# Covariance Matrix Adaptation in Multiobjective Optimization

Dimo Brockhoff
INRIA Lille – Nord Europe

October 30, 2014

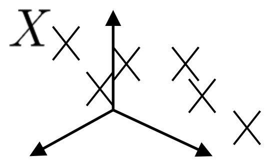
PGMO-COPI'2014, Ecole Polytechnique, France

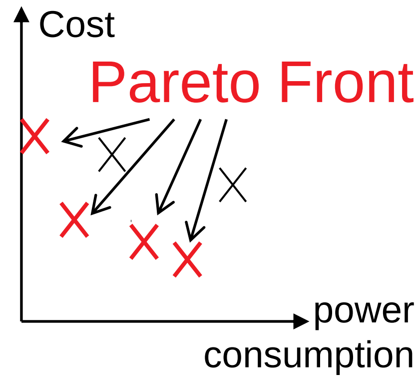
INVENTORS FOR THE DIGITAL WORLD

# Scenario: Multiobjective Optimization

## Most problems are multiobjective in nature...

$$\min_{x \in X} f(x) = (f_1(x), \ldots, f_k(x)) \in \mathbb{R}^k$$




## Set-based optimization view [Zitzler et al. 2010]:

- interested in finding a set of solutions ("Pareto front approxim.")
- human decision maker can "learn about the problem"
- mathematically: unary quality indicator $I$ transforms problem into a single-objective set problem
- randomized, set-based algorithms well suited for difficult (blackbox) multiobjective problems: field of Evolutionary Multiobjective Optimization (EMO)

**Goal of my talk:**
introduce the idea behind one algorithm (class): MO-CMA-ES

# The Multiobjective CMA-ES

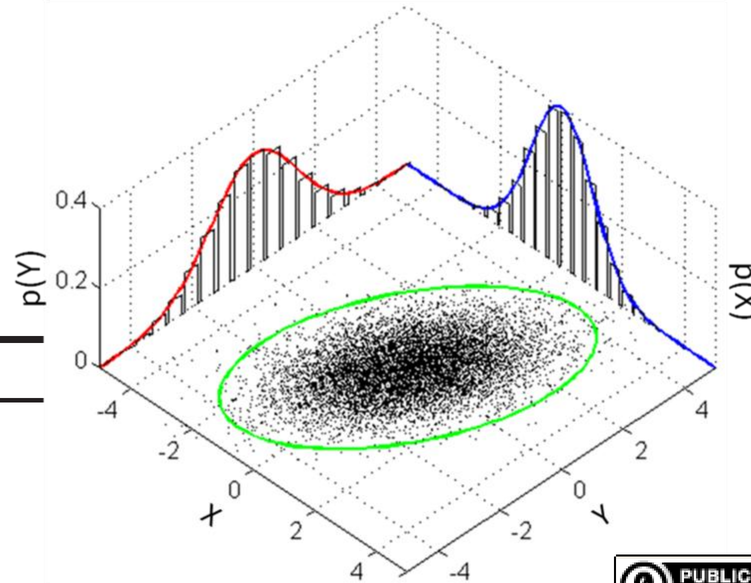
**CMA-ES** [remember talk of Nikolaus Hansen yesterday]

- Covariance Matrix Adaptation Evolution Strategy [Hansen and Ostermeier 1996, 2001]
- the state-of-the-art numerical black box optimizer for large budgets and difficult functions [Hansen et al. 2010]

**CMA-ES for multiobjective optimization**

- "THE" MO-CMA-ES does not exist
- original one of [Igel et al. 2007] in ECJ
- improved success definition [Voß et al. 2010] at GECCO 2010
- recombination between solutions [Voß et al. 2009] at EMO 2009
- all based on combination of μ single (1+1)-CMA-ES

$$a^{(g)} = \left(\mathbf{x}^{(g)}, \bar{p}_{succ}^{(g)}, \sigma^{(g)}, \mathbf{p}_c^{(g)}, \mathbf{C}^{(g)}\right)$$






**Algorithm 1:** (1+λ)-CMA-ES

$g = 0$, initialize $a_{\text{parent}}^{(g)}$

**repeat**

  $a_{\text{parent}}^{(g+1)} \leftarrow a_{\text{parent}}^{(g)}$

  **for** $k = 1, \ldots, \lambda$ **do**

    $x_k^{(g+1)} \sim \mathcal{N}\left(x_{\text{parent}}^{(g)}, {\sigma^{(g)}}^2 C^{(g)}\right)$

  $\texttt{updateStepSize}\left(a_{\text{parent}}^{(g+1)}, \frac{\lambda_{\text{succ}}^{(g+1)}}{\lambda}\right)$

  **if** $f\left(x_{1:\lambda}^{(g+1)}\right) \leq f\left(x_{\text{parent}}^{(g)}\right)$ **then**

    $x_{\text{parent}}^{(g+1)} \leftarrow x_{1:\lambda}^{(g+1)}$

    $\texttt{updateCovariance}\left(a_{\text{parent}}^{(g+1)}, \frac{x_{\text{parent}}^{(g+1)} - x_{\text{parent}}^{(g)}}{\sigma_{\text{parent}}^{(g)}}\right)$

  $g \leftarrow g + 1$

**until** stopping criterion is met

# (1+λ)-CMA-ES: Updates

**Procedure** `updateStepSize`$(a = [x, \bar{p}_{\text{succ}}, \sigma, p_c, C], p_{\text{succ}})$

$$\bar{p}_{\text{succ}} \leftarrow (1 - c_p)\,\bar{p}_{\text{succ}} + c_p p_{\text{succ}}$$

$$\sigma \leftarrow \sigma \cdot \exp\left(\frac{1}{d}\,\frac{\bar{p}_{\text{succ}} - p_{\text{succ}}^{\text{target}}}{1 - p_{\text{succ}}^{\text{target}}}\right)$$

**Procedure** `updateCovariance`$(a = [x, \bar{p}_{\text{succ}}, \sigma, p_c, C], x_{\text{step}} \in \mathbb{R}^n)$

**if** $\bar{p}_{\text{succ}} < p_{\text{thresh}}$ **then**

$$p_c \leftarrow (1 - c_c)p_c + \sqrt{c_c(2 - c_c)}\; x_{\text{step}}$$

$$C \leftarrow (1 - c_{\text{cov}})C + c_{\text{cov}} \cdot p_c {p_c}^{\mathrm{T}}$$

