CPS with a Surrogate Model and Active CMA

Alexander Fabisch – DFKI GmbH, Robotics Innovation Center

Empirical Evaluation of Contextual Policy Search with a Comparison-based Surrogate Model and Active Covariance Matrix Adaptation

Contextual Function Optimization

Contextual policy search, an extension to the original problem formulation of policy search:

$$\arg\max_{\omega} \int_{s} p(\boldsymbol{s}) \int_{\theta} \pi_\omega(\boldsymbol{\theta}|\boldsymbol{s}) \mathbb{E}\left[R(\boldsymbol{\theta}, \boldsymbol{s})\right] d\boldsymbol{\theta} d\boldsymbol{s},$$

where $\boldsymbol{s} \in S$ is a context, π_ω is a stochastic upper-level policy parameterized by $\boldsymbol{\omega}$ that defines a distribution of policy parameters for a given context (Deisenroth et al., 2013). The return R is extended to take into account the context. During the learning process, we optimize $\boldsymbol{\omega}$, observe the current context $\boldsymbol{s}$, and select $\boldsymbol{\theta}_i \sim \pi_\omega(\boldsymbol{\theta}|\boldsymbol{s})$.

Contextual black-box optimization, the corresponding deterministic problem formulation:

$$\arg\min_{\omega} \int_{s} f_s(g_\omega(\boldsymbol{s})) d\boldsymbol{s},$$

where f_s is a parameterized objective and we want to find an optimal function g_ω from a parameterized class of functions.

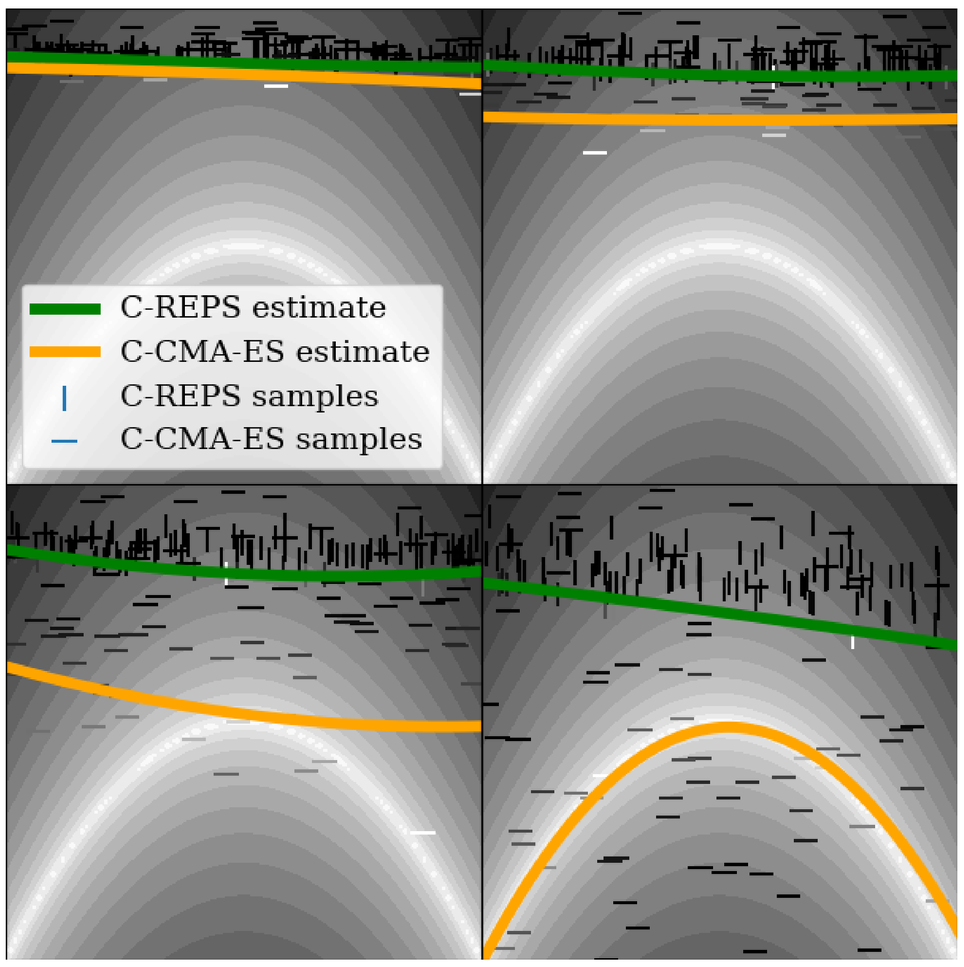


Illustration of 1D contextual function optimization.

