A New Acquisition Function for Robust Bayesian Optimization of Unconstrained Problems

Sibghat Ullah
LIACS, Leiden University
Leiden, The Netherlands
s.ullah@liacs.leidenuniv.nl

Hao Wang
LIACS, Leiden University
Leiden, The Netherlands
h.wang@liacs.leidenuniv.nl

Stefan Menzel
Honda Research Institute Europe
Offenbach/Main, Germany
stefan.menzel@honda-ri.de

Bernhard Sendhoff
Honda Research Institute Europe
Offenbach/Main, Germany
bernhard.sendhoff@honda-ri.de

Thomas Bäck
LIACS, Leiden University
Leiden, The Netherlands
t.h.w.Baeck@liacs.leidenuniv.nl

ABSTRACT

A new acquisition function is proposed for solving robust optimization problems via Bayesian Optimization. The proposed acquisition function reflects the need for the robust instead of the nominal optimum, and is based on the intuition of utilizing the higher moments of the improvement. The efficacy of Bayesian Optimization based on this acquisition function is demonstrated on four test problems, each affected by three different levels of noise. Our findings suggest the promising nature of the proposed acquisition function as it yields a better robust optimal value of the function in 6/12 test scenarios when compared with the baseline.

CCS CONCEPTS

• **Computing methodologies** → *Search with partial observations.*

KEYWORDS

Bayesian Optimization, Kriging , Robust Optimization

ACM Reference Format:
Sibghat Ullah, Hao Wang, Stefan Menzel, Bernhard Sendhoff, and Thomas Bäck. 2021. A New Acquisition Function for Robust Bayesian Optimization of Unconstrained Problems. In *2021 Genetic and Evolutionary Computation Conference Companion (GECCO '21 Companion), July 10–14, 2021, Lille, France.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3449726.3463206

1 INTRODUCTION

Real-world scenarios of robust design optimization (RDO) can encompass some of the most complicated optimization setups due to many shapes and forms of uncertainty [1, 3], variety of fitness landscapes and high dimensionality [2]. The famous Bayesian Optimization (BO) algorithm [1] has been adapted to efficiently solve RDO problems, and is referred to as Robust Bayesian Optimization (RDO) [4] in this paper. The performance of the RBO algorithm is greatly determined by the acquisition function (AF) which balances the trade-off of exploration and exploitation. Following the intuition of utilizing the higher moment of the improvement [5], we propose a new AF to solve RDO problems. The proposed AF is based on the moment-generating function of the improvement (MGFI) [5], and is compatible with RBO algorithm [4] to solve RDO problems.

2 ROBUST BAYESIAN OPTIMIZATION

We aim to minimize a continuous black-box function, i.e., $f\colon \mathcal{S} \subseteq \mathbb{R}^D \to \mathbb{R}$, where uncertainty—denoted by Δ—is presenting on the decision variables. To formulate robustness, we employ the so-called minimax principle which tries to minimize the maximum possible realization of the objective with respect to the uncertainty set $\mathrm{U} \subseteq \mathbb{R}^D$ [3]. Effectively, this refers to minimizing the following objective function: $g(\mathbf{x}) = \max_{\Delta \in \mathrm{U}} f(\mathbf{x} + \Delta_{\mathbf{x}})$.

The MGFI [5] is an AF to solve nominal optimization problems via BO [1], and utilizes the best-so-far observed value of the function—denoted by $f_{\min}$—to formulate the improvement. However, to adapt the MGFI to the robust scenario, this reference value becomes meaningless as it does not convey enough information on the best robust optimal value (ROV) achieved thus far. Therefore, one should consider the current ROV predicted by the Kriging model, namely $r_{\mathcal{K}} = \min_{\mathbf{x} \in \mathcal{S}} \max_{\Delta_{\mathbf{x}} \in \mathrm{U}} \mathcal{K}_f(\mathbf{x} + \Delta_{\mathbf{x}})$, as a substitute for $f_{\min}$. Note that $\mathcal{K}_f$ refers to the Kriging model. Furthermore, one should consider the non-trivial problem of defining the improvement over $r_{\mathcal{K}}$ since for each point $\mathbf{x}$, the minimax principle considers effectively the worst improvement that can be realized by a stochastic process $\{Y(\mathbf{x} + \Delta_{\mathbf{x}})\colon \Delta_{\mathbf{x}} \in \mathrm{U}\}$ over $r_{\mathcal{K}}$, which is an optimization task under uncertainty itself and thereby difficult to solve. Practically, we consider a point $\mathbf{x}_{\max}$ which corresponds to the worst predicted value concerning the uncertainty set U, i.e., $\mathbf{x}_{\max} = \mathbf{x} + \arg\max_{\Delta \in \mathrm{U}} \mathcal{K}_f(\mathbf{x} + \Delta_{\mathbf{x}})$, and take the random response $Y(\mathbf{x}_{\max})$ at this point to define the improvement, namely the robust improvement as: $\mathcal{I}_r(\mathbf{x}_{\max}) = \max\{0, r_{\mathcal{K}} - Y(\mathbf{x}_{\max})\}$ [4]. Together, these two modifications allow us to adapt the nominal MGFI [5] to the robust MGFI, which is denoted by RMGFI. For more details on the RBO and the MGFI, please refer to [4] and [5] respectively.