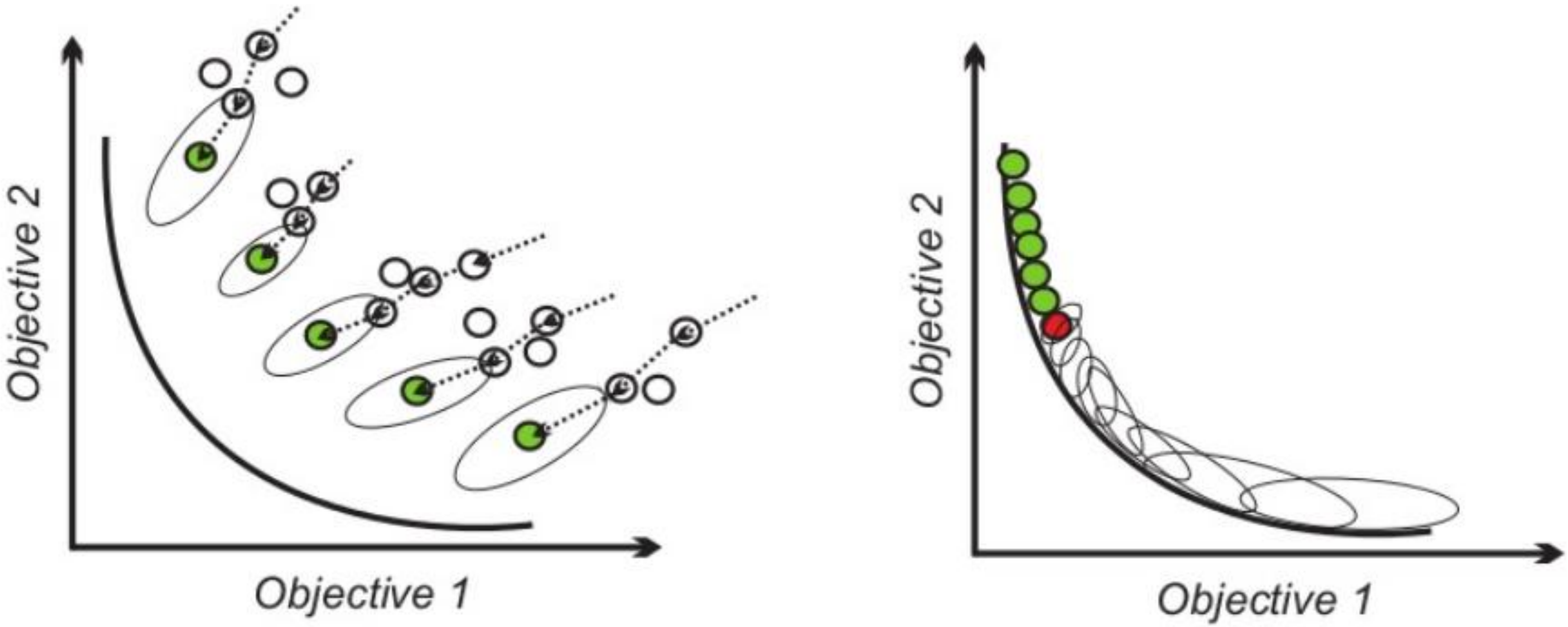
**else**

$$p_c \leftarrow (1 - c_c)p_c$$

$$C \leftarrow (1 - c_{\text{cov}})C + c_{\text{cov}} \cdot \left(p_c {p_c}^{\mathrm{T}} + c_c(2 - c_c)C\right)$$

# MO-CMA-ES: Basic Idea

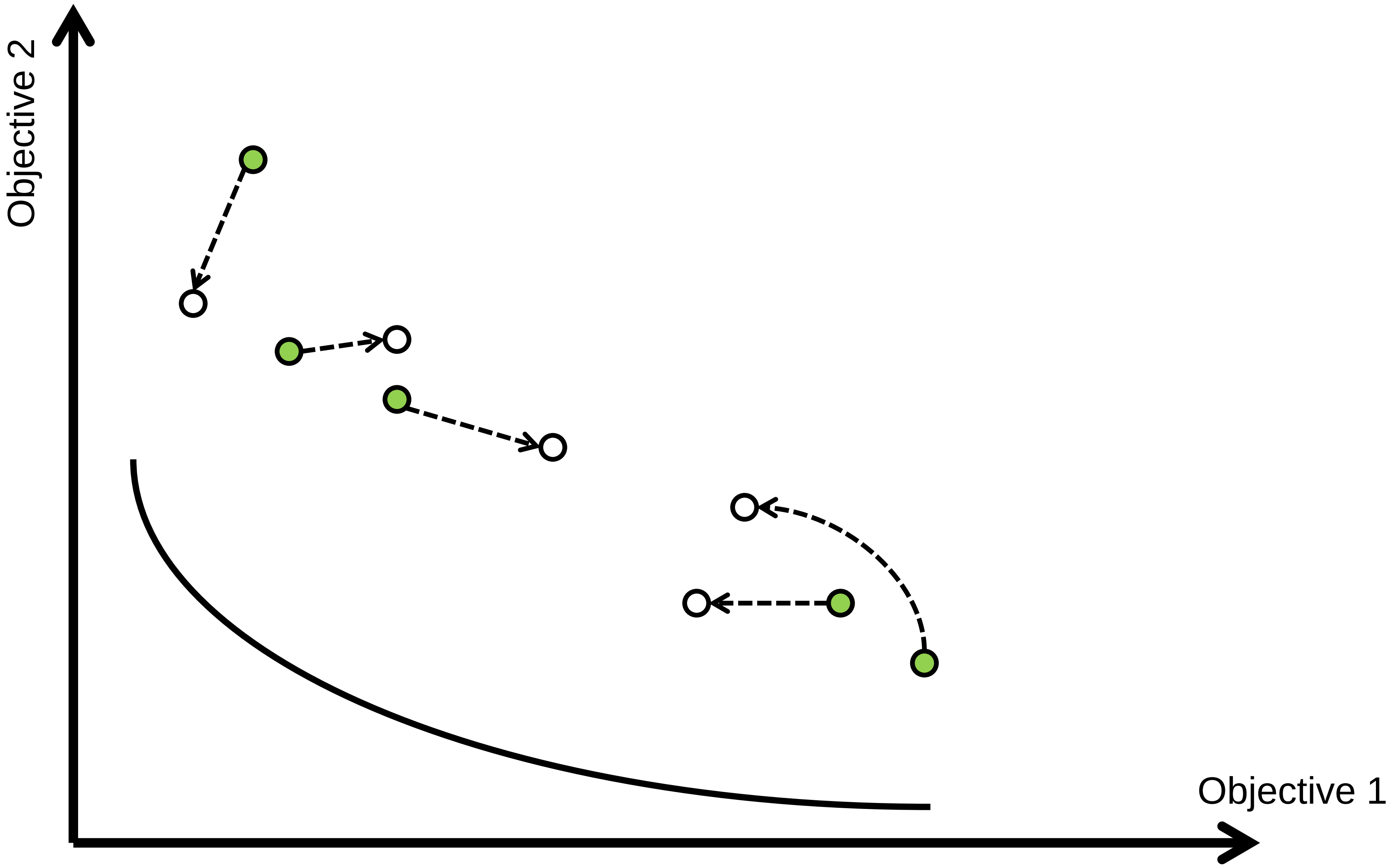
**Illustration in objective space:** intuition not accurate!




