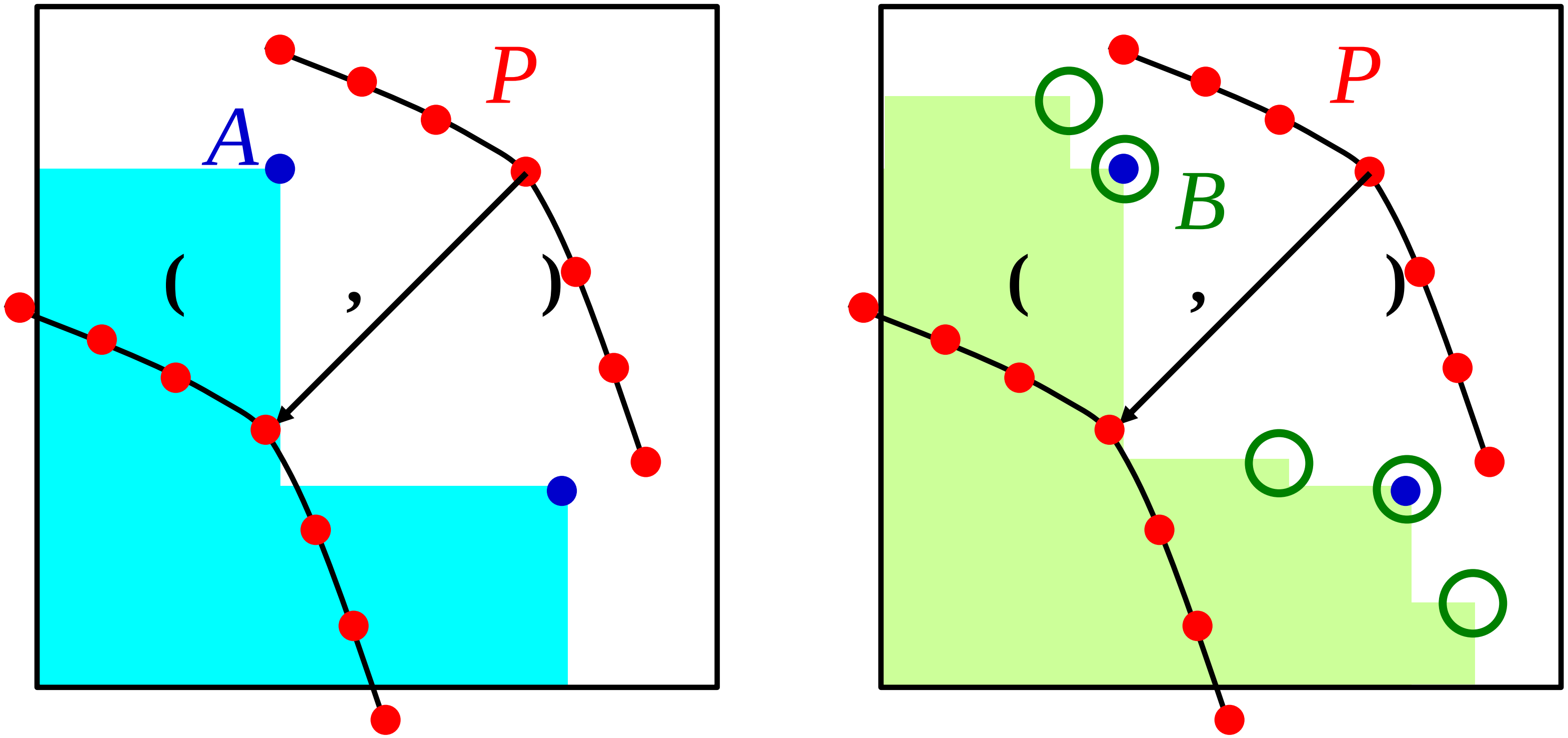
-indicator (Additive epsilon)

$$I_{\varepsilon+}(A,P) = \min\{\varepsilon \in \Re \mid P_{\varepsilon-} \preceq A\} \quad P \quad = \{\boldsymbol{p}_1 \quad , \boldsymbol{p}_2 \quad , \ldots\}$$

Basic Idea:

The minimum change (the minimum shift) of P so that the solution set A weakly dominates the reference set P.

All reference points in P should be moved into the dominated region by the solution set A. ==> **Weak Pareto Compliant.**



-indicator (Additive epsilon)

$$I_{\varepsilon+}(A,P) = \min\{\varepsilon \in \Re \mid P_{\varepsilon-} \preceq A\} \quad P \quad = \{\boldsymbol{p}_1 \quad , \boldsymbol{p}_2 \quad , \ldots\}$$

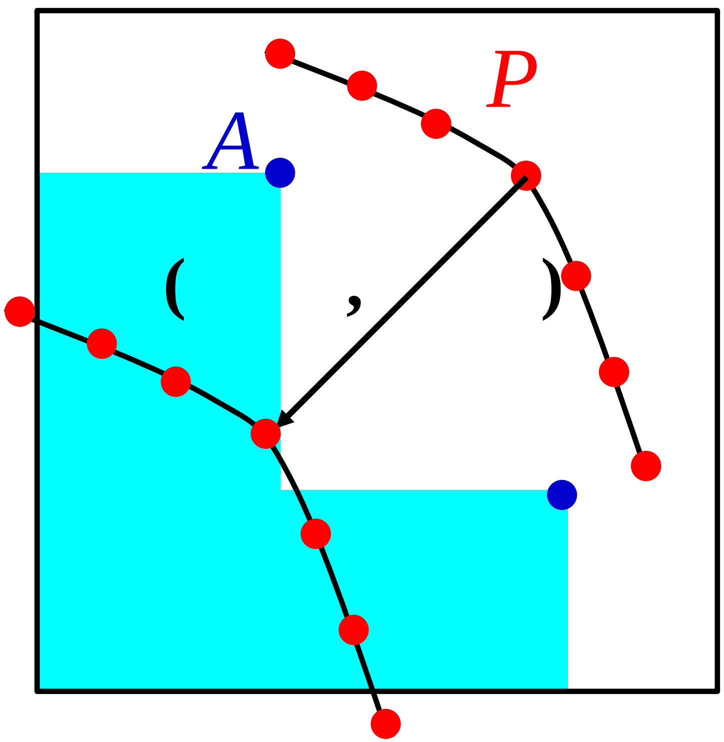
Some difficulties:

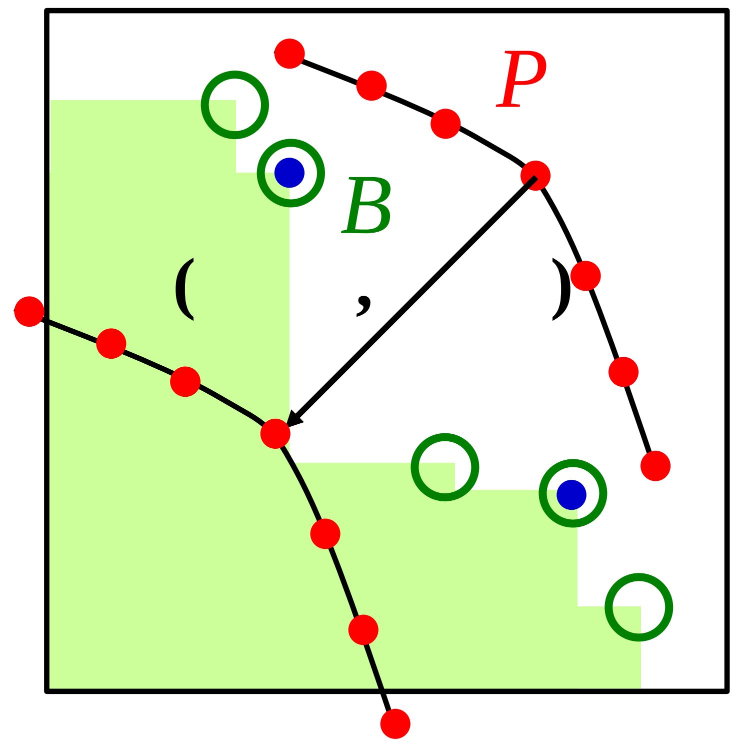
(i) I $_{+}$ depends on the reference set (as IGD and IGD^{+})

(ii) Weak Pareto compliant (not Pareto compliant as IGD^{+})

==> Better solution sets can be evaluated as the same.

This is severe since the calculation is based on only a single point.

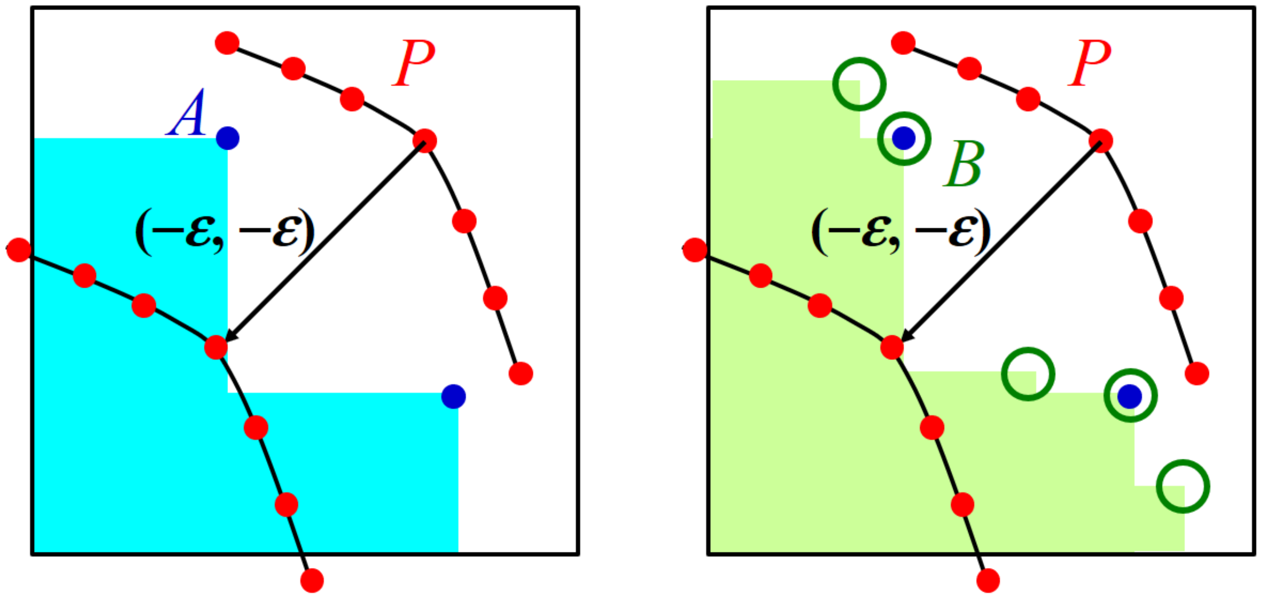




Computational Experiments

Step 1: Solution sets A and B ($A \lhd B$) are generated.
A and B are intermediate populations in a run of an EMO algorithm.

Step 2: It is examined which is hold among
- Pareto compliant: $I(A) > I(B)$,
- Weakly Pareto compliant: $I(A) = I(B)$,
- Inconsistent: $I(A) < I(B)$.



H. Ishibuchi, H. Masuda and Y. Nojima: A Study on Performance Evaluation Ability of a Modified Inverted Generational Distance Indicator. GECCO 2015.

Table 1. Results on the 2193 pairs of ordered non-dominated solution sets of the two-objective DTLZ2 problem.

Indicator	Pareto Compliant $I(A) > I(B)$	Weakly Compliant $I(A) = I(B)$	Inconsistent $I(A) < I(B)$
GD	98.04%	0.00%	**1.96%**
IGD	97.95%	0.00%	**2.05%**
IGD^+	100.00%	0.00%	0.00%
Additive	91.24%	**8.97%**	0.00%

Table 2. Results on the 145 pairs of ordered non-dominated solution sets of the four-objective DTLZ2 problem.

Indicator	Pareto Compliant $I(A) > I(B)$	Weakly Compliant $I(A) = I(B)$	Inconsistent $I(A) < I(B)$
GD	97.93%	0.00%	**2.07%**
IGD	100.00%	0.00%	0.00%
IGD^+	100.00%	0.00%	0.00%
Additive	93.79%	**6.21%**	0.00%

Table 3. Results on the 47993 pairs of ordered non-dominated solution sets of the two-objective 500-item knapsack problem.

Indicator	Pareto Compliant $I(A) > I(B)$	Weakly Compliant $I(A) = I(B)$	Inconsistent $I(A) < I(B)$
GD	81.80%	0.00%	**18.20%**
IGD	77.22%	**17.58%**	**5.21%**
IGD^+	82.41%	**17.59%**	0.00%
Additive	14.83%	**85.17%**	0.00%

Table 4. Results on the 4704 pairs of ordered non-dominated solution sets of the four-objective 500-item knapsack problem.

Indicator	Pareto Compliant $I(A) > I(B)$	Weakly Compliant $I(A) = I(B)$	Inconsistent $I(A) < I(B)$
GD	98.15%	0.00%	**1.85%**
IGD	99.30%	**0.60%**	**0.11%**
IGD^+	99.36%	**0.64%**	0.00%
Additive	53.93%	**46.07%**	0.00%

Overall Performance Indicators

IGD (inverted generational distance) indicator

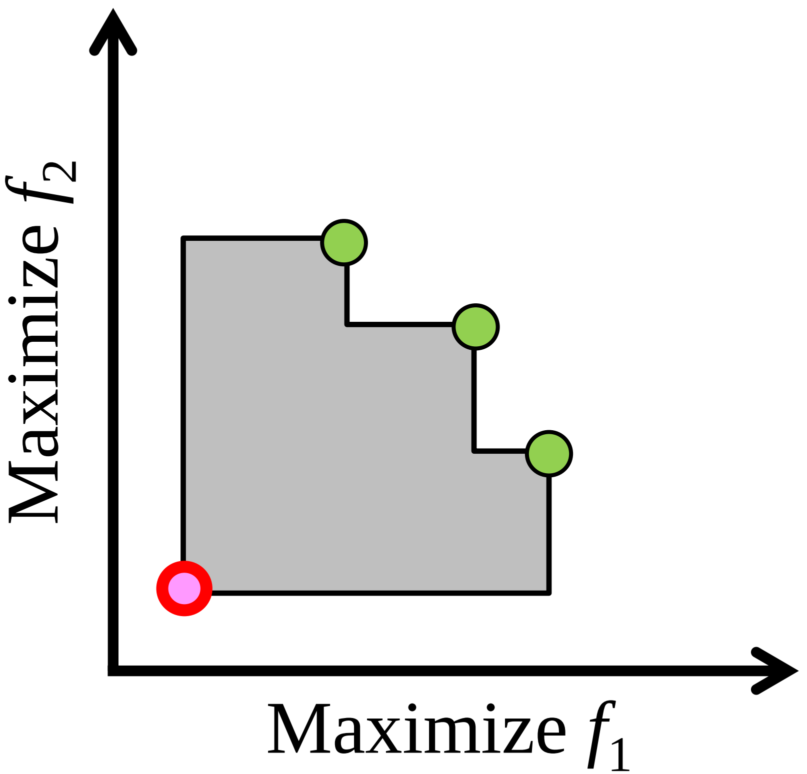
IGD$^+$ (inverted generational distance plus) indicator

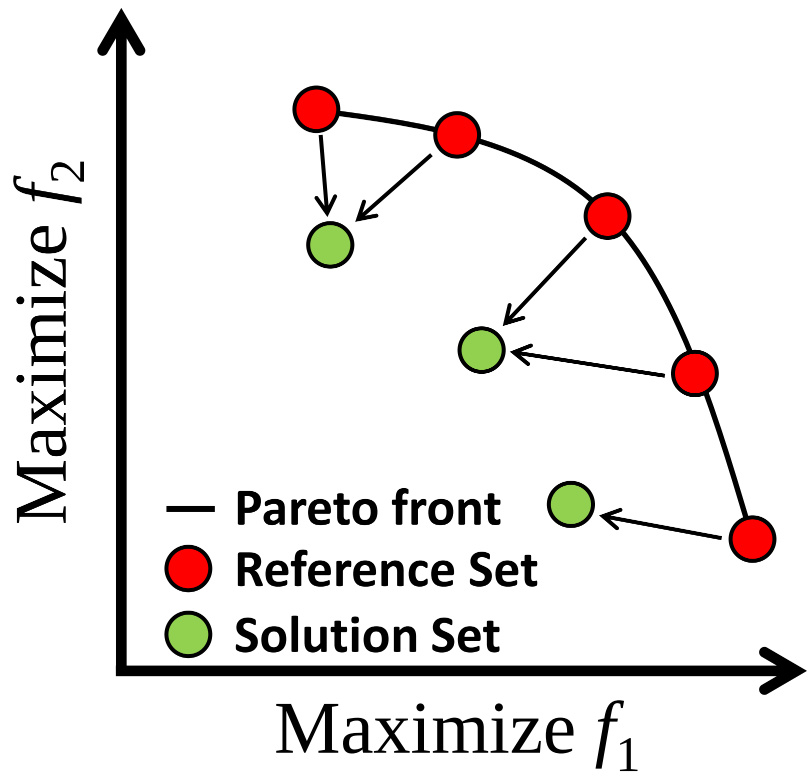
-indicator (Additive epsilon)

HV (hypervolume) indicator

Hypervolume

Hypervolume is the volume of the dominated region by the obtained solutions bound by the reference point.

