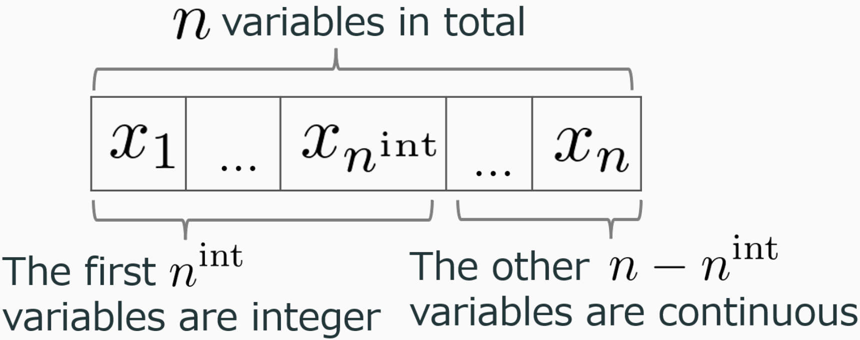
# Benchmarking Parameter Control Methods in DE for Mixed-Integer Black-Box Optimization

GECCO 2024 at Melbourne

Ryoji Tanabe

Yokohama National University
Yokohama city, Japan

# Mixed-integer black-box optimization in this work




Note: Categorical variables are not considered

# Mixed-integer BBOB (bbob-mixint) [Tušar 19]

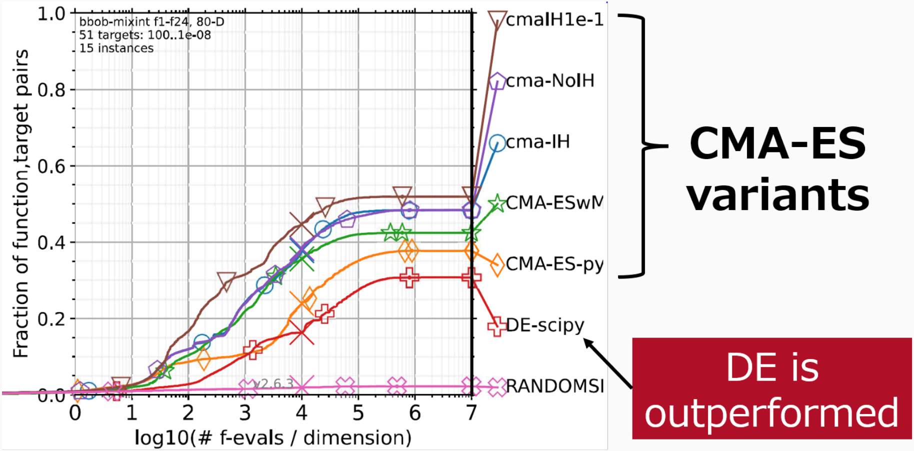
- bbob-mixint consists of 24 mixed-integer functions
  - They are mixed-integer versions of the 24 bbob functions
  - The first $4n/5$ variables are integer

Value ranges for $n = 10$

| $x$ | Range | |
|---|---|---|
| $x_1, x_2$ | $\{0, 1\}$ | The first 8 variables are integer |
| $x_3, x_4$ | $\{0, 1, 2, 3\}$ | |
| $x_5, x_6$ | $\{0, 1, \ldots, 7\}$ | |
| $x_7, x_8$ | $\{0, 1, \ldots, 15\}$ | |
| $x_9, x_{10}$ | $[-5, 5]$ | The other 2 variables are continuous |

Tea Tušar, Dimo Brockhoff, Nikolaus Hansen: Mixed-integer benchmark problems for single- and bi-objective optimization. GECCO 2019: 718-726

# CMA-ES variants outperform the Scipy imple. of differential evolution (DE) on `bbob-mixint` [Marty 23]




Tristan Marty, Yann Semet, Anne Auger, Sébastien Héron, Nikolaus Hansen : Benchmarking CMA-ES with Basic Integer Handling on a Mixed-Integer Test Problem Suite. GECCO Companion 2023: 1628-1635

## Contribution: This work benchmarks parameter control methods (PCMs) in DE on `bbob-mixint`

- Two critical parameters in DE
  - Scale factor $s > 0$ for differential mutation
  - Crossover rate $c \in [0, 1]$ for crossover
  - Best settings of $s$ and $c$ depend on a problem
- DE requires a PCM that automatically adjusts $s$ and $c$
  - Their importance is well known for continuous optimization
  - All state-of-the-art DEs (e.g., L-SHADE) use them
- Their effectiveness is unknown for mixed-integer opt.
  - The performance of DE can possibly be improved