[copyright by Ilya Loshchilov]

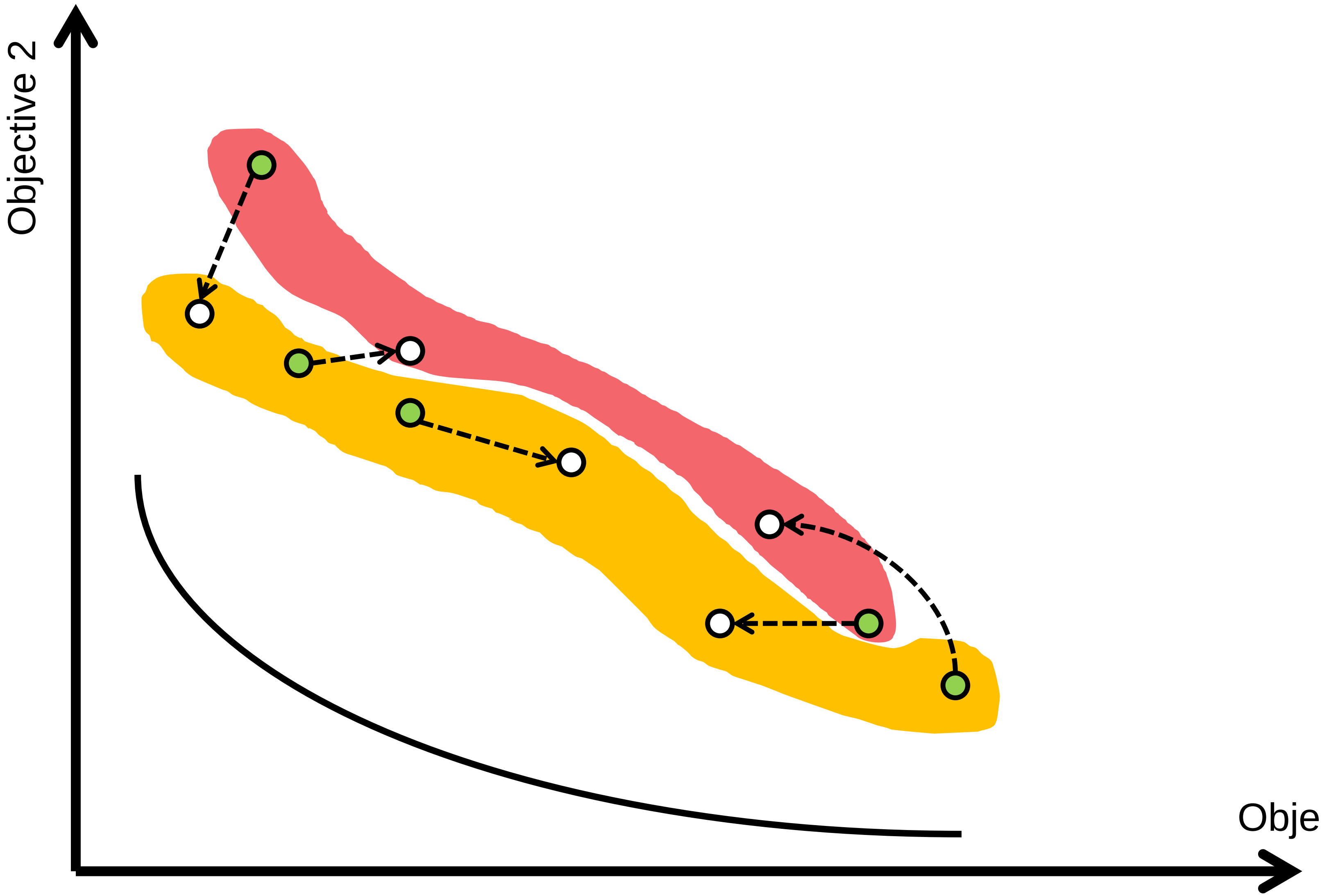
# Concrete MO-CMA-ES Baseline Algorithm

μ x (1+1)-CMA-ES: $a_i^{(g)} = (\mathbf{x}_i^{(g)}, \overline{p}_{succ,i}^{(g)}, \sigma_i^{(g)}, \mathbf{p}_{c,i}^{(g)}, \mathbf{C}_i^{(g)})$



