Multi Objective Regionalized Bayesian Optimization via Entropy Search

Thomas James 7HOMASJAMES@GMAIL.COM
Sinnu Susan Thomas SINNU.THOMAS@DUK.AC.IN
Digital University Kerala, India

Abstract

Line search optimization methods fail with multiple objective functions whose gradients are unavailable. The center of a crowded, trusted region is typically chosen as the point on the Pareto front with the highest hypervolume contribution. The proposed approach uses an entropy selection procedure to search the entire Pareto front, avoiding the computation of the Pareto front samples via cheap multi-objective optimization. By reducing uncertainty in each region, the algorithm directs its search towards areas with the highest potential for Pareto improvement. We tested the proposed method on the DTLZ test suite and other real-world applications, such as the welded beam design problem and the trajectory planning rover problem. The proposed approach yields results at par with state-of-the-art methods for exploring the Pareto front.
Keywords: Bayesian Optimization; Multi-Objective Optimization.

1. Introduction

Pursuing optimal outcomes needs explicit mathematical formulations, which are costly to analyze and require understanding the smoothness of the objective functions. This is where multi-objective Bayesian optimization proves to be an efficient strategy for tackling complex optimization problems. Bayesian optimization [15–17] excels in its ability to work with limited data. Moreover, it leverages existing knowledge, making it particularly well-suited for multi-objective scenarios [6, 18] where balancing tradeoffs between competing objectives is crucial. The acquisition function in Bayesian optimization was used for one-dimensional unconstrained optimization problems. However, most real-world problems must meet multiple objectives, resulting in multiple optimal solutions [3]. The optimal solutions for these problems form a Pareto Optimal Set [7], depending on the decision maker's preference [8]. To evaluate Pareto Optimal Solutions, the hypervolume metric was introduced, measuring the volume of non-dominated points [20]. Numerous Simulated Annealing techniques were proposed using hypervolume as a measure, though few methods address high-dimensional settings with a high cost [5]. However, these methods require strong assumptions about the problem's structure and often need to be improved if these assumptions are valid. This issue is predominant in multi-objective optimization, where all objectives must share the same assumed structure, which is often challenging to follow [9]. State-of-the-art approaches select the Pareto point that contributes the most to the hypervolume as the trust region center. However, this approach can result in poor coverage of the Pareto front, potentially neglecting regions with high potential for Pareto improvement. It prioritizes exploitation over achieving a balanced tradeoff between exploration and exploitation. We propose an Entropy-based search mechanism that computes the entropies of the Trusted Regions. By reducing uncertainty in each region, the algorithm directs its search towards areas with the highest potential for Pareto improvement. Our method ensures scalability to high dimensions, distributing the computational load across various regions of the solution space and optimizing complex objective functions in a parallel and scalable manner.

2. Multi Objective Regionalized Bayesian Optimization via Entropy Search

The proposed Multi-objective regionalized Bayesian optimization via entropy search (MORBES) utilizes a collaborative multi-trust region (multi-TR) approach for constrained higher-dimensional multi-objective Bayesian optimization. Unlike Multi-objective regionalized Bayesian optimization (MORBO) [5], which traditionally depends on hypervolume indicators, the proposed method employs entropy to evaluate the exploration potential of a Pareto point. It uses Tchebycheff's scalarization function to effectively identify solutions in non-convex regions of the Pareto front. Entropy-based methods such as MESMO and PESMO [9] maximize information gain, improving model predictions globally, ensuring that each evaluation is highly informative. Maximizing information gain is equivalent to minimizing entropy. By reducing entropy, MORBES effectively gathers more information about the objective space, particularly in unexplored regions. Information gain $I(\mathbf{x})$ is

$$I(\mathbf{x}) = I(f_{\mathbf{x}}; \{p^*\} \mid \mathcal{D}) \tag{1}$$

The information gain between the function values at $\mathbf{x}$ and the Pareto set p^*, given the observed data $\mathcal{D}$, denoted as $I(f_{\mathbf{x}}; \{p^*\} \mid \mathcal{D})$, quantifies the reduction in uncertainty of the Pareto frontier upon observing the objective values at $\mathbf{x}$ given the current dataset. This value captures how much knowledge about the Pareto frontier is enhanced by additional objective evaluations at $\mathbf{x}$ in the context of the existing data $\mathcal{D}$.