# Scale factor $s$ and crossover rate $c$ are critical in DE

- For each iteration of DE, the $\mu$ individuals $\mathbf{x}_1, \ldots, \mathbf{x}_\mu$ generate the $\mu$ children $\mathbf{u}_1, \ldots, \mathbf{u}_\mu$
    - In environmental selection, $\mathbf{u}_i$ is compared only to $\mathbf{x}_i$
        - For each $i \in \{1, \ldots, \mu\}$, if $f(\mathbf{u}_i) \leq f(\mathbf{x}_i)$, $\mathbf{x}_i \leftarrow \mathbf{u}_i$
- Scale factor $s$ determines the magnitude of mutation

$$\mathbf{v}_i = \mathbf{x}_{r_1} + s \left(\mathbf{x}_{r_2} - \mathbf{x}_{r_3}\right)$$

- Crossover rate $c$ determines the number of elements inherited from each parent $\mathbf{x}$ to a child $\mathbf{u}$

$$u_{i,j} = \begin{cases} v_{i,j} & \text{if } \mathrm{rand}[0,1] \leq c \text{ or } j = j_{\mathrm{rand}} \\ x_{i,j} & \text{otherwise} \end{cases}$$

# This work focuses only on PCMs in DE

I am interested in “the PCM in jDE”, not “jDE”

**DE instances**

| DE template | jDE | CoDE |
| --- | --- | --- |
| Mutation strategy | rand/1 | rand/1, rand/2, curr-to-rand/1 |
| Crossover method | binomial | binomial |
| PCM | PCM in jDE | PCM in CoDE |
| : | : | : |

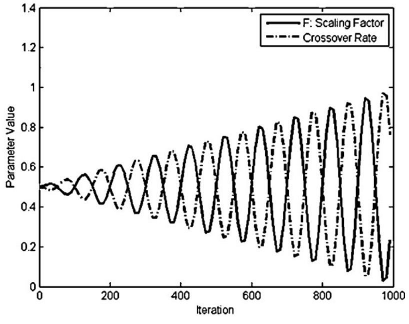
cDE, SaDE, EPSDE, SinDE, JADE, ...

# This work benchmarks 9 PCMs in DE

- 3 deterministic PCMs
  - PCM in CoDE [Wang 11]
  - PCM in SinDE [Draa 15]
  - PCM in CaRS+S [Molina-Pérez 24]
- 6 adaptive PCMs
  - PCM in jDE [Brest 06]
  - PCM in cDE [Tvrdık 06]
  - PCM in JADE [Zhang 09]
  - PCM in EPSDE [Mallipeddi 11]
  - PCM in SHADE [Tanabe 13]
  - PCM in CoBiDE [Wang 14]

# Deterministic parameter control methods in DE

- Generating $s$ and $c$ values based on a simple rule
  - without using any information obtained by the search
- Example: The PCM in SinDE [Draa 15]
  - $s$ and $c$ are determined based on the sinusoidal functions



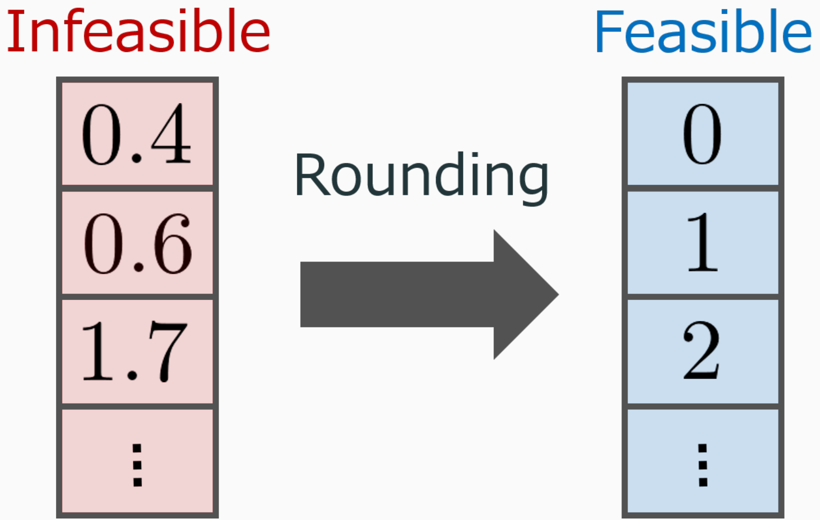

Amer Draa, Samira Bouzoubia, Imene Boukhalfa: A sinusoidal differential evolution algorithm for numerical optimisation. Appl. Soft Comput. 27: 99-126 (2015)