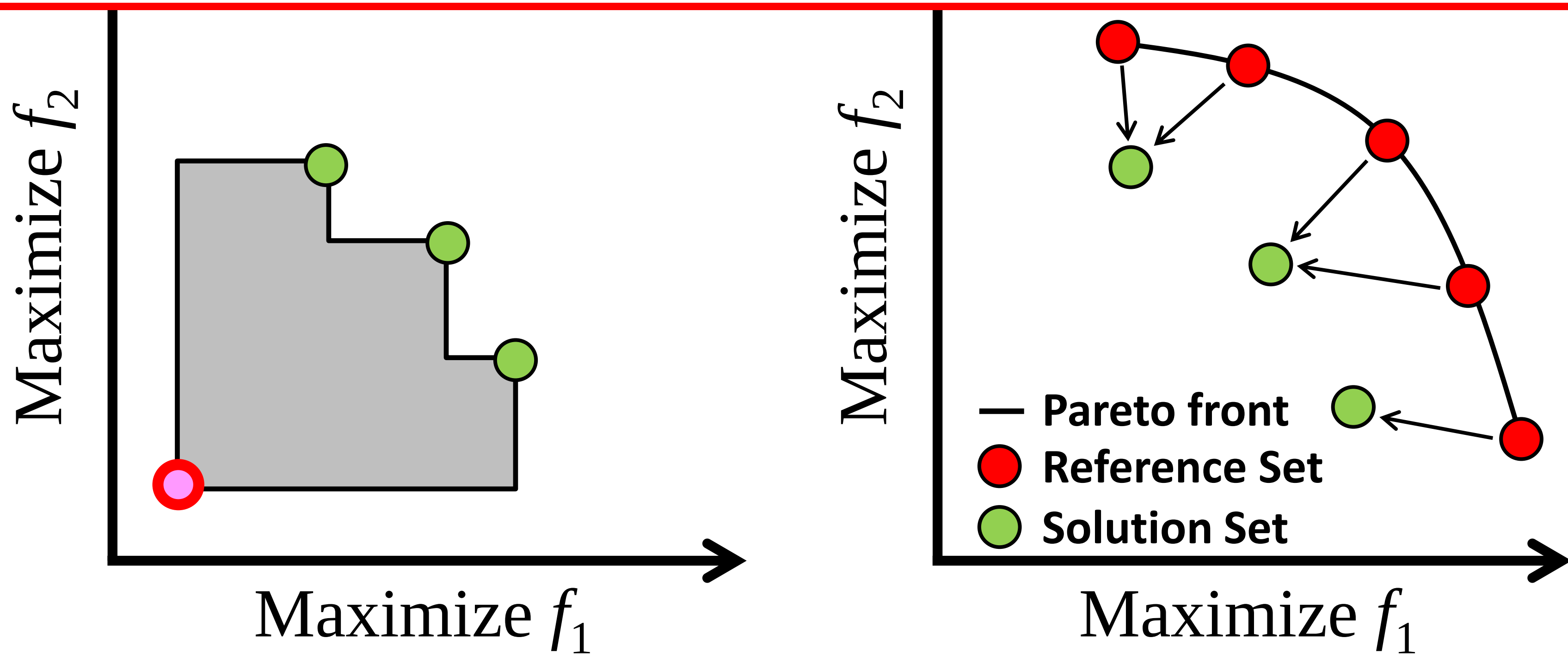




Hypervolume

Hypervolume is the volume of the dominated region by the obtained solutions bound by the reference point.

Question: Is hypervolume a Pareto compliant indicator ?
Your answer:________.

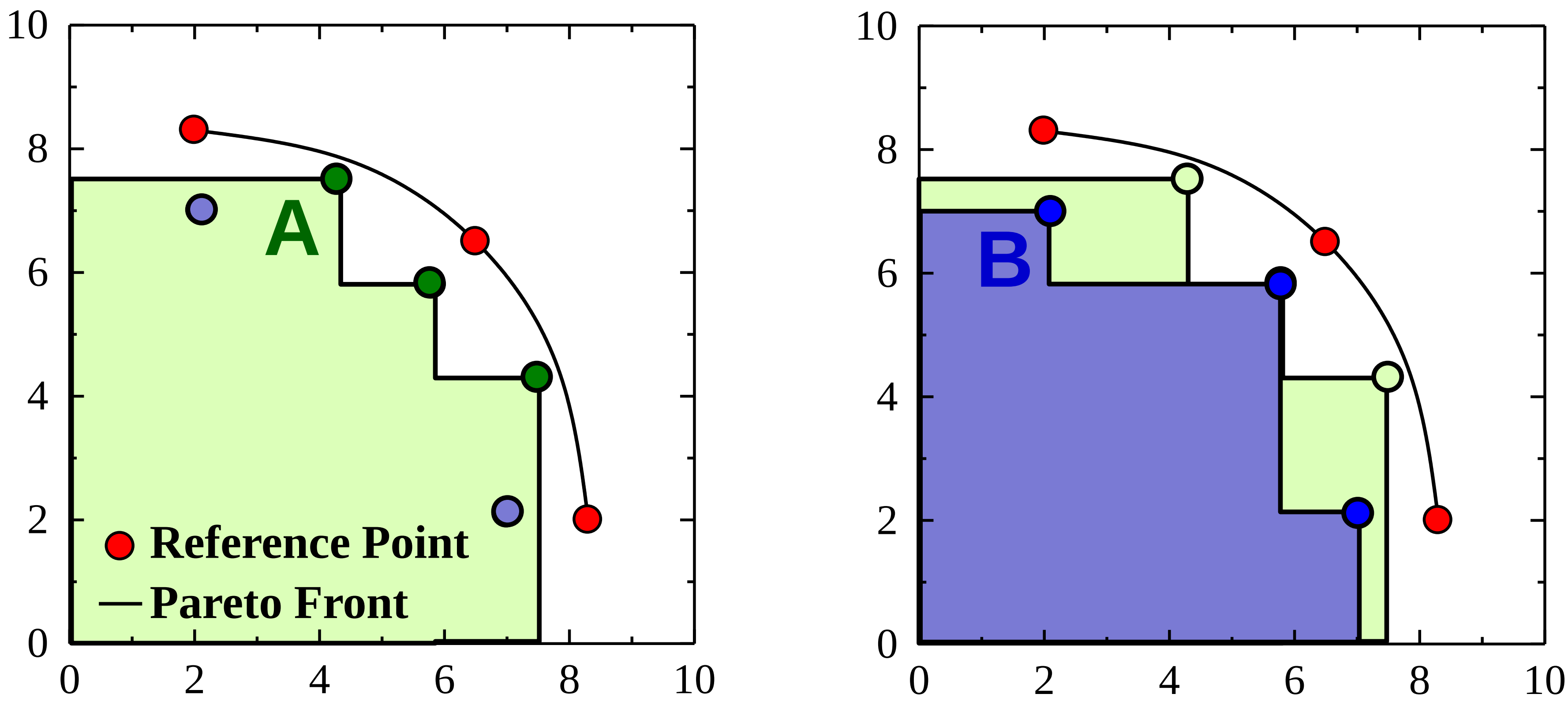


Theoretical Result

When solution set A is better than B

Hypervolume(A) > Hypervolume(B) always holds.

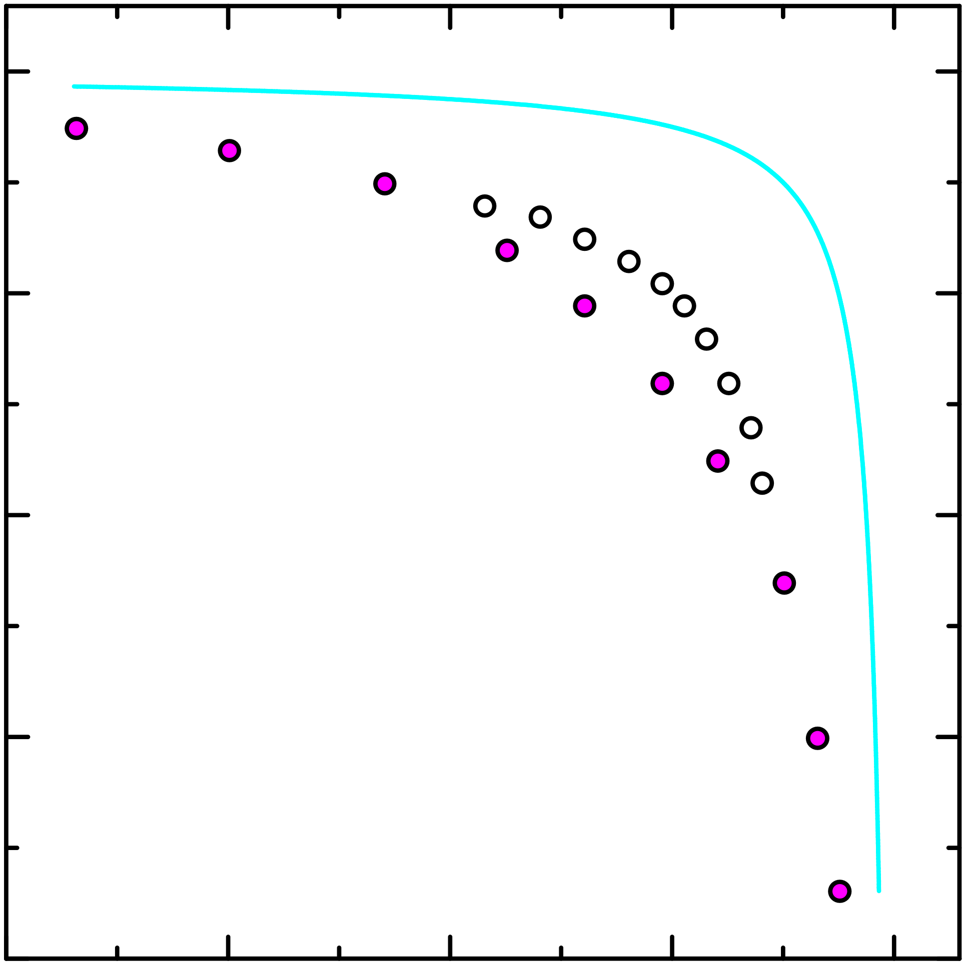
Well-Known Property: Hypervolume is Pareto compliant.



Hypervolume(A) > Hypervolume(B)

Two Solution Sets:

Which do you like?

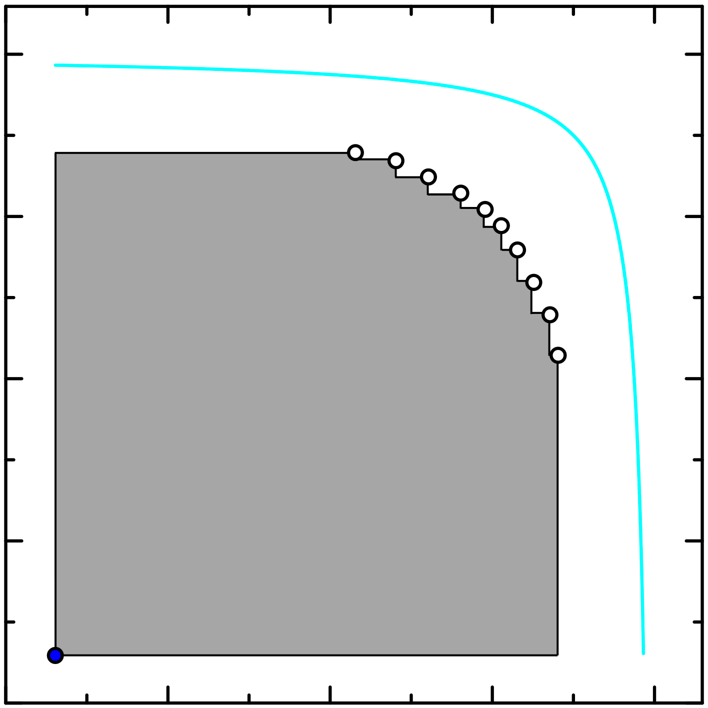


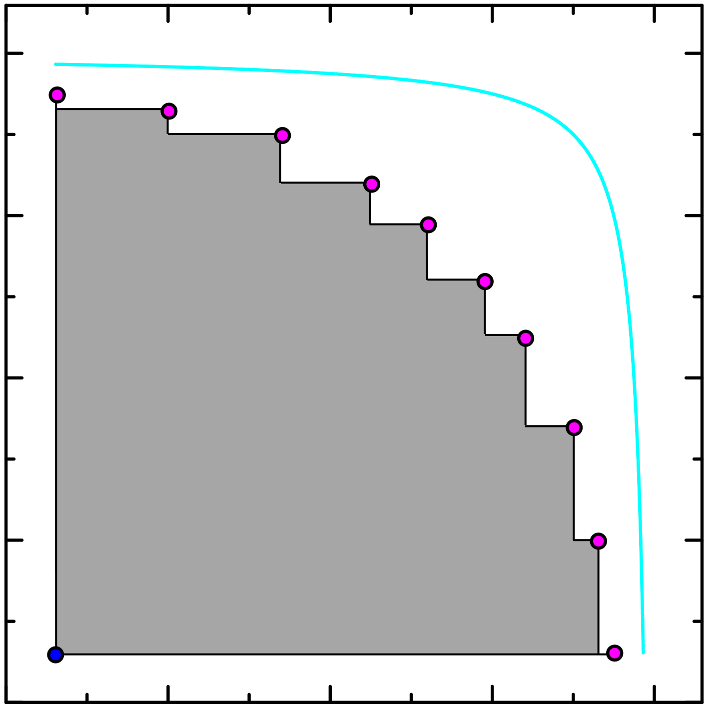
This is not an easy problem even for a two-objective case.

