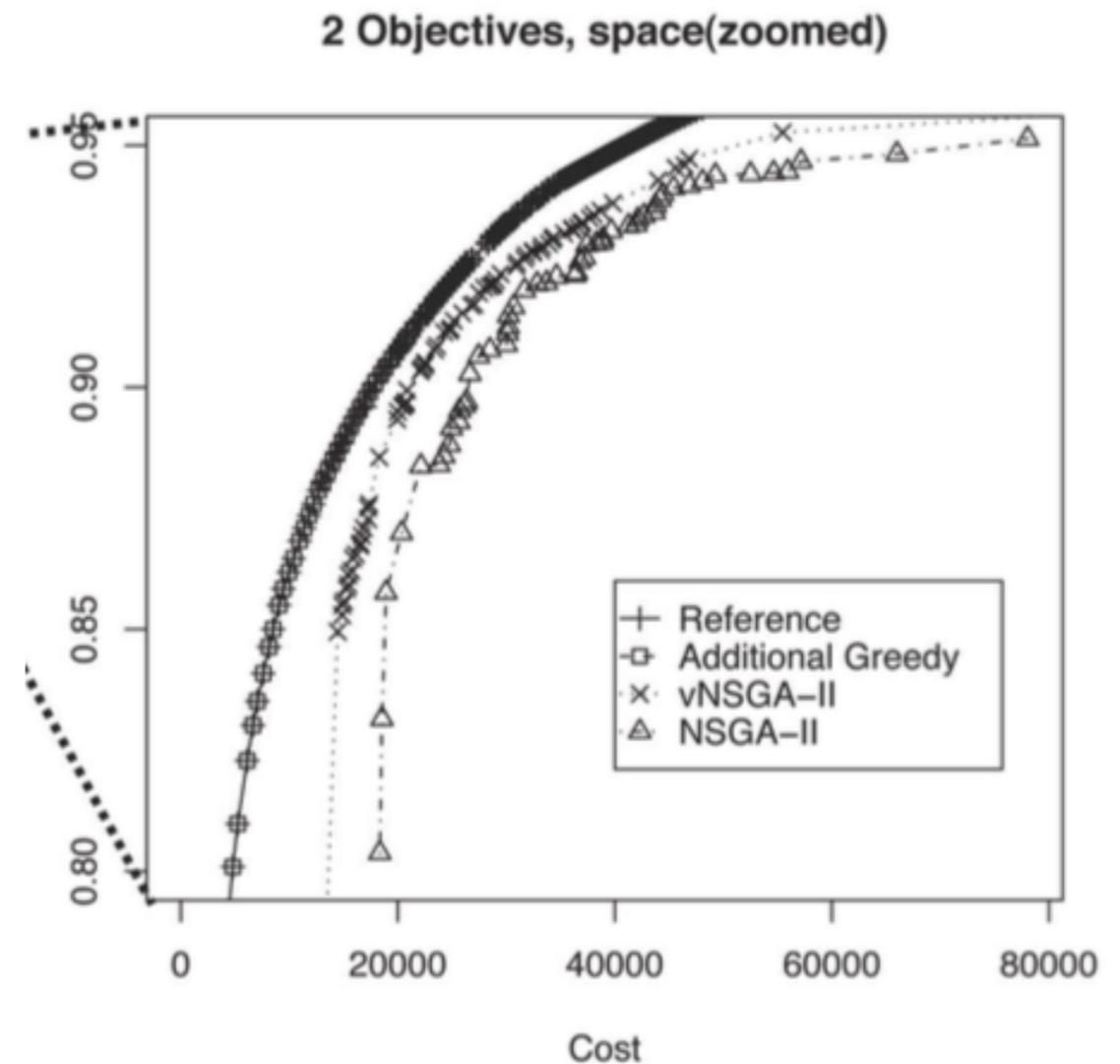


# MOEA Evaluation Metrics

SEP592, Summer 2021  
Shin Yoo

# Comparing Pareto Fronts

- Comparison itself is not as straightforward as comparing two scalar values.
- There is no reference point, as the true Pareto front is usually not known.



