BI-population CMA-ES Algorithms with Surrogate Models and Line Searches

Ilya Loshchilov[1], Marc Schoenauer[2] and Michèle Sebag [2]

[1]LIS, École Polytechnique Fédérale de Lausanne

[2]TAO, INRIA — CNRS — Université Paris-Sud

July 6th, 2013

Historical overview: BBOB'2012

Expensive Optimization

- Self-adaptive surrogate-assisted CMA-ES (IPOP-saACM-ES and BIPOP-saACM-ES) on noiseless[1] and noisy testbeds[2].
- BIPOP-saACM-ES demonstrates one of the best performance among the algorithms of the BBOB-2009, 2010 and 2012.

Multimodal Optimization

- Alternative restart strategies (NBIPOP-aCMA-ES and NIPOP-aCMA-ES) on noiseless testbed[3].
- NBIPOP-aCMA-ES is TOP-3 algorithm of the CEC'2013 (preliminary results).

[1] [Loshchilov, Schoenauer and Sebag; GECCO-BBOB 2012] "Black-box optimization benchmarking of IPOP-saACM-ES and BIPOP-saACM-ES on the BBOB-2012 noiseless testbed"

[2] [Loshchilov, Schoenauer and Sebag; GECCO-BBOB 2012] "Black-box optimization benchmarking of IPOP-saACM-ES on the BBOB-2012 noisy testbed"

[3] [Loshchilov, Schoenauer and Sebag; GECCO-BBOB 2012] "Black-box Optimization Benchmarking of NIPOP-aCMA-ES and NBIPOP-aCMA-ES on the BBOB-2012 Noiseless Testbed"

This talk: BBOB'2013

Expensive Optimization

- saACM with intensive surrogate model exploitation (BIPOP-saACM-ES-k) on noiseless testbed[4].
- BIPOP-saACM-ES-k **further improves** BIPOP-saACM-ES.

Optimization of separable and non-separable functions

- BIPOP-aCMA-STEP: a hybrid of BIPOP-aCMA and STEP algorithm.
- BIPOP-aCMA-STEP **demonstrates a cheap way** to identify and exploit the separability.

Efficient Optimization

- HCMA: a hybrid of BIPOP-saACM-ES-k, STEP and NEWUOA algorithms.

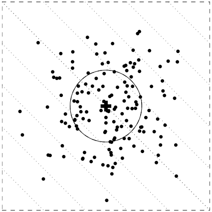
[4] [Loshchilov, Schoenauer and Sebag; GECCO 2013] "Intensive Surrogate Model Exploitation in Self-adaptive Surrogate-assisted CMA-ES (saACM-ES)"

Content

1. State-of-the-art
   - Covariance Matrix Adaptation Evolution Strategy (CMA-ES)
   - ^{s*}ACM-ES: Self-Adaptive Surrogate-Assisted CMA-ES
2. Contribution
   - Intensive surrogate model exploitation
   - Optimization of separable and non-separable functions

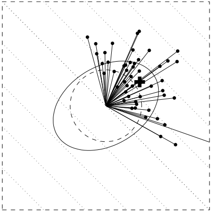
(μ, λ)-Covariance Matrix Adaptation Evolution Strategy

Rank-μ Update [5] [6]



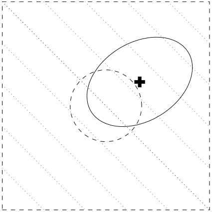
$\boldsymbol{y}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{C})\,, \quad \mathbf{C} = \mathbf{I}$
$\boldsymbol{x}_i = \boldsymbol{m} + \sigma\, \boldsymbol{y}_i\,, \quad \sigma = 1$

sampling of λ solutions



$\mathbf{C}_\mu = \frac{1}{\mu} \sum \boldsymbol{y}_{i:\lambda} \boldsymbol{y}_{i:\lambda}^{\mathrm{T}}$
$\mathbf{C} \leftarrow (1 - 1) \times \mathbf{C} + 1 \times \mathbf{C}_\mu$

calculating $\mathbf{C}$ from best μ out of λ



$\boldsymbol{m}_{\text{new}} \leftarrow \boldsymbol{m} + \frac{1}{\mu} \sum \boldsymbol{y}_{i:\lambda}$

new distribution

The adaptation increases the probability of successful steps to appear again.