Hypervolume (HV)

Comparison results depend on the reference point

When the reference point is close to the Pareto front:



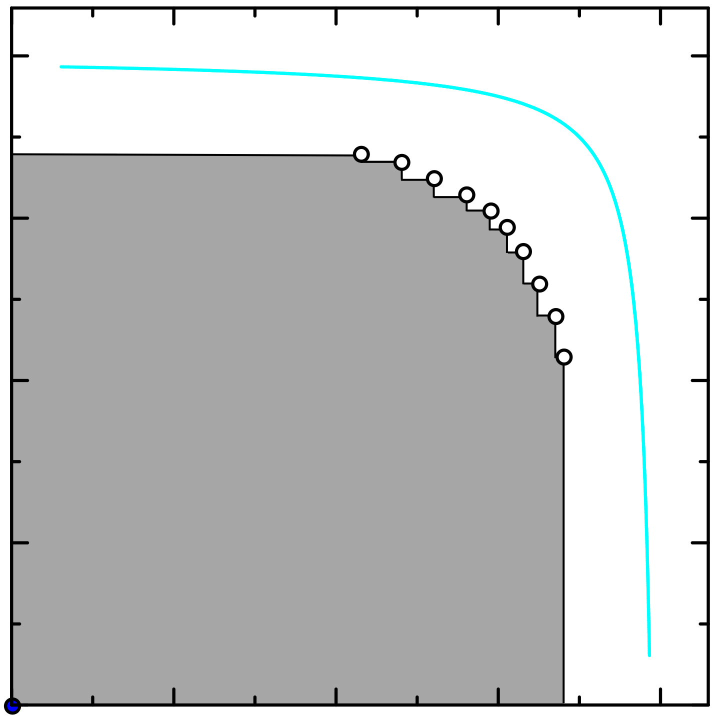


Better Solution Set

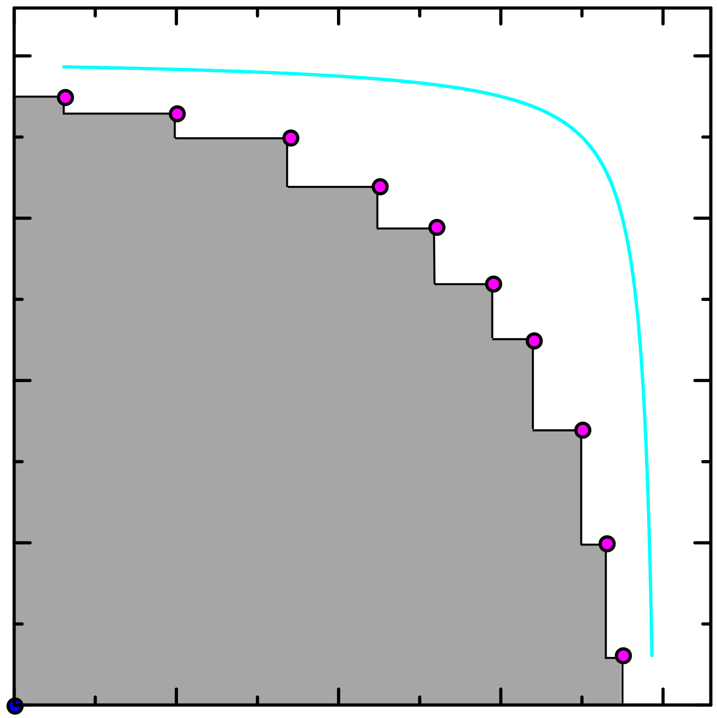
Hypervolume (HV)

Comparison results depend on the reference point

When the reference point is far from the Pareto front:

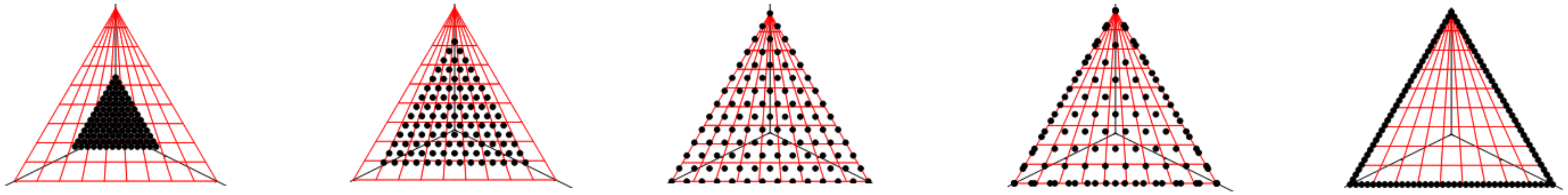


<



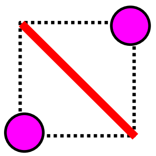
Better Solution Set

Which is the Best Solution Set by Hypervolume-Based Comparison?

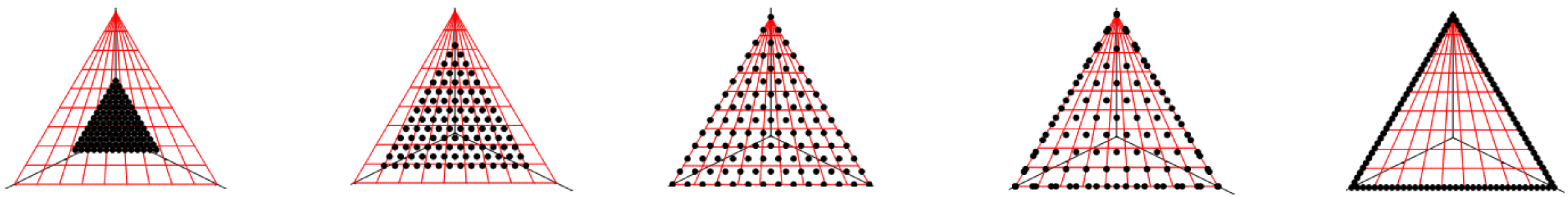


Which is the Best Solution Set by Hypervolume-Based Comparison?

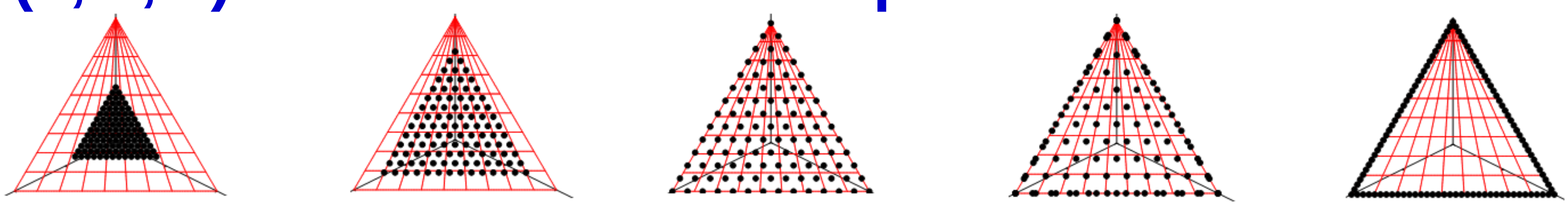
When the reference point is the nadir point:



(0, 0, 0) for the maximization problem

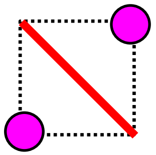


(1, 1, 1) for the minimization problem

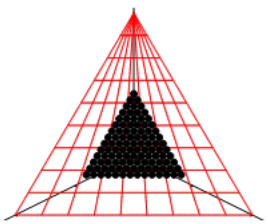


Which is the Best Solution Set by Hypervolume-Based Comparison?

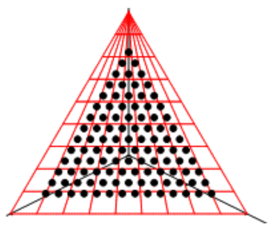
When the reference point is the nadir point:



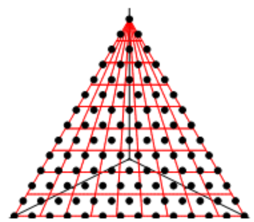
(0, 0, 0) for the maximization problem



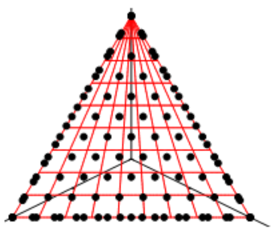
4



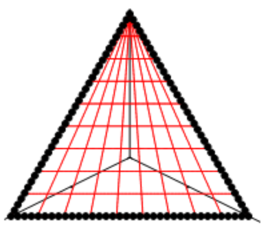
1



2

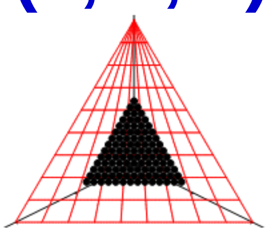


3

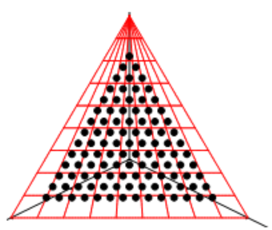


5

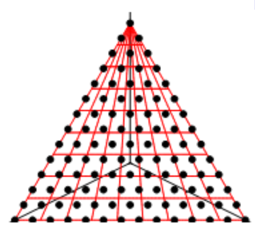
(1, 1, 1) for the minimization problem



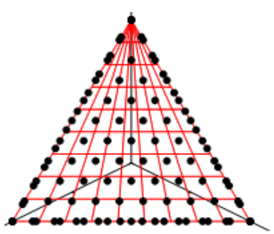
5



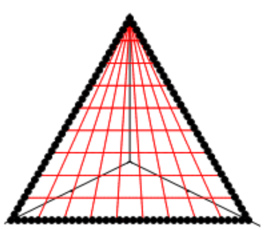
4



1

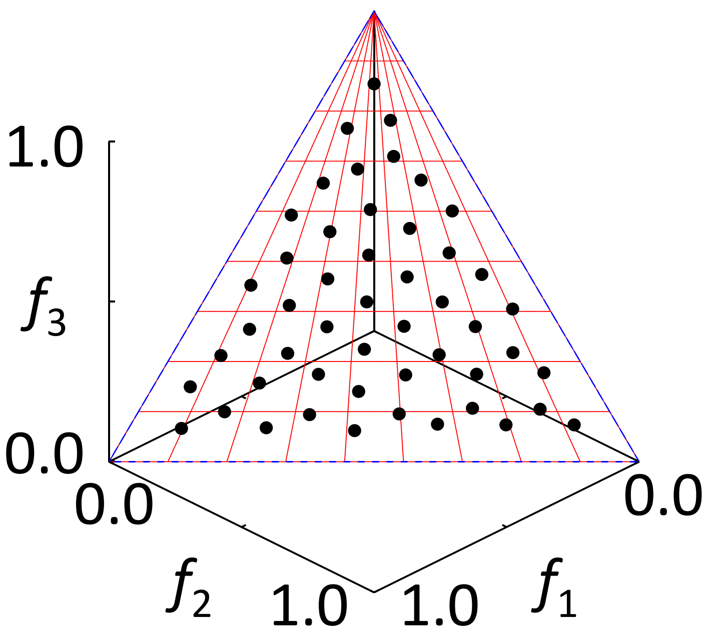


2

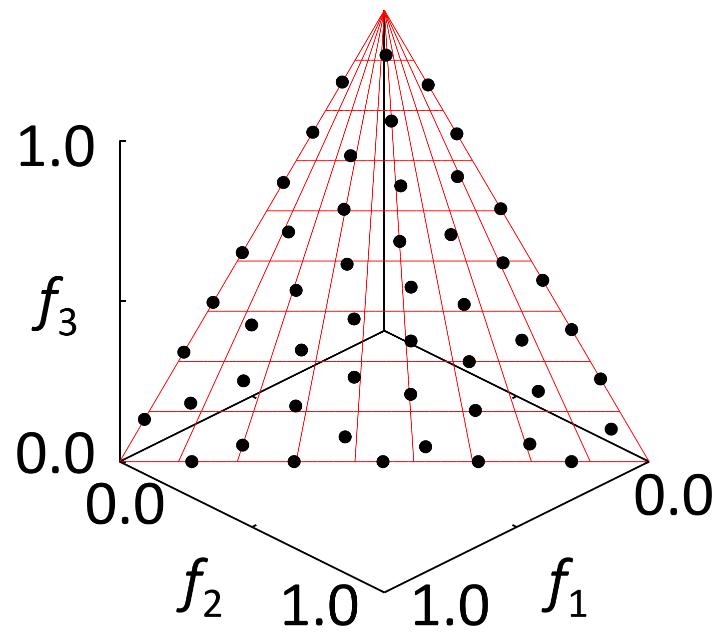


3

Near Optimal Distribution when the reference point is the nadir point



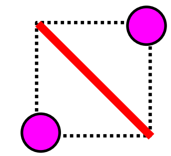
Maximization Problem



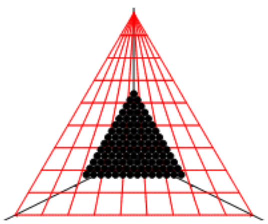
Minimization Problem

Which is the Best Solution Set by Hypervolume-Based Comparison?

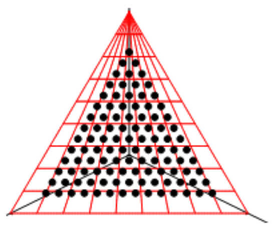
When the reference point is the nadir point:



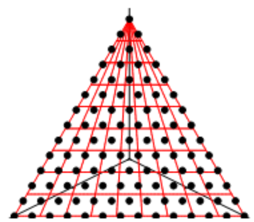
(0, 0, 0) for the maximization problem



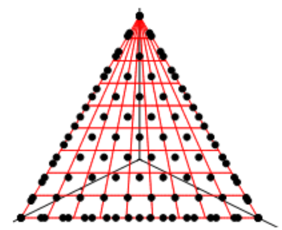
4



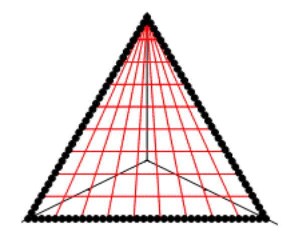
1



2

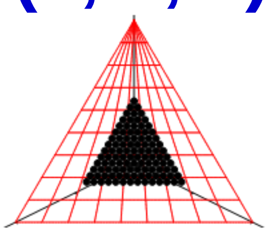


3

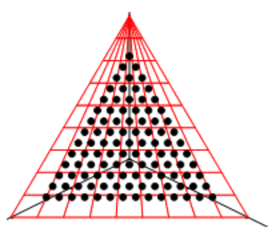


5

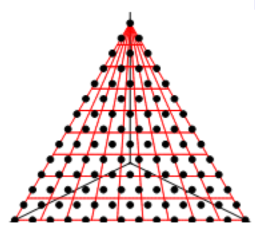
(1, 1, 1) for the minimization problem



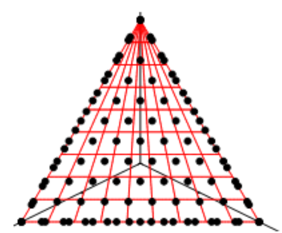
5



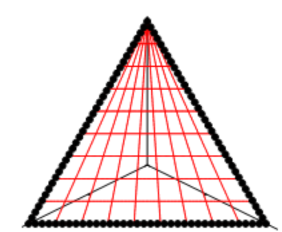
4



1

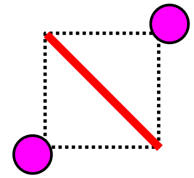


2



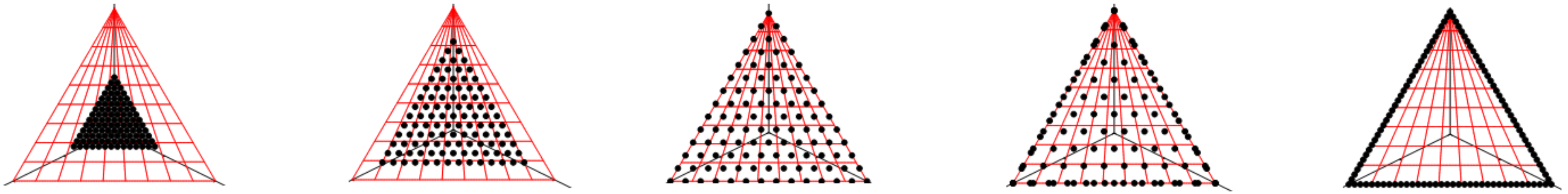
3

Which is the Best Solution Set by Hypervolume-Based Comparison?

