Are we Forgetting about Compositional Optimisers in Bayesian Optimisation?

Antoine Grosnit[*,1] **Alexander I. Cowen-Rivers**[*,1] **Rasul Tutunov**[*,1]
Ryan-Rhys Griffiths[1, 2] **Jun Wang**[1, 3] **Haitham Bou-Ammar**[1, 3]

[1]Huawei Noah's Ark Lab
[2]University of Cambridge
[3]University College London
[*]Equal contributions

Abstract Bayesian optimisation presents a sample-efficient methodology for global optimisation. Within this framework, a crucial performance-determining subroutine is the maximisation of the acquisition function, a task complicated by the fact that acquisition functions tend to be non-convex and thus nontrivial to optimise. In this paper, we undertake a comprehensive empirical study of approaches to maximise the acquisition function. Additionally, by deriving novel, yet mathematically equivalent, compositional forms for popular acquisition functions, we recast the maximisation task as a compositional optimisation problem, allowing us to benefit from the extensive literature in this field. We highlight the empirical advantages of the compositional approach to acquisition function maximisation across 3958 individual experiments comprising synthetic optimisation tasks as well as tasks from Bayesmark. Given the generality of the acquisition function maximisation subroutine, we posit that the adoption of compositional optimisers has the potential to yield performance improvements across all domains in which Bayesian optimisation is currently being applied. An open-source implementation is made available at `https://github.com/huawei-noah/noah-research/tree/CompBO/BO/HEBO/CompBO`.

1 Reproducibility Checklist

1. For all authors...

   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes] Our abstract and introduction clearly state our contributions (introduction of compositional form of acquisition functions, and extensive empirical study of the acquisition function maximisation step) and the scope of our paper (impact on Bayesian Optimisation framework).

   (b) Did you describe the limitations of your work? [Yes] We do describe the limitations of our work, showing that compositional forms is not always superior to standard form, discussing the memory-efficiency, as well as the time efficiency of the compositional optimisation with respect to their non-compositional counterpart.

   (c) Did you discuss any potential negative societal impacts of your work? [No] We did not include such section, but we wrote one (the one sent along with this checklist) that we would be happy to include in our paper if it is accepted at AutoML

   (d) Have you read the ethics author's and review guidelines and ensured that your paper conforms to them? `https://automl.cc/ethics-accessibility/` [Yes] We read the review guidelines and ensured that our paper conforms to them. The reported figures are of high-quality, with contrasted colors and with curves enriched with different line styles and markers. We adopted an inclusive language, and did not include any toxic speech.

AutoML Conference 2022 © 2022 the authors, released under CC BY 4.0

2. If you are including theoretical results...

   (a) Did you state the full set of assumptions of all theoretical results? [Yes] The assumptions are always fully provided, either in the main paper or in the Appendix.

   (b) Did you include complete proofs of all theoretical results? [Yes] We provide the proofs with detailed computations for all of our theoretical results (notably in Appendix A. and Appendix F.)

3. If you ran experiments...

   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results, including all requirements (e.g., `requirements.txt` with explicit version), an instructive `README` with installation, and execution commands (either in the supplemental material or as a URL)? [Yes] We put in the abstract a URL pointing to a public Github repository containing the code we developped to run our experiments. This repository comprises a `requirements.txt` file with explicit version, and a detailed `README` providing installation instructions and snippets to rerun our large set of experiments on synthetic tasks easily. We also provide a set of bash scripts to retune the algorithm hyperparameters as described in the paper, and run the BO with the best found hyperparameters. Even though the provided code cover the majority of the experiments carried out for our paper, we could not include in our repo the code we run for some baseline (BOHB) and tasks (based on `Bayesmark`) as they come from a projects under less permissive licence than MIT.

   (b) Did you include the raw results of running the given instructions on the given code and data? [No] Due to the large amount of outputs produced (we ran more than 3900 experiments) we did not include them in our repo but provide instructions to regenerate them. We also include one notebook indicating how to retrieve the results once the experiments are run.