Unlike Max-value entropy search for multi-objective Bayesian optimization [1], MORBES does not require calculating the entire Pareto front. Instead, it needs to identify a point present in the Pareto front with the highest information gain relative to the referred point. So, Information gain $I(\mathbf{x})$ is redefined as

$$\begin{aligned} I(\mathbf{x}) &= I(p^*; p^* - \{X, Y\} \mid \mathcal{D}) && (2) \\ &= H(p^* \mid \mathcal{D}) - H(p^* - \{X, Y\} \mid \mathcal{D}) && (3) \end{aligned}$$

where $H(p^* - \{X, Y\} \mid \mathcal{D})$ represents the uncertainty of the Pareto Front, excluding the specified data point $\{X, Y\}$. The first mathematical term in Eq. 3 is

$$H(p^* \mid \mathcal{D}) = \frac{1}{2}\left[M + (\ln(2\pi)M) + N + (\ln(2\pi)N)\right] + \sum_{i=1}^{M} \ln(\sigma_{fi}^2(\mathbf{x})) + \sum_{j=1}^{N} \ln(\sigma_{ck}^2(\mathbf{x})) \tag{4}$$

The conditional entropy $H(p^* \mid \mathcal{D})$ quantifies the information content of the Pareto set p^* given the observed data $\mathcal{D}$. A Pareto point is identified as

$$\mathbf{x} = \arg\max_{\mathbf{x} \in \mathbf{X}^*} \left(H(p^* - \{X, Y\} \mid \mathcal{D})\right) \tag{5}$$

Using a sequential greedy approach, the proposed method chooses TR centers, omitting places previously chosen as the center for another TR with the help of a mask. While MORBES carries out local optimization inside a TR, a method based on hypervolume scalarization is used to

Algorithm 1: Multi Objective Regionalized Bayesian Optimization via Entropy Search

Data: n_{TR} (# of trust regions), Len_{init} (initial trust region length), Len_{max} (maximum trust region length), Len_{min} (minimum trust region length), f (objective function)

Result: Approximate Pareto Frontier

Initialize the trust regions $T_1, \ldots, T_{n_{TR}}$ and evaluate a starting set of points

Use the Entropy search center selection procedure to locate the centers.

while *budget remains* **do**

  Build a Gaussian model inside each trusted region Select m candidates using hypervolume improvement Evaluate candidates and obtain new data points

  for $i = 1$ **to** n_{TR} **do**

    Add new data points to trust regions Increment success/failure counters Update edge length Len_j for T_j

    if $Len_j < Len_{min}$ **then**

      Use Reinitialization Algorithm

    end

  end

end

return Approximate Pareto Frontier

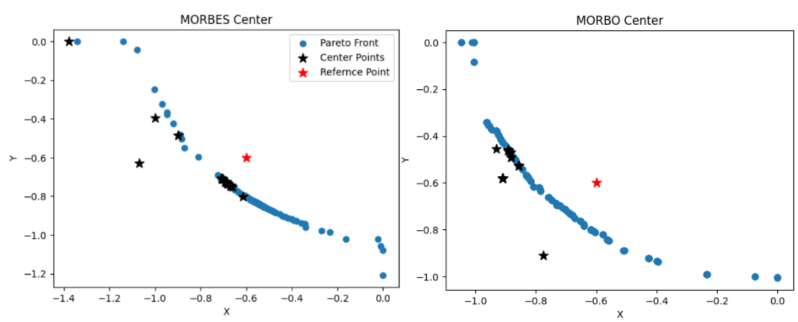


Figure 1: Exploration of Pareto Front concerning DTLZ2.