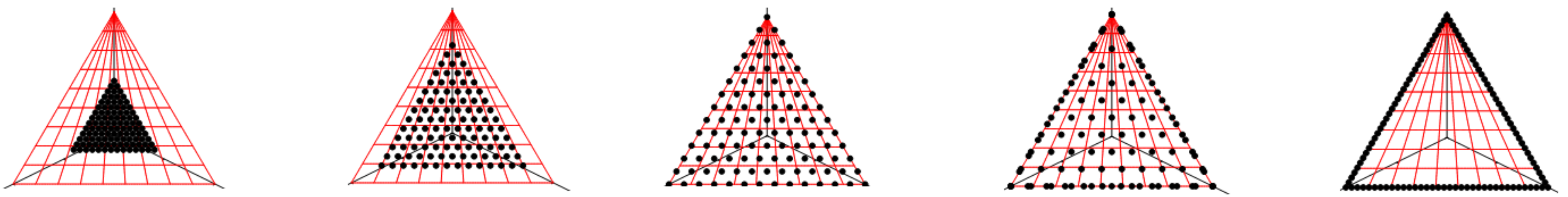


When the reference point is a little bit far:

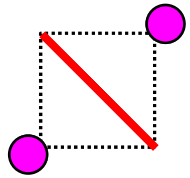
(-0.1, -0.1, -0.1) for the maximization problem



(1.1, 1.1, 1.1) for the minimization problem

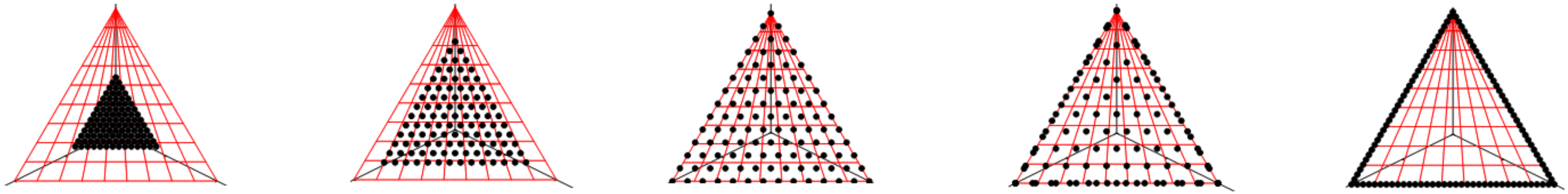


Which is the Best Solution Set by Hypervolume-Based Comparison?



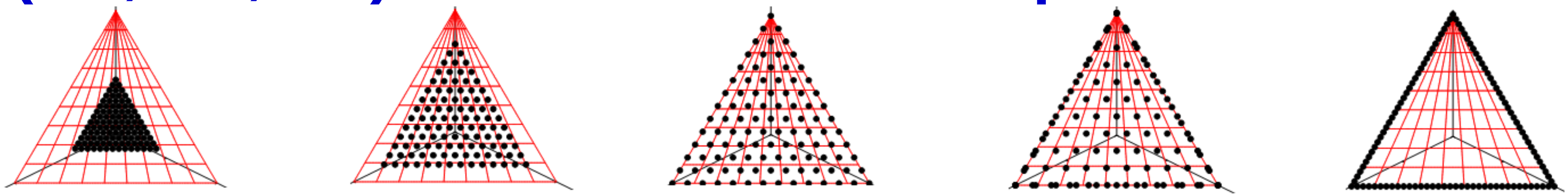
When the reference point is a little bit far:

(-0.1, -0.1, -0.1) for the maximization problem



4 3 1 2 5

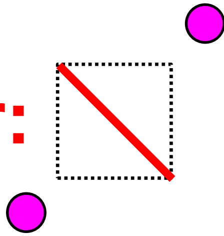
(1.1, 1.1, 1.1) for the minimization problem



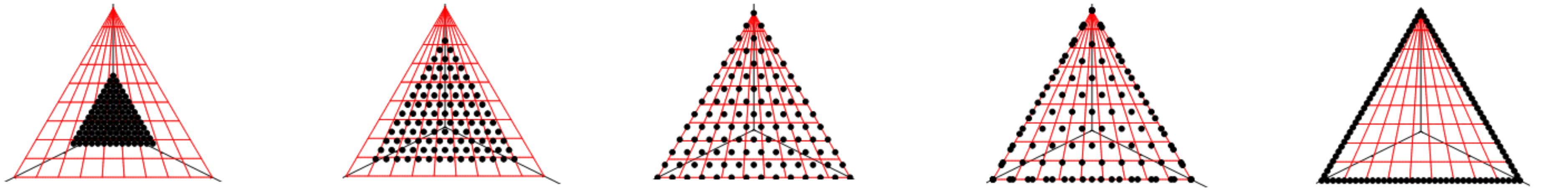
5 4 1 2 3

Which is the Best Solution Set by Hypervolume-Based Comparison?

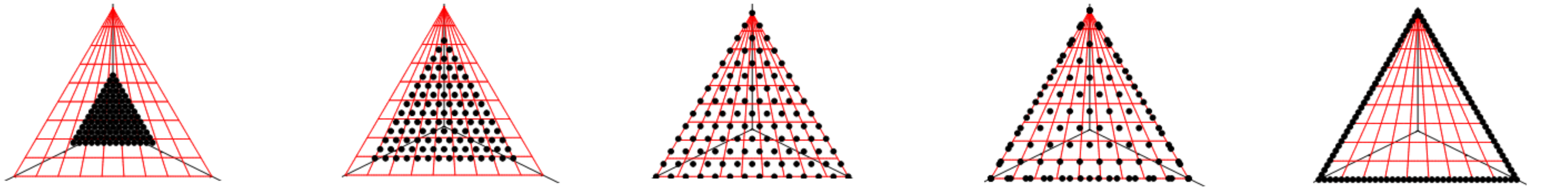
When the reference point is a little bit more far:



(-0.5, -0.5, -0.5) for the maximization problem

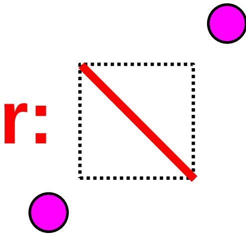


(1.5, 1.5, 1.5) for the minimization problem

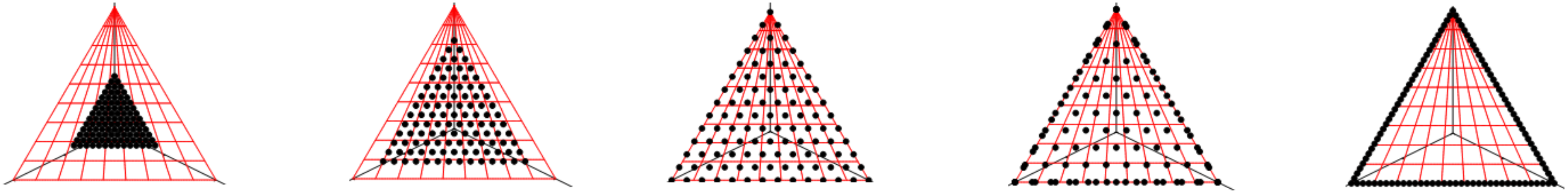


Which is the Best Solution Set by Hypervolume-Based Comparison?

When the reference point is a little bit more far:

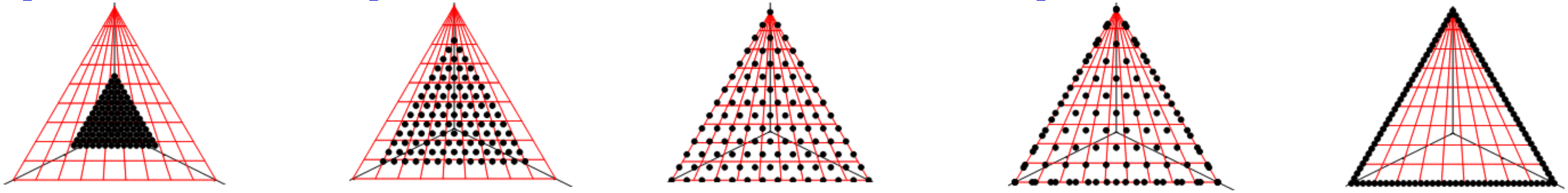


(-0.5, -0.5, -0.5) for the maximization problem



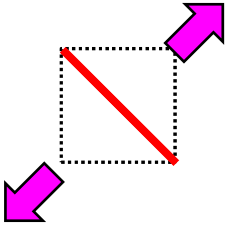
5 4 2 1 3

(1.5, 1.5, 1.5) for the minimization problem



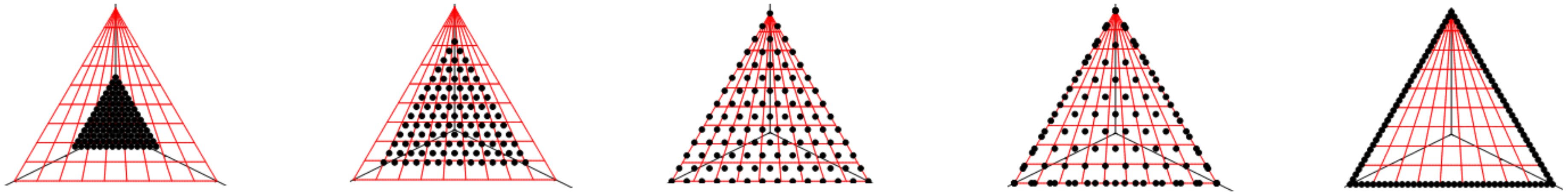
5 4 1 2 3

Which is the Best Solution Set by Hypervolume-Based Comparison?

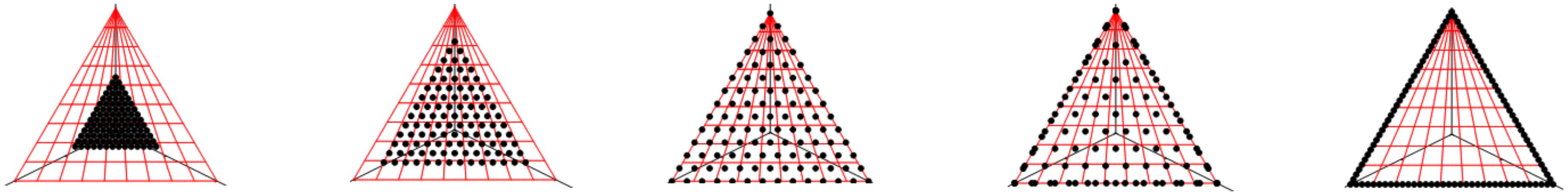


When the reference point is far away:

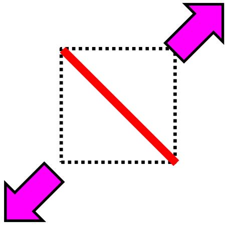
(-20, -20, -20) for the maximization problem



(20, 20, 20) for the minimization problem

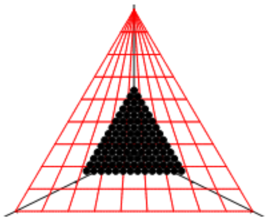


Which is the Best Solution Set by Hypervolume-Based Comparison?

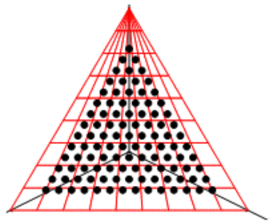


When the reference point is far away:

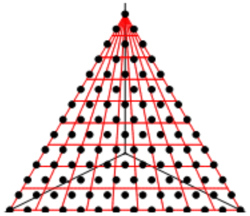
(-20, -20, -20) for the maximization problem



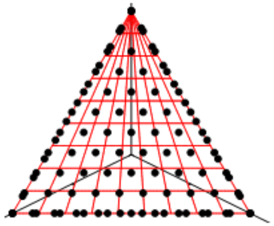
5



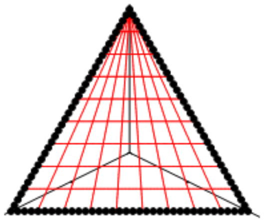
4



3

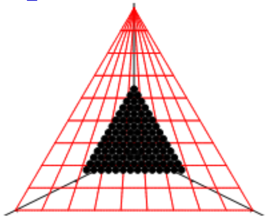


2

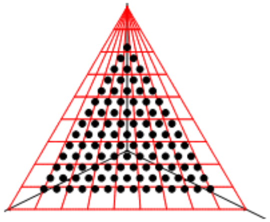


1

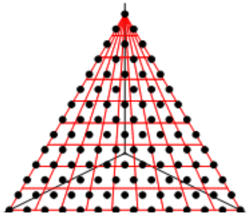
(20, 20, 20) for the minimization problem



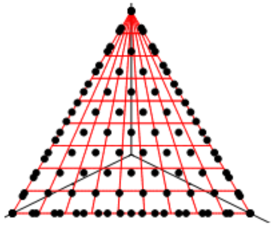
5



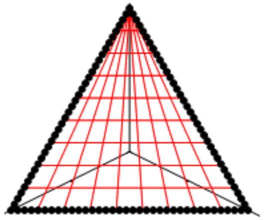
4



1



2



3

Dependency of the optimal distribution of solutions on the location of the reference point (r = 1.0: Nadir Point)

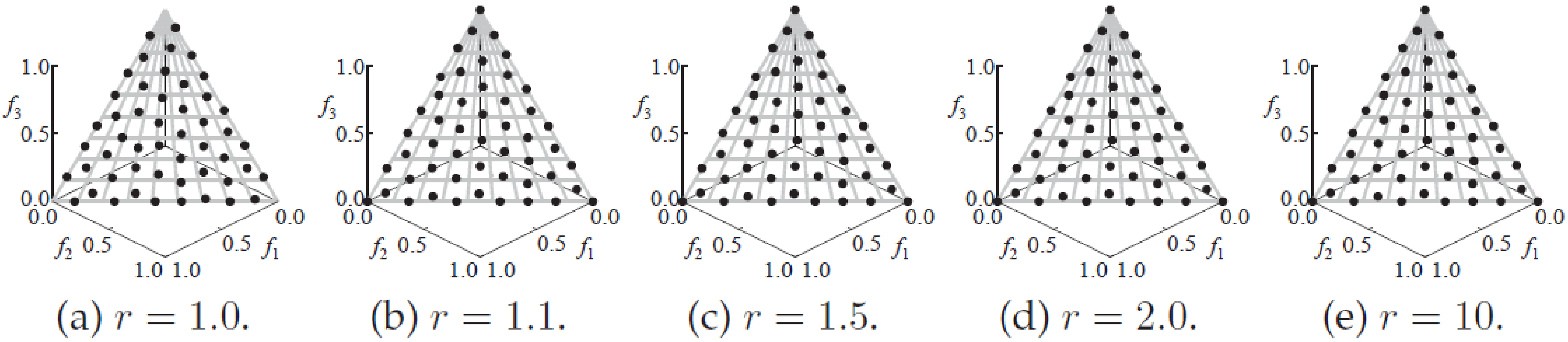


Figure 1: Obtained solution sets for the three-objective normalized DTLZ1.