   (c) Did you include scripts and commands that can be used to generate the figures and tables in your paper based on the raw results of the code, data, and instructions given? [Yes] We provide a Notebook in our repo (as described in the README.md) showing how to generate the main figures of our paper from raw results, and the way to build tables from the raw data is made straightforward.

   (d) Did you ensure sufficient code quality such that your code can be safely executed and the code is properly documented? [Yes] The code has been tested on a new machine (that we had not used for our experiments before), and we made efforts to make the code easily understandable thanks to a clear structure and a clear documentation for all the important functions and classes defined in the code.

   (e) Did you specify all the training details (e.g., data splits, pre-processing, search spaces, fixed hyperparameter settings, and how they were chosen)? [Yes] All these details are explicitly specified in the Experiment Section 4. and in the Appendix H.

   (f) Did you ensure that you compared different methods (including your own) exactly on the same benchmarks, including the same datasets, search space, code for training and hyperparameters for that code? [Yes] We made sure to compare methods fairly, running them on the same benchmarks, using the same seeds, and using the same procedure for the tuning of hyperparameters (always using Bayesian optimisation with the same number of iterations).

   (g) Did you run ablation studies to assess the impact of different components of your approach? [Yes] We conduct an extensive study to assess the impact of the choice of the acquisition function (Figure 6), the optimiser order, and the use of the compositional form (Figure 1), as well as the use of the memory efficient computation scheme (Figure 8).

(h) Did you use the same evaluation protocol for the methods being compared? [Yes] For each experiment, we granted all methods the same batch size, number of initial observations, and total number of acquisition steps.

(i) Did you compare performance over time? [Yes] We do compare the performance obtained when using compositional and non-compositional forms (Section 4.4 on Runtime Efficiency).

(j) Did you perform multiple runs of your experiments and report random seeds? [Yes] As reported in the paper, we ran each experiment using at least 5 seeds.

(k) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] We always report the average performance and standard deviation associated to the use of multiple random seeds.

(l) Did you use tabular or surrogate benchmarks for in-depth evaluations? [Yes] We used standard benchmarks widely accepted by the community (synthetic functions from `BoTorch` library, and hyperparameter tuning tasks from the `Bayesmark` library.)

(m) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] For the experiments involving comparison of the execution time, we specify which and how many devices were used.

(n) Did you report how you tuned hyperparameters, and what time and resources this required (if they were not automatically tuned by your AutoML method, e.g. in a NAS approach; and also hyperparameters of your own method)? [Yes] We provide details on the choice of hyperparameters in Appendix H., indicating that we use Bayesian optimisation and providing the list of tuned hyperparameters along with their tuning domains. Moreover we specify the number of acquisitions we did to tune the hyperparameters which allows to extrapolate the time spent on hyperparameter tuning.

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets…

(a) If your work uses existing assets, did you cite the creators? [Yes] We did cite the creators of the assets we used. We notably cited the creators the `PyTorch`, `GPyTorch`, and `BoTorch` libraries that we used to implement our compositional optimisers and acquisition functions. We also cited the creators of the Diabetes UCI dataset that we used in our experiments.

(b) Did you mention the license of the assets? [No] We cited the creators of the library and dataset we used but did not additionally specify under which license each asset is (though we made sure each was compatible with our project). If it appears to be necessary, we will make sure these appear in the final version. Besides, we ran a code-dependency check before releasing the code, to ensure it could be released under MIT license.

(c) Did you include any new assets either in the supplemental material or as a URL? [Yes] We provided the link to our implementation of the compositional optimisers and acquisition functions.

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A] We did not use any sensitive dataset and relied on synthetic function optimisation or on widely known open-source anonymous datasets (Diabetes UCI dataset).

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A] We did not use any sensitive dataset and relied on synthetic function optimisation or on widely known open-source anonymous datasets (Diabetes UCI dataset).

5. If you used crowdsourcing or conducted research with human subjects…

   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A] We did not use crowdsourcing nor conducted research with human subjects

   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A] We did not use crowdsourcing nor conducted research with human subjects

   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A] We did not use crowdsourcing nor conducted research with human subjects