A Framework for Bayesian Optimization in Mixture Spaces

Paola Driza *1 **Benjamin Sanchez-Lengeling** *1

*Equal contribution [1]*University of Toronto, Toronto, Canada M5S 3E5.* Correspondence to: Paola Driza paola.driza@mail.utoronto.ca.

1. Introduction

The formulation, synthesis, and optimization of mixtures is an integral facet of materials science and engineering. With a plethora of applications in olfaction [1], drug discovery [2], and consumer goods [3], many industrial, and more broadly, scientific challenges involve the selection and proportioning of mixture accords for downstream use. Traditionally reliant on extensive human-driven experimentation, the evaluation of these problems often necessitates costly laboratory measurements and manual workflows. The volume of the throughput is inherently lower, evidently highlighting the bottlenecks in manual experimentation, including: labour-intensive techniques, variability in reproducibility, data & analytical constraints, and limitations in ingenuity & creativity.

Bayesian Optimization (BO) has emerged as a powerful technique for optimizing objective functions, especially in cases where the function to optimize is ill-defined, or, more colloquially, a "black box" [4]. In a standard construction, BO consists of a stochastic surrogate model—used to approximate the typically-unknown objective function—and an acquisition function—a heuristic designed to search for optimal solutions in such a way that strikes some balance between uncertainty (exploration) and reward (exploitation). It is the acquisition function that guides the selection of the next recommendation(s) by quantitatively evaluating the information gained by proposing that point. (In the context of this paper, that point would indeed be a mixture formulation.) Iteratively selecting and evaluating new candidates in a closed-loop makes BO-powered systems ideal for high-throughput screening and autonomous experimentation, and is often coupled with a robotic platform or self-driving laboratory (SDL) to probe a large search space, generate data, update models, and overall refine experimentation with minimal human intervention with drastically-reduced experimental work.

However, the optimization of mixtures is governed by structural characteristics, chemical principles, and physical properties that are not intrinsic to standard BO pipelines. All of the aforementioned features that define mixtures give rise to constrained optimization; however, there remains a clear lack of a modelling hierarchy with imposed constraints for BO in the mixture formulation problem space, which compromises the efficacy of the optimization algorithm. In this work, we present a structured framework for BO in the mixture space. Our contributions are:

(1) Formalizing mixture design as the conjunction of accord selection and compositional allocation under practical and chemically-informed constraints.

(2) Introducing a taxonomy of constraint classes common to mixture formulation problems in the literature.

We provide a taxonomy for characterizing mixture formulation optimization problems into effective BO algorithm constraints, with additional considerations for autonomous experimental systems.

2. Substantial section

In this section, relevant literature at the intersection of the mixture formulation space and BO space will be presented to highlight the novelty of our work. Then, we will introduce a preliminary yet detailed constraint taxonomy to facilitate BO algorithm decision-making for mixtures.

2.1 Related work

Prior work on adapting BO pipelines for the mixture domain has made meaningful progress, yet a clear gap remains. Recent work focuses on the more sophisticated techniques of combining models with BO.

Variational autoencoders are a type of generative model featuring a latent space of condensed information from input data. Combining this with BO led to the development of the Latent Consistent Aware-Acquisition Function (LCA-AF) [5]. It achieves high sample-efficiency and effective exploration of the problem search space whilst penalizing latent inconsistencies. However, when these latent inconsistencies are nonetheless encountered, the points queried by the BO are less-than-optimal selections. The Categorical Optimisation With Belief of underlYing Structure (COWBOYS) approach addresses the coupling of the variational autoencoder and the BO (responsible for sub-optimal performance) [6]. The decoupled approach that COWBOYS employs trains the surrogate model and generative model separately, ameliorating the selection of queried points, however, this is at the expense of the sampling algorithms, with compromised efficiency.

Efforts in these papers are focused on modelling strategies and incorporating them with the BO, comprising highly complex restructuring of BO processes. Our work targets the component which modelling-focused BO algorithms often neglect to account for, which is constraint-based construction design for optimization.

2.2 Constraint Taxonomy

The following serves as an important constraint taxonomy that exists in the scope of mixture formulations, as a way to inform the implementation and tuning of optimization loops. The primary consideration that will influence the scope of the problem is the assumption of whether the mixture is linear or not. Strictly from a chemistry perspective, it is the case that mixtures are generally not linear, save for some edge cases that may exist, and which should be analysed thoroughly and confirmed before proceeding with subsequent constraint considerations.