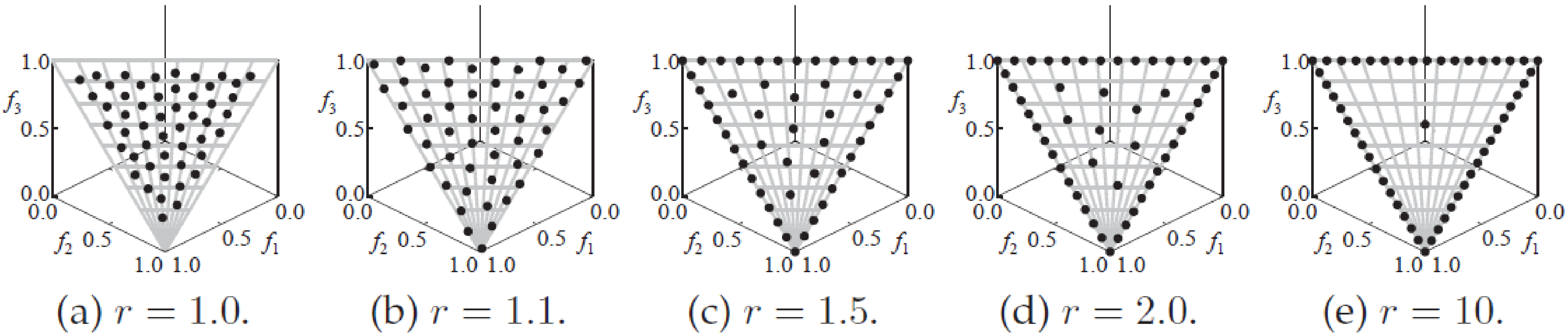
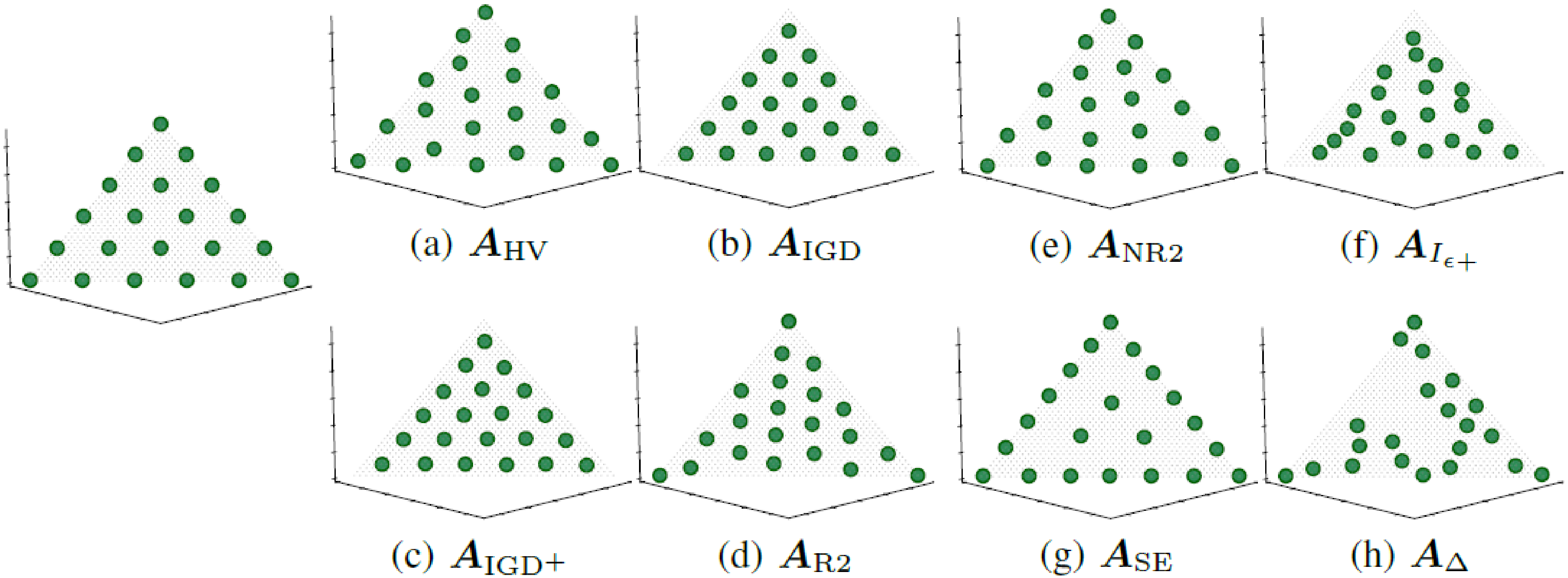


Figure 2: Obtained solution sets for the three-objective normalized Minus-DTLZ1.

H. Ishibuchi et al., How to specify a reference point in hypervolume calculation for fair performance comparison," *Evolutionary Computation* (2018).

One Additional Note on Performance Indicators

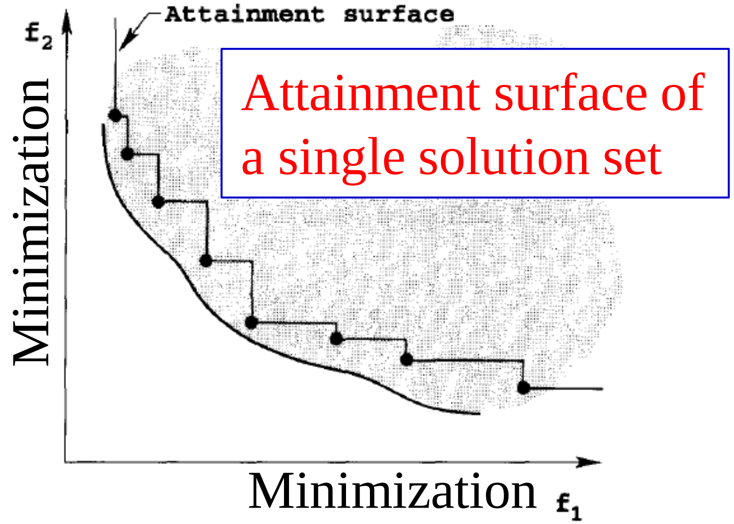
Uniformly-distributed solutions over the entire Pareto front (including the boundary) are usually not optimal for any performance indicators even for linear Pareto fronts.

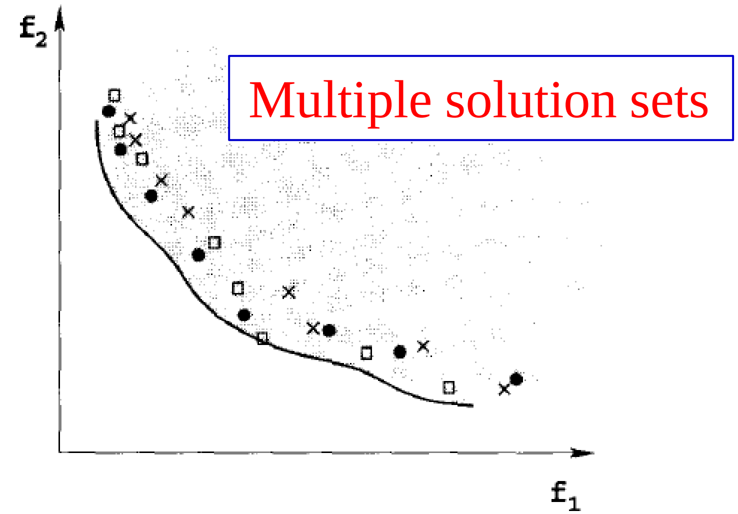


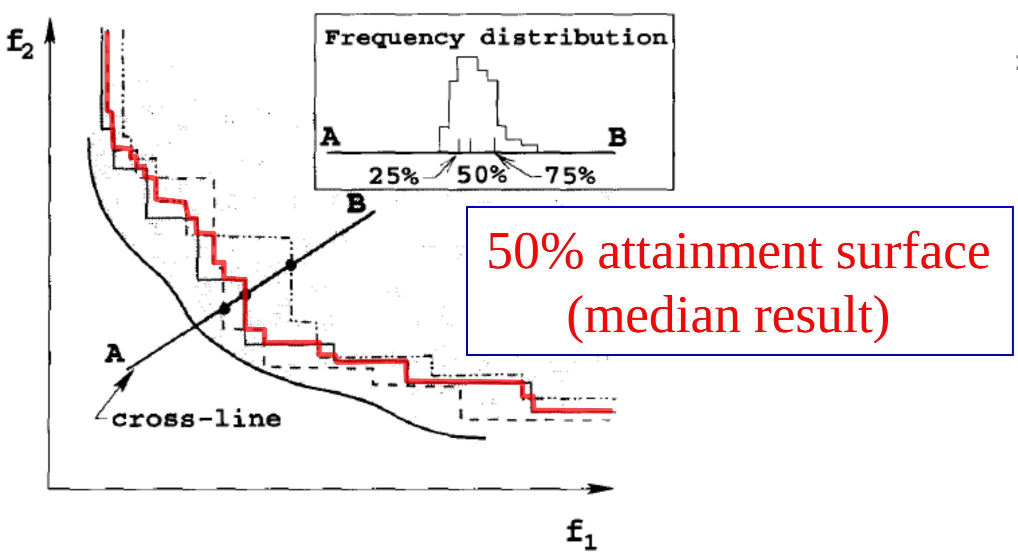
R. Tanabe and H. Ishibuchi, **"An analysis of quality indicators using approximated optimal distributions in a three-dimensional objective space,"** *IEEE Trans. on Evolutionary Computation* (Early Access).

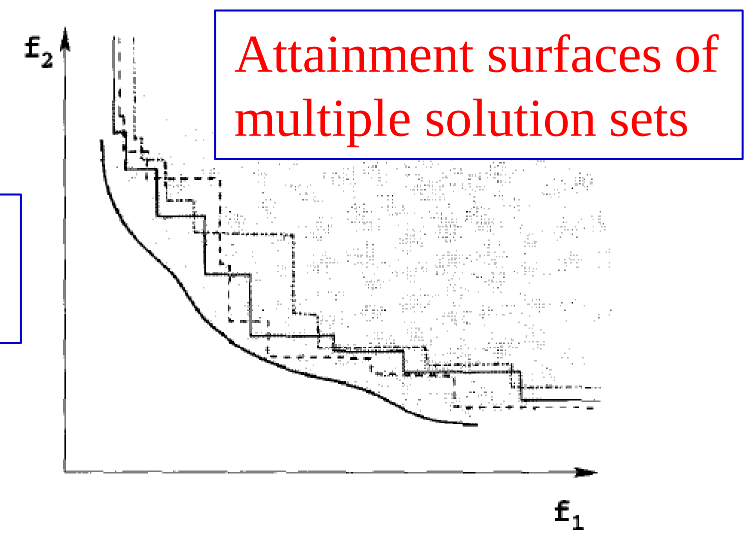
Other Performance Evaluation Method (Visual Evaluation)

(1) Attainment Surface for two-objective problems







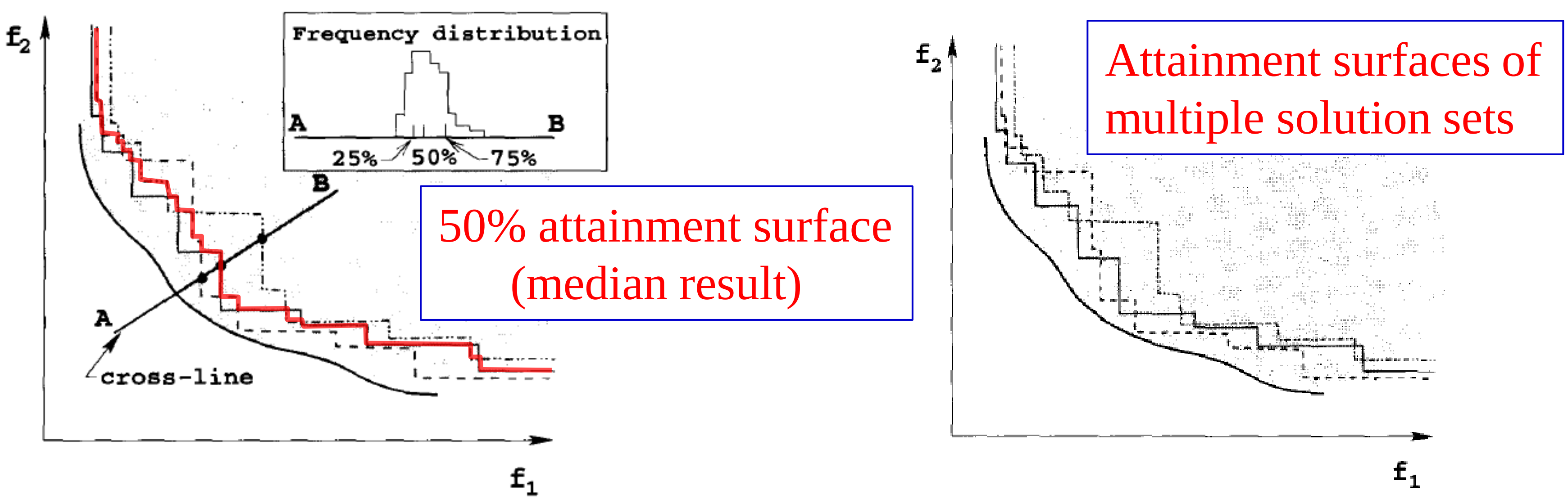


K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*, John Wiley & Sons, 2001.

Other Performance Evaluation Method (Visual Evaluation)

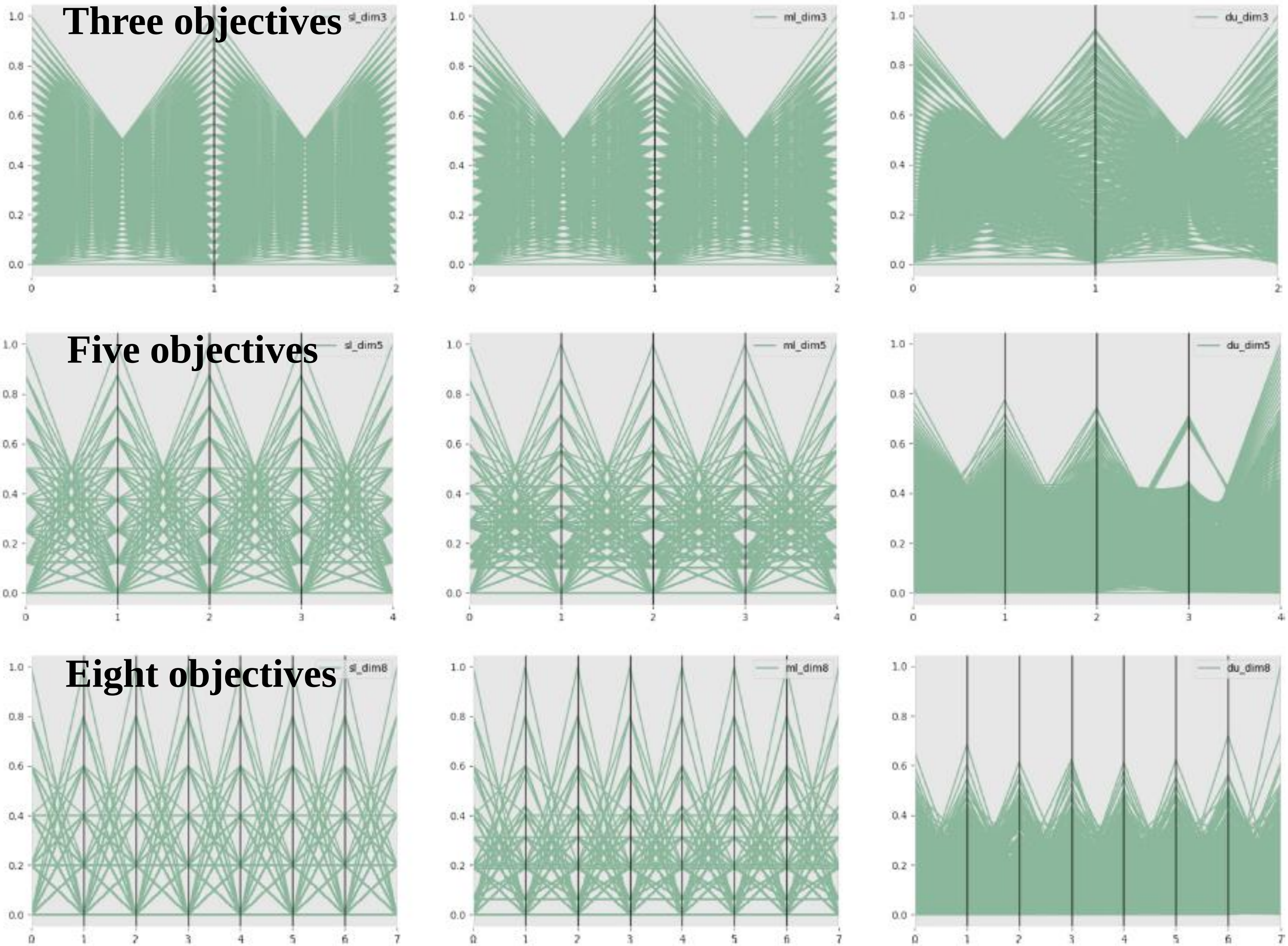
(1) Attainment Surface for two-objective problems

For a two-objective problem, it is often recommended (by reviewers for a journal paper, by the session chair for a conference paper) to use the 50% attainment surface for visual performance comparison of the proposed algorithm with other existing algorithms. By using the 50% attainment surface, you can visually compare experimental results of different algorithms in a somewhat statistically solid manner.