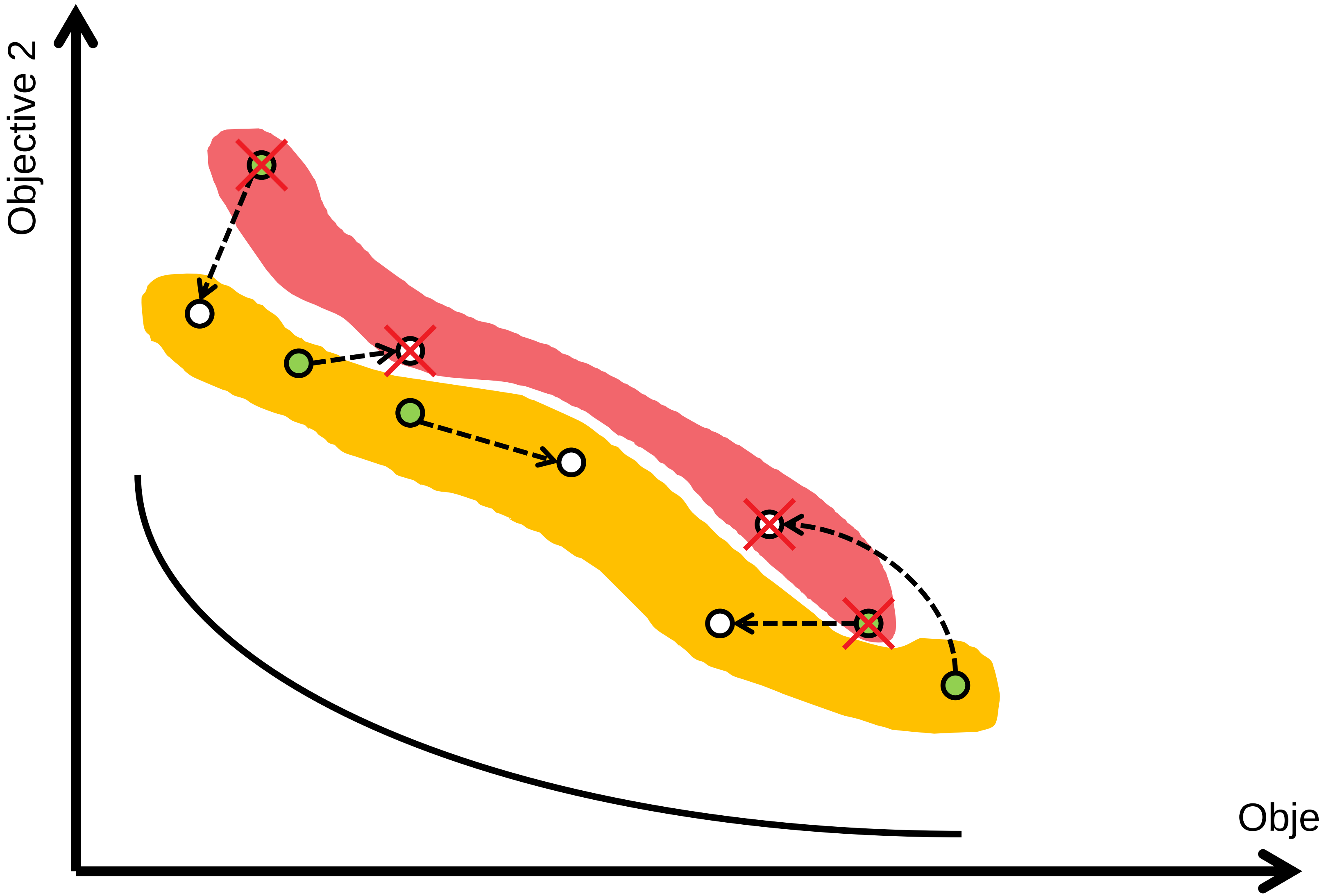
# Concrete MO-CMA-ES Baseline Algorithm

μ x (1+1)-CMA-ES: $a_i^{(g)} = (\mathbf{x}_i^{(g)}, \overline{p}_{succ,i}^{(g)}, \sigma_i^{(g)}, \mathbf{p}_{c,i}^{(g)}, \mathbf{C}_i^{(g)})$



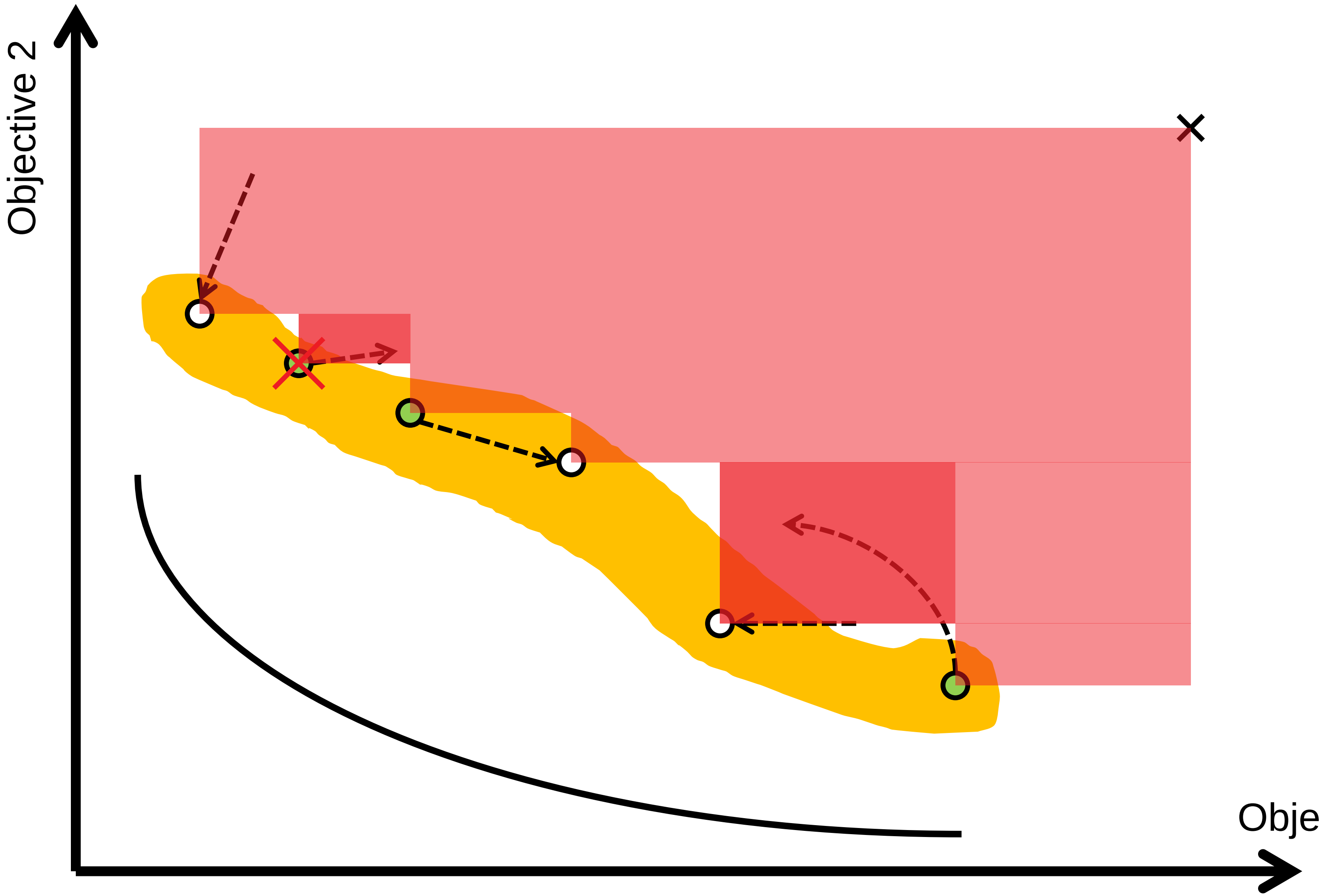
# Concrete MO-CMA-ES Baseline Algorithm

μ x (1+1)-CMA-ES: $a_i^{(g)} = (\mathbf{x}_i^{(g)}, \overline{p}_{succ,i}^{(g)}, \sigma_i^{(g)}, \mathbf{p}_{c,i}^{(g)}, \mathbf{C}_i^{(g)})$