Other components of CMA-ES: *step-size adaptation*, *evolution path*.

---

[5] [Hansen *et al.*, ECJ 2003] "Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES)"

[6] the slide adopted by courtesy of Nikolaus Hansen

Invariance: Guarantee for Generalization

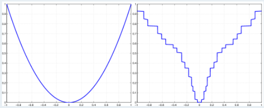
Invariance properties of CMA-ES

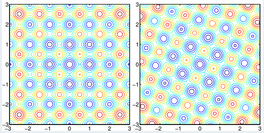
- Invariance to order-preserving transformations in function space

  true for all comparison-based algorithms

- Translation and rotation invariance

  thanks to $\mathbf{C}$





CMA-ES is almost parameterless (as a consequence of invariances)

- Tuning on a small set of functions — Hansen & Ostermeier 2001
- Default values generalize to whole classes
- Exception: population size for multi-modal functions [a] [b]

---

[a] [Auger & Hansen, CEC 2005] "A restart CMA evolution strategy with increasing population size"

[b] [Loshchilov *et al.*, PPSN 2012] "Alternative Restart Strategies for CMA-ES"

BIPOP-CMA-ES

BIPOP-CMA-ES: [7] (BIPOP-aCMA-ES [8])

Regime-1 (large populations, IPOP part):
Each restart: $\lambda_{large} = 2 * \lambda_{large}$, $\sigma^0_{large} = \sigma^0_{default}$

Regime-2 (small populations):
Each restart:

$$\lambda_{small} = \left\lfloor \lambda_{default} \left(\frac{1}{2} \frac{\lambda_{large}}{\lambda_{default}} \right)^{U[0,1]^2} \right\rfloor, \quad \sigma^0_{small} = \sigma^0_{default} \times 10^{-2U[0,1]}$$

where $U[0,1]$ stands for the uniform distribution in $[0,1]$.

BIPOP-CMA-ES launches the first run with default population size and initial step-size. In each restart, it **selects the restart regime with less function evaluations** used so far.

[7] Hansen (GECCO BBOB 2009). "Benchmarking a BI-population CMA-ES on the BBOB-2009 function testbed"

[8] Loshchilov, Schoenauer and Sebag (GECCO BBOB 2012). "Black-box Optimization Benchmarking of NIPOP-aCMA-ES and NBIPOP-aCMA-ES on the BBOB-2012 Noiseless Testbed"

Contents

^{s*}ACM-ES: Self-Adaptive Surrogate-Assisted CMA-ES

Using Ranking SVM as the surrogate model

- Build a global model using Ranking SVM [9]

$$\mathbf{x_i} \succ \mathbf{x_j} \text{ iff } \hat{\mathcal{F}}(\mathbf{x_i}) < \hat{\mathcal{F}}(\mathbf{x_j})$$

✓ Comparison-based surrogate models $\rightarrow$ invariance to rank-preserving transformations of $\mathcal{F}(x)$

How to choose an appropriate Kernel?

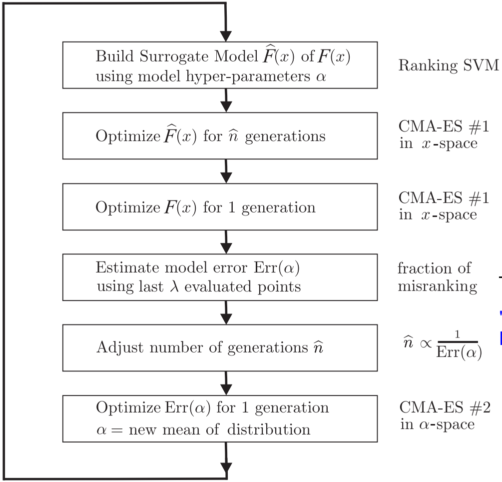
- Use covariance matrix C adapted by CMA-ES in Gaussian kernel[10]

$$K(x_i, x_j) = e^{-\frac{(x_i - x_j)^T (x_i - x_j)}{2\sigma^2}}; \quad K_C(x_i, x_j) = e^{-\frac{(x_i - x_j)^T C^{-1} (x_i - x_j)}{2\sigma^2}}$$

