Parallelized bayesian optimization for problems with expensive evaluation functions

Rebolledo, Margarita; Rehbach, Frederik; Eiben, A. E.; Bartz-Beielstein, Thomas

published in
GECCO '20
2020

DOI (link to publisher)
10.1145/3377929.3390017

document version
Publisher's PDF, also known as Version of record

citation for published version (APA)
Rebolledo, M., Rehbach, F., Eiben, A. E., & Bartz-Beielstein, T. (2020). Parallelized bayesian optimization for problems with expensive evaluation functions. In *GECCO '20: Proceedings of the 2020 Genetic and Evolutionary Computation Conference Companion* (pp. 231-232). Association for Computing Machinery, Inc. https://doi.org/10.1145/3377929.3390017

Parallelized Bayesian Optimization for Problems with Expensive Evaluation Functions

Margarita Rebolledo
Institute for Data Science, Engineering, and Analytics
TH Köln, Gummersbach, Germany
margarita.rebolledo@th-koeln.de

Frederik Rehbach
Institute for Data Science, Engineering, and Analytics
TH Köln, Gummersbach, Germany
frederik.rehbach@th-koeln.de

A.E. Eiben
Department of Computer Science,
Vrije Universiteit Amsterdam
a.e.eiben@vu.nl

Thomas Bartz-Beielstein
Institute for Data Science, Engineering, and Analytics
TH Köln, Gummersbach, Germany
thomas.bartz-beielstein@th-koeln.de

ABSTRACT

Many black-box optimization problems rely on simulations to evaluate the quality of candidate solutions. These evaluations can be computationally expensive and very time-consuming. We present and approach to mitigate this problem by taking into consideration two factors: The number of evaluations and the execution time. We aim to keep the number of evaluations low by using Bayesian optimization (BO) – known to be sample efficient– and to reduce wall-clock times by executing parallel evaluations. Four parallelization methods using BO as optimizer are compared against the inherently parallel CMA-ES. Each method is evaluated on all the 24 objective functions of the Black-Box-Optimization-Benchmarking test suite in their 20-dimensional versions. The results show that parallelized BO outperforms the state-of-the-art CMA-ES on most of the test functions, also on higher dimensions.

CCS CONCEPTS

• **Mathematics of computing** → *Bayesian computation*; **Probabilistic algorithms**; • **Theory of computation** → **Timed and hybrid models**; **Parallel computing models**; • **Computing methodologies** → **Modeling and simulation**;

KEYWORDS

Parallel Optimization, Bayesian optimization, CMAES, BBOB

ACM Reference Format:
Margarita Rebolledo, Frederik Rehbach, A.E. Eiben, and Thomas Bartz-Beielstein. 2020. Parallelized Bayesian Optimization for Problems with Expensive Evaluation Functions. In *Genetic and Evolutionary Computation Conference Companion (GECCO '20 Companion), July 8–12, 2020, Cancún, Mexico.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3377929.3390017

1 INTRODUCTION

The evaluation of candidate solutions in real world optimization problems is often expensive and time-consuming. Therefore, the maximum number of evaluations is usually limited to only a few hundred evaluations for such problems.

In principle, it is possible to reduce the total running time by executing multiple evaluations in parallel. This, however, requires more computational resources. Population-based algorithms, like evolutionary algorithms (EAs), which can handle several candidate solutions per iteration can easily benefit from parallel computation. Among EAs the covariance matrix adaptation evolution strategy (CMA-ES) is the state-of-the-art for real-valued optimization [5] .

Reducing the number of function evaluations is another important improvement for the optimization problem's efficiency. Bayesian optimization (BO)[2, 8] achieves these reductions by implementing a data-driven surrogate model. Other than CMA-ES, BO is not directly applicable for multiple parallel evaluations but several adaptations exists [4]. Considering this, two research questions arise:

- **RQ-1** Can parallel variants of BO outperform inherently parallel algorithms like CMA-ES if the evaluation budget is severely constrained?
- **RQ-2** Which parallel variants of BO show the best performance?

To answer these questions we perform an empirical study. To make execution times practicable, we do not work with expensive real-world problems. Instead, we use the Black-Box-Optimization-Benchmarking test suite (BBOB) [6] and maintain a maximum of 100 evaluations.

2 METHODOLOGY

Two optimization algorithms were implemented and compared: BO and CMA-ES.
BO is an iterative global optimization framework useful for expensive black-box derivative free problems. Four parallelization methods were tested: investment portfolio improvement (IPI) [9], multi-point expected improvement (q-EI) [3], multi-objective infill criteria (MOI) [1], and multi-kernel Bayesian optimization (mK-BO). mK-BO refers to an approach where a different BO configuration is run on each free core of the parallel environment. For example, one core implements BO with expected improvement as an acquisition function while another core runs BO with predicted value as its acquisition function.

The CMA-ES is a state-of-the-art EA for optimizing non-linear non convex black-box functions. Given its population-based nature

GECCO '20 Companion, July 8–12, 2020, Cancún, Mexico

ACM ISBN 978-1-4503-7127-8/20/07…$15.00
https://doi.org/10.1145/3377929.3390017

it is easily implementable in parallel environments without any notable modification.

3 EXPERIMENTS

The algorithms were extensively tested on the BBOB benchmarking test suite. The test suite contains 24 single objective test functions organized into five groups: Separable, low to moderate conditioning, unimodal and high conditioned, multi-modal with adequate global structure, and lastly, multi-modal with weak global structure.

