Intensive Surrogate Model Exploitation in Self-adaptive Surrogate-assisted CMA-ES (saACM-ES)

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 9th, 2013

Historical overview: PPSN'2010, GECCO'2012

PPSN'2010

- CMA-ES assisted with comparison-based surrogate models (ACM algorithm) [1].
- Good performance but exploitation is independent on surrogate model quality, surrogate hyper-parameters are fixed.

GECCO'2012

- Self-adaptive surrogate-assisted CMA-ES (IPOP-saACM-ES and BIPOP-saACM-ES) on noiseless[2] and noisy testbeds[3].
- BIPOP-saACM-ES demonstrates good performance w.r.t. all algorithms tested during the BBOB-2009, 2010 and 2012.

[1] [Loshchilov, Schoenauer and Sebag; PPSN 2010] "Comparison-based optimizers need comparison-based surrogates"

[2] [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"

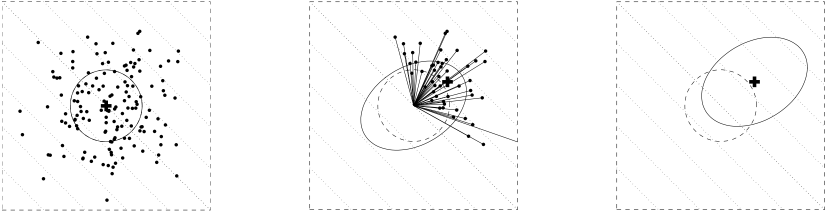
[3] [Loshchilov, Schoenauer and Sebag; GECCO-BBOB 2012] "Black-box optimization benchmarking of IPOP-saACM-ES on the BBOB-2012 noisy testbed"

Content

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

(μ, λ)-Covariance Matrix Adaptation Evolution Strategy

Rank-μ Update [4] [5]



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

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

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

sampling of λ solutions

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

new distribution

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

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

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

[5] the slide adopted by courtesy of Nikolaus Hansen

Invariance: Guarantee for Generalization

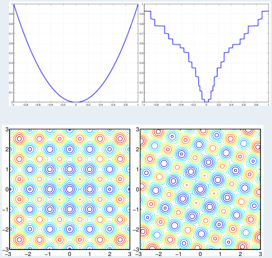
Invariance properties of CMA-ES

- 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"

Contents

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

Using Ranking SVM as the surrogate model

- Build a global model using Ranking SVM [6]

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

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

How to choose an appropriate Kernel?

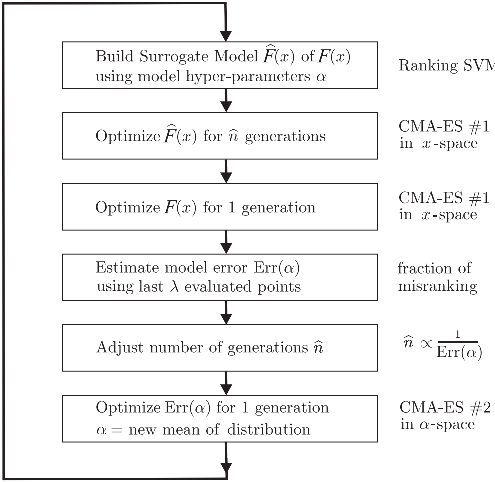
- Use covariance matrix C adapted by CMA-ES in Gaussian kernel[7]

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

✓ Invariance to rotation of the search space thanks to C

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

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

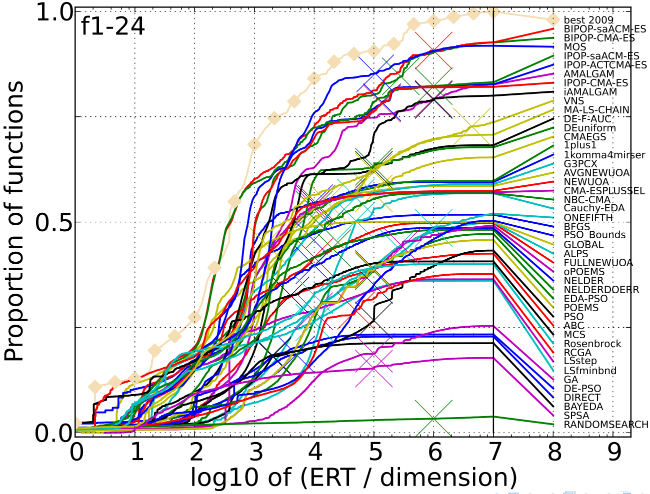


Surrogate-assisted CMA-ES with online adaptation of model hyper-parameters [a]

