Non-elitist Evolutionary Multi-objective Optimizers Revisited

Ryoji Tanabe and Hisao Ishibuchi (SUSTech)

Slides and source code are available at my website

Revisiting non-elitist EMO algorithms (EMOAs)

1985 (VEGA)	1999 (SPEA)	2019 (Our work)
Non-elitist EMOAs	Elitist EMOAs	Non-elitist EMOAs

Common belief: Elitist EMOAs always outperform non-elitist EMOAs

- Since 1999, only elitist EMO algorithms have been studied
- NSGA-II, SPEA2, IBEA, MOEA/D, SMS-EMOA, ...

We revisit non-elitist EMOAs for the first TIE time in 20 years

- COMO-CMA-ES [Touré GECCO'19] presented in the EMO1 session
- Target problem domain: Bi-objective continuous optimization
  - Bi-objective BBOB problems [Tusar 16]
- We show a counter-example to the common belief
  - Non-elitist EMOAs can outperform elitist EMOAs under some conditions
- Our results significantly expand the design possibility of EMOAs

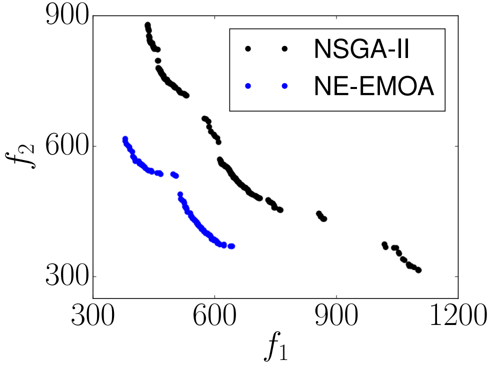
T. Tusar, D. Brockhoff, N. Hansen, and A. Auger. 2016. COCO: The Bi-objective Black Box Optimization Benchmarking (bbob-biobj) Test Suite. CoRR abs/1604.00359 (2016).

Comparison of NSGA-II and a non-elitist EMOA on the 40-dimensional f_{46} in the bi-objective BBOB problem suite

f_{46}: the rotated-Rastrigin (f_1) and the rotated-Rastrigin (f_2)

- The final populations in a single run are shown

The non-elitist EMOA finds a better approximation than NSGA-II



Please do not get angry at my presentation

One reviewer was extremely angry!

Summary of Reviews of pap218s2: Non-elitist Evolutionary Multi-objective Optimizers Revisited

Reviewer	rel	sig	orig	ach	writ	rep	tech	rec	conf
Reviewer 1	5	5	4	4	4	5	5	4-probably accept as full paper (4)	5
Reviewer 2	5	4	5	4	5	5	5	5-definitely accept as full paper (5)	5
Reviewer 3	4	3	3	2	4	3	1	2-probably accept as poster (2)	5
Reviewer 4	5	4	4	4	5	4	3	4-probably accept as full paper (4)	4
Averages:	**4.8**	**4.0**	**4.0**	**3.5**	**4.5**	**4.3**	**3.5**	**3.8**	**4.8**

Bad

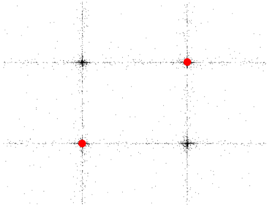
- Grrrrrrrrr! Elitist EMOAs must outperform non-elitist EMOAs!
- Terrible! A was not done! B was not done! ... Z was not done!

Good

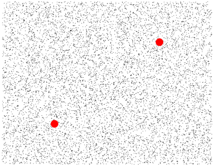
- Elitist EMOAs may be outperformed by non-elitist EMOAs
- Blue ocean! I have a lot to do! Homework for GECCO2020!