# Empirical Evaluation

- Empirical evaluation of different MOEAs becomes a meta-comparison: it is not only about the domain-specific quality, but it is also about the quality of the front itself.
- Properties that we want to evaluate:
  - Closeness to the true Pareto front
  - Diversity of the solutions on the Pareto front

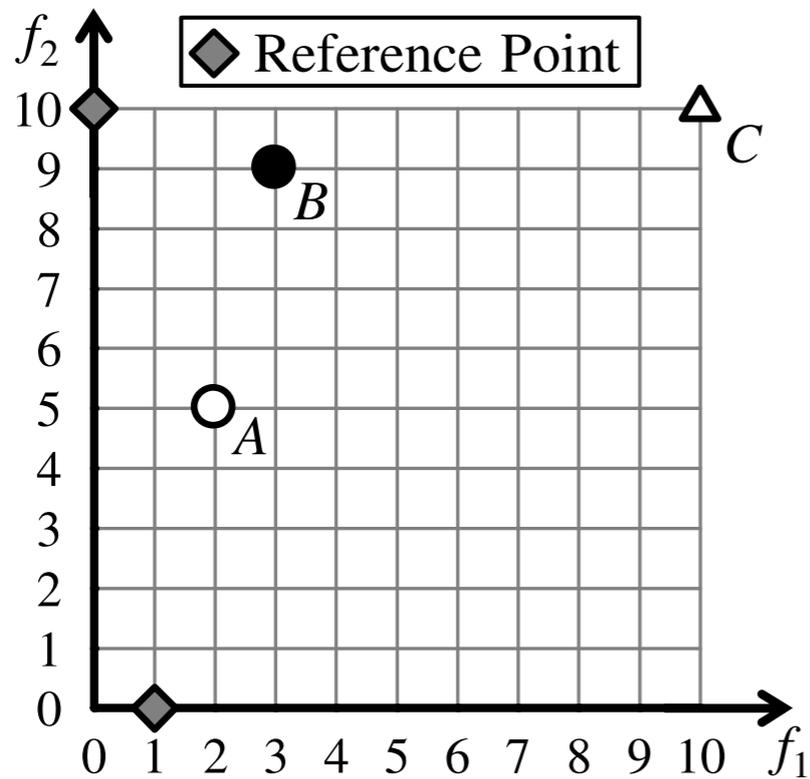
# Closeness to true front

- There are cases where the true Front is known:
  - for example, benchmark optimisation problems.
- For cases where the true front is not known, we use what's called "reference front":
  - Collect all known solutions from all MOEAs involved.
  - Extract a single Pareto front from the collected solutions.
- Reference Pareto Front will include solutions contributed by different MOEAs.