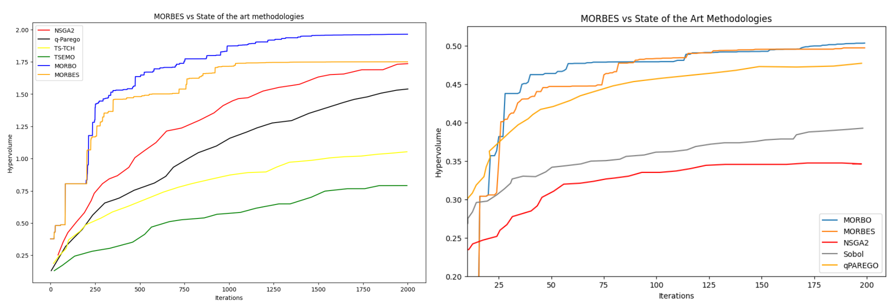
reinitialize the TRs to guarantee global optimization. This technique reduces a multi-objective optimization problem to a single-objective problem. As a result, maximizing the hypervolume is the same as maximizing the randomized single-objective scalarization problem [19]. The simplest form of scalarized function is the weighted sum scalarization. However, this locates the Pareto front's convex hull $f(M_{p^*})$ [10]. The weighted Tchebycheff's function $ch_c(\mathbf{y} \mid \lambda)$ with weights λ

$$\min_{\mathbf{y} \in \mathcal{X}} ch_c(\mathbf{y} \mid \lambda) = \min_{\mathbf{y} \in \mathcal{X}} \max_{1 \leq i \leq m} \{\lambda_i(f_i(\mathbf{y})) - \lambda_i(l_i^*) + \lambda_i(\epsilon)\}. \quad (6)$$

Its goal is to find a solution $\mathbf{y}$ inside the feasible set $\mathcal{X}$ that minimizes the maximum deviation from a reference point l_i^*. The vector-valued function $\mathbf{f}(\mathbf{y})$ evaluated at $\mathbf{y}$ has the i-th component represented by $f_i(\mathbf{y})$, λ_i is the weight assigned to the i-th component, and ϵ is a small positive value that ensures feasibility. The solution $\mathbf{y}$ is optimized to balance closeness to l_i^* while preserving the priority suggested by the weights λ. This is achieved by minimizing $ch_c(\mathbf{y} \mid \lambda)$.

Test Function / Problem	MORBES Convergence	MORBO Convergence	Observation
DTLZ2	$\sim$ 250 iterations	$\sim$ 400 iterations	MORBES converges faster, offering higher hypervolume and better non-dominated front.
DTLZ5	$\sim$ 1000 iterations	$>$ 2000 iterations	MORBES achieves convergence faster, reducing computational cost significantly.
DTLZ7	Same as MORBO	Same as MORBES	Both algorithms show similar performance in convergence and hypervolume for continuous/discontinuous regions.
Welded Beam Design Problem	$\sim$ 80 iterations	$\sim$ 120 iterations	MORBES converged earlier, performing at the same level as MORBO regarding hypervolume.
Rover Problem (Trajectory Planning)	$\sim$ 1100 iterations	$\sim$ 1500 iterations	MORBES converge faster, advantageous in scenarios focusing on computational efficiency.

Table 1: Comparison of MORBES and MORBO performance across different test functions.



Rover Problem Beam Problem

Figure 2: MORBES comparison concerning Bayesian and Non-Bayesian approaches.

Consider the collection of restart points to be $\mathcal{D}_{t-1}$. Using a sample from a global Gaussian Process [13], conditioned on $\mathcal{D}_{t-1}$ as a condition, one finds the center point $\mathbf{x}_t$ of the new trust region given $\mathcal{D}_{t-1}$ by maximizing a random hypervolume Tchebycheff scalarization of the objectives under a sample from a global Gaussian Process. This is known as the Reinitialization algorithm. MORBES algorithm is detailed in Algorithm 1.

The cumulative regret for entropy-based methods is bounded by $R_T \leq \mathcal{O}(T \cdot \gamma_T)$, where γ_T is the maximum information gain following T observations. As γ_T decreases over time, entropy-based methods achieve a logarithmic regret bound: $R_T \leq \mathcal{O}(\log(T))$. This shows that entropy-based methods reduce uncertainty more efficiently [14], leading to faster convergence than other optimization techniques.

3. Experimental Results

The experimental results of MORBES on diverse synthetic and real-world benchmarks are described, with experimentation conducted on an NVIDIA DGX A100 server.