Firstly, mixtures are compositional with non-negative proportions summing up to some fixed total. Secondly, a prototypical mixture formulation is sparse: of an entire library of ingredients, only a small subset is selected for the accord. As a consequence of this, having a fixed library of accords is another constraint due to availability of resources. Thirdly, mixture size—in terms of number of ingredients, concentrations, and volumetric considerations—is variable across solutions, and the way the problem is designed could lead to a discrete or continuous optimization task. Moreover, practical constraints govern industrial formulation problems, including: cost limitations, ingredient availability, solubility and volatility properties, and regulatory legislation (often inconsistent across differing jurisdictions). Furthermore, in the context of SDLs, the main constraint is minimizing the number of operations, and thus time, that the robot spends on programmed physical manipulations.

3. Results

One of the main problems in the space is optimizing mixture similarity of a collection of accords against one target formulation. Using constrained BO based on the taxonomy presented in 2.2, and comparing it to a random search algorithm, we use a standard benchmark function, `griewank`, to perform Griewank optimization on a challenging landscape.

The objective of this ten-dimensional problem was to maximize the output of this function, ergo the similarity score, with constrained bounds. Each algorithm was run for 20 iterations, and seeded with five randomly-sampled initial points. Each iteration resulted in five new candidates being proposed. The results displayed in Figure 1 demonstrate the superior performance and efficiency of the BO, relative to the random search algorithm. After a few iterations, the power of the constrained BO is evident, converging towards much higher similarity scores; in contrast, the random search stagnates with minimal performance gains over the course of the 20 iterations. Effectively, implementing a BO algorithm with constraints proves to be highly effective, achieving a similarity score of nearly 80%, while the random search without constraints achieves a similarity score of about 10%.

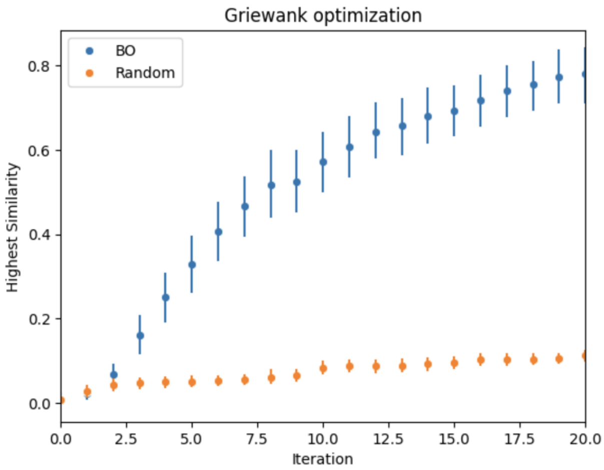

Fig. 1: Experimental results of Griewank benchmarking on constrained BO versus random search.

4. Conclusion

In conclusion, we propose a constraint taxonomy for informing constraint-based BO algorithms in the mixture formulation space, by formalizing mixture design as compositional accords under practical and chemical constraints. We report on a standard benchmark objective function comparing constrained BO against a random search algorithm, achieving a score of about 80% for the BO with constraints from 2.2.

Acknowledgments

The authors would like to acknowledge G. Tom for resources relating to BO with mixtures.

References

[1] Gary Tom, Cher Tian Ser, Ella M. Rajaonson, Stanley Lo, Hyun Suk Park, Brian K. Lee, and Benjamin Sanchez-Lengeling. From molecules to mixtures: Learning representations of olfactory mixture similarity using inductive biases, 2025. Available at: `https://arxiv.org/abs/2501.16271`.

[2] Alex K Chew, Mohammad Atif Faiz Afzal, Zachary Kaplan, Eric M Collins, Suraj Gattani, Mayank Misra, Anand Chandrasekaran, Karl Leswing, and Mathew D Halls. Leveraging high-throughput molecular simulations and machine learning for the design of chemical mixtures. *npj Computational Materials*, 11(1):72, 2025.

[3] Giacomo Squeo, Davide De Angelis, Riccardo Leardi, Carmine Summo, and Francesco Caponio. Background, applications and issues of the experimental designs for mixture in the food sector. *Foods*, 10(5):1128, 2021.

[4] Peter I. Frazier. A tutorial on bayesian optimization, 2018. Available at: `https://arxiv.org/abs/1807.02811`.

[5] Onur Boyar and Ichiro Takeuchi. Latent space bayesian optimization with latent data augmentation for enhanced exploration, 2024. Available at: `https://arxiv.org/abs/2302.02399`.

[6] Henry B. Moss, Sebastian W. Ober, and Tom Diethe. Return of the latent space cowboys: Rethinking the use of vaes for bayesian optimisation of structured spaces, 2025. Available at: `https://arxiv.org/abs/2507.03910`.