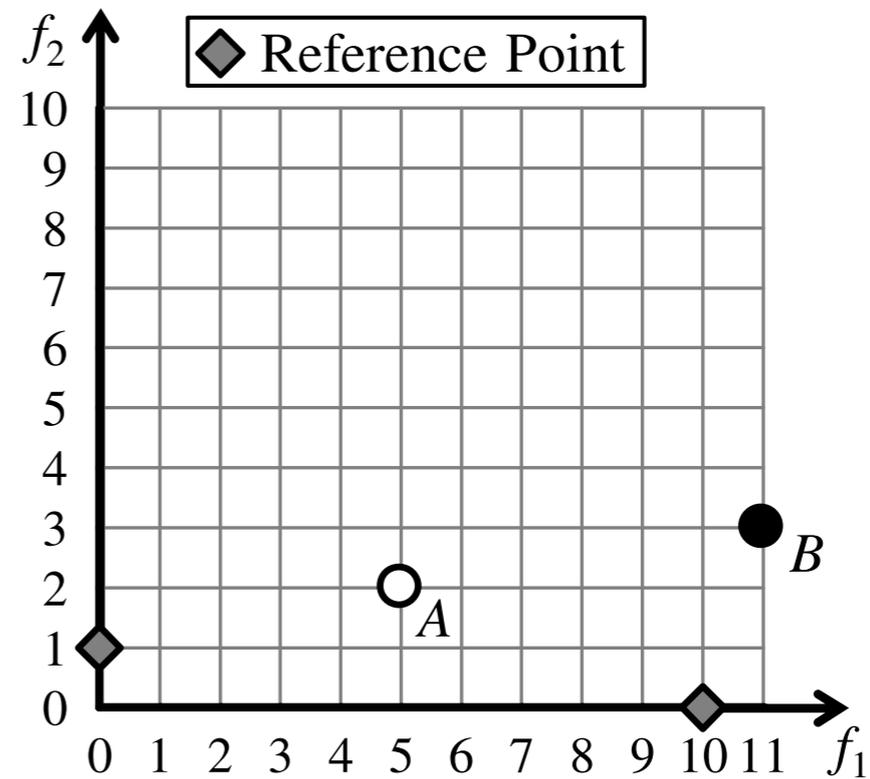
# Generational Distance (GD) and Inverted Generational Distance (IGD)

- GD: average of distance from each solution to its closest reference point.
- IGD: average of distance from each reference point to its closest solution
- The smaller, the better.

# Weaknesses of GD



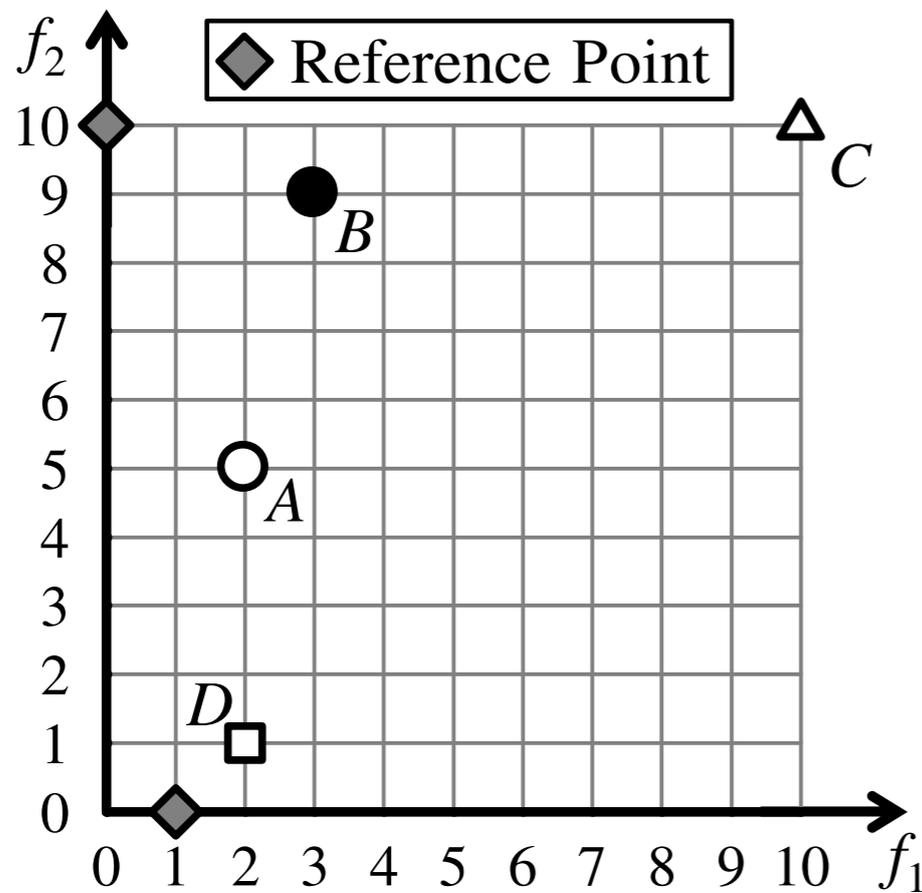
**Fig. 1.** Example 1 (Zitzler et al. [26])



**Fig. 2.** Example 2 (Schütze et al. [18])

B has the shortest distance to its closest reference point, but arguably A is a better solution.

