Bayesian multi-objective optimization for stochastic simulators

S. Basak[†,1,2], J. Bect[§,2], L. Guillier[§,1], F. Tenenhaus-Aziza[§,3], E. Vazquez[§,2]

[†] PhD student (presenting author). [§] PhD supervisor

PhD expected duration: Jan. 2021 – Dec. 2023

[1] Agence nationale de sécurité sanitaire (ANSES)
{prénom}.{nom}@anses.fr

[2] Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes (L2S)
{prénom}.{nom}@centralesupelec.fr

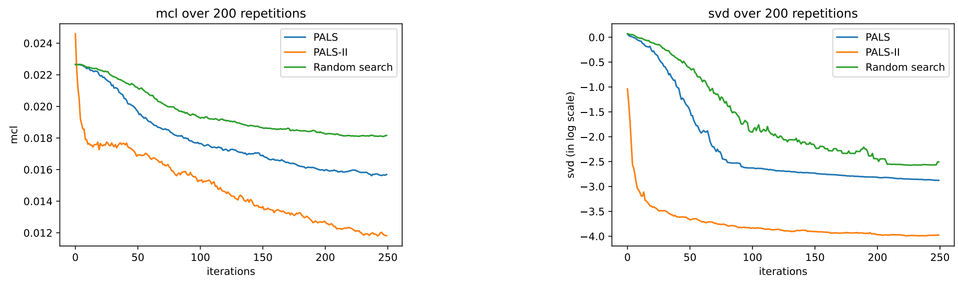
[3] Centre National Interprofessionnel de l'Economie Laitière (CNIEL)
ftenenhaus@cniel.com

Abstract

We consider the problem of multi-objective optimization for stochastic simulators, where each evaluation of the objectives $f = (f_1, f_2, \ldots, f_q)$ at a given point x in a discrete finite search domain $\mathbb{X}$ yields random responses $Z_i = f_i(x) + \epsilon_i, 1 \leq i \leq q$, with possibly heteroscedastic noise ϵ_i. The best possible trade-off between the conflicting objectives is determined by the Pareto domination rule : $f(x) \prec f(x')$ (x Pareto dominates x'), when $f_i(x) \leq f_i(x')$ for all i, with at least one of the inequalities being strict. The goal is to estimate the set of Pareto optimal solutions $\mathcal{P} = \{x \in \mathbb{X} : \nexists x' \in \mathbb{X}, f(x') \prec f(x)\}$, from a limited number of evaluations of the objective functions.

In this work, we propose a modified version of the PALS (Pareto Active Learning for Stochastic simulators) algorithm [1] called PALS-II, that significantly improves the optimization performance in a high noise setting and we perform a numerical benchmark (see Figure 1) comparing its performance with several other competing approaches. The idea is to construct an efficient, inexpensive and easy to implement algorithm, unlike other Bayesian approaches in the literature based on computationally-intensive criteria for point selection. The metrics proposed to assess the performance are the volume of the symmetric difference, which assesses the global error on Pareto front estimates (in the objective space), and the misclassification rate, which measures the global error on the Pareto set estimates (in the input space). PALS builds a Gaussian Process surrogate model over the computationally expensive stochastic simulator and iteratively classifies each point in $\mathbb{X}$ as either potentially Pareto optimal (P_n), non Pareto optimal (N_n) or unclassified (U_n), using confidence hyper-rectangles of the form $\{z \in \mathbb{R}^q : \mu_n(x) - \beta^{1/2}\sigma_n(x) \prec z \prec \mu_n(x) + \beta^{1/2}\sigma_n(x)\}$, where β is a parameter controlling the size of the confidence region, and where $\mu_n(x)$ and $\sigma_n^2(x)$ stand for the GP posterior mean and variance at the n-th iteration. The point with the highest uncertainty (maximum rectangle diameter) is evaluated at the next iteration. PALS suffers from one major drawback: the tuning of the parameter β can be difficult. It is not directly related to a probabilistic interpretation in terms of Pareto domination and it affects the performance in a significant way.

To address this drawback, we propose a new rectangle-free classification rule that takes into account the global uncertainty and the correlation between the points. At the n-th iteration,we



(a) Misclassification rate (b) Volume of symmetric difference

Figure 1: Figure comparing misclassification rate (mcl) and volume of symmetric difference (svd) averaged over 200 repetitions for PALS, PALS-II and a baseline random search approach, for a test problem over a grid of size 21×21.

simulate sample paths from the GP, conditional on the observed evaluations. The conditional sample paths are used to obtain empirical estimates of the probabilities of being Pareto optimal for each point in $\mathbb{X}$. These probabilities are used to classify points in P_n, N_n and U_n. The rule for selecting the next evaluation point remains the same as PALS. We carry out a numerical comparison of PALS-II with other algorithms from the literature of Bayesian ranking and selection where the main idea is to allocate the simulation budget to the input design points such that the Pareto set can be estimated with high confidence. In particular, we compare PALS-II to MOCBA (Multi-objective Optimal Computing Budget Allocation [3]) that perform allocation by minimizing the asymptotic upper bounds of misclassification error of the design points.

Application case: In the context of the project ArtiSaneFood, we aim to find optimal process intervention parameters in cheese production, which translates into a multi-objective optimization problem of a stochastic simulator [2], based on a Quantitative Risk Assessment model.

Implementation details: PALS-II is coded in Python, using the open source packages `GPmp` (Gaussian process micro package) and `GPmp-contrib`. Our code will be published as a part of these packages, available at `https://github.com/gpmp-dev`.

Short biography (PhD student)

Subhasish Basak is a third-year doctoral student at the Laboratoire des Signaux et Systèmes (L2S), Université Paris-Saclay and at Agence nationale de sécurité sanitaire (ANSES). His thesis is a part of the project ArtiSaneFood project (grant number : ANR-18-PRIM-0015) which is part of the PRIMA program supported by the European Union. This project aims to establish methodological recommendations to cheese producers in France, in order to reduce the risk of Haemolytic Uremic Syndrome (HUS) in children, from the consumption of raw-milk soft cheese.

References

[1] B. Barracosa, J. Bect, H. Dutrieux Baraffe, J. Morin, J. Fournel, and E. Vazquez. Extension of the pareto active learning method to multi-objective optimization for stochastic simulators. In *SIAM Conference on Computational Science and Engineering (CSE21)*, Mar 2021.

[2] Subhasish Basak, Julien Bect, Laurent L. Guillier, Fanny Tenenhaus-Aziza, Janushan Christy, and Emmanuel Vazquez. Bayesian multi-objective optimization for quantitative risk assessment in microbiology. MASCOT-NUM 2022, June 2022. Poster.

[3] L. H. Lee, E. P. Chew, S. Teng, and D. Goldsman. Finding the non-dominated pareto set for multi-objective simulation models. *IIE Transactions*, 42(9):656–674, 2010.