✓ Invariance to rotation of the search space thanks to C

[9] [Runarsson *et al.*, PPSN 2006] "Ordinal Regression in Evolutionary Computation"
[10] [Loshchilov *et al.*, PPSN 2010] "Comparison-based optimizers need comparison-based surrogates"

s^*ACM-ES: Self-Adaptive Surrogate-Assisted CMA-ES

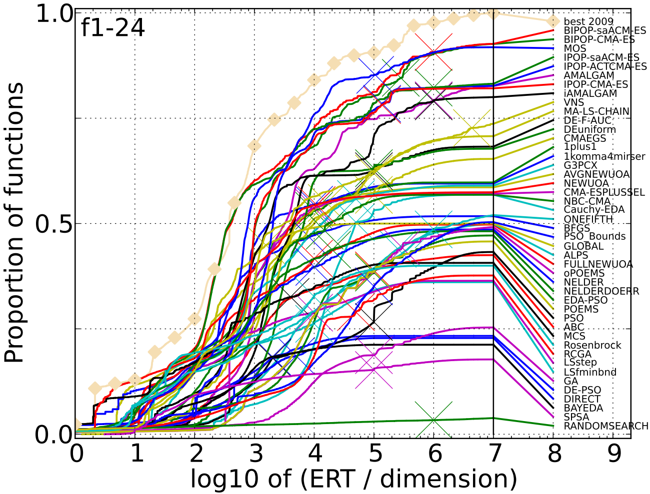


Surrogate-assisted CMA-ES with online adaptation of model hyper-parameters [a]

[a] [Loshchilov *et al.*, GECCO 2012] "Self-Adaptive Surrogate-Assisted Covariance Matrix Adaptation Evolution Strategy"

Results on Black-Box Optimization Competition

BIPOP-^{s*}aACM and IPOP-^{s*}aACM (with restarts) on 24 noiseless 20 dimensional functions

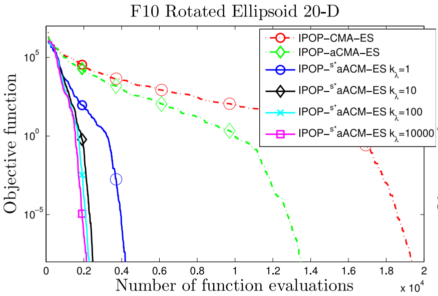


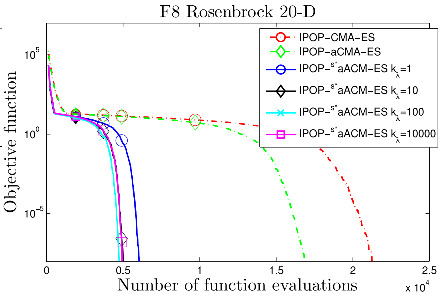
Contents

Intensive surrogate model exploitation

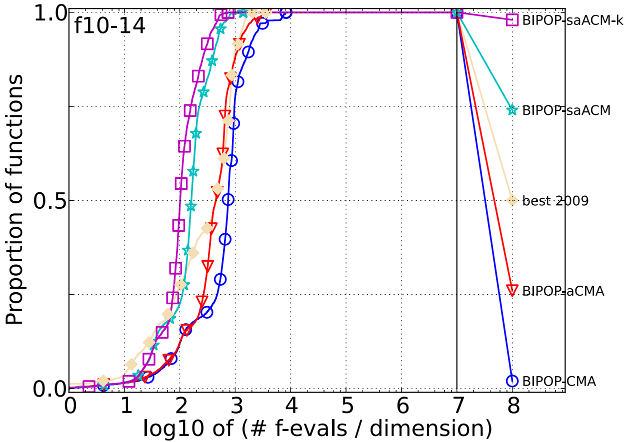
The only difference between **BIPOP-saACM-k** and **BIPOP-saACM**:

- Intensive exploitation: when optimizing $\hat{F}$, $\lambda = k_\lambda \lambda_{def}$, $\mu = \mu_{def}$. $k_\lambda = 1$ for D<10 and $k_\lambda = 10, 100, 1000$ for 10, 20, 40-D.
- Divergence Prevention: $k_\lambda > 1$ is used only of $\hat{n} \geq \hat{n}_{k_\lambda}$, $\hat{n}_{k_\lambda} = 4$.



