NEURAL INFORMATION PROCESSING SYSTEMS

Multi Objective Regionalized Bayesian Optimization Via Entropy Search

Thomas James, Sinnu Susan Thomas

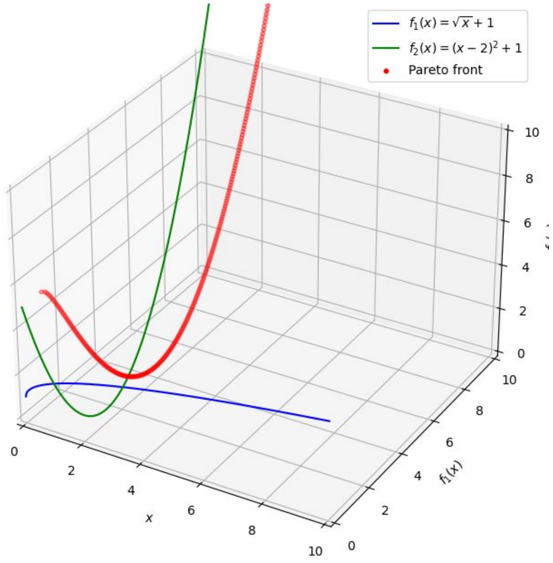
Digital University Kerala

India

Introduction

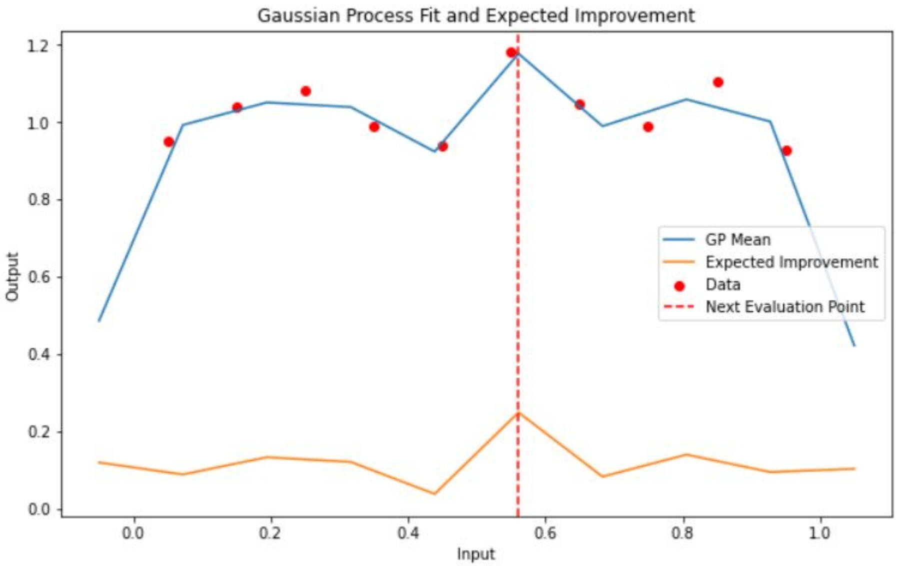
Multi Objective Optimization

- Multiple Conflicting Objectives
- Pareto Frontier



Bayesian Optimization

- Surrogate Model $f_\ell^{(j)} \sim \mathrm{GP}_\ell^{(j)}(\mu_\ell(y), k_\ell(y, y')),$
- Acquisition Function