# Most adaptive PCMs in DE use a success criterion

- If the child $\mathbf{u}_i$ is better than its parent $\mathbf{x}_i$, the pair of $s_i$ and $c_i$ are said to be *successful*
    - Assumption: The better child was generated because the pair of $s_i$ and $c_i$ was suitable for a target problem
- Example: The PCM in jDE [Brest 06]


Janez Brest, Saso Greiner, Borko Boskovic, Marjan Mernik, Viljem Zumer: Self-Adapting Control Parameters in Differential Evolution: A Comparative Study on Numerical Benchmark Problems. IEEE Trans. Evol. Comput. 10(6): 646-657 (2006)

# Rounding operator to repair infeasible solutions



# The Lamarckian and Baldwinian repair methods

- Let $\mathbf{y}$ be a repaired feasible version of an infeasible $\mathbf{x}$
- Both of them use $f(\mathbf{y})$ as $f(\mathbf{x})$, but ...
  - The Lamarckian one *replaces* $\mathbf{x}$ with $\mathbf{y}$
  - The Baldwinian one *doesn't* replace $\mathbf{x}$ with $\mathbf{y}$
- All individuals in the population are ...
  - always feasible for the Lamarckian one
  - likely to be infeasible for the Baldwinian one
- There is no clear winner between them [Salcedo-Sanz 09]

---

Sancho Salcedo-Sanz: A survey of repair methods used as constraint handling techniques in evolutionary algorithms. Comput. Sci. Rev. 3(3): 175-192 (2009)

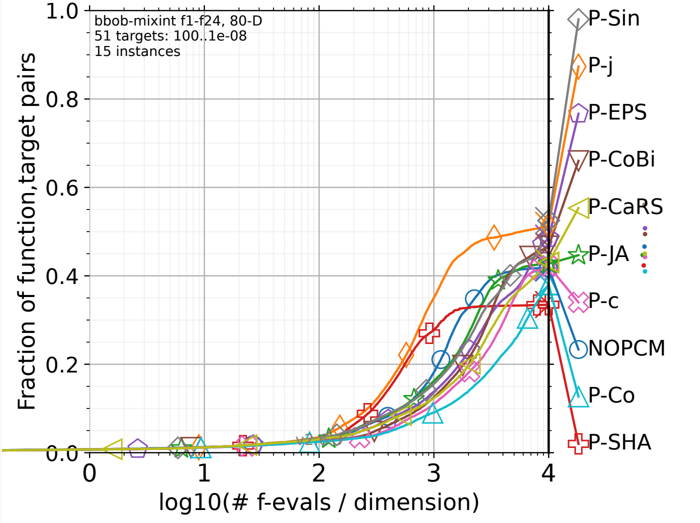
# Experimental setup

- The COCO platform [Hansen 21]
  - The 24 `bbob-mixint` functions
  - Num. variables $n \in \{5, 10, 20, 40, 80, 160\}$
- Settings for DE: $(9 + 1) \times 8 \times 2 = 160$ configurations
  - The max. fun. evals. was set to $10^4 \times n$
  - The population size $\mu$: $100$
  - Hyper-par. in the 9 PCMs: default
  - For DE with no PCM, $s = 0.5$ and $c = 0.9$
  - Eight representative mutation strategies
  - The Lamarckian and Baldwinian repair methods

Nikolaus Hansen, Anne Auger, Raymond Ros, Olaf Mersmann, Tea Tušar, Dimo Brockhoff: COCO: a platform for comparing continuous optimizers in a black-box setting. Optim. Methods Softw. 36(1): 114-144 (2021)

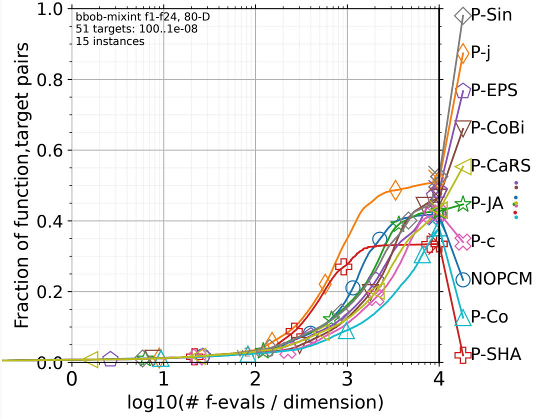
# Results for $n = 80$ (rand/1, Baldwinian)