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

Results on Black-Box Optimization Competition

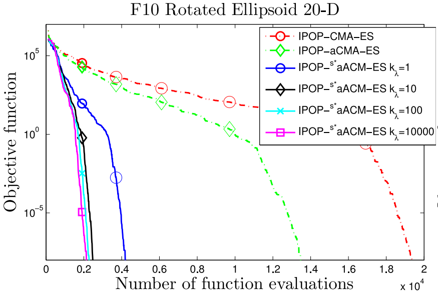
BIPOP-^{s*}aACM and IPOP-^{s*}aACM (with restarts) on 24 noiseless 20 dimensional functions

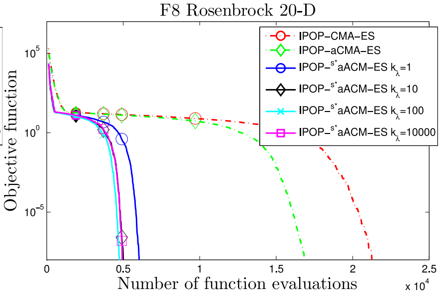


Contents

Intensive surrogate model exploitation

- When optimizing $\hat{F}$, $\lambda = k_\lambda \lambda_{def}$, $\mu = \mu_{def}$
- Parameter settings: $k_\lambda = 1$ for D<10 and $k_\lambda = 10, 100, 1000$ for 10, 20, 40-D.





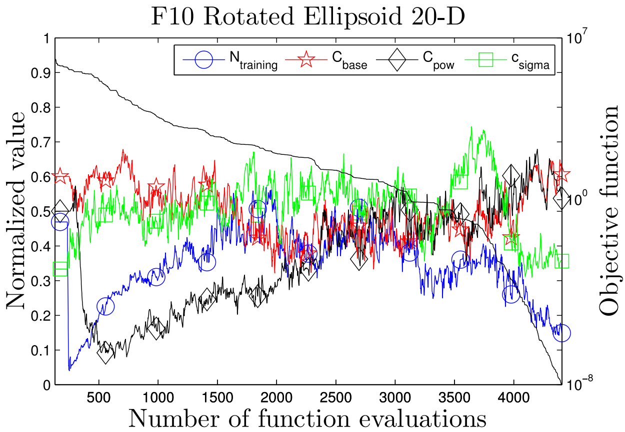
An interpretation: trust-region search.

Divergence prevention

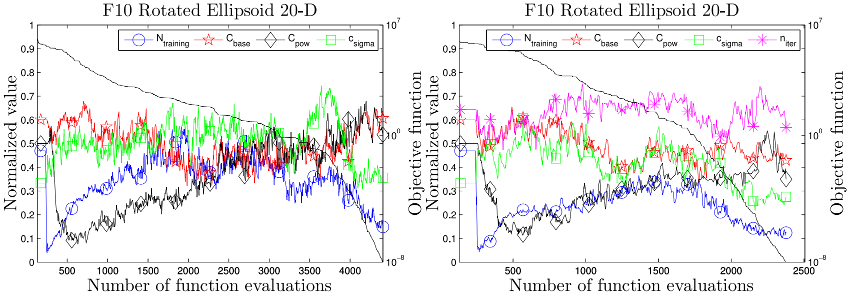
Surrogate model error estimation can be imprecise
The proposed simple fix:

- Exploit surrogate only if the model is "reasonably" precise: $k_\lambda > 1$ is used only of $\hat{n} \geq \hat{n}_{k_\lambda}$
- Parameter settings: $\hat{n}_{k_\lambda} = 4$.