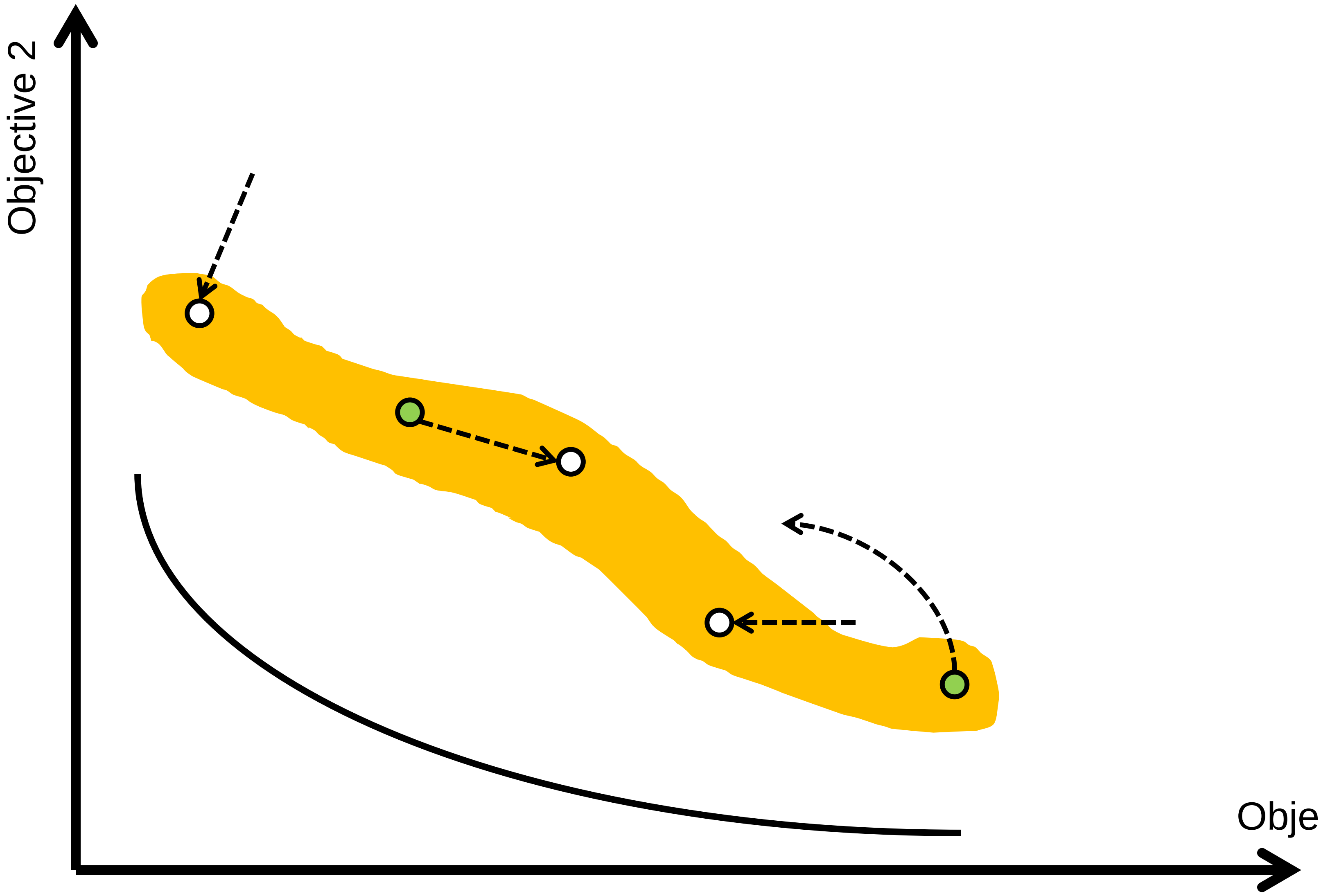
# Concrete MO-CMA-ES Baseline Algorithm

μ x (1+1)-CMA-ES: $a_i^{(g)} = (\mathbf{x}_i^{(g)}, \overline{p}_{succ,i}^{(g)}, \sigma_i^{(g)}, \mathbf{p}_{c,i}^{(g)}, \mathbf{C}_i^{(g)})$



# Concrete MO-CMA-ES Baseline Algorithm

μ x (1+1)-CMA-ES: $a_i^{(g)} = (\mathbf{x}_i^{(g)}, \overline{p}_{succ,i}^{(g)}, \sigma_i^{(g)}, \mathbf{p}_{c,i}^{(g)}, \mathbf{C}_i^{(g)})$



# MO-CMA-ES baseline algorithm

- μ x (1+1)-CMA-ES
- hypervolume-based selection
- update of CMA strategy parameters based on different success notions

**Update of parameters:**

- step size of parents and offspring based on success
- covariance matrix only for offspring

**Success Definitions:**

- original success [Igel et al. 2007]: if offspring dominates parent
- improved success [Voß et al. 2010]: if offspring selected into new population