# Weaknesses of GD



IGD is more sensitive to gaps.

$$IGD(A) = \frac{1}{2} \left( \sqrt{(2-0)^2 + (5-10)^2} + \sqrt{(2-1)^2 + (5-0)^2} \right) = 5.24,$$

$$IGD(D) = \frac{1}{2} \left( \sqrt{(2-0)^2 + (1-10)^2} + \sqrt{(2-1)^2 + (1-0)^2} \right) = 5.32.$$

**Fig. 3.** Example 3 with a new solution set  $D$

# Weaknesses of GD/IGD

- When the shape of the true Front is known as a continuous function: reference Pareto front, i.e. a set of points, is sampled from the function.
- How you sample can affect distances
  - For 11 reference points,  $IGD(Z, A) < IGD(Z, B)$
  - For 21 reference points,  $IGD(Z, A) > IGD(Z, B)$

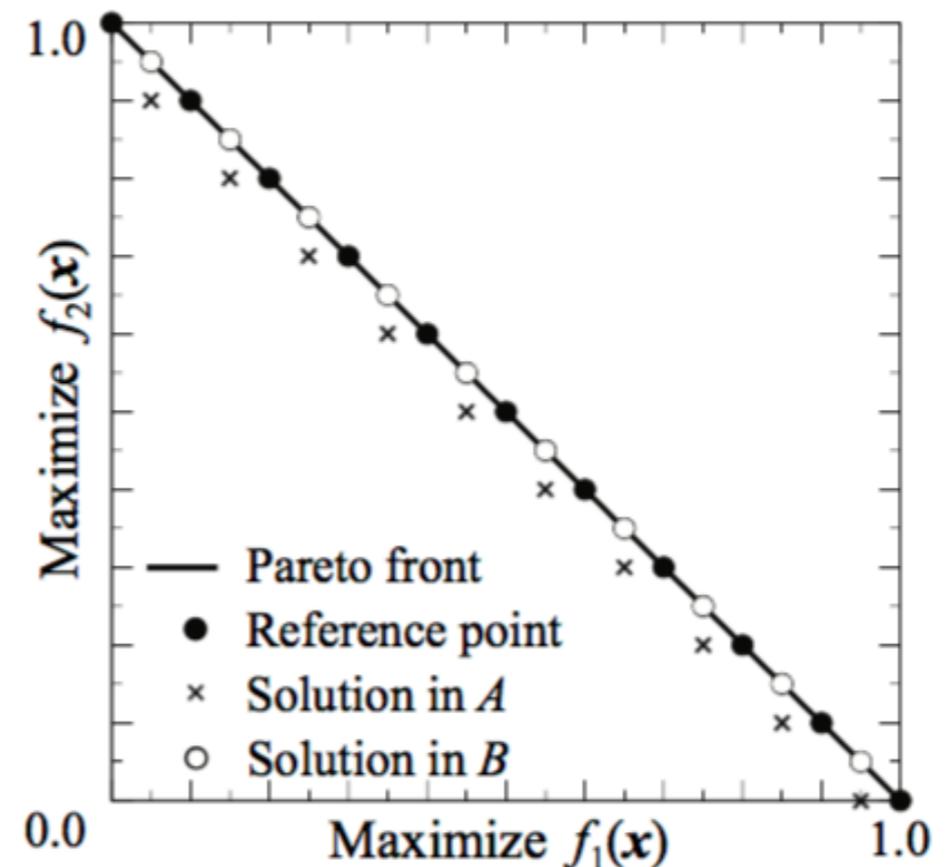


Fig. 3. Reference points ( $H = 10$ ) and two solution sets  $A$  and  $B$ .