Five crossover methods in GAs for continuous optimization

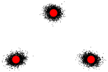
(a) SBX+PM (Deb 95)



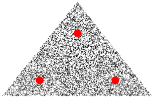
(b) BLX (Eshelman 92)



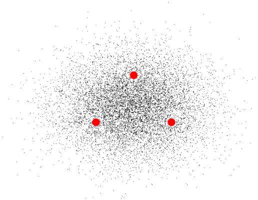
(c) PCX (Deb 02)



(d) SPX (Tsutsui 99)



(e) REX (Akimoto 10)



A "simple" EMO framework analyzed in this work

Initialize the population $\boldsymbol{P} = \{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(\mu)}\}$;
while Not happy **do**
 $\boldsymbol{R} \leftarrow$ Randomly select k parents from $\boldsymbol{P}$;
 $\boldsymbol{Q} \leftarrow$ Generate λ children by applying the crossover method to $\boldsymbol{R}$;
 $\boldsymbol{P} \leftarrow$ Apply the environmental selection $(\boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{R})$;

1. Best-all: the traditional elitist $(\mu + \lambda)$-selection
   - The best μ individuals are selected from $\boldsymbol{P} \cup \boldsymbol{Q}$
2. Best-family: An elitist restricted selection
   - The selection is performed only among the "family" $\boldsymbol{R} \cup \boldsymbol{Q}$
   - The best k individuals are selected from the $k + \lambda$ individuals
3. Best-children: A non-elitist restricted selection (not (μ, λ)-selection)
   - An extended version of JGG [Akimoto 10] for single-obj. opt.
   - The k parents in $\boldsymbol{R}$ are always removed from $\boldsymbol{P}$
   - The best k individuals are selected from the λ children in $\boldsymbol{Q}$

Youhei Akimoto. Design of Evolutionary Computation for Continuous Optimization. Ph.D. Dissertation. Tokyo Institute of Technology (2010)

The "simple" EMO framework analyzed in this work (continued)

Summary of the three environmental selections

	Elitism?	Restricted?	Max. replacements
Best-all	Yes	No	Pop. size μ
Best-family	Yes	Yes	Num. parents k
Best-children	No	Yes	Num. parents k

The EMOA requires a ranking method to select the best individuals

- The EMOA can be combined with any ranking method
  - Similar to MO-CMA-ES
- Ranking methods in NSGA-II, SMS-EMOA, SPEA2, and IBEA
  - Their results are similar

The ranking method in NSGA-II

1. Individuals are ranked based on their non-domination levels
2. Ties are broken by the crowding distance

Experimental settings

Problem suite

- Experiments were performed using the COCO platform [Hansen 16]
- 55 bi-objective BBOB problems [Tusar 16]
- Number of decision variables $n \in \{2, 3, 5, 10, 20, 40\}$
- Number of function evaluations: $10^4 \times n$

Performance measure in COCO

- Roughly speaking, hypervolume value of non-dominated solutions in the unbounded external archive

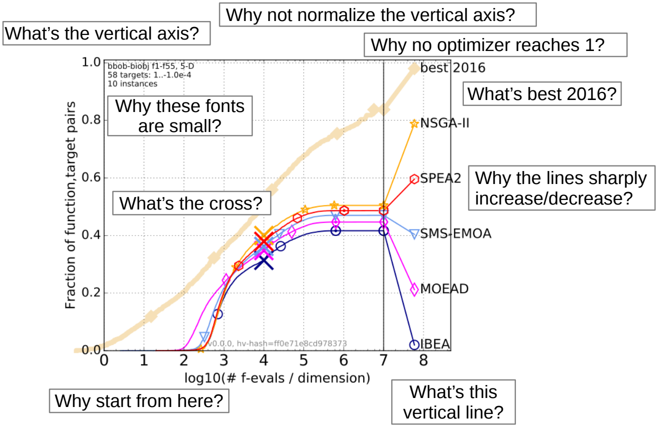
EMO algorithms

- EMOAs were implemented using jMetal [Durillo 11]
- Population size $\mu = \lfloor 100 \ln(n) \rfloor$
- Number of children $\lambda = 10 \times n$
- Number of parents $k = 2$ for SBX and BLX
  - $k = n + 1$ for PCX, SPX, and REX

N. Hansen, A. Auger, O. Mersmann, T. Tusar, and D. Brockhoff. COCO: A Platform for Comparing Continuous Optimizers in a Black-Box Setting. CoRR abs/1603.08785 (2016).