BO uses Gaussian Process [7] as a surrogate model. The radial basis function $\Sigma_i = \exp(-\sum_{j=1}^{n} \theta_j |x_j - x'|^{p_j})$, where p_j determines the correlation function smoothness and θ_j the extend of a point's x_i influence, was selected as the Kernel for all BO implementations. Two variants are tested: P2, where $p = 2$, and FitP, where p is part of the optimization loop. Three different acquisition functions are implemented: expected improvement (EI), lower confidence bound (LCB), and predicted value (PV) [8]. In total six different implementations are tested: Three acquisition functions are combined with the two different variants of kernel configuration. These combinations make up the mK-BO approach. Since there are six processors available for the experiments,CMA-ES is initialized with a population size, $\lambda = 6$. The default step-size, $\sigma = 0.5$, is maintained. The BO variants are started with an initial latin hypercube design of 10 points. Since q-EI and MOI require initial values larger than input dimensionality both algorithms are started with the next multiple of six number of initial samples.

All BBOB functions are tested on their 20 dimensional versions. Each experiment has a maximum evaluation budget of 100 and is repeated 30 times for statistical analysis.

4 RESULTS

To determine whether a significant difference between the algorithms is present the Kruskal-Wallis rank sum test is used. If the test is positive, a post-hoc test according to Conover, for pairwise multiple comparisons, checks for differences in each algorithm pair. The comparisons further used to rank the algorithms.

Parallel-BO implementations seemed to converge faster and have better performance than CMA-ES, even on higher dimensions. Figure 1 shows an example of this behavior on the Sharp Ridge function. Here it can be seen that for the 20-dimensional case MOI outperforms all other methods, followed by mK-BO. Interestingly CMA-ES is only better than random search (RS). Convergence plots and ranked Box-plots for all functions can be found on the supplementary material. Only for functions $Weierstrass$ and $Katsura$ did BO not achieve better or equal performance than CMA-ES. Both functions are highly rugged and repetitive.

Between the different parallel implementations, it was observed that the approaches MOI and mK-BO often perform better than q-EI and IPI.

5 CONCLUSIONS

To answer our research questions the performance of different parallel-BO implementations was compared against CMA-ES and random search on a simulated expensive black box optimization problem with a maximum of 100 function evaluations.

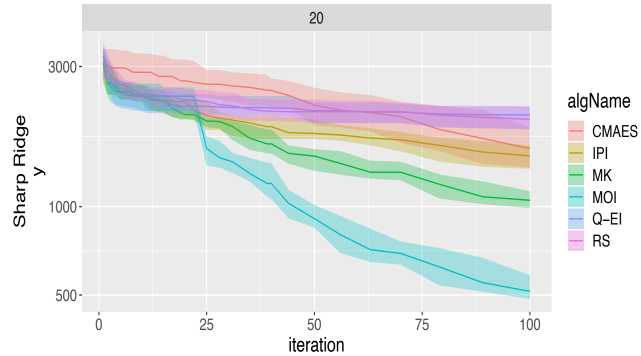

Figure 1: Convergence plots for the 20-dimensional Sharp Ridge function. Y-axis: function value (lower is better). Median, upper, and, lower quartiles are shown.

RQ-1: Can parallel variants of BO outperform inherently parallel algorithms like CMA-ES if the evaluation budget is severely constrained? We demonstrated on several function landscapes that parallel-BO can outperform inherently parallel CMA-ES on problems with limited function evaluations.

RQ-2: Which parallel variants of BO show the best performance on the tested problem landscapes? Parallelization methods MOI and mK-BO showed the better performance. Of both approaches, MOI outperformed the other methods on most cases and is our preferred approach for problems similar to the ones tested in this work.

Our ongoing and future work is aimed at testing the parallelization strategies with a real world simulation problem. Furthermore, the effect of mixing different regression models on the parallelization methods remains to be tested.

REFERENCES

[1] Bernd Bischl, Simon Wessing, Nadja Bauer, Klaus Friedrichs, and Claus Weihs. 2014. MOI-MBO: Multiobjective Infill for Parallel Model-Based Optimization. In *Learning and Intelligent Optimization.*

[2] Peter I. Frazier. 2018. A Tutorial on Bayesian Optimization. (2018). arXiv:stat.ML/1807.02811

[3] David Ginsbourger, Rodolphe Le Riche, and Laurent Carraro. 2010. *Kriging Is Well-Suited to Parallelize Optimization.* Springer Berlin Heidelberg, Berlin, Heidelberg, 131–162. https://doi.org/10.1007/978-3-642-10701-6_6

[4] Raphael T Haftka, Diane Villanueva, and Anirban Chaudhuri. 2016. Parallel surrogate-assisted global optimization with expensive functions–a survey. *Structural and Multidisciplinary Optimization* 54, 1 (2016), 3–13.

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

[6] Hansen Nikolaus, Finck Steffen, Ros Raymond, and Anne Auger. 2009. *Real-Parameter Black-Box Optimization Benchmarking 2009: Noiseless Functions Definitions.* Research Report inria- 00362633v2. INRIA.

[7] CE. Rasmussen and CKI. Williams. 2006. *Gaussian Processes for Machine Learning.* MIT Press, Cambridge, MA, USA. 248 pages.

[8] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas. 2016. Taking the Human Out of the Loop: A Review of Bayesian Optimization. *Proc. IEEE* 104, 1 (Jan 2016), 148–175. https://doi.org/10.1109/JPROC.2015.2494218

[9] Rasmus K. Ursem. 2014. From Expected Improvement to Investment Portfolio Improvement: Spreading the Risk in Kriging-Based Optimization. In *Parallel Problem Solving from Nature – PPSN XIII*, Thomas Bartz-Beielstein, Jürgen Branke, Bogdan Filipič, and Jim Smith (Eds.). Springer International Publishing, Cham, 362–372.