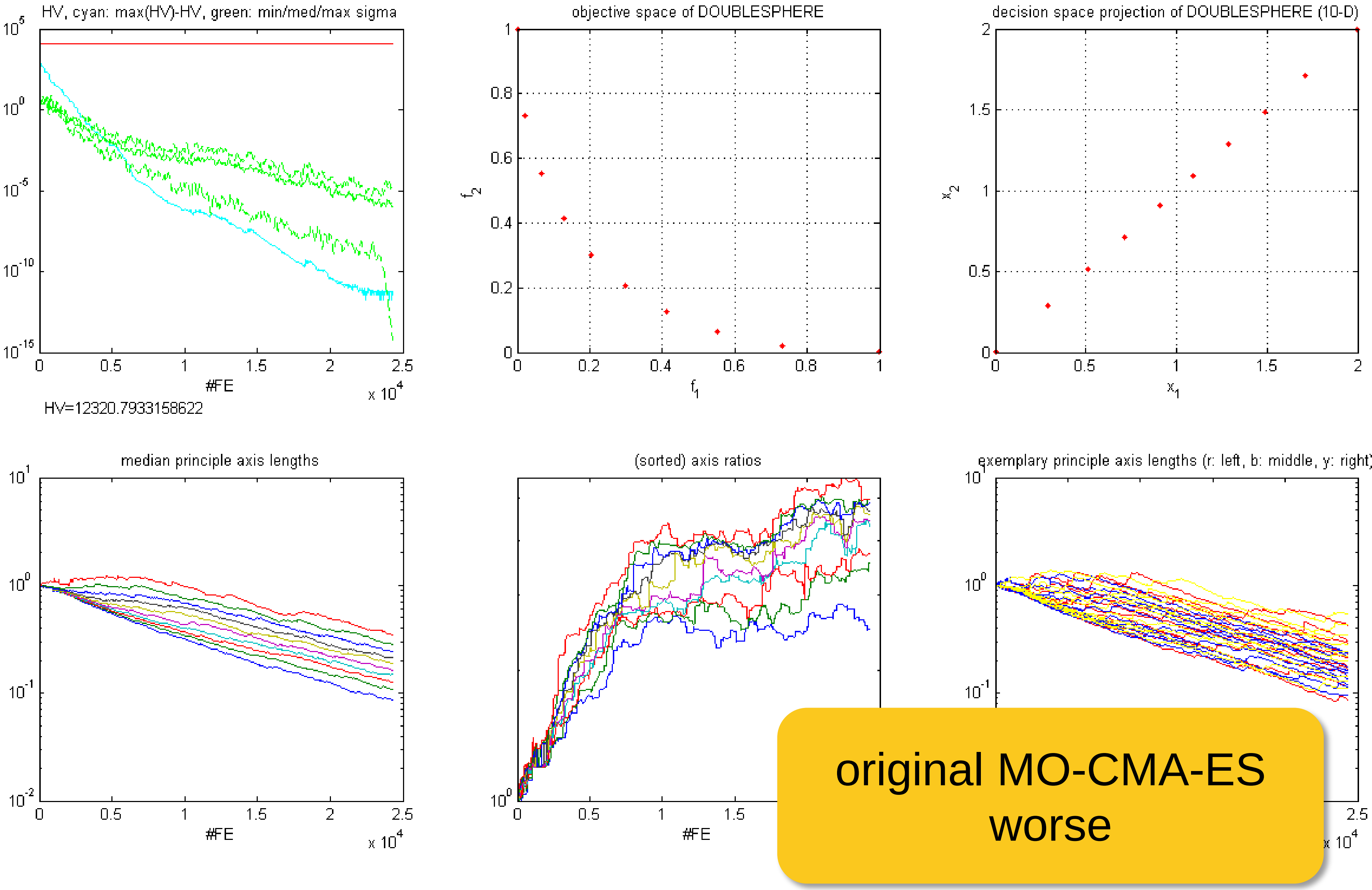
# Available Implementations

**Baseline implementation in Shark machine learning library**

- http://image.diku.dk/shark/
- C++

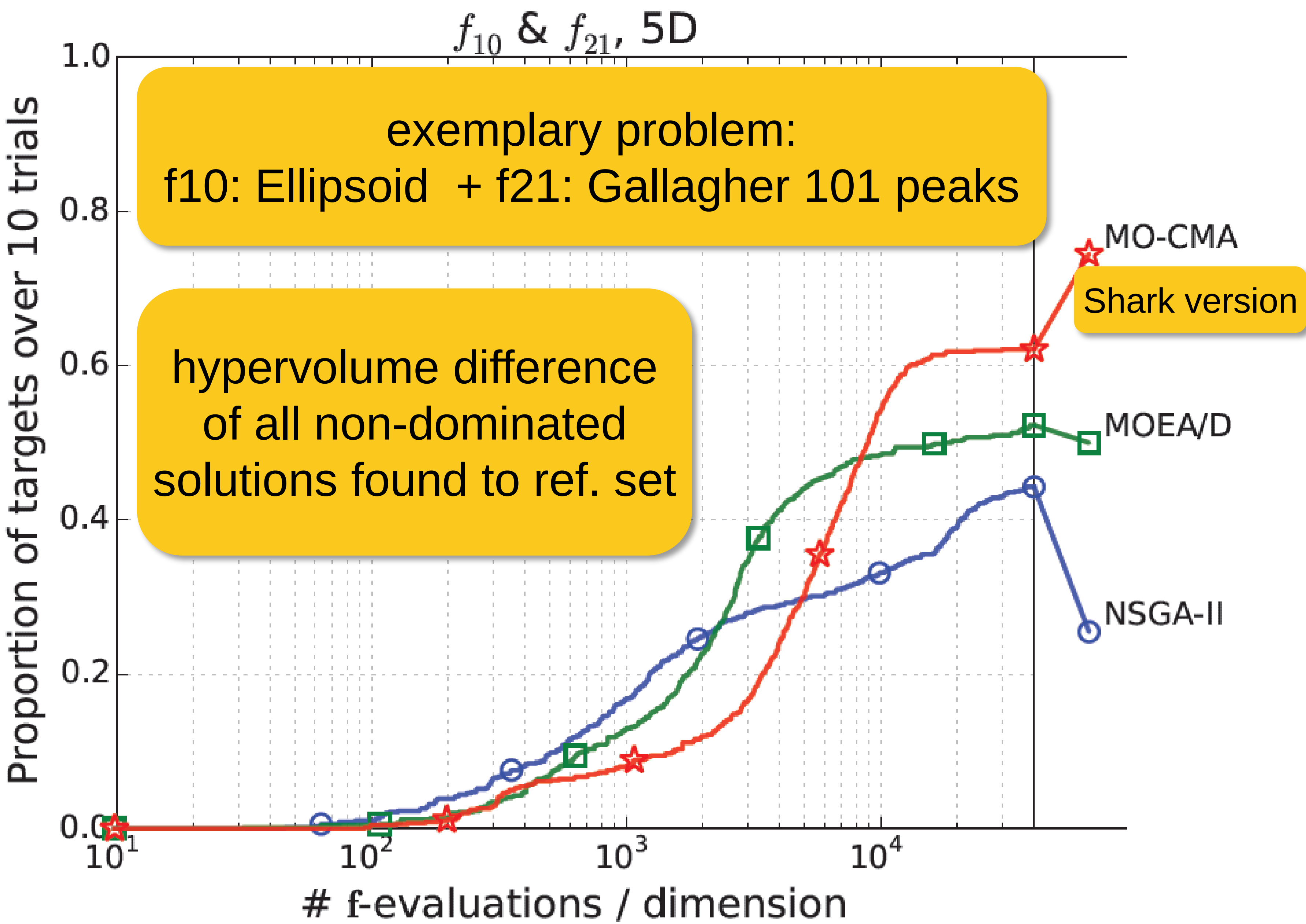
**Now also available in MATLAB**

- easy prototyping of new ideas
- visualization of algorithm's state variables (similar to CMA-ES)

