Gaussian Process Surrogate Models for the CMA-ES (Extended Abstract)

Lukáš Bajer
Cisco Systems, Prague, Czech Republic, and
Czech Academy of Sciences, Institute of Computer Science
Pod Vodárenskou věží 2, Prague, Czech Republic
bajeluk@gmail.com

Zbyněk Pitra
Czech Academy of Sciences, Institute of Computer Science,
and Czech Technical University
Břehová 7, Prague, Czech Republic
z.pitra@gmail.com

Jakub Repický
Czech Academy of Sciences, Institute of Computer Science,
and Charles University
Malostran. nám. 25, Prague, Czech Republic
j.repicky@gmail.com

Martin Holeňa
Czech Academy of Sciences, Institute of Computer Science
Pod Vodárenskou věží 2, Prague, Czech Republic
martin@cs.cas.cz

ABSTRACT

This extended abstract previews the usage of Gaussian processes in a surrogate-model version of the CMA-ES, a state-of-the-art black-box continuous optimization algorithm. The proposed algorithm DTS-CMA-ES exploits the benefits of Gaussian process uncertainty prediction, especially during the selection of points for the evaluation with the surrogate model. Very brief results are presented here, while much more elaborate description of the methods, parameter settings and detailed experimental results can be found in the original article *Gaussian Process Surrogate Models for the CMA Evolution Strategy* [2], to appear in the *Evolutionary Computation*[1].

CCS CONCEPTS

• **Computing methodologies → Continuous space search**; *Model development and analysis*; Uncertainty quantification;

KEYWORDS

black-box optimization, evolutionary optimization, surrogate modelling, Gaussian process

ACM Reference Format:
Lukáš Bajer, Zbyněk Pitra, Jakub Repický, and Martin Holeňa. 2019. Gaussian Process Surrogate Models for the CMA-ES (Extended Abstract). In *Genetic and Evolutionary Computation Conference Companion (GECCO '19 Companion), July 13–17, 2019, Prague, Czech Republic.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3319619.3326764

1 INTRODUCTION

The CMA-ES [5] stands for one of the most successful evolutionary continuous black-box optimizers of the last two decades. It iteratively samples λ points from a Gaussian distribution $\mathcal{N}(\mathbf{m}, \sigma^2\mathbf{C})$ and then recalculates the distribution parameters $\mathbf{m}$, σ and $\mathbf{C}$ based on the μ best fitness-evaluated points.

The CMA-ES learns some core characteristics of the fitness via perturbations of its covariance matrix $\mathbf{C}$ and the step-size σ. Nevertheless, the exact information from the passed evaluations can be used more intensively via *surrogate modeling*. Such acceleration was previously shown, for example, in [7] or [8]. Our contribution is in more intensive exploitation of the Gaussian process uncertainty prediction in the algorithm DTS-CMA-ES (Doubly Trained Surrogate CMA-ES). The algorithm uses a GP surrogate model for evaluation of the most promising points every iteration, and the selection of these points relies on the GP ability to estimate the whole distribution of the predicted fitness values.

This abstract introduces the proposed evolution control and the DTS-CMA-ES algorithm and briefly overviews the experimental results of Gaussian process surrogate CMA-ES algorithms on the COCO single-objective benchmark testing set.

2 DOUBLY TRAINED EVOLUTION CONTROL

As the so-far existing *evolution controls*—methods of combining evaluations from the original and model fitness—makes the exploitation of the GP predictive uncertainty difficult, we have proposed another solution called *doubly trained* evolution control. Each generation of this control can be summarized in the following steps:

(1) sample a new population of size λ (CMA-ES offspring),
(2) train the *first* surrogate model on the points from the archive,
(3) select $\lceil \alpha\lambda \rceil$ point(s) wrt. a criterion C and the *first* model,
(4) evaluate these point(s) with the original fitness,
(5) re-train the surrogate model (into the *second* model),
(6) predict the fitness for the non-original evaluated points with this *second* model.

Employing the doubly trained evolution control in the CMA-ES results in the Doubly trained surrogate CMA-ES (DTS-CMA-ES, [9]) whose pseudocode is shown in Algorithm 1. The algorithm is parametrized by the parameter $\alpha^{(0)}$—the initial value of the ratio of original-evaluated points in a population, by the criterion C for the selection of these points, and by the surrogate model and its parameters. The ratio α adapts itself throughout the run in the self-adaptive version of the algorithm.