Surrogate model hyper-parameters adaptation: the original saACM

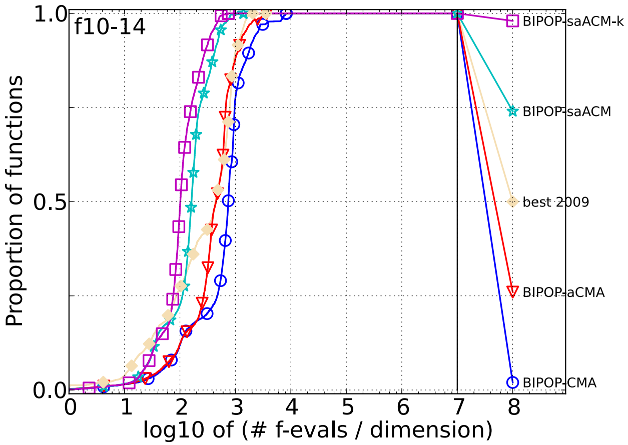


Surrogate model hyper-parameters adaptation: the proposed saACM-k

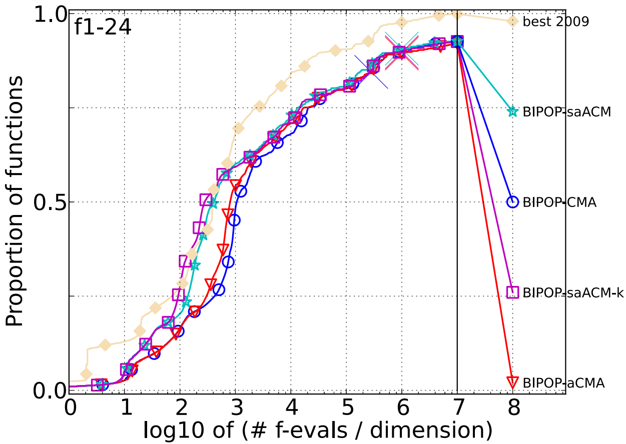


- CPU time per function evaluation: 1.22 sec. ($k_\lambda = 1$), 0.313 sec. ($k_\lambda = 10$), 0.311 sec. ($k_\lambda = 100$), 0.367 sec. ($k_\lambda = 1000$), 1.52 sec. ($k_\lambda = 10000$).
- Here, for $k_\lambda = 100$ it is even computationally cheaper!

Ill-conditioned problems



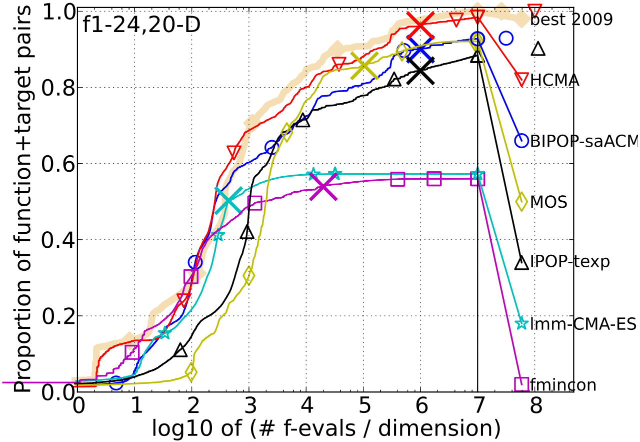
All BBOB problems



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

Hybridization

- HCMA = NEWUOA (first $10n$ evaluations) + STEP (to exploit separability if any) + BIPOP-saACM-k.



Conclusion

Pros:

- Faster convergence, especially on ill-condition problems.
- Potentially smaller CPU per function evaluation.

Cons:

- Decrease of performance if surrogate error estimation is imprecise.

Perspective

- Kullback-Leibler divergence measure for surrogate model control.[8]

[8] [Loshchilov, Schoenauer and Sebag; CAP 2013] "KL-based Control of the Learning Schedule for Surrogate Black-Box Optimization"

Thank you for your attention!

Questions?