J. José Durillo and A. J. Nebro. jMetal: A Java framework for multi-objective optimization. Adv. Eng. Softw. 42, 10 (2011), 760771.

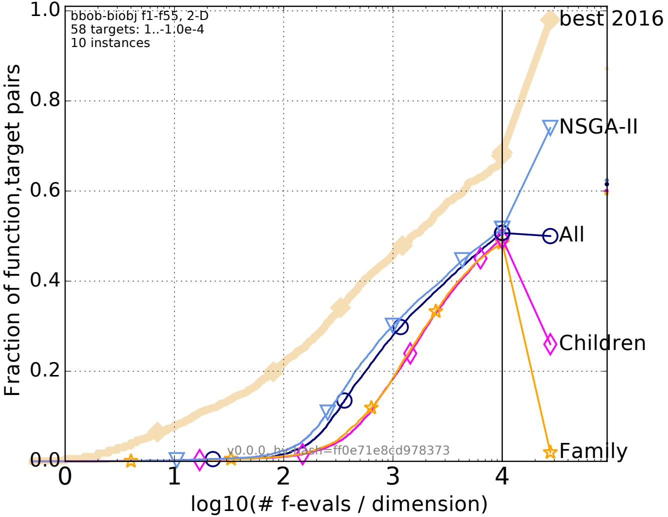
I am tired of explaining how to read ECDF figures



- It is tooooo time-consuming
- Please see the guideline [Hansen 16] after this presentation
  - Don't think now. Feel

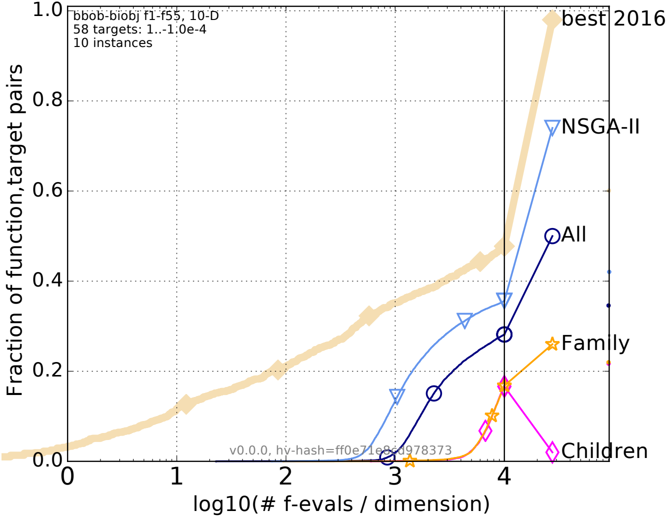
Comparison on all the 55 bi-objective BBOB problems (SBX, $n = 2$)

NSGA-II outperforms best-all, best-family, and best-children



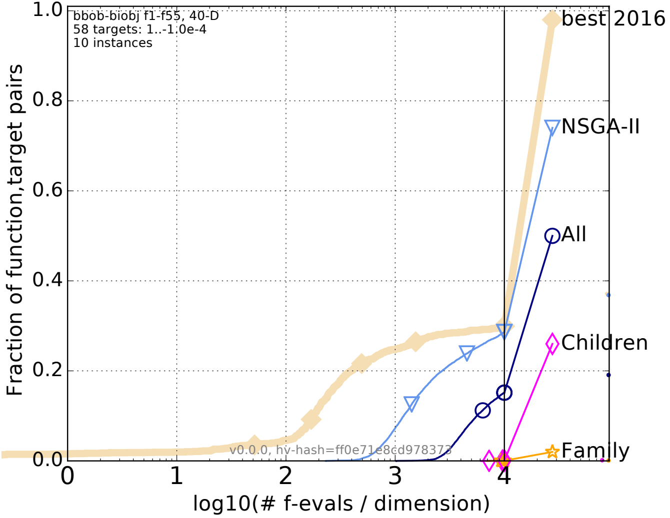
Comparison on all the 55 bi-objective BBOB problems (SBX, $n = 10$)