[1] Article available at https://www.mitpressjournals.org/doi/abs/10.1162/evco_a_00244

GECCO '19 Companion, July 13–17, 2019, Prague, Czech Republic

ACM ISBN 978-1-4503-6748-6/19/07…$15.00
https://doi.org/10.1145/3319619.3326764

3 IMPLEMENTATION DETAILS AND EXPERIMENTAL RESULTS

The original article [2] provides the details of the proposed algorithms, their parameter tuning and the chosen parameter settings. Particularly, it includes:

- GP model-training procedure,
- GP model parameter tuning (training set selection methods, training-set size, GP covariance function, hyperparameter initialization, etc.) and implementation details,
- DTS-CMA-ES parameter tuning (ratio α, population size λ, original re-evaluation criteria C),
- DTS-CMA-ES self-adaptation method and its parameter settings.

The article evaluates six algorithms based on the CMA-ES and a GP model: the S-CMA-ES [1] (the predecessor of DTS-CMA-ES), DTS-CMA-ES in both the fixed and self-adaptive version, the MA-ES [11], GPOP [3] and SAPEO [12]. For the sake of comparison with other state-of-the-art optimizers, we present also the results of the IPOP-CMA-ES with both the recommended and the doubled population size, two other surrogate-assisted CMA-ES algorithms lmm-CMA [7] and ^{s*}ACM-ES [8], a Bayesian optimizer called SMAC [6], and two local-search numerical optimization algorithms based on a trust region method: the BOBYQA algorithm [10] and the interior-point method [4] from the Matlab `fmincon` function.

Figure 1 asses the set of 13 algorithms in 10-D, using the COCO-provided ECDF graphs[2]. Based on the much more detailed results in the original work, we conclude that the fixed-α-ratio DTS-CMA-ES represents an algorithm of choice for multimodal functions with weak global structure and is very eligible for unimodal landscapes, too, especially in lower dimensions. The self-adaptive version of the DTS-CMA-ES, on the other hand, excels on the globally decreasing multimodal functions where it outperforms other compared algorithms.

Acknowledgements. This research was supported by the Czech Science Foundation grants No. 17-01251 and 18-18080S. Access to the MetaCentrum cloud under the "Projects of Large Research, Development, and Innovations Infrastructures" (CESNET LM2015042) is greatly appreciated.

[2] A raw COCO results dataset is already available at the COCO website: http://coco.gforge.inria.fr/doku.php?id=algorithms-bbob

Algorithm 1: DTS-CMA-ES (simplified pseudocode)

Input : initial value of $\alpha^{(0)}$, C, GP model

$\mathcal{A} \leftarrow \emptyset$; λ, $\sigma^{(0)}$, $\mathbf{m}^{(0)}$, $\mathbf{C}^{(0)} \leftarrow$ CMA-ES initialize

for *generation* $g = 0, 1, 2, \ldots$ *until stopping conditions met* **do**

- $\mathbf{x}_k \sim \mathcal{N}(\mathbf{m}^{(g)}, (\sigma^{(g)})^2\mathbf{C}^{(g)})$ for $k = 1, \ldots, \lambda$
- $f_{\mathcal{M}1} \leftarrow$ trainModel$(\mathcal{A}, \sigma^{(g)}, \mathbf{C}^{(g)})$ /* 1^{st} *model train* */
- $(\hat{\mathbf{y}}, \hat{\mathbf{s}}^2) \leftarrow f_{\mathcal{M}1}([\mathbf{x}_1, \ldots, \mathbf{x}_\lambda])$ /* *model-fitness eval* */
- $\mathbf{X}_{\text{orig}} \leftarrow$ select $\lceil \alpha^{(g)}\lambda \rceil$ best points according to $C(\hat{\mathbf{y}}, \hat{\mathbf{s}}^2)$
- $\mathbf{y}_{\text{orig}} \leftarrow f(\mathbf{X}_{\text{orig}})$, $\mathcal{A} = \mathcal{A} \cup \{(\mathbf{X}_{\text{orig}}, \mathbf{y}_{\text{orig}})\}$
- $f_{\mathcal{M}2} \leftarrow$ trainModel$(\mathcal{A}, \sigma^{(g)}, \mathbf{C}^{(g)})$
- $\mathbf{y} \leftarrow f_{\mathcal{M}2}([\mathbf{x}_1, \ldots, \mathbf{x}_\lambda])$ /* 2^{nd} *model prediction* */
- $\alpha^{(g+1)} \leftarrow$ selfAdaptation$(\alpha^{(g)}, \hat{\mathbf{y}}, \mathbf{y})$
- $\sigma^{(g+1)}, \mathbf{m}^{(g+1)}, \mathbf{C}^{(g+1)} \leftarrow$ update based on sorted $\mathbf{x}_{1:\lambda}$