3 EXPERIMENTAL SETUP

To gauge the ability of the proposed AF, we compare it against the baseline, namely the robust expected improvement criterion (REIC)

GECCO '21 Companion, July 10–14, 2021, Lille, France

ACM ISBN 978-1-4503-8351-6/21/07…$15.00
https://doi.org/10.1145/3449726.3463206

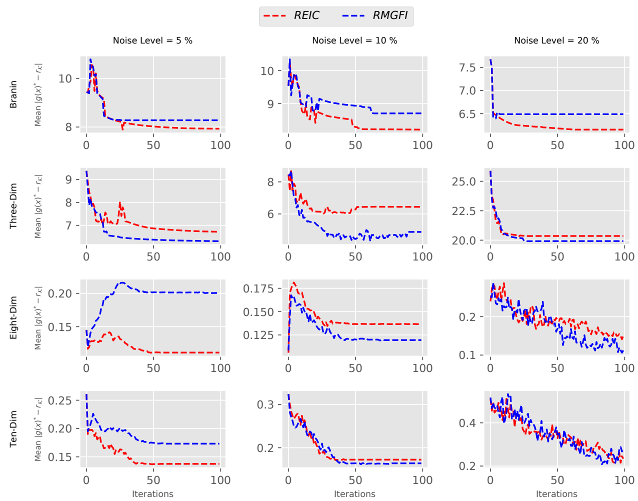

Figure 1: Mean Absolute Difference to the globally robust optimal function value based on the proposed RMGFI and the baseline REIC.

proposed in [4]. The comparison involves four test problems in total, all taken from the existing literature [4]. Three of these problems, namely the "Three", the "Eight" and the "Ten-Dimensional" problems are taken from [4], in addition to a two-dimensional "Branin" function. The comparison also involves three noise levels based on 5, 10 and 20 % deviation in the nominal values of the decision variables, giving rise to a total of 12 test scenarios for comparison. In addition to the baseline comparison, we alter the configuration of the initial temperature to comprehend the role it plays in the performance of the proposed AF. To this end, we only choose the "Branin" and the "Three-Dimensional" problems alongside three different configuration for the initial temperature. The results originating from the baseline comparison and the variation of the initial temperature are presented in the next section.

4 RESULTS

Graphs pertaining to the baseline comparison are presented in figure 1. Note that each column of plots in this figure corresponds to a specific noise level, whereas the rows distinguish between different problem instances. Each subplot in this figure presents two curves based on the two AFs. Each of these curves indicates the mean absolute difference (MAD) to the globally robust optimal function value (GROFV) based on 25 independent runs. From figure 1, we observe that in 6/12 cases, the RMGFI yields a better optimal function value whereas the implementation with the REIC performs superior in 5/12 cases. In particular, the REIC exceeds the RMGFI for all three test scenarios related to the "Branin" function whereas the RMGFI is better on the "Three-Dimensional" problem. Additionally, it can be observed that both acquisition functions perform competitively on the third noise level. Note that in the test scenarios of the "Eight" and the "Ten-Dimensional" problems concerning the first noise level, the RMGFI undershoots the GROFV. Next, the results based on the variation of the initial temperature are presented in Table 1. In this table, the first column reads the optimization problem at hand, whereas the next three columns describe the noise level, initial temperature, and the MAD to the GROFV accompanied with the SE. An important observation from Table 1 suggests that for the "Branin" function, the best performance is achieved for the

Table 1: Mean Absolute Difference (MAD) to the globally robust optimal function value based on three different settings of the initial temperature.

Problem	Noise Level	Initial Tempera..	MAD
Branin	5 %	1	8.08 ± 1.44
		3	8.87 ± 1.59
		5	8.31 ± 1.48
	10 %	1	7.78 ± 2.23
		3	8.46 ± 2.22
		5	7.81 ± 2.18
	20 %	1	5.88 ± 0.85
		3	6.20 ± 1.01
		5	6.09 ± 1.01
Three-Dimensional	5 %	1	6.58 ± 1.09
		3	5.66 ± 0.90
		5	5.83 ± 0.95
	10 %	1	5.61 ± 0.79
		3	5.27 ± 0.55
		5	5.75 ± 0.75
	20 %	1	21.01 ± 2.02
		3	20.15 ± 1.81
		5	19.63 ± 1.91

low initial temperature. On the other hand, the higher settings for the initial temperature giver rise to a better performance on the "Three-Dimensional" problem.

5 CONCLUSION AND OUTLOOK

In this paper, we presented and evaluated a new AF to efficiently solve RDO problems via BO. The observations from the previous section suggest the promising nature of the proposed AF as it yields a better GROFV when compared with the baseline. Additionally, it can be observed that the choice of initial temperature is problem-dependent, which could be solved via hyper-parameter optimization. The limitations of the current study suggest that future research is necessary to validate these findings on more complex problems.

ACKNOWLEDGMENTS

This research has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement number 766186 (ECOLE).

REFERENCES

[1] Donald R. Jones, Matthias Schonlau, and William J. Welch. 1998. Efficient Global Optimization of Expensive Black-Box Functions. *J. Glob. Optim.* 13, 4 (1998), 455–492. https://doi.org/10.1023/A:1008306431147

[2] Sibghat Ullah, Duc Anh Nguyen, Hao Wang, Stefan Menzel, Bernhard Sendhoff, and Thomas Bäck. 2020. Exploring Dimensionality Reduction Techniques for Efficient Surrogate-Assisted Optimization. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2965–2974.

[3] Sibghat Ullah, Hao Wang, Stefan Menzel, Bernhard Sendhoff, and Thomas Back. 2019. An empirical comparison of meta-modeling techniques for robust design optimization. In *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 819–828.

[4] Samee ur Rehman, Matthijs Langelaar, and Fred van Keulen. 2014. Efficient Kriging-based robust optimization of unconstrained problems. *J. Comput. Sci.* 5, 6 (2014), 872–881. https://doi.org/10.1016/j.jocs.2014.04.005

[5] Hao Wang, Bas van Stein, Michael Emmerich, and Thomas Back. 2017. A new acquisition function for Bayesian optimization based on the moment-generating function. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 507–512.