Literature Review

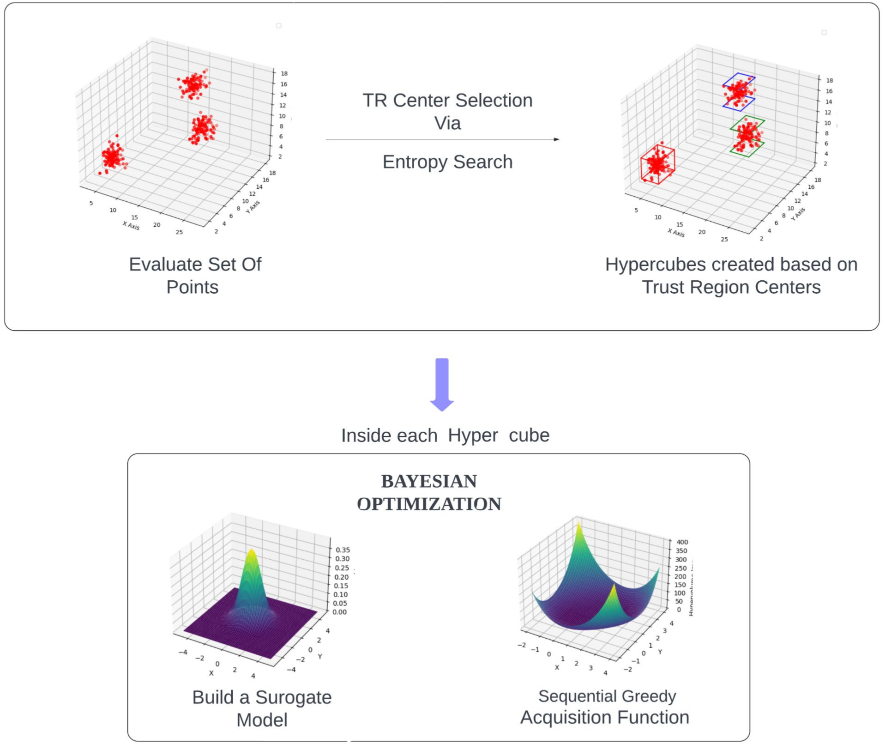
- Trusted Region Bayesian Optimization (TuRBO) [1] builds ‘n’ separate local Gaussian Process models inside each TR, but it is not suitable for optimization problems having more than one objective functions.

- Max Value Entropy search for Multi Objective Optimization (MESMO) [2] uses an output-space entropy based acquisition function to select the sequence of inputs for evaluation, but it is only employed to calculate the acquisition function.

- Multi Objective Regionalized Bayesian Optimization (MoRBO) [3] is specially designed for multiobjective problems by collaboratively optimizing the points inside ‘n’ trusted regions, but it has high tendency to select centers from crowded regions.

Methodology

Center Selection Mechanism Via Entropy Search



Information Gain is a metric used to quantify the reduction in uncertainty or entropy about a random variable when new information is obtained.

Mathematically Information Gain can be defined as :

$$
\begin{aligned}
I(\mathbf{x}) &= I(f_{\mathbf{x}}; \text{Pareto Set} \mid \text{Datapoints}) \\
&= I(f_{\mathbf{x}}; \{p^*\} \mid \mathcal{D}) \\
&= I(p^*; p^* - \{X, Y\} \mid \mathcal{D}) \\
&= H(p^* \mid \mathcal{D}) - H(p^* - \{X, Y\} \mid \mathcal{D})
\end{aligned}
$$

Methodology

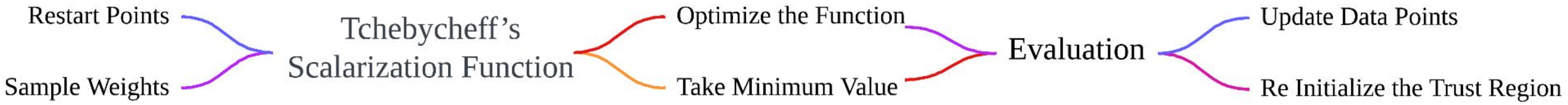
NEURAL INFORMATION PROCESSING SYSTEMS

Tchebycheff's Scalarization

- Maximizing the hypervolume is the same as maximizing the randomized single-objective scalarization
- Enhancing the exploration of non-convex Pareto regions improves the understanding and optimization of trade-offs in complex multi-objective problems.
- To find a solution y inside the feasible set X that minimizes the maximum deviation from a reference point .