Extending C-CMA-ES to aC-ACM-ES

C-CMA-ES (Abdolmaleki et al., 2017) is based on CMA-ES (Hansen and Ostermeier, 2001). We transfer two extensions of CMA-ES to C-CMA-ES: active CMA-ES (Jastrebski and Arnold, 2006) and ACM-ES (Loshchilov et al., 2010), which uses a surrogate model.

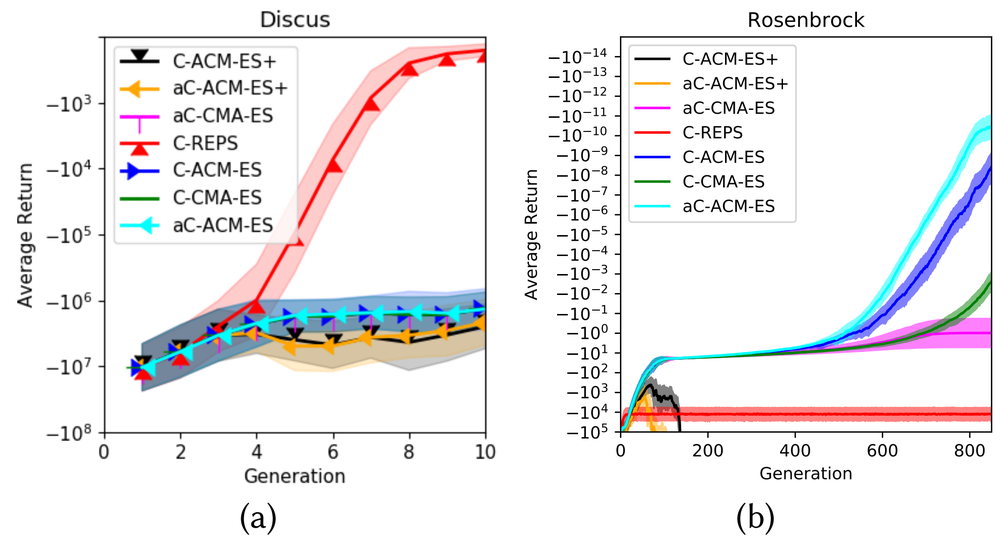
Hyperparameters: We have two configurations. Standard and aggressive exploitation of the surrogate model. We set the number of samples the surrogate after the model is accurate enough to be used to $n_{start} = 3000$ or $n_{start} = 100$. The number of samples tested with the surrogate model is set to $\lambda' = 3\lambda$ and $\lambda' = 10\lambda$ respectively. The population size is $\lambda = 50$. Larger values for n_{iter}, the number of iterations to train the surrogate model, improve the result. As a compromise between computational overhead and sample-efficiency, we select $n_{iter} = 1000$. c_{pow} is a parameter of the ranking SVM objective. Although in the original ACM-ES (Loshchilov et al., 2010) the default value is 2, $c_{pow} = 1$ works better for C-ACM-ES.

Evaluation

Experiments are similar to the ones of Abdolmaleki et al. (2017) with additional objectives. We make standard benchmarks contextual by defining $f_{\boldsymbol{s}}(\boldsymbol{\theta}) = f(\boldsymbol{\theta} + \boldsymbol{G}\boldsymbol{s})$, where components of the matrix $\boldsymbol{G}$ are sampled iid from $\mathcal{N}(0, 1)$. In our case $\boldsymbol{\theta} \in \mathbb{R}^{20}$ and $\boldsymbol{s} \in \mathbb{R}^{n_s}$. Components of $\boldsymbol{s}$ are sampled from $[1, 2)$. To make results comparable to the one of Abdolmaleki et al. (2017), we use the same sphere and Rosenbrock functions. In addition, we use the Ackley function and ellipsoidal, discus, and different powers from the COCO platform (Hansen et al., 2016).

Algorithms are compared in Table 1. We use C-REPS with $\epsilon = 1$ and C-CMA-ES as baselines. aC-CMA-ES refers to active C-CMA-ES, C-ACM-ES uses the surrogate model, and aC-ACM-ES combines both. "+" indicates aggressive exploitation of the surrogate model.