H. Ishibuchi, H. Masuda, Y. Tanigaki, and Y. Nojima. Difficulties in specifying reference points to calculate the inverted generational distance for many-objective optimization problems. In Computational Intelligence in Multi-Criteria Decision-Making (MCDM), 2014 IEEE Symposium on, pages 170–177, Dec 2014.

# Weaknesses of IGD

- When using collected reference front, different MOEAs will contribute solutions with different characteristics:
  - some may show strong convergence
  - others may show greater diversity
- Again, this may result in sampling bias when evaluating an MOEA.

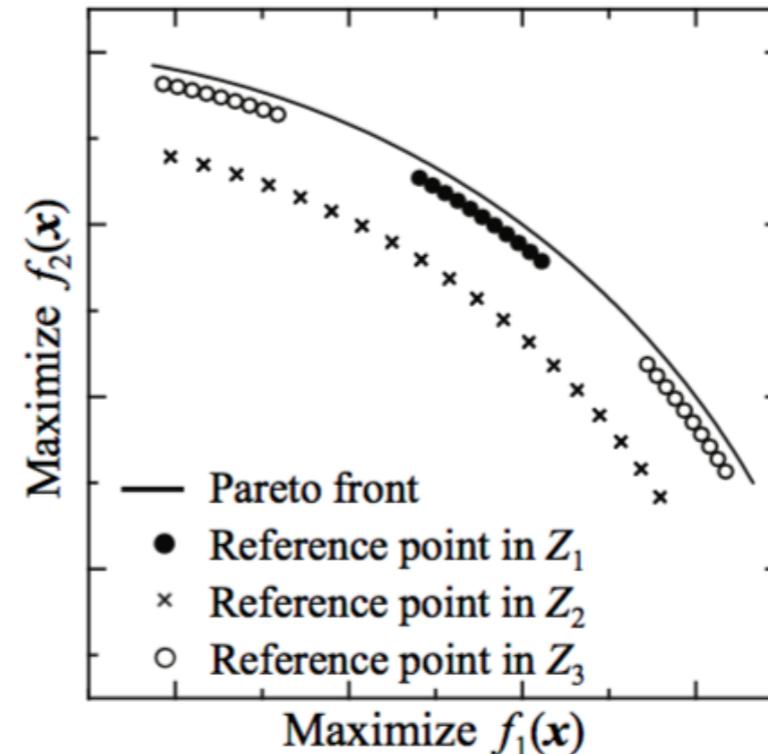


Fig. 7. Three typical situations of reference points: Concentration on the center of the Pareto front ( $Z_1$ ), uniform distribution over the entire Pareto front ( $Z_2$ ), and emphasis on the edges of the Pareto front ( $Z_3$ ).

H. Ishibuchi, H. Masuda, Y. Tanigaki, and Y. Nojima. Difficulties in specifying reference points to calculate the inverted generational distance for many-objective optimization problems. In Computational Intelligence in Multi-Criteria Decision-Making (MCDM), 2014 IEEE Symposium on, pages 170–177, Dec 2014.

# Epsilon

- Binary indicator,  $I_\epsilon(A, B)$ : intuitively, the amount of “shift” required to change  $B$  so that it is weakly dominated by  $A$  (i.e.  $A$  is not worse than  $B$  in all objectives).

$$I_\epsilon(A, B) = \inf_{\epsilon \in \mathbb{R}} \{ \forall z^2 \in B \exists z^1 \in A : z^1 \preceq_\epsilon z^2 \}$$