K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*, John Wiley & Sons, 2001.

(2) Parallel coordinates plots for many-objective problems



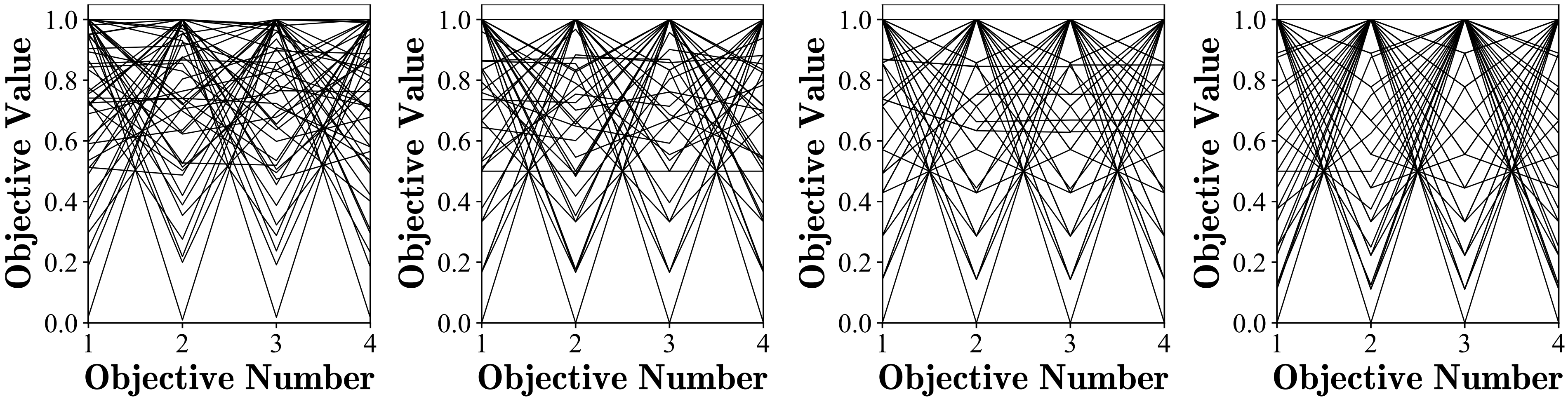
References for Visualization of Solution Sets

[1] T. Tušar and B. Filipič, "**Visualization of Pareto front approximations in evolutionary multiobjective optimization**: A critical review and the prosection method" *IEEE Transactions on Evolutionary Computation*, 2015.

[2] M. Li, L. Zhen, and X. Yao, "**How to read many-objective solution sets in parallel coordinates**," *IEEE Computational Intelligence Magazine*, 2017.

Visualization is not easy for the case of four or more objectives

- Can you explain the difference among the four solution sets?
- Which do you think the best solution set?

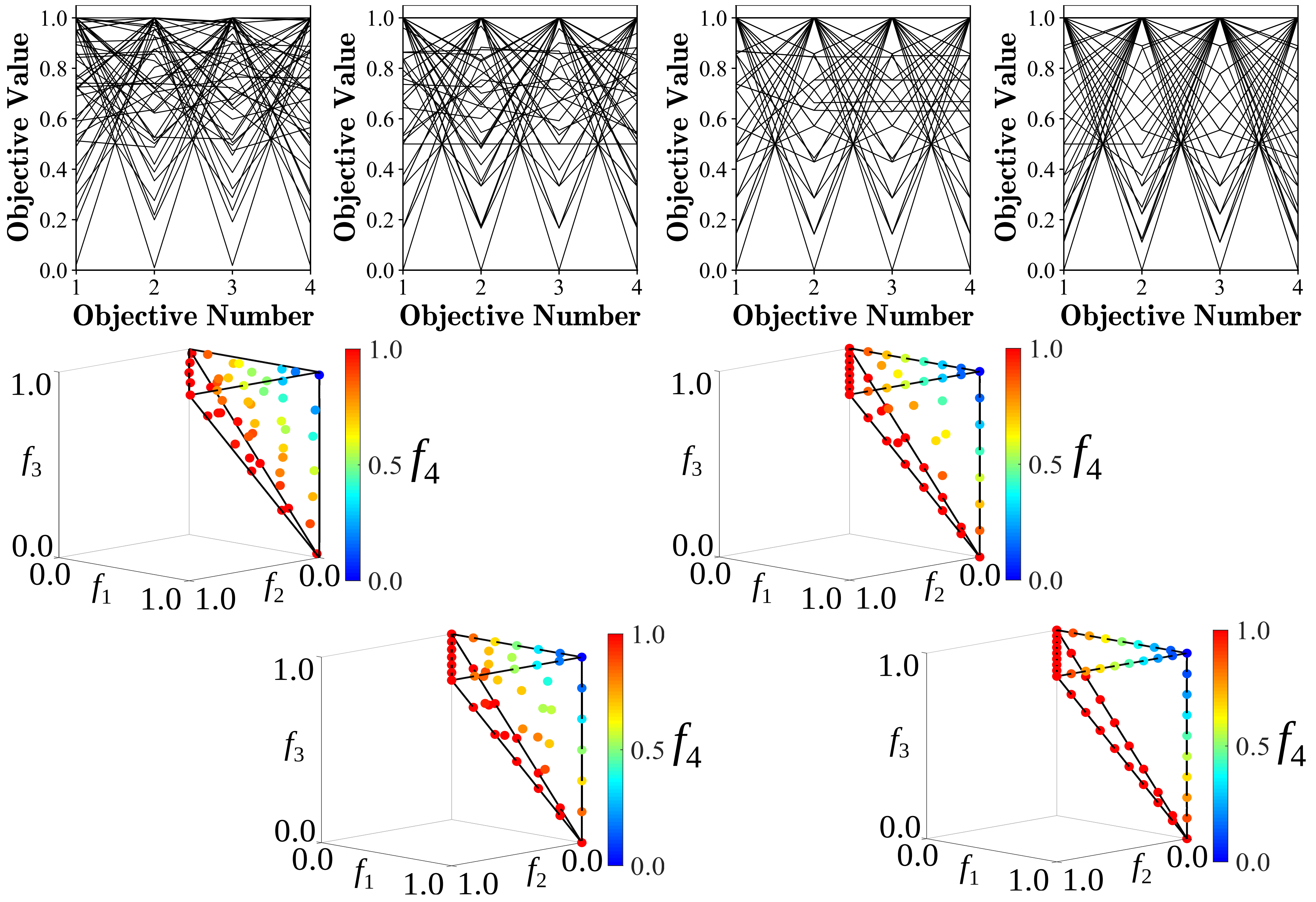


(a) *r* = 1.2. **(b) *r* = 1.4.** **(c) *r* = 2.0.** **(d) *r* = 100**

Optimal distributions of solutions of the four-objective minus-DTLZ1 problem for HV maximization with respect to different specifications of the reference point $\boldsymbol{r} = (r, r, r, r)$ in the normalized objective space.

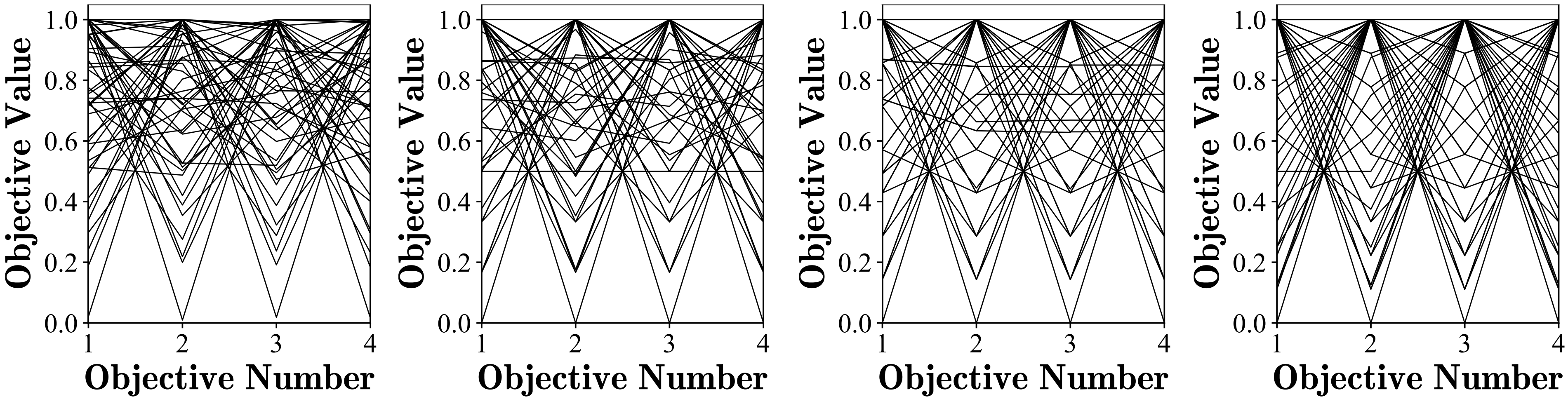
H. Ishibuchi, T. Matsumoto, N. Masuyama, and Y. Nojima, "Optimal distributions of solutions for hypervolume maximization on triangular and inverted triangular Pareto fronts of four-objective Problems," IEEE SSCI 2019.

Visualization is not easy for the case of four or more objectives

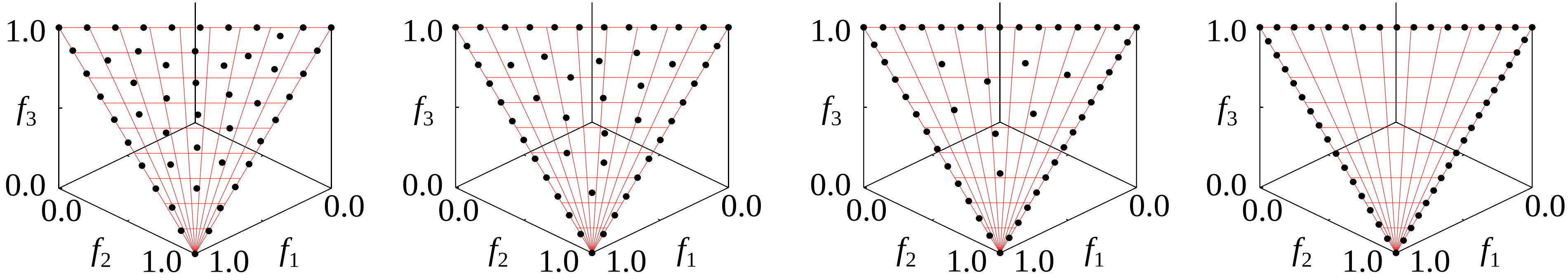


Visualization is not easy for the case of four or more objectives

Results on Four-Objective Minus-DTLZ1

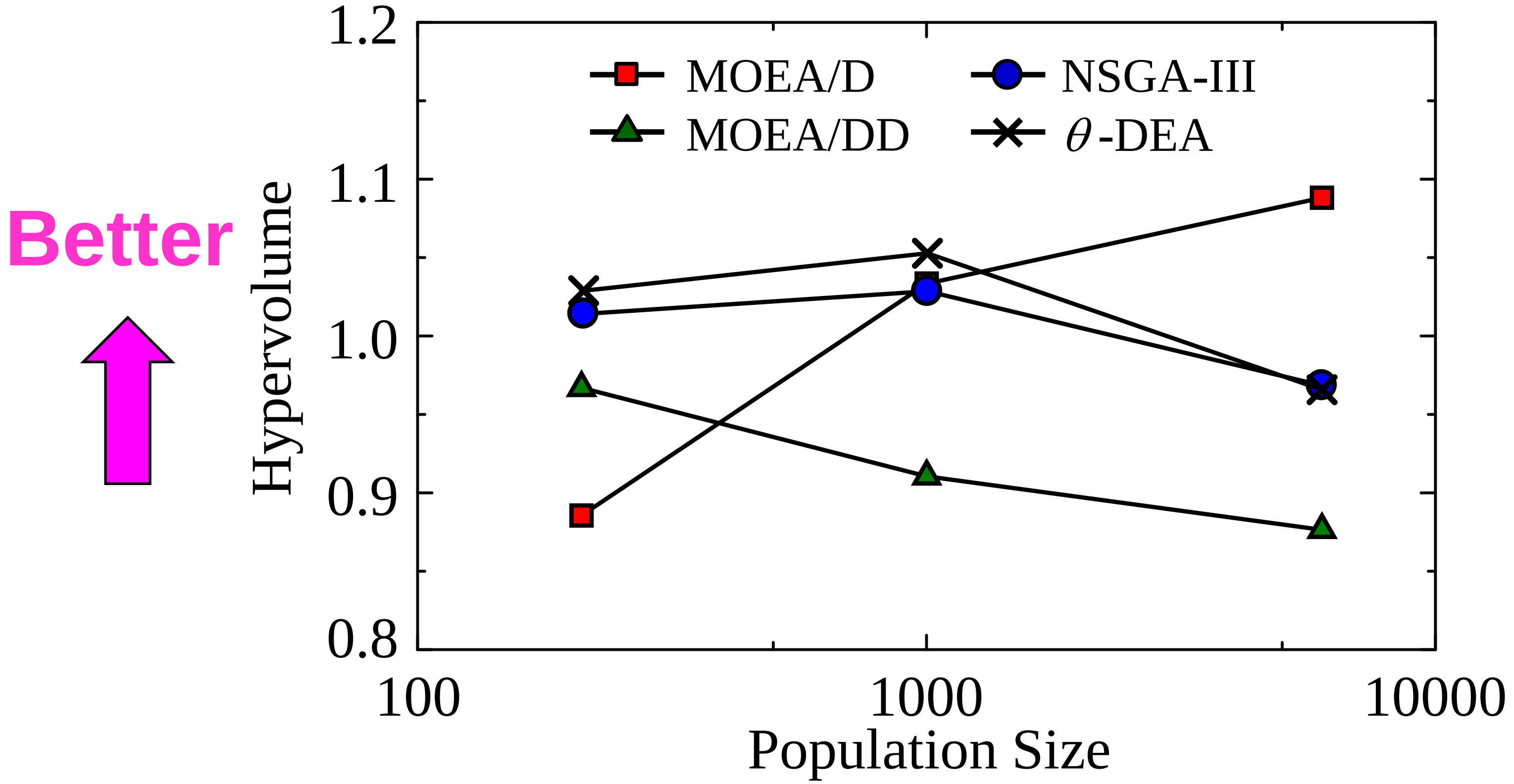


Results on Three-Objective Minus-DTLZ1



Another Difficulty related to Performance Evaluation

Performance comparison results strongly depend on the specification of the population size.