NSGA-II outperforms best-all, best-family, and best-children



Comparison on all the 55 bi-objective BBOB problems (SBX, $n = 40$)

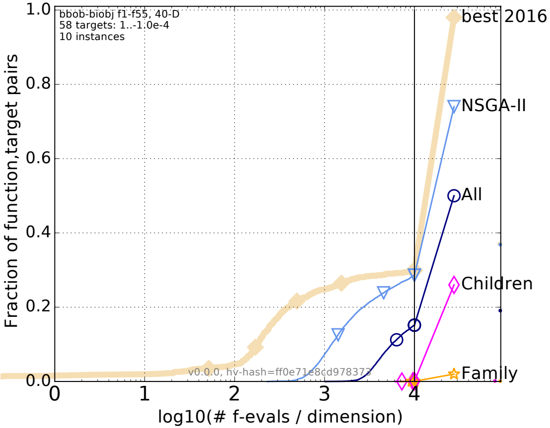
NSGA-II outperforms best-all, best-family, and best-children



Summary of the results when using SBX

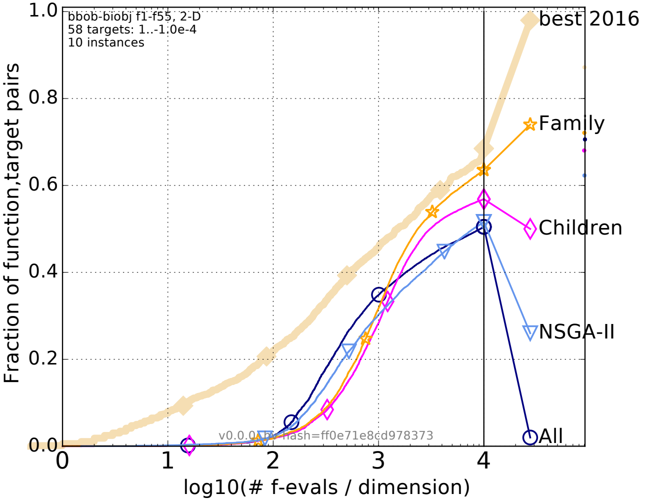
Restricted best-family and best-children perform the worst

- NSGA-II performs the best
- Results are consistent with previous studies
- Results of SBX, BLX, and PCX are similar



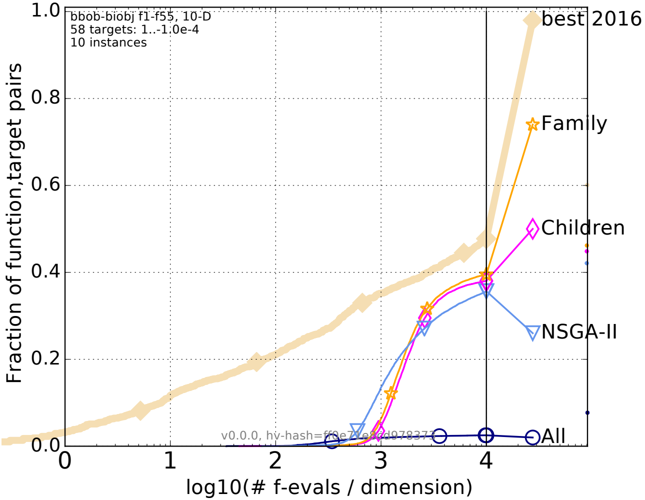
Comparison on all the 55 bi-objective BBOB problems (SPX, $n = 2$)

- Best-family performs the best after $10^3 \times n$ function evaluations
- Best-children performs the second best