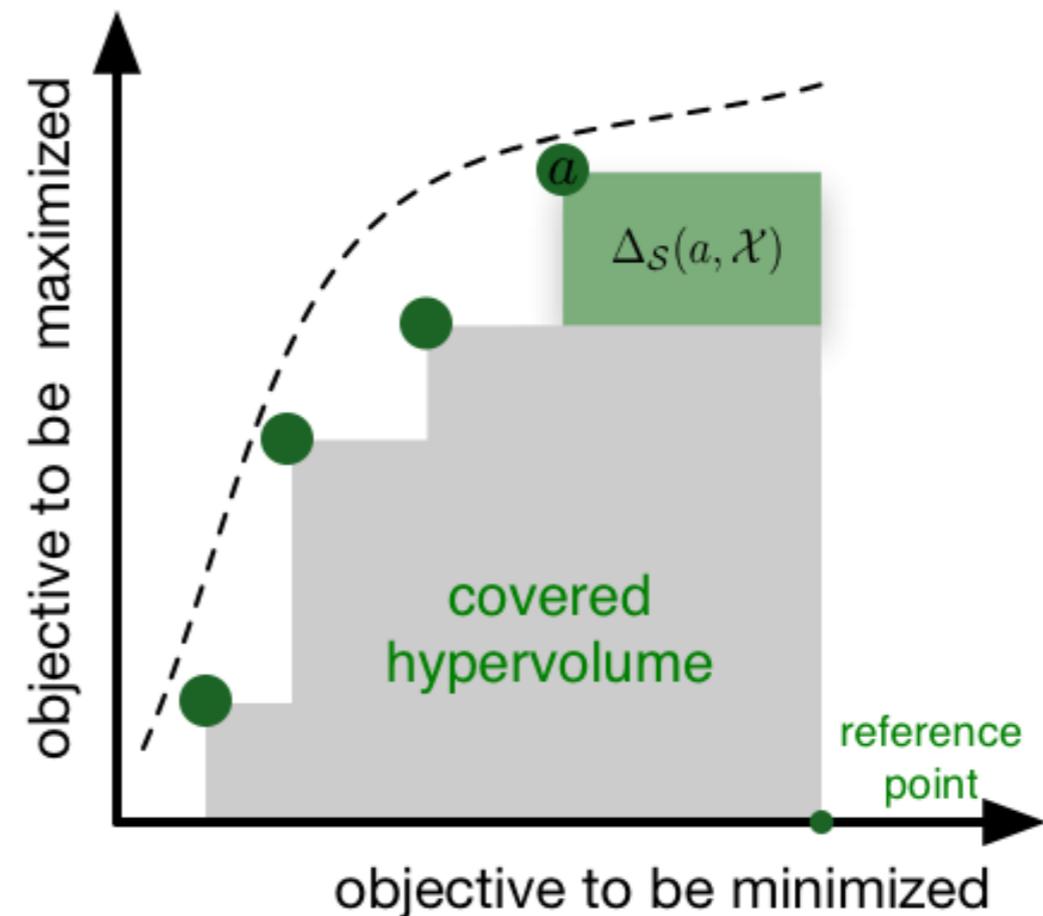
$$z^1 \preceq_\epsilon z^2 \iff \forall i \in 1..n : z_i^1 \leq \epsilon \cdot z_i^2$$

$$z^1 \preceq_{\epsilon+} z^2 \iff \forall i \in 1..n : z_i^1 \leq \epsilon + z_i^2.$$

$$I_\epsilon^1(A) = I_\epsilon(A, R).$$

# Hypervolume

- Intuitively, hypervolume measures the area (space) dominated by a given Pareto front.
- An unary indicator: does not need a reference front.
- Can be thought to measure both convergence and diversity.



# Scaling and Normalisation

- The concept of Pareto optimality itself is independent from scale and normalisation: it is strictly based on partial order only.
- For quality indicators, normalisation may be necessary:
  - so that multiple objectives contribute equally to the indicators.

# Simple Linear Scaling

- ... may be applicable, if the bounds are known.
- If not, there are other normalisation methods.
- ... which leads into the next topic :)

$$z'_i = \frac{z_i - z_i^{\min}}{z_i^{\max} - z_i^{\min}}$$