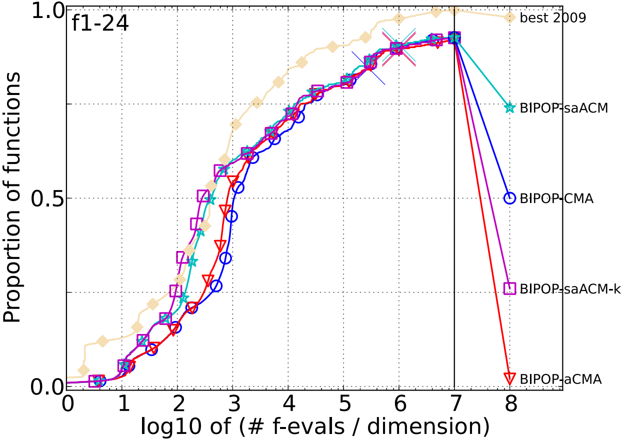


Intensive surrogate model exploitation



* smaller budget for surrogate-assisted search: $10^4 D$ for BIPOP-saACM-k versus $10^6 D$ for BIPOP-saACM.

Intensive surrogate model exploitation



* smaller budget for surrogate-assisted search: $10^4 D$ for BIPOP-saACM-k versus $10^6 D$ for BIPOP-saACM.

Contents

Optimization of separable and non-separable functions

Select the easiest point (**STEP**) [11] [12]

- Simple line search method based on iterative interval division.
- Great optimizer of one-dimensional multimodal functions.

An extension to multi-dimensional (sequential) search

+ simple idea: **sequentially** optimize one dimension after another.
- some **stopping criteria** should be **set a priori**, e.g., number of evaluations or target precision.
- **no hint** whether the problem is **separable or not** is available.

[11] [Swarzberg *et al.*, CEC 1994] "The easiest way to optimize a function"
[12] [Posík *et al.*, ECJ 2012] "Restarted local search algorithms for continuous black box optimization"

Optimization of separable and non-separable functions

Parallel multi-dimensional STEP

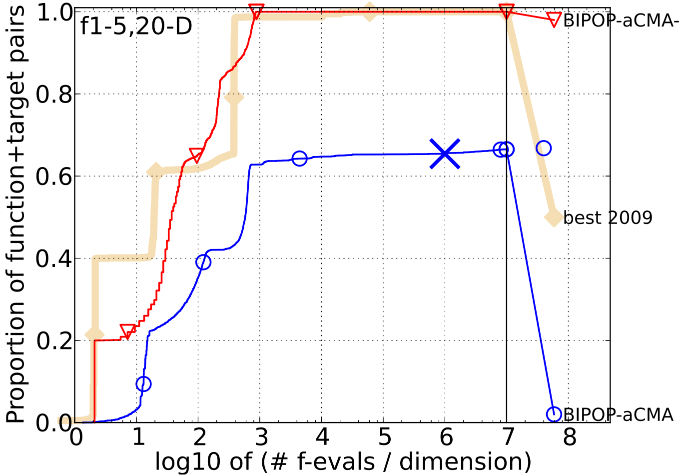
1. Check one new STEP point **per** each dimension.
2. Current estimate of the optimum x^* = a solution composed of **best** x_i^*-values from all variables.
3. If the current estimate is worse than the previous one, then the problem **is not separable**.