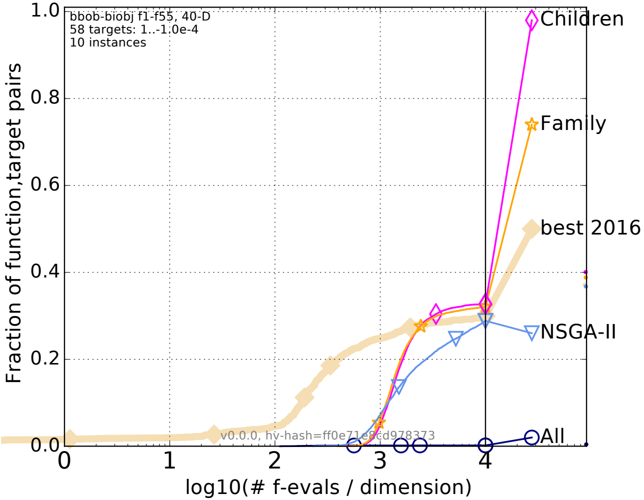
Comparison on all the 55 bi-objective BBOB problems (SPX, $n = 10$)

- Best-family performs the best after $2 \times 10^3 \times n$ function evaluations
- Difference between best-family and best-children is small



Comparison on all the 55 bi-objective BBOB problems (SPX, $n = 40$)

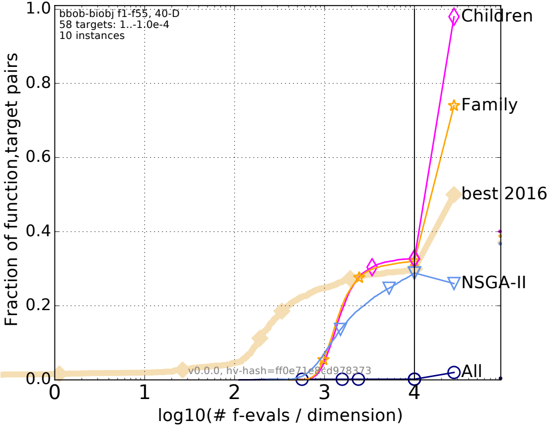
- Best-children performs the best after $2 \times 10^3 \times n$ function evaluations
- Best-family performs the second best



Summary of the results when using SPX

Non-elitist best-children performs best on the 40-dimensional problems

- Best-family performs the best for $n < 40$
- Two restricted selections (best-family and best-children) work well
- Results using SPX and REX are similar



Why SPX is unsuitable for the traditional $(\mu + \lambda)$ best-all?
Why SPX is suitable for the restricted best-family and best-children?

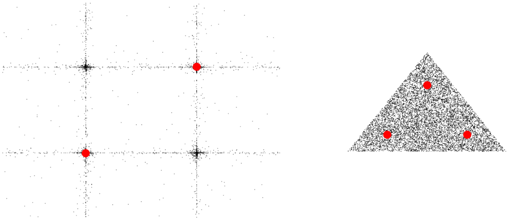
Recall the property of the three environmental selections

	Elitism?	Restricted?	Max. replacements
Best-all	Yes	No	Pop. size μ
Best-family	Yes	Yes	Num. parents k
Best-children	No	Yes	Num. parents k

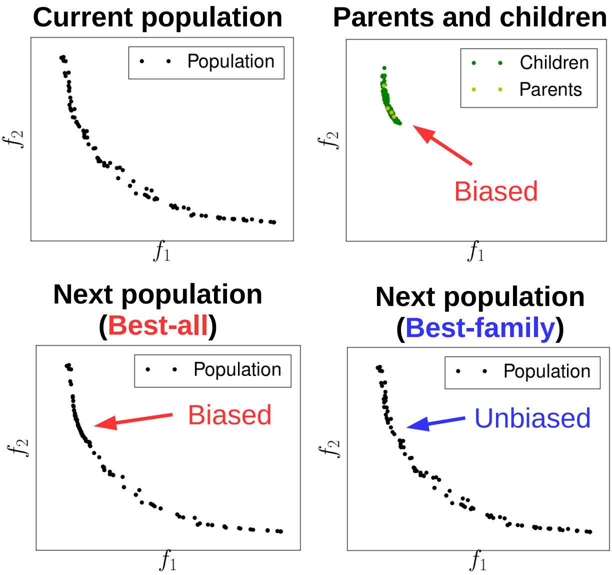
Answer: Restricted selection can prevent the premature convergence