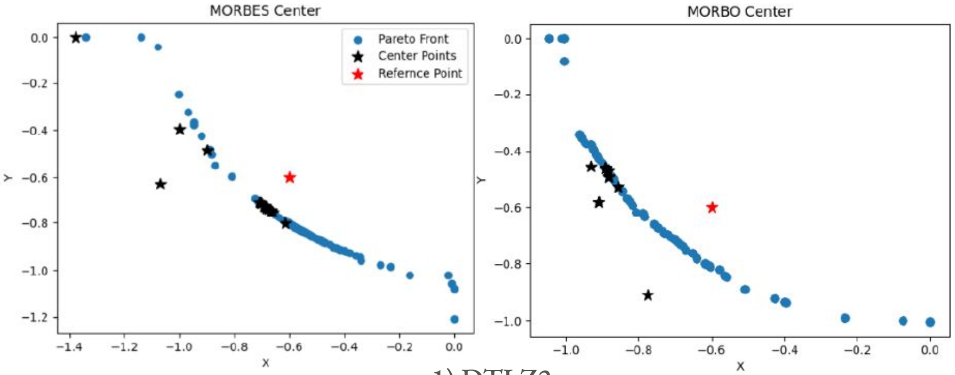
$$\min_{\mathbf{y}\in\mathcal{X}} ch_c(\mathbf{y} \mid \lambda) = \min_{\mathbf{y}\in\mathcal{X}} \max_{1\leq i\leq m} \{\lambda_i(f_i(\mathbf{y})) - \lambda_i(l_i^*) + \lambda_i(\epsilon)\}.$$

Re initialization Strategy

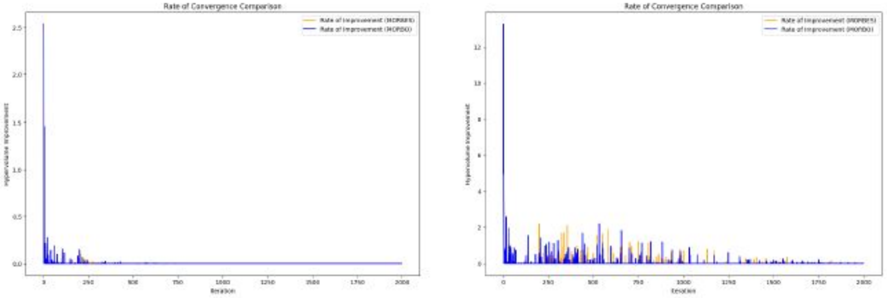


Results

Exploration Of Pareto Front



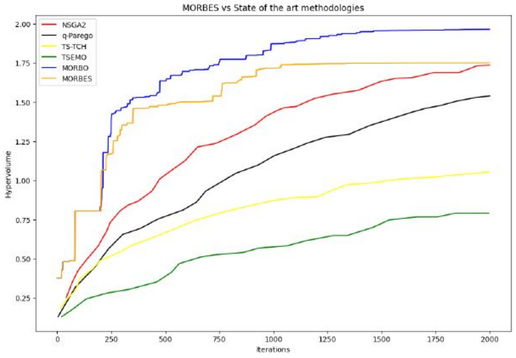
1) DTLZ2



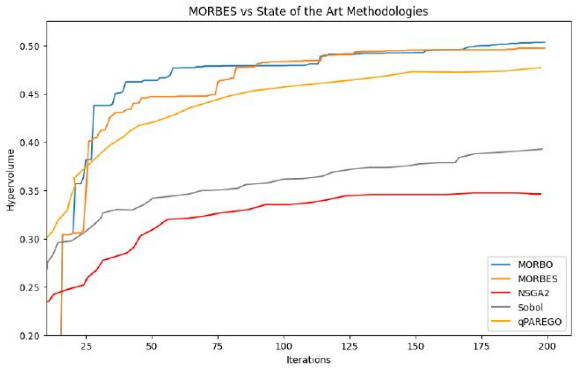
DTLZ2 DTLZ7

2) Rate Of Convergence

Convergence Property Of MORBES



3) Rover Problem



4)Welded Beam

References

[1] D. Eriksson, M. Pearce, J. Gardner, R. D. Turner, and M. Poloczek, “Scalable global optimization via local bayesian optimization,” in NeurIPS, vol. 32, 2019, pp. 5496–5507.

[2] S. Belakaria, A. Deshwal, and J. R. Doppa, “Max-value entropy search for multi-objective bayesian optimization,” NeurIPS, vol. 32, pp. 9025–9036, 2019.

[3] S. Daulton, D. Eriksson, M. Balandat, and E. Bakshy, “Multi-objective bayesian optimization over high dimensional search spaces,” in UAI, vol. 180, 2022, pp. 507–517.