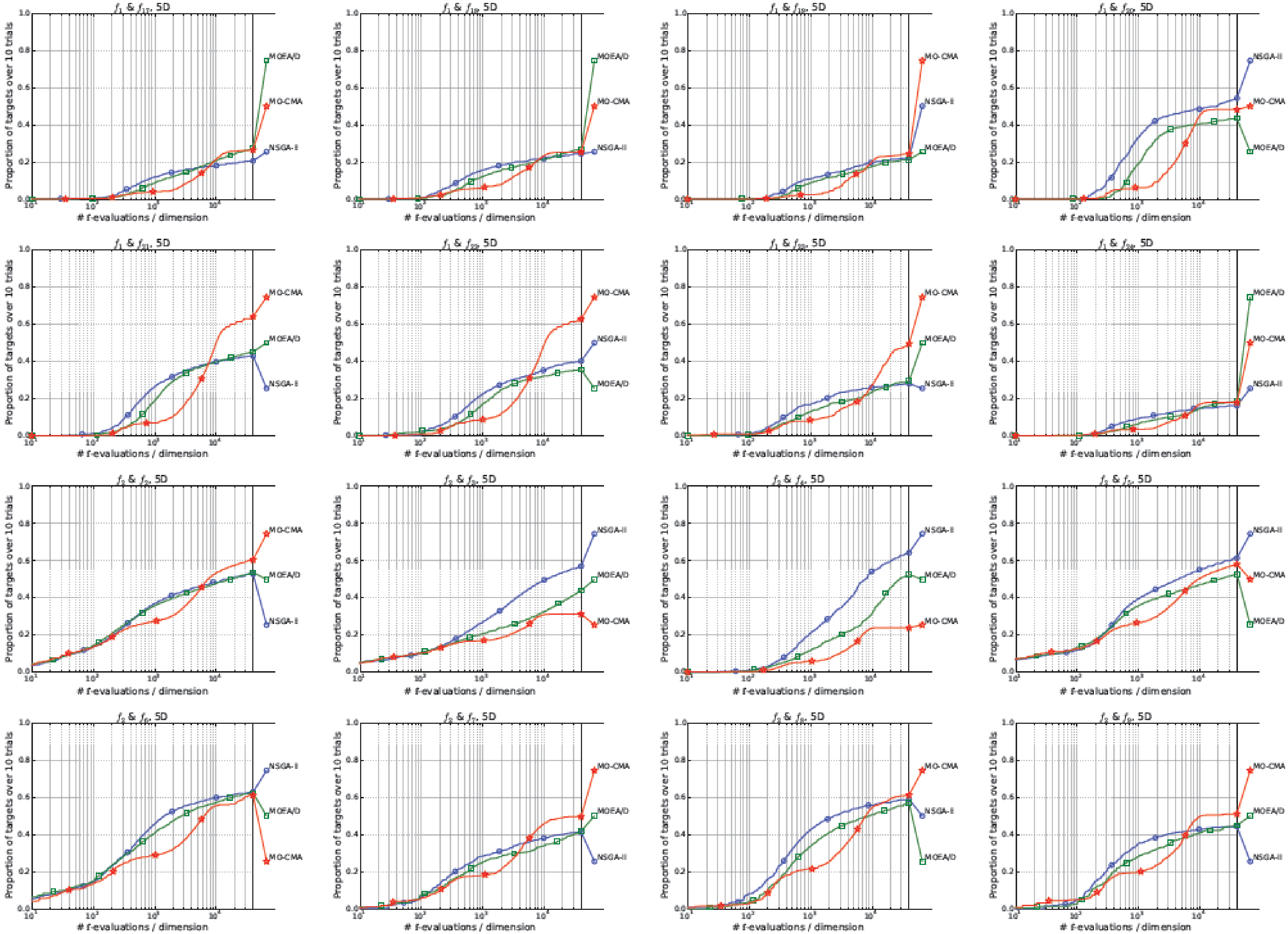



original MO-CMA-ES worse

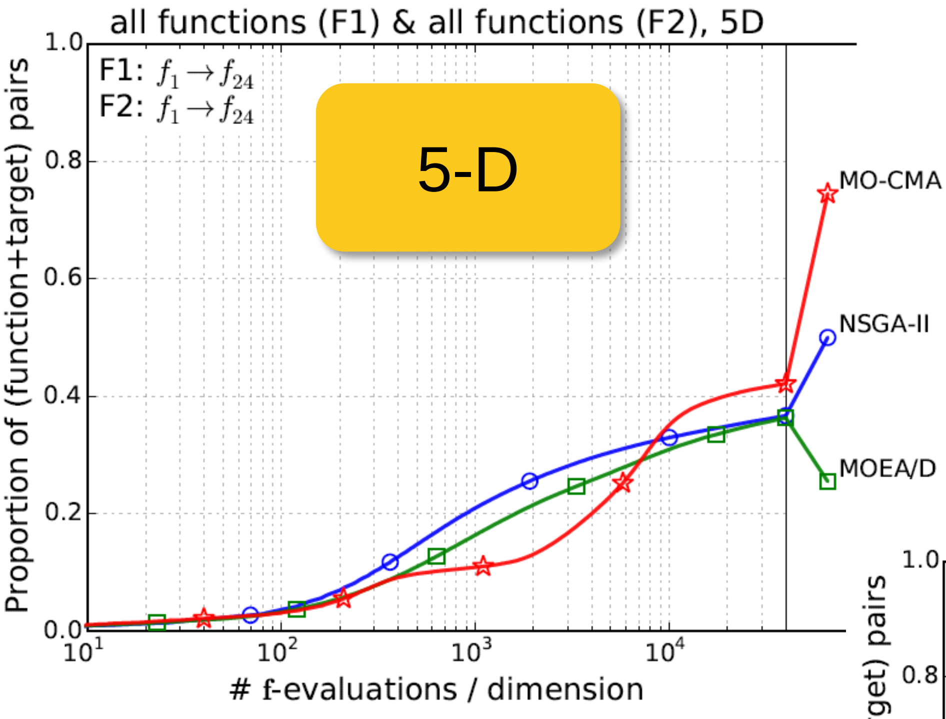
# Comparison with Other Well-Known Algorithms