- SPX can generate children near the parents when λ is enough large
  - This causes the premature convergence in best-all

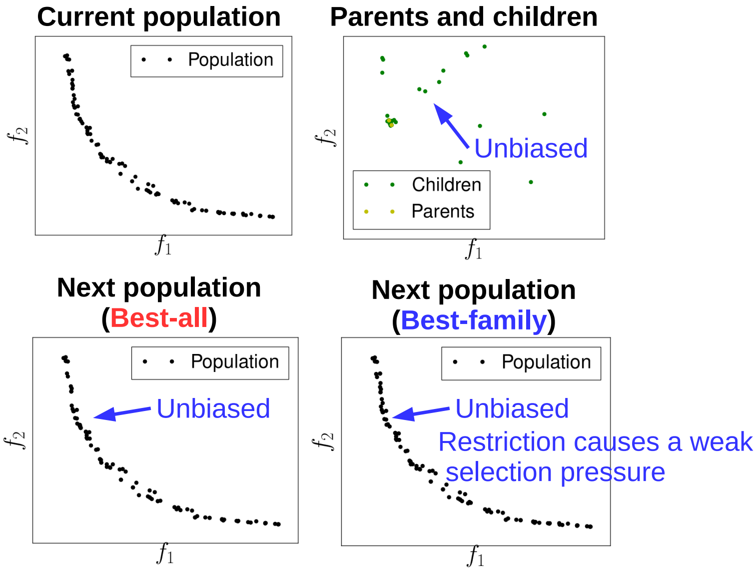
(a) SBX+PM (Deb 95) (b) SPX (Tsutsui 99)



Effect of the restricted selection on the 3-dim f_1 (when using SPX)



Bad effect of the restricted selection on the 3-dim f_1 (when using SBX)

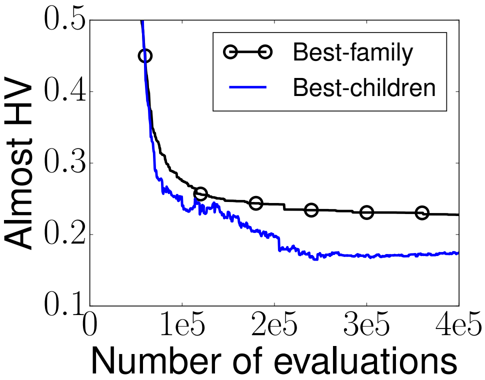


Advantage of the non-elitist best-children selection

Non-elitist selections can accept "uphill" moves as in simulated annealing

- Elitist selections (best-family) can accept only "downhill" moves
- Uphill moves help the population to escape from local optima
- Benefit of the non-elitist selection is consistent with [Akimoto 10]

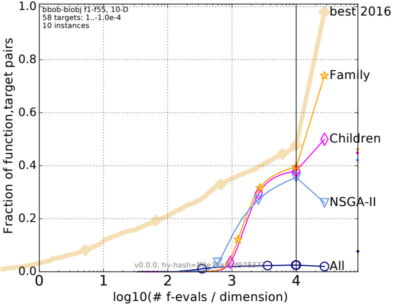
Results on the 40-dimensional f_{46} function (Rast./Rast.)



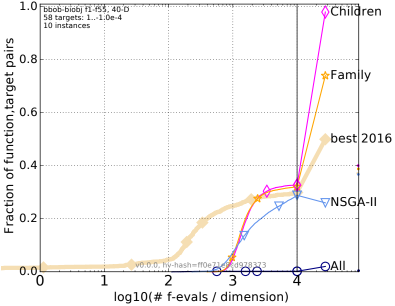
Disadvantages of the non-elitist best-children

1. Poor performance on problems $n < 40$
   - Best-children is outperformed by best-family for $n < 40$
   - Reason is unclear
2. Slow convergence
   - Best-children performs worse than best-family at the early stage
   - Reason is similar to the relation between $(1, \lambda)$-ES and $(1 + \lambda)$-ES