Performance of the final population with different population size (Comparison results on the five-objective WFG3)

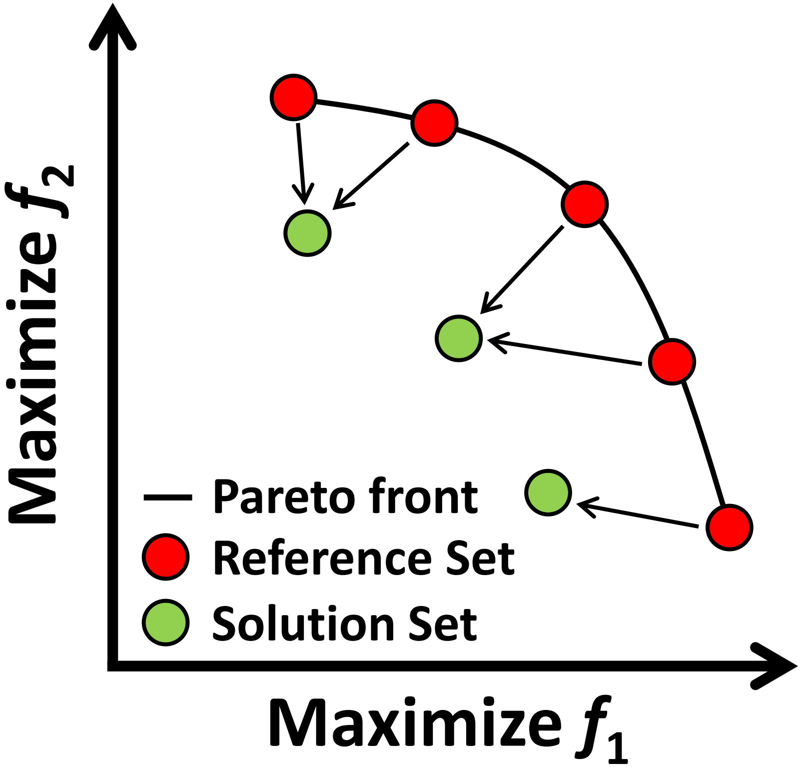
Ishibuchi et al. (IEEE CEC 2016)

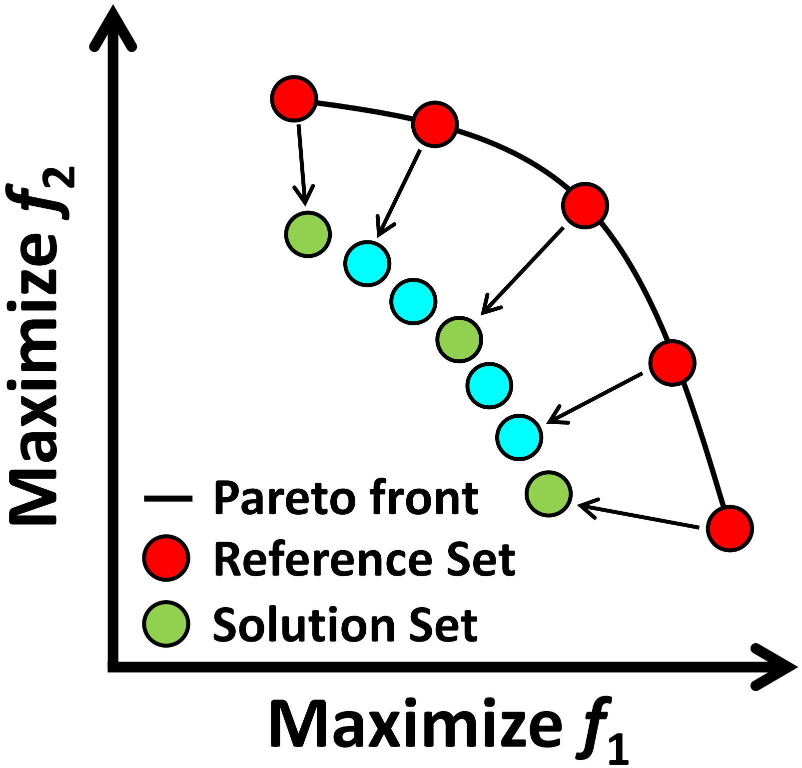
Another Difficulty related to Performance Evaluation

Performance comparison results strongly depend on the specification of the population size.

The same size solution sets are needed for fair comparison.

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

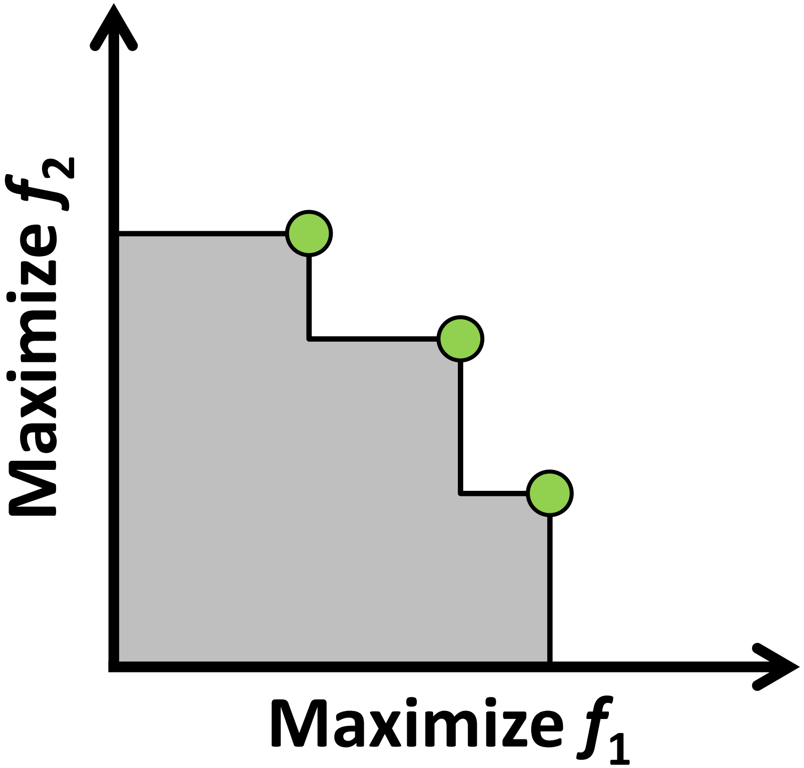


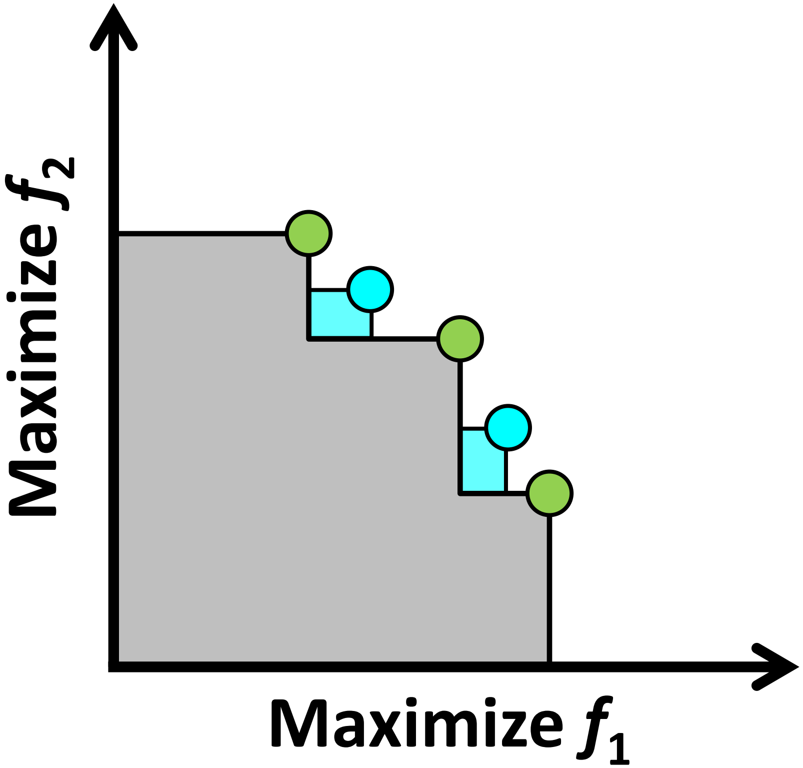


Another Difficulty related to Performance Evaluation

Performance comparison results strongly depend on the specification of the population size.
The same size solution sets are needed for fair comparison.

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

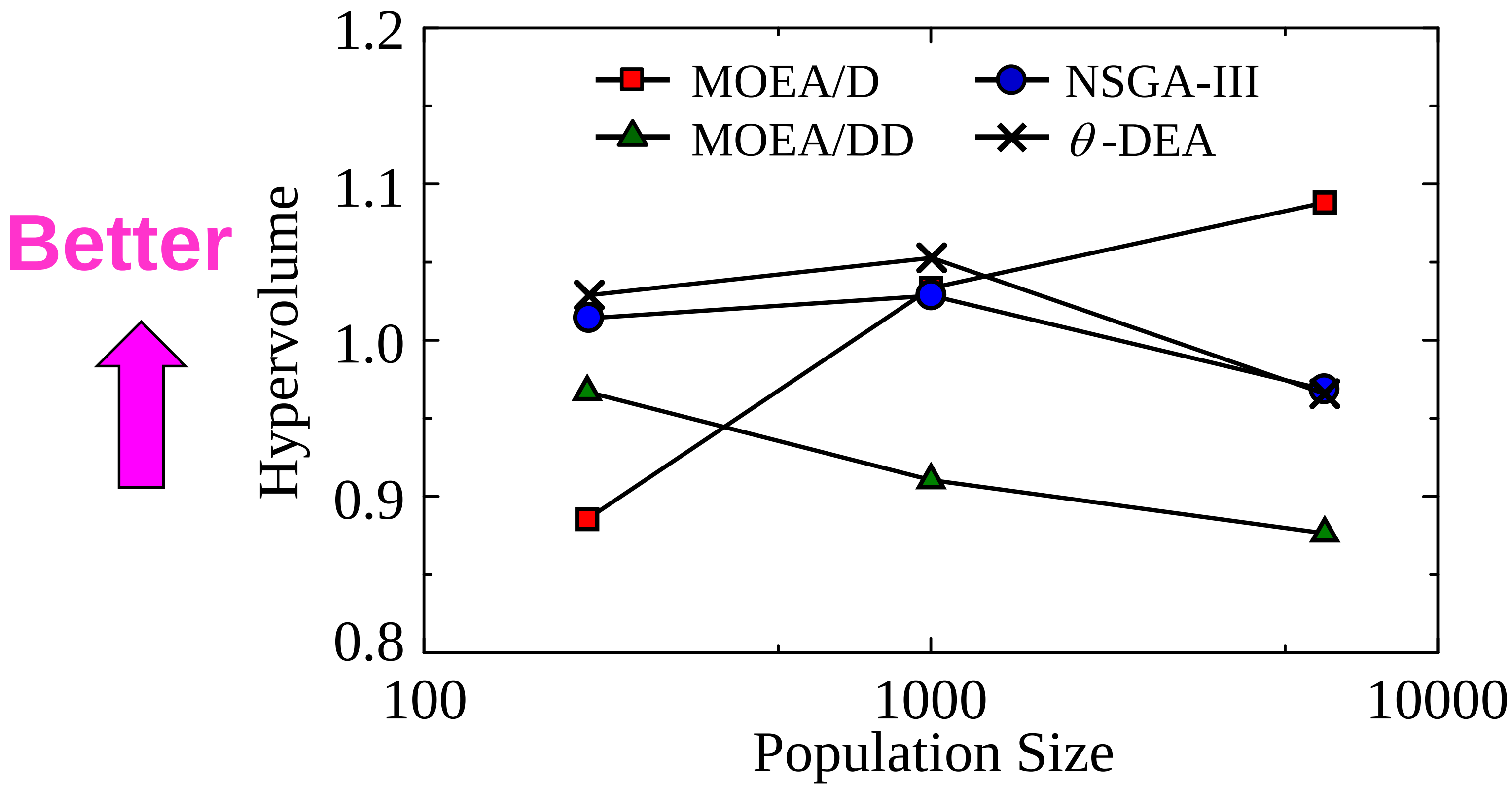




Another Difficulty related to Performance Evaluation

Performance comparison results strongly depend on the specification of the population size.
The same size solution sets are needed for fair comparison.

Q: How to choose the same population size for all algorithms.



Another Difficulty related to Performance Evaluation

Performance comparison results strongly depend on the specification of the population size.
The same size solution sets are needed for fair comparison.

Recent Trend: Selection of a pre-specified number of solutions from all the examined solutions. Population size can be different as long as the same number of solutions are selected.

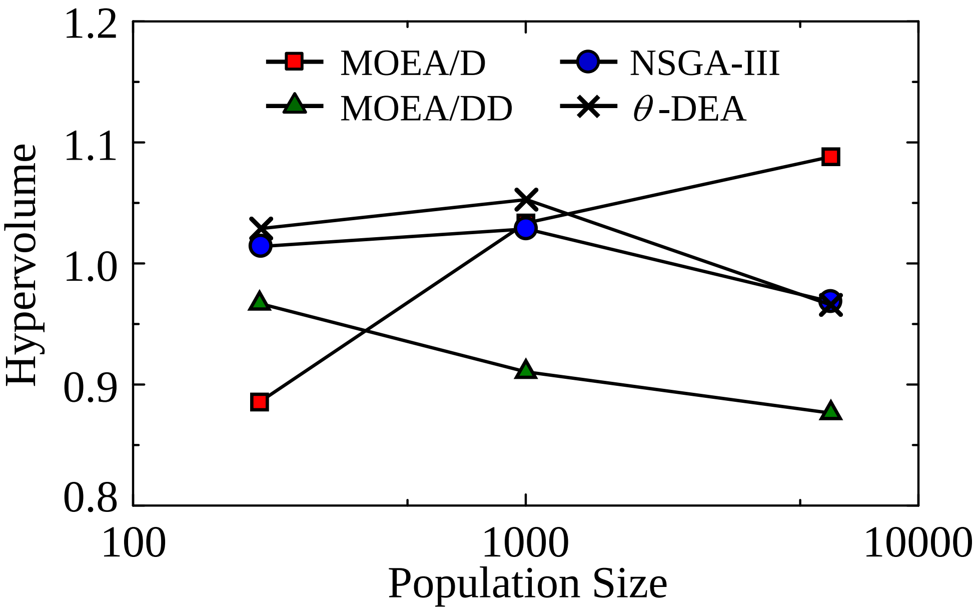
H. Ishibuchi, Y. Setoguchi, H. Masuda, and Y. Nojima, **"How to compare many-objective algorithms under different settings of population and archive sizes,"** *Proc. of 2016 IEEE Congress on Evolutionary Computation*, pp. 1149-1156, Vancouver, Canada, July 24-29, 2016.

R. Tanabe, H. Ishibuchi, and A. Oyama, **"Benchmarking multi- and many-objective evolutionary algorithms under two optimization scenarios,"** IEEE Access, vol. 5, pp. 19597-19619, December 2017.

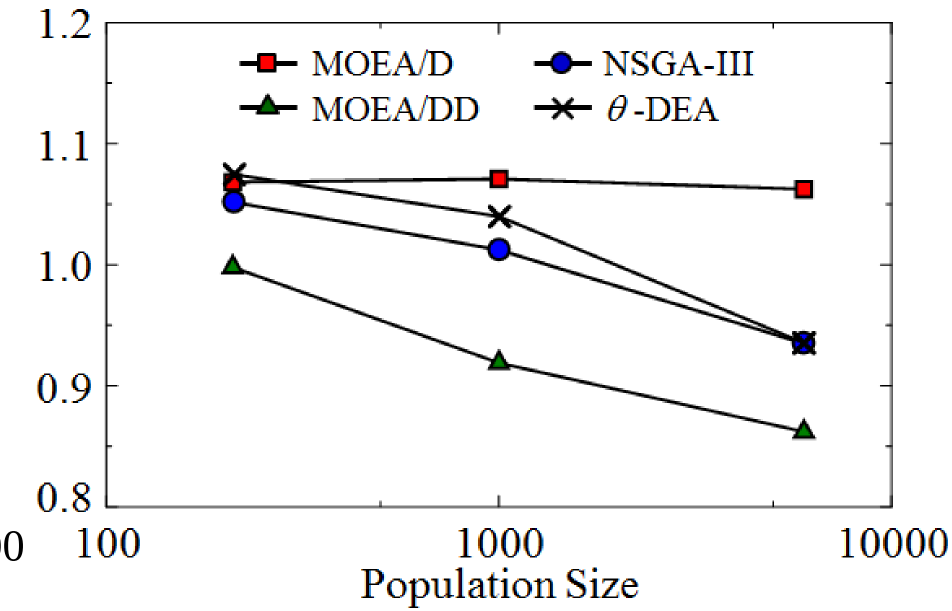
H. Ishibuchi, L. M. Pang, and K. Shang, **"A new framework of evolutionary multi-objective algorithms with an unbounded external archive,"** *Proc. of 24th European Conference on Artificial Intelligence* (ECAI 2020), Santiago, Spain, August 29 - September 02, 2020 (Accepted).