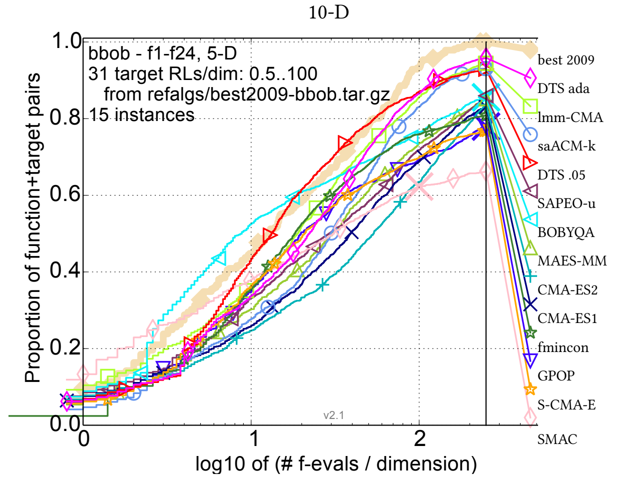

Figure 1: Aggregated results of 13 algorithms, COCO's ECDF performance graphs in 10-D. Bootstrapped empirical cumulative distribution of the number of objective function evaluations divided by dimension (FEs/D).

REFERENCES

[1] Lukáš Bajer, Zbyněk Pitra, and Martin Holeňa. 2015. Benchmarking Gaussian Processes and Random Forests Surrogate Models on the BBOB Noiseless Testbed. In *Proceedings of the Companion Publication of the 2015 on Genetic and Evolutionary Computation Conference (GECCO Companion '15)*. ACM, New York, NY, USA, 1143–1150. https://doi.org/10.1145/2739482.2768468

[2] Lukáš Bajer, Zbyněk Pitra, Jakub Repický, and Martin Holeňa. 2019. Gaussian Process Surrogate Models for the CMA Evolution Strategy. *Evolutionary Computation* to appear (2019), 30. https://doi.org/10.1162/evco_a_00244

[3] D. Büche, N. N. Schraudolph, and P. Koumoutsakos. 2005. Accelerating evolutionary algorithms with Gaussian process fitness function models. *IEEE Transactions on Systems, Man, and Cybernetics, Part C* 35, 2 (2005), 183–194.

[4] Richard H. Byrd, Jean Charles Gilbert, and Jorge Nocedal. 2000. A trust region method based on interior point techniques for nonlinear programming. *Mathematical Programming* 89, 1 (Nov. 2000), 149–185.

[5] Nikolaus Hansen and Andreas Ostermeier. 2001. Completely Derandomized Self-Adaptation in Evolution Strategies. *Evolutionary Computation* 9, 2 (June 2001), 159–195.

[6] Frank Hutter, Holger Hoos, and Kevin Leyton-Brown. 2013. An Evaluation of Sequential Model-based Optimization for Expensive Blackbox Functions. In *Proceedings of the 15th GECCO Companion*. ACM, 1209–1216.

[7] Stefan Kern, Nikolaus Hansen, and Petros Koumoutsakos. 2006. Local Meta-models for Optimization Using Evolution Strategies. In *Parallel Problem Solving from Nature – PPSN IX (LNCS)*, Vol. 4193. Springer, 939–948.

[8] Ilya Loshchilov, Marc Schoenauer, and Michele Sebag. 2013. Intensive Surrogate Model Exploitation in Self-adaptive Surrogate-assisted CMA-ES (saACM-ES). In *Proceedings of the 15th Annual Conference (GECCO '13)*. ACM, New York, 439–446.

[9] Zbyněk Pitra, Lukáš Bajer, and Martin Holeňa. 2016. Doubly trained Evolution Control for the Surrogate CMA-ES. In *Parallel Problem Solving from Nature – PPSN XIV (LNCS)*, Vol. 9921. Springer, 59–68.

[10] Michael J. D. Powell. 2009. *The BOBYQA algorithm for bound constrained optimization without derivatives.* Technical Report NA2009/06. University of Cambridge, Cambridge, UK.

[11] H. Ulmer, F. Streichert, and A. Zell. 2003. Evolution strategies assisted by Gaussian processes with improved preselection criterion. In *The 2003 Congress on Evolutionary Computation, 2003. CEC '03*, Vol. 1. 692–699 Vol.1.

[12] Vanessa Volz, Günter Rudolph, and Boris Naujoks. 2017. Investigating Uncertainty Propagation in Surrogate-assisted Evolutionary Algorithms. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO '17)*. ACM, New York, 881–888.