The proposed method's architecture dynamics revolve around the TR center selection mechanism. As Fig.1 shows, MORBO selects its centers from more crowded regions. In contrast, the

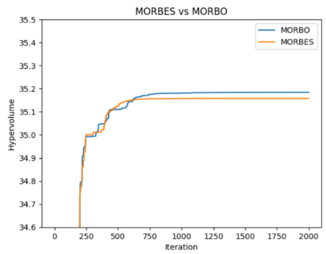


Figure 3: DTLZ2

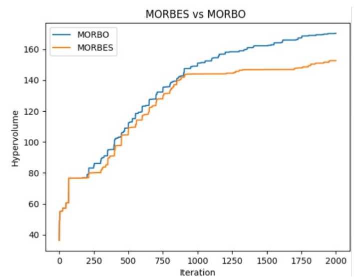


Figure 4: DTLZ5

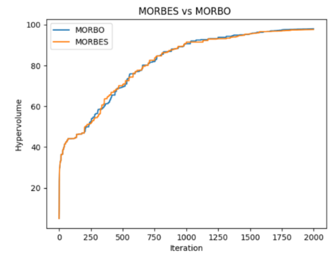
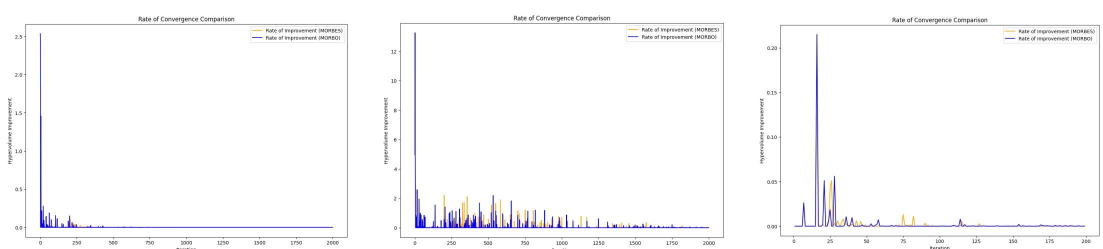


Figure 5: DTLZ7

proposed approach explores the Pareto front concerning the reference point, providing a broader and more conclusive approach.

Our method is applied to real-world problems such as the welded beam design and trajectory planning of the Rover problem (Fig.2). MORBES is compared to both Bayesian approaches such as MORBO and non-Bayesian multi-objective optimization techniques such as qPAREGO [4], NSGA2 [6], TS-TCH [12], TSEMO [2] and Sobol [11], where the superiority of MORBES is well established regarding the hypervolume indicator.

Additionally, MORBES is tested on benchmark problems such as DTLZ2 (Fig.3), DTLZ5 (Fig.4), and DTLZ7 (Fig.5) to evaluate various aspects such as convergence, scalability, and the ability to handle many-objective scenarios. Table 1 presents a comprehensive analysis of the results and their implications. Based on the convergence rate shown in Fig. 6, MORBO has an advantage in the early optimization stages, with its higher peaks indicating faster initial gains. However, MORBES demonstrates a steadier and more consistent approach throughout, which is beneficial in reaching a more reliable outcome. It outperforms MORBO by having a higher median hypervolume.



DTLZ2 DTLZ7 Welded Beam

Figure 6: Convergence Analysis of MORBES concerning MORBO.

4. Conclusion

The proposed approach evaluates the true objective function at selected points, dynamically adapting the Trusted Regions. This iterative refinement process enhances the search, allowing MORBES to efficiently identify the Pareto set, even in scenarios where the objective function is analytically complex or computationally expensive. We propose enhancing the architecture by creating an acquisition function based on Output Space Entropy. This approach aims to overcome the limitations associated with hypervolume improvement, offering a more robust solution that eliminates the need for a cheaply calculated Pareto front.

References

[1] Syrine Belakaria, Aryan Deshwal, and Janardhan Rao Doppa. Max-value entropy search for multi-objective bayesian optimization. *NeurIPS*, 32:9025–9036, 2019.

[2] Eric Bradford, Artur M Schweidtmann, and Alexei Lapkin. Efficient multiobjective optimization employing gaussian processes, spectral sampling and a genetic algorithm. *Journal of global optimization*, 71(2):407–438, 2018.