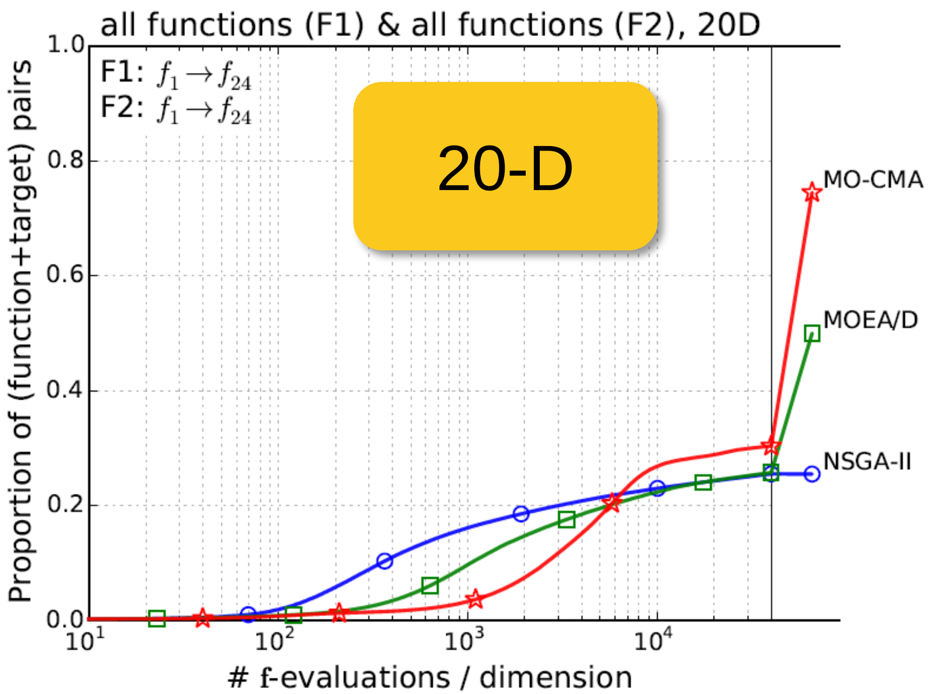


# More Problems

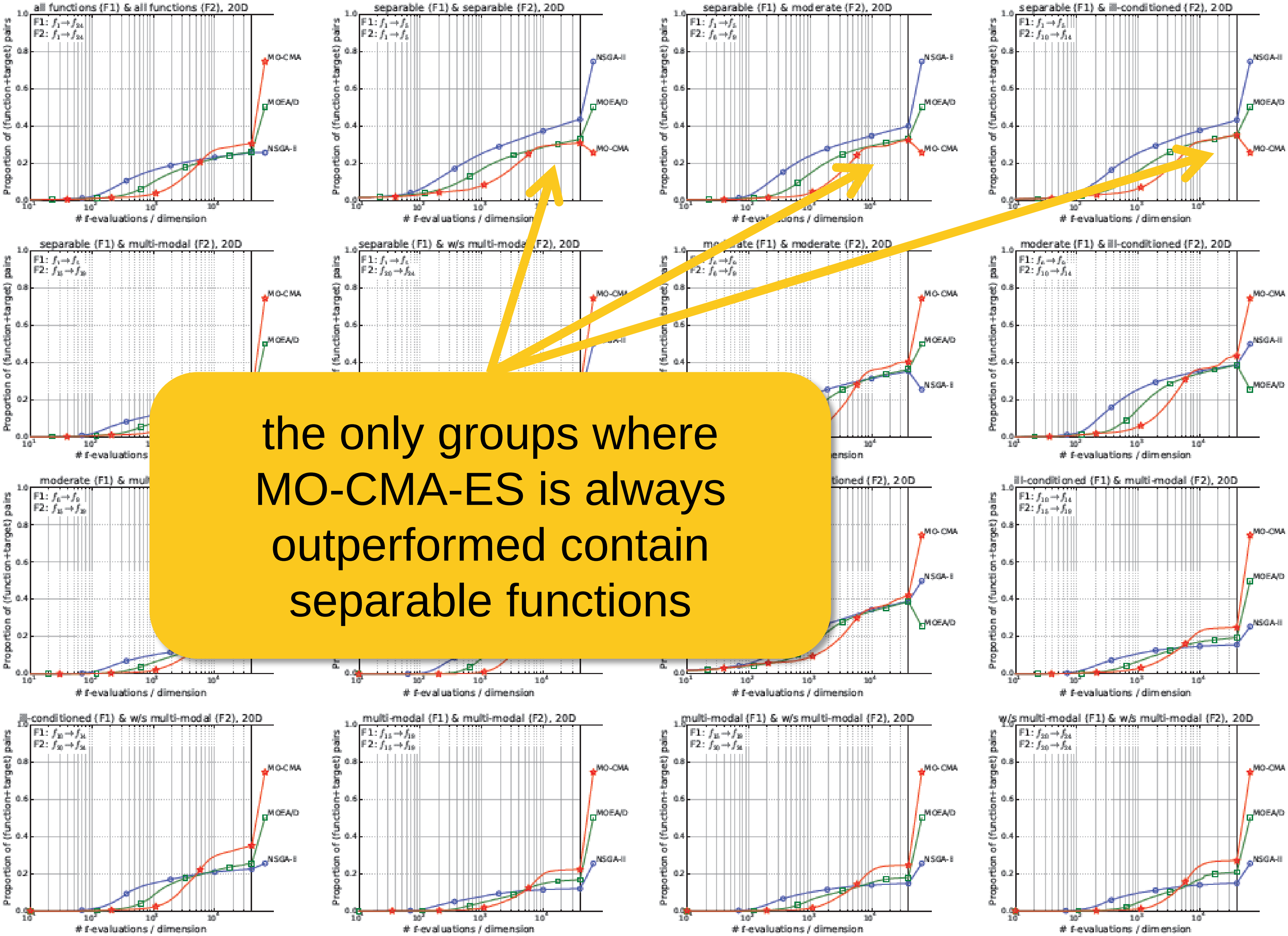


# Aggregation Over All BBOB Problems (300 total)



# Aggregation Over Function Groups




the only groups where MO-CMA-ES is always outperformed contain separable functions

# Other MO-CMA-ES Variants

## Strategy Parameter Recombination in MO-CMA-ES (EMO'2009)

- learning of probability distribution also based on neighbors
- neighbors = influence weighted by Mahalanobis distance
- performance difference to original MO-CMA-ES less strong than between original and improved success criterion

## Towards Integrating Komma-Strategies

- (1+1)-CMA-ES not optimal for noisy problems
- "Problem" with integrating more robust (μ/μ,λ)-CMA-ES:
  - how to rank with respect to parent?
- First ideas there, but algorithm still in progress

# Conclusions

- MO-CMA-ES as a multiobjective extension of the prominent single-objective CMA-ES
- Several variants
- Shark (C++) and Matlab implementations available
- shows superiority for large(r) budgets on non-separable bi-objective functions wrt archive of non-dominated solutions found
- therefore probably a good alternative over NSGA-II if you have difficult multiobjective problems to solve in practice

**Questions?**

# Publications


[Hansen and Ostermeier 1996] N. Hansen and A. Ostermeier. **Adapting arbitrary normal mutation distributions in evolution strategies: the covariance matrix adaptation**. In *Congress on Evolutionary Computation (CEC 1996)*, pages 312–317, Piscataway, NJ, USA, 1996. IEEE. [476 —

[Hansen and Ostermeier 2001] N. Hansen and A. Ostermeier. **Completely Derandomized Self-Adaptation in Evolution Strategies**. *Evolutionary Computation*, 9(2):159–195, 2001.

[Hansen et al 2010] N. Hansen, A. Auger, R. Ros, S. Finck, and P. Posik. **Comparing Results of 31 Algorithms from the Black-Box Optimization Benchmarking BBOB-2009**. In *Genetic and Evolutionary Computation Conference (GECCO 2010)*, pages 1689–1696, 2010. ACM

[Igel et al. 2007] C. Igel, N. Hansen, and S. Roth. **Covariance Matrix Adaptation for Multi-objective Optimization**. *Evolutionary Computation*, 15(1):1–28, 2007.

[Voß et al. 2009] T. Voß, N. Hansen, and C. Igel. **Recombination for Learning Strategy Parameters in the MO-CMA-ES**. In *Evolutionary Multi-Criterion Optimization (EMO 2009)*, pages 155–168. Springer, 2009.

[Voß et al. 2010] T. Voß, N. Hansen, and C. Igel. **Improved Step Size Adaptation for the MO-CMA-ES**. In J. Branke et al., editors, *Genetic and Evolutionary Computation Conference (GECCO 2010)*, pages 487–494. ACM, 2010.