Results: Variants of C-ACM-ES outperform vanilla C-CMA-ES. C-REPS is often much faster in the early phase (see Figure 2 (a)). In the first 10 generations C-REPS outperforms all algorithms by orders of magnitude. This phase is interesting, e.g., for optimization in robotics. However, C-REPS converges too early while variants of C-CMA-ES will make progress (see Figure 2 (b)).



Learning curves (mean and standard deviation of 20 experiments) for (a) Discus and (b) Rosenbrock function.

Objective	Sphere	Rosenbrock	Ackley	Ellipsoidal	Diff. Powers	Discus
n_s	2	1	1	1	1	1
Test after generation	200	850	1100	800	600	850
Method	Average objective function value over contexts: $\frac{1}{\lvert S\rvert}\sum_{s\in S} f_s(x)$					
C-REPS	$-4.509 \cdot 10^{+01}$	$-1.255 \cdot 10^{+04}$	$-1.947 \cdot 10^{+01}$	$-2.944 \cdot 10^{+05}$	$-9.088 \cdot 10^{+02}$	$-1.288 \cdot 10^{+02}$
C-CMA-ES	$-1.815 \cdot 10^{-05}$	$-2.328 \cdot 10^{-03}$	$-8.762 \cdot 10^{-07}$	$-2.337 \cdot 10^{+02}$	$-1.562 \cdot 10^{-07}$	$-2.995 \cdot 10^{-10}$
aC-CMA-ES	$-1.348 \cdot 10^{-05}$	$-9.736 \cdot 10^{-01}$	$-8.773 \cdot 10^{-07}$	$-1.524 \cdot 10^{+02}$	$-3.038 \cdot 10^{-07}$	$-3.838 \cdot 10^{-10}$
C-ACM-ES+	$-1.294 \cdot 10^{-08}$	$-1.445 \cdot 10^{+15}$	NaN	$-1.300 \cdot 10^{+16}$	$-7.111 \cdot 10^{+74}$	$-8.297 \cdot 10^{+27}$
aC-ACM-ES+	$-1.506 \cdot 10^{-01}$	$-3.227 \cdot 10^{+19}$	NaN	$-2.407 \cdot 10^{+18}$	$-8.717 \cdot 10^{+82}$	$-1.250 \cdot 10^{+24}$
C-ACM-ES	$-6.257 \cdot 10^{-04}$	$-3.656 \cdot 10^{-09}$	$\underline{-3.995 \cdot 10^{-09}}$	$-1.039 \cdot 10^{-10}$	$-2.464 \cdot 10^{-14}$	$\underline{-8.877 \cdot 10^{-12}}$
aC-ACM-ES	$-2.309 \cdot 10^{-04}$	$\underline{-3.899 \cdot 10^{-11}}$	$-1.813 \cdot 10^{-08}$	$\underline{-2.388 \cdot 10^{-11}}$	$\underline{-1.284 \cdot 10^{-14}}$	$-1.684 \cdot 10^{-11}$

Results on several objective functions. Best results are underlined.

Conclusion

We demonstrated that active C-CMA-ES, C-ACM-ES and its combination yield impressive results on contextual function optimization problems in comparison to C-CMA-ES. We have shown, however, that these results are actually not directly transferable to the domains where we would like to learn successful contextual policies in 100–1000 episodes at maximum.

Literature

Abdolmaleki, A., Price, B., Lau, N., Reis, L., and Neumann, G. (2017). Contextual covariance matrix adaptation evolutionary strategies. In *IJCAI*, pages 1378–1385.

Deisenroth, M., Neumann, G., and Peters, J. (2013). A survey on policy search for robotics. *Foundations and Trends in Robotics*, 2(1–2):1–142.

Hansen, N., Auger, A., Mersmann, O., Tusar, T., and Brockhoff, D. (2016). COCO: A platform for comparing continuous optimizers in a black-box setting. *CoRR*, abs/1603.08785.

Hansen, N. and Ostermeier, A. (2001). Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195.

Jastrebski, G. and Arnold, D. (2006). Improving Evolution Strategies through Active Covariance Matrix Adaptation. In *CEC*, pages 2814–2821.

Loshchilov, I., Schoenauer, M., and Sebag, M. (2010). Comparison-Based Optimizers Need Comparison-Based Surrogates. In *PPSN*, pages 364–373. Springer.

Universität Bremen

Contact:
Alexander Fabisch
alexander.fabisch@dfki.de