Final Population vs. Selected Solutions

(50 solutions are selected)



Final Population Results



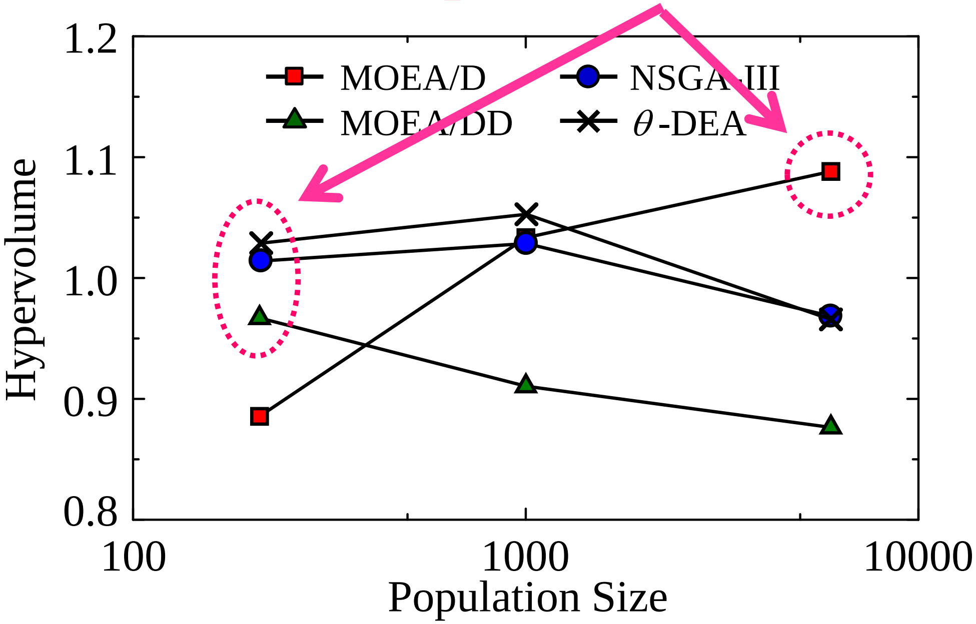
Selected Solutions Results

When the four algorithms are compared using the final population with the standard population size (about 200), MOEA/D is the worst. However, when they are compared by the selected solutions, MOEA/D is the best.

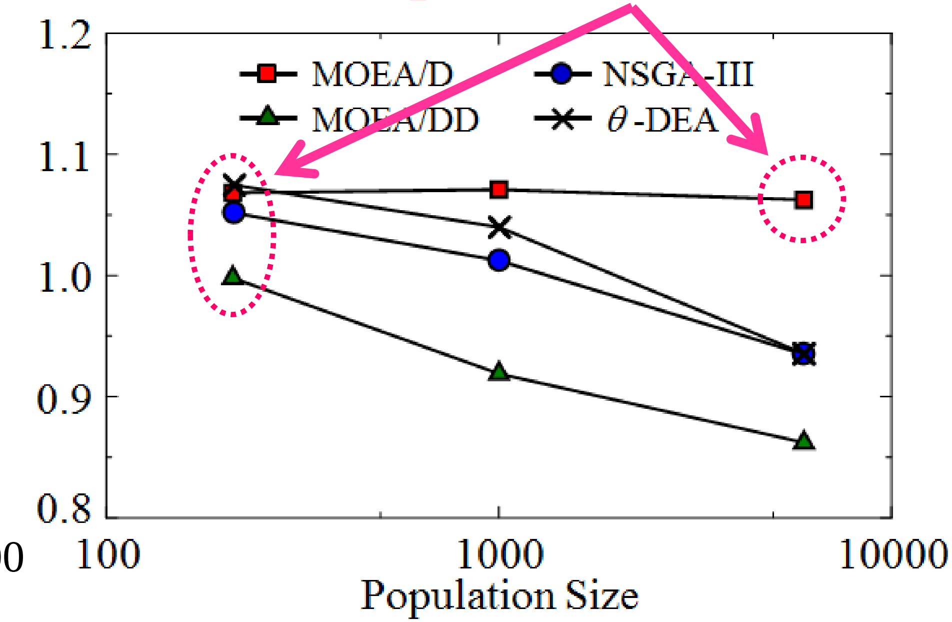
Final Population vs. Selected Solutions

(50 solutions are selected)

We cannot compare these results but we can compare these results.



Final Population Results



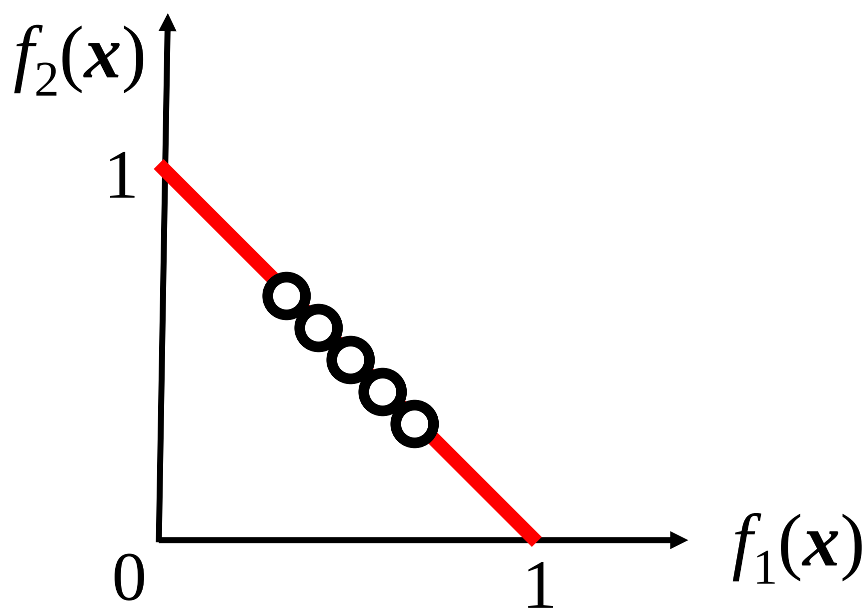
Selected Solutions Results

When the four algorithms are compared using the final population with the standard population size (about 200), MOEA/D is the worst. However, when they are compared by the selected solutions, MOEA/D is the best.

Lab Session Task 1:

Two-Objective Problem: Minimize $f_1(\boldsymbol{x})$ and $f_2(\boldsymbol{x})$

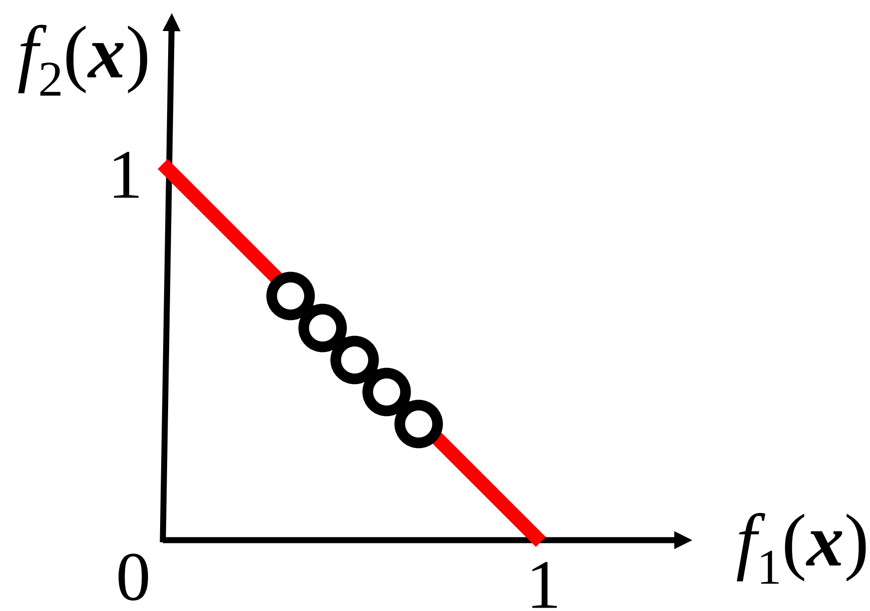
Let us assume that the Pareto front of this problem is the line between (1, 0) and (0, 1) including these two points. We use an infinitely large number of reference points on the Pareto front: (1, 0), (1 ,), (1 2 , 2), ..., (0, 1) where is an infinitely small positive number. We also assume that we have five solutions on the Pareto front, which have the minimum GD value (among all sets of five solutions on the Pareto front). Please show those five solutions.



Lab Session Task 2:

Two-Objective Problem: Minimize $f_1(\boldsymbol{x})$ and $f_2(\boldsymbol{x})$

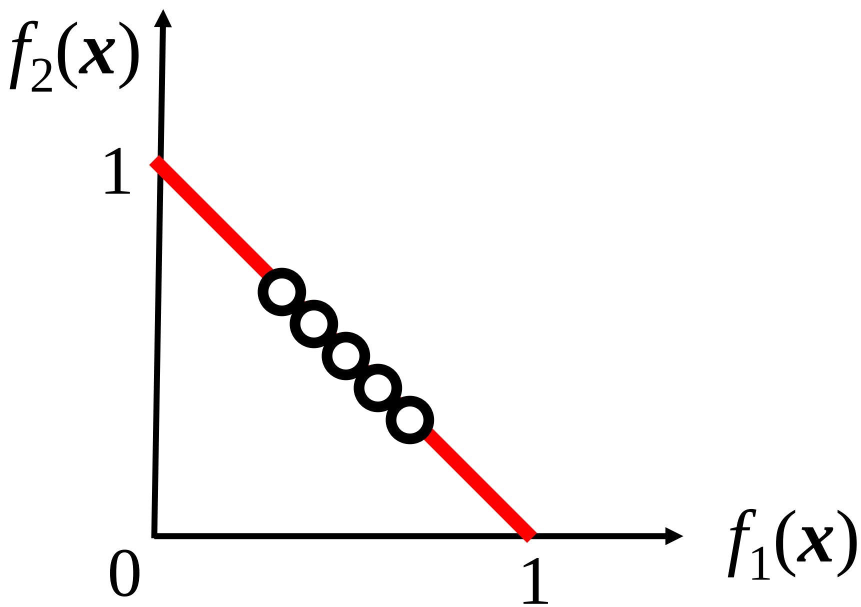
Let us assume that the Pareto front of this problem is the line between (1, 0) and (0, 1) including these two points. We use an infinitely large number of reference points on the Pareto front: (1, 0), (1 ,), (1 2 , 2), ..., (0, 1) where is an infinitely small positive number. We also assume that we have five solutions on the Pareto front, which have the minimum IGD value (among all sets of five solutions on the Pareto front). Please show those five solutions.



Lab Session Task 3:

Two-Objective Problem: Minimize $f_1(\boldsymbol{x})$ and $f_2(\boldsymbol{x})$

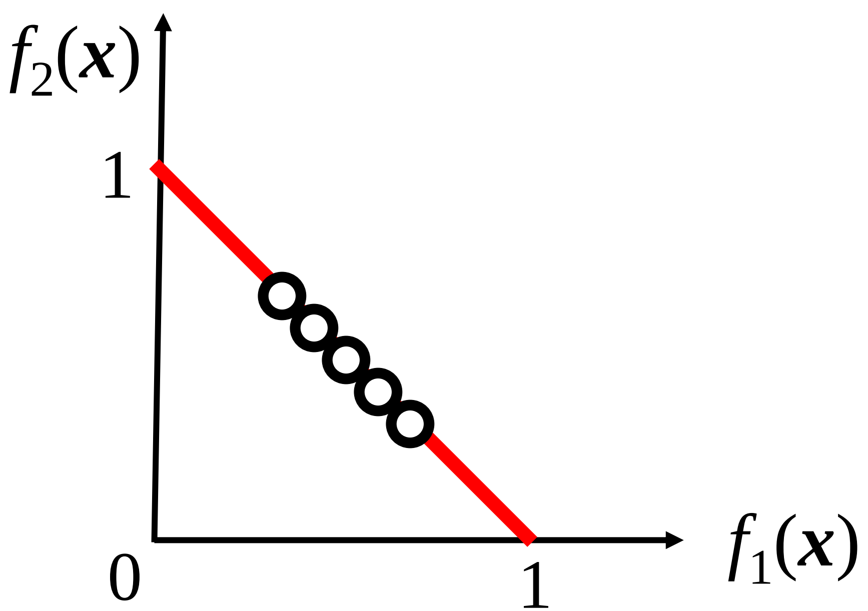
Let us assume that the Pareto front of this problem is the line between (1, 0) and (0, 1) including these two points. We use (1, 1) as the reference point for hypervolume calculation. We also assume that we have five solutions on the Pareto front, which have the maximum hypervolume value (among all sets of five solutions on the Pareto front). Please show those five solutions.



Lab Session Task 4:

Two-Objective Problem: Minimize $f_1(\boldsymbol{x})$ and $f_2(\boldsymbol{x})$

Let us assume that the Pareto front of this problem is the line between (1, 0) and (0, 1) including these two points. We use (2, 2) as the reference point for hypervolume calculation. We also assume that we have five solutions on the Pareto front, which have the maximum hypervolume value (among all sets of five solutions on the Pareto front). Please show those five solutions.



Lab Session Task 5 (Optional):

Three-Objective Problem: Minimize $f_1(\boldsymbol{x})$, $f_2(\boldsymbol{x})$, and $f_3(\boldsymbol{x})$

Let us consider the following two Pareto fronts in $[0, 1]^3$:

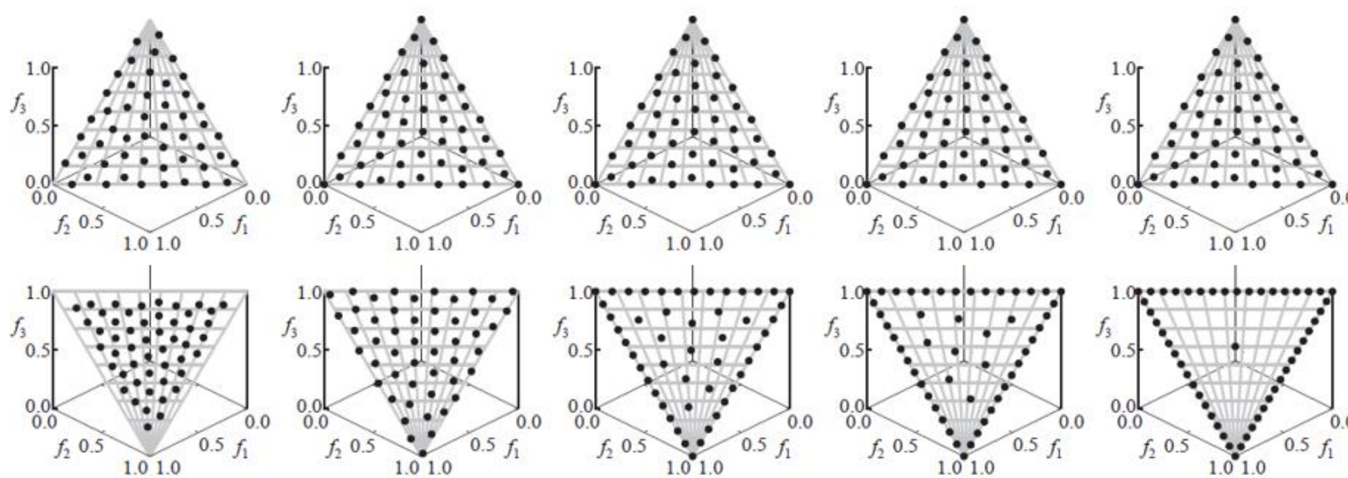
(i) $f_1(\boldsymbol{x}) + f_2(\boldsymbol{x}) + f_3(\boldsymbol{x}) = 1$. (ii) $f_1(\boldsymbol{x}) + f_2(\boldsymbol{x}) + f_3(\boldsymbol{x}) = 2$.

The following distribution is to maximize the hypervolume value for each Pareto front when the reference point is (1000, 1000):

(i) Solutions are well distributed over the entire Pareto front.

(ii) All (almost all) solutions are on the boundary of the Pareto front.

Please explain why these totally different solution sets maximize the hypervolume value for the two Pareto fronts.



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