(a) $n = 10$



(b) $n = 40$



Conclusion

We revisited the performance of non-elitist EMOAs

- We examined the three environmental selections
  - Two elitist selections: best-all and best-family
  - One non-elitist selection: best-children
- Results show that best-children performs better than best-all and best-family on the bi-objective BBOB problems with $n = 40$
  - When using rotational invariant SPX and REX
  - Similar results are found in NSGA-II, SPEA2, IBEA, and SMS
- A counter-example to the common belief
  - Non-elitist EMOAs can outperform elitist EMOAs
  - Not claim: non-elitist EMOAs always outperform elitist ones

Many future works

- Investigating the scalability with respect to the num. of objectives
- Designing a non-elitist decomposition-based EMOA
- Designing a non-elitist MO-CMA-ES (but difficult)

Q. Why no one has tried to design non-elitist EMOAs for 20 years?
A. Because of DTLZ and WFG

DTLZ and WFG have produced many elitist EMOAs

1. Only the DTLZ and WFG test problems have been available
2. Only SBX+PM works well on DTLZ and WFG
    - Because of the position and distance variables [Ishibuchi 17]
3. Only elitist EMOAs fit for SBX
4. Only elitist EMOAs with SBX have been studied

BBOB may produce many non-elitist EMOAs

1. BBOB is now available
2. Rotational invariant operators (e.g., SPX and REX) work well on most BBOB problems
3. Non-elitist EMOAs fit for some rotational invariant operators
4. Non-elitist EMOAs may be studied

H. Ishibuchi, Y. Setoguchi, H. Masuda, Y. Nojima: Performance of Decomposition-Based Many-Objective Algorithms Strongly Depends on Pareto Front Shapes. IEEE Trans. Evolutionary Computation 21(2): 169-190 (2017)

Traditional benchmarking scenario for EMOAs

The traditional benchmarking scenario

- Nondominated solutions in the population P are used
- E.g., the hypervolume value of P is the performance of an EMOA

Issues of the traditional benchmarking scenario

1. Difficulty in comparing EMOAs with different population sizes [Ishibuchi 16]
   - The appropriate population size differ depending on EMOAs
   - Solution set with different sizes cannot be compared in a fair manner
2. Difficulty in maintaining good solutions
   - Good potential solutions found so far are likely to be discarded from the population

Benchmarking scenario with an unbounded external archive

Benchmarking scenario

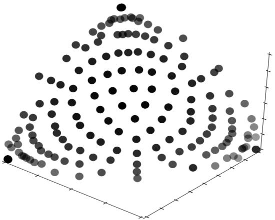
- The unbounded external archive $\boldsymbol{A}$ stores all nondominated solutions found so far
- All nondominated solutions in $\boldsymbol{A}$ are used
- E.g., the hypervolume value of $\boldsymbol{A}$ is the performance of an EMOA
- The unbounded external archive addresses the issues of the traditional benchmarking scenario

Post-processing methods for decision making

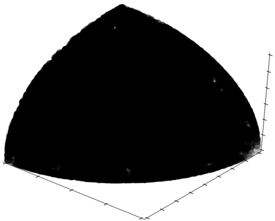
- If the decision maker wants to examine a small number of solutions, a post-processing method can be applied to

Three solution sets (MOEA/D with the Tchebycheff function, WFG4)

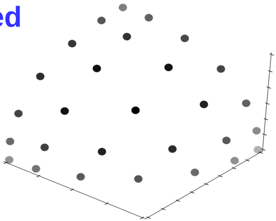
(i) 210 solutions in the population



(ii) 8,120 solutions in the unbounded external archive



(iii) 28 solutions selected from the unbounded external archive



1. Best-all: An elitist $(\mu + \lambda)$-selection

The best individual is repeatedly added to the next $\boldsymbol{P}$

1. Assign ranks to $\mu + \lambda$ individuals in $\boldsymbol{P} \cup \boldsymbol{Q}$
2. Let $\boldsymbol{S}$ be $\boldsymbol{P} \cup \boldsymbol{Q}$
3. Remove all μ individuals from $\boldsymbol{P}$
4. Until $|\boldsymbol{P}| = \mu$, repeatedly select the best $\boldsymbol{x}$ from $\boldsymbol{S}$, adding $\boldsymbol{x}$ to $\boldsymbol{P}$