Optimization of separable and non-separable functions

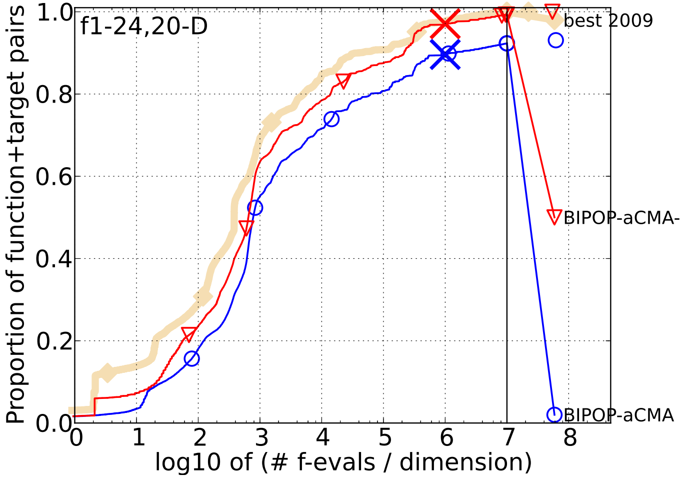
BIPOP-aCMA-STEP

1. BIPOP-aCMA-STEP and STEP are running in parallel, a fraction $\rho_{STEP} = 0.5$ of function evaluations is allocated to STEP.
2. At each iteration after $n_{MinIterSTEP} = 10$ iterations the STEP can be stopped if its best solution is worse than the one of BIPOP-aCMA-ES.

Optimization of separable and non-separable functions



Intensive surrogate model exploitation



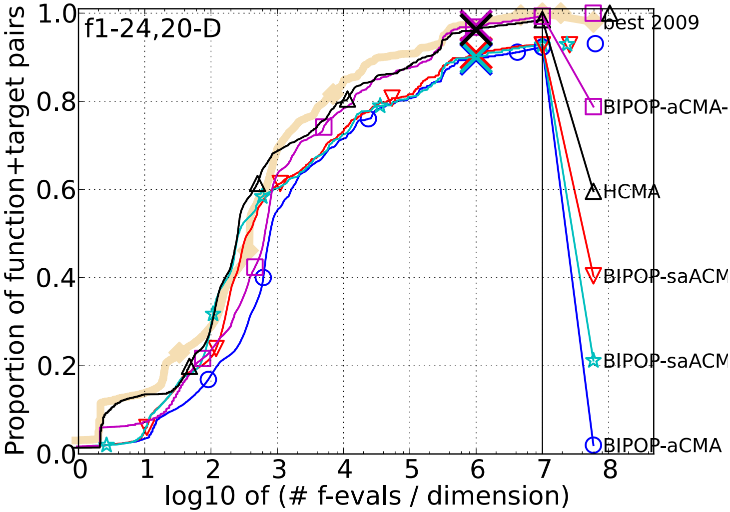
Efficient Optimization

HCMA = BIPOP-saACM-ES-k + STEP + NEWUOA[13]

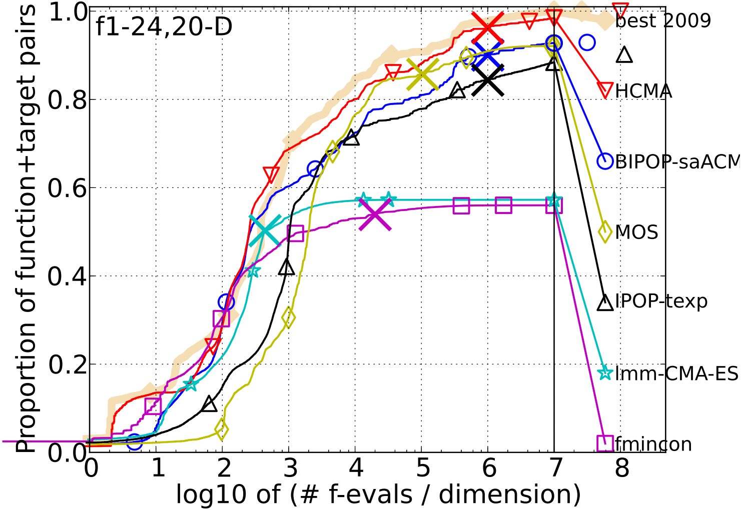
1. NEWUOA with $m = 2n + 1$ for $10n$ functions evaluations.
2. BIPOP-saACM-ES-k and STEP with $n_{MinIterSTEP} = 10$ (e.g., $10n$ evaluations).

[13] [Powell, 2006] "The NEWUOA software for unconstrained optimization without derivatives"

Efficient Optimization



Efficient Optimization



Conclusion

- Intensive surrogate model exploitatiom improves the performance on unimodal functions.
- STEP algorithm is a cheap tool to deal with separable problems.
- HCMA demonstrates the best overall performance.

Perspective

- Implement NEWUOA-like search within saACM-ES.
- Use alternative restart strategies (NBIPOP and NIPOP) in HCMA.

Thank you for your attention!

Questions?