[3] Giorgio Chiandussi, Marco Codegone, Simone Ferrero, and Federico Erminio Varesio. Comparison of multi-objective optimization methodologies for engineering applications. *Computers & mathematics with applications*, 63(5):912–942, 2012.

[4] Samuel Daulton, Maximilian Balandat, and Eytan Bakshy. Differentiable expected hypervolume improvement for parallel multi-objective bayesian optimization. *Advances in Neural Information Processing Systems*, 33:9851–9864, 2020.

[5] Samuel Daulton, David Eriksson, Maximilian Balandat, and Eytan Bakshy. Multi-objective bayesian optimization over high-dimensional search spaces. In *UAI*, volume 180, pages 507–517, 2022.

[6] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and TAMT Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE transactions on evolutionary computation*, 6 (2):182–197, 2002.

[7] Mark Fleischer. The measure of pareto optima applications to multi-objective metaheuristics. In *EMO*, pages 519–533, 2003.

[8] Andreia P Guerreiro, Carlos M Fonseca, and Luís Paquete. The hypervolume indicator: Computational problems and algorithms. *ACM computing surveys*, 54(6):1–42, 2021.

[9] Daniel Hernández-Lobato, Jose Hernandez-Lobato, Amar Shah, and Ryan Adams. Predictive entropy search for multi-objective bayesian optimization. In *ICML*, pages 1492–1501, 2016.

[10] Xi Lin, Zhiyuan Yang, Xiaoyuan Zhang, and Qingfu Zhang. Pareto set learning for expensive multi-objective optimization. In *NeurIPS*, volume 35, pages 19231–19247, 2022.

[11] Art B Owen. Quasi-monte carlo sampling. *Monte Carlo Ray Tracing: Siggraph*, 1:69–88, 2003.

[12] Biswajit Paria, Kirthevasan Kandasamy, and Barnabás Póczos. A flexible framework for multi-objective bayesian optimization using random scalarizations. In *Uncertainty in Artificial Intelligence*, pages 766–776. PMLR, 2020.

[13] Carl Edward Rasmussen. Gaussian processes in machine learning. In *Summer school on machine learning*, pages 63–71. Springer, 2003.

[14] Niranjan Srinivas, Andreas Krause, Sham M Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting:no regret and experimental design. In *ICML*, volume 10, page 1015–1022, 2009.

[15] Sinnu Susan Thomas, Jacopo Palandri, Mohsen Lakehal-Ayat, Punarjay Chakravarty, Friedrich Wolf-Monheim, and Matthew B. Blaschko. Designing macpherson suspension architectures using bayesian optimization. In *Benelearn*, volume 2491, 2019.

[16] Sinnu Susan Thomas, Jacopo Palandri, Mohsen Lakehal-Ayat, Punarjay Chakravarty, Friedrich Wolf-Monheim, and Matthew B. Blaschko. Kinematics design of a macpherson suspension architecture based on bayesian optimization. *IEEE Transactions on Cybernetics*, 53(4):2261–2274, 2023.

[17] Sinnu Susan Thomas, Guillaume Lamine, Jacopo Palandri, Mohsen Lakehal-ayat, Punarjay Chakravarty, Friedrich Wolf-Monheim, and Matthew B. Blaschko. Mitigating bias in bayesian optimized data while designing macpherson suspension architecture. *IEEE Transactions on Artificial Intelligence*, 5(2):904–915, 2024.

[18] Kaifeng Yang, Michael Emmerich, André Deutz, and Thomas Bäck. Multi-objective bayesian global optimization using expected hypervolume improvement gradient. *Swarm and evolutionary computation*, 44:945–956, 2019.

[19] Richard Zhang and Daniel Golovin. Random hypervolume scalarizations for provable multi-objective black box optimization. In *ICML*, volume 119, pages 11096–11105, 2020.

[20] Eckart Zitzler and Lothar Thiele. Multiobjective evolutionary algorithms: a comparative case study and the strength pareto approach. *IEEE transactions on evolutionary computation*, 3(4): 257–271, 1999.