2. Best-family: An elitist restricted selection

The selection is performed only among the so-called “family” ($\boldsymbol{R} \cup \boldsymbol{Q}$)

1. Assign ranks to $\mu + \lambda$ individuals in $\boldsymbol{P} \cup \boldsymbol{Q}$
2. Let $\boldsymbol{S}$ be $\boldsymbol{R} \cup \boldsymbol{Q}$
3. Remove all k individuals in $\boldsymbol{R}$ from $\boldsymbol{P}$ (i.e., $\boldsymbol{P} \leftarrow \boldsymbol{P} \backslash \boldsymbol{R}$)
4. Until $|\boldsymbol{P}| = \mu$, repeatedly select the best $\boldsymbol{x}$ from $\boldsymbol{S}$, adding $\boldsymbol{x}$ to $\boldsymbol{P}$

3. Best-children: A non-elitist restricted selection ($\lambda > k$)

The selection is performed only among the children

1. Remove all k individuals in $\boldsymbol{R}$ from $\boldsymbol{P}$ (i.e., $\boldsymbol{P} \leftarrow \boldsymbol{P} \backslash \boldsymbol{R}$)
2. Assign ranks to $\mu - k + \lambda$ individuals in $\boldsymbol{P} \cup \boldsymbol{Q}$
3. Until $|\boldsymbol{P}| = \mu$, repeatedly select the best $\boldsymbol{x}$ from $\boldsymbol{Q}$, adding $\boldsymbol{x}$ to $\boldsymbol{P}$

Best-children is an extended version of JGG [Akimoto 10]

Just generation gap (JGG)

- A non-elitist selection in GA for single-objective optimization
- When using SPX, GA with JGG significantly outperforms elitist GAs

JGG assigns the rank to each child absolutely

- Children $\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(\lambda)}$ are ranked based on their objective values $f(\boldsymbol{x}^{(1)}), ..., f(\boldsymbol{x}^{(\lambda)})$

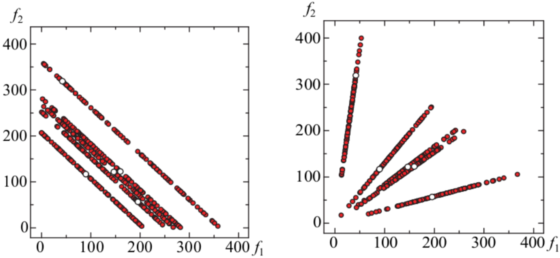
Best-children assigns the rank to each child relatively

- Children $\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(\lambda)}$ are ranked based on their objective vectors $\boldsymbol{f}(\boldsymbol{x}^{(1)}), ..., \boldsymbol{f}(\boldsymbol{x}^{(\lambda)})$ and objective vectors of individuals in $\boldsymbol{P}$
- Individuals in $\boldsymbol{P}$ do not directly participate in the selection process
- But, they indirectly contribute to assign ranks to the children

Youhei Akimoto. Design of Evolutionary Computation for Continuous Optimization. Ph.D. Dissertation. Tokyo Institute of Technology (2010)

Q. Why has only SBX been used in the EMO community?
A. Because SBX specially works well on DTLZ and WFG [Ishibuchi 17]

$$\boldsymbol{x} = (\underbrace{x_1, \ldots, x_{m-1}}_{\text{Position variables}}, \underbrace{x_m, \ldots, x_n}_{\text{Distance variables}})$$



H. Ishibuchi, Y. Setoguchi, H. Masuda, Y. Nojima: Performance of Decomposition-Based Many-Objective Algorithms Strongly Depends on Pareto Front Shapes. IEEE Trans. Evolutionary Computation 21(2): 169-190 (2017)

Traditional $(\mu + \lambda)$ best-all selection is in a dilemma

- A large λ value is helpful to exploit the current search area
  - But it causes premature convergence
- A small λ value can prevent from the premature convergence
  - But it is not sufficiently large to exploit the current search area