PCMs can improve the performance of DE



# Results for $n = 80$ (rand/1, Baldwinian)

- DE with the PCM in SHADE performs poorly
  - P-SHA is SOTA for continuous bbob
  - Rankings for bbob and bbob-mixint are different



# Best PCM depends on the type of mutation strategy

Results of the Baldwinian repair method

| Strategy | $n = 5$ | $n = 10$ | $n = 20$ | $n = 40$ | $n = 80$ | $n = 160$ |
|---|---|---|---|---|---|---|
| rand/1 | NOPCM | P-CoBi | P-c | P-Sin | P-Sin | P-Sin |
| rand/2 | P-Sin | P-Sin | P-Sin | P-Sin | P-j | P-j |
| best/1 | P-Co | P-Co | P-Co | P-Co | P-Co | P-JA |
| best/2 | P-CoBi | P-CoBi | P-EPS | P-CoBi | P-c | P-c |
| ctb/1 | P-CoBi | P-Co | P-Co | P-CoBi | P-Co | P-JA |
| ctr/1 | P-CoBi | P-CoBi | P-CoBi | P-CoBi | P-CoBi | P-CoBi |
| ct$p$/1 | P-Co | P-CoBi | P-Co | P-Co | P-Co | P-Co |
| rt$p$/1 | P-Co | P-Co | P-Co | P-Co | P-Co | P-Co |

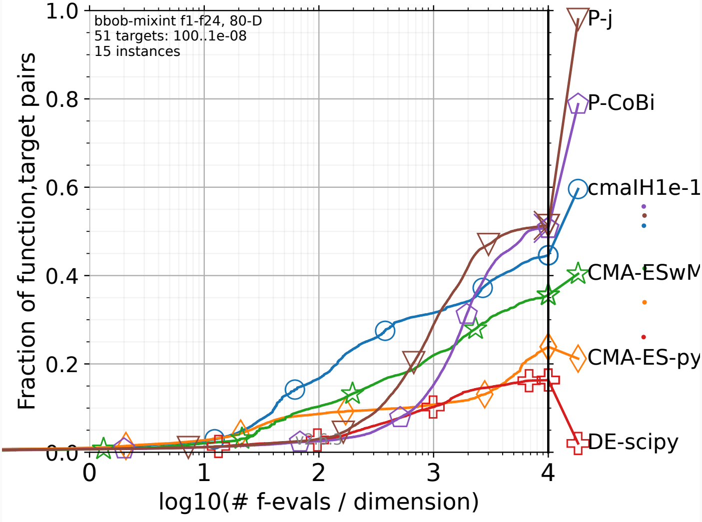
# Best PCM depends on the type of repair method

Results of the Lamarckian repair method

| Strategy | $n = 5$ | $n = 10$ | $n = 20$ | $n = 40$ | $n = 80$ | $n = 160$ |
|---|---|---|---|---|---|---|
| rand/1 | P-CoBi | P-CoBi | P-CoBi | P-CaRS | P-CoBi | P-JA |
| rand/2 | P-Sin | P-Sin | P-CoBi | P-Sin | P-CoBi | P-CoBi |
| best/1 | P-Co | P-Co | P-Co | P-Co | P-Co | P-Co |
| best/2 | P-CoBi | P-CoBi | P-Co | P-CaRS | P-Co | P-Co |
| ctb/1 | P-Co | P-CoBi | P-Co | P-Co | P-Co | P-Co |
| ctr/1 | P-CoBi | P-CoBi | P-Co | P-CoBi | P-CoBi | P-CoBi |
| ct$p$/1 | P-CoBi | P-CoBi | P-CoBi | P-Co | P-Co | P-Co |
| rt$p$/1 | P-CoBi | P-SHA | P-Co | P-Co | P-Co | P-Co |

# Comparison with CMA-ES variants ($n = 80$)

DE with PCMs perform well for larger budgets of fevals



# Conclusion: parameter control methods (PCMs)

- 9 PCMs in DE were benchmarked on `bbob-mixint`
    - Using a PCM can improve the performance of DE
    - The best PCM depends on the type of
        - mutation strategy (e.g., rand/1, rand/2, best/1, etc.)
        - repair method (Lamarckian and Baldwinian)
    - The best PCMs on `bbob` and `bbob-mixint` are different
        - PCM in SHADE does not work for `bbob-mixint`
        - PCM in CoDE, CoBiDE, jDE, and SinDE work well
    - DE with a suitable PCM performs better than CMA-ES on `bbob-mixint` for larger budgets of fevals
- Many future works: analysis of "DE" and "PCMs"
    - Nobody can explain how the results happened
    - Even analysis of DE for mix-int. opt. has not been done