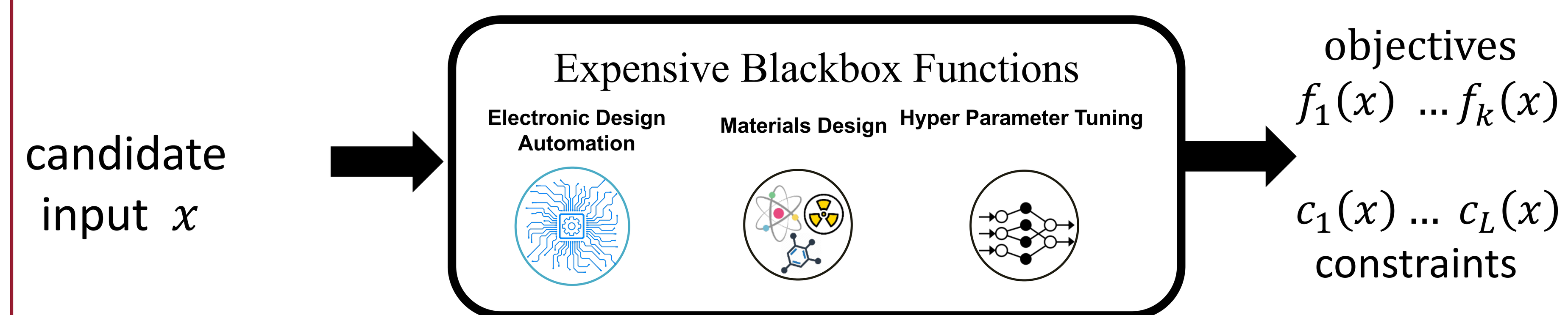
WASHINGTON STATE UNIVERSITY

Max-value Entropy Search for Multi-ObjectiveBayesian Optimization with Constraints

Syrine Belakaria, Aryan Deshwal, and Jana Doppa

Constrained Multi-Objective Bayesian Optimization

❑ Bayesian optimization (BO) is a framework to **maximize expensive black-box functions** using the following elements:



- **Statistical models** as a prior for the functions: Gaussian processes (GPs) can provide prediction $\mu(x)$ and uncertainty via variance $\sigma(x)$
- **Acquisition function** to score the utility of evaluating input x
- **Optimization procedure** to select the best input x for evaluation

Prior Work and Our Contributions

❑ Drawbacks of existing methods

- *Does not handle constraints*
- *Genetic algorithms assume that the functions are cheap to evaluate and require a very large number of evaluations*
- *Scalarization*: relies on random scalars that can be sub-optimal
- *Hypervolume improvement*: not scalable for high-dimensional input spaces and large number of objective functions
- *Uncertainty and information theory:* They either maximize information gain about the optimal Pareto set X^* and rely on approximating a very expensive and high-dimensional distribution or minimize the uncertainty over a finite set of points.

Our Approach:

❑ **MESMOC framework** selects the candidate input x for evaluation that maximizes the information gain about the optimal Pareto front $\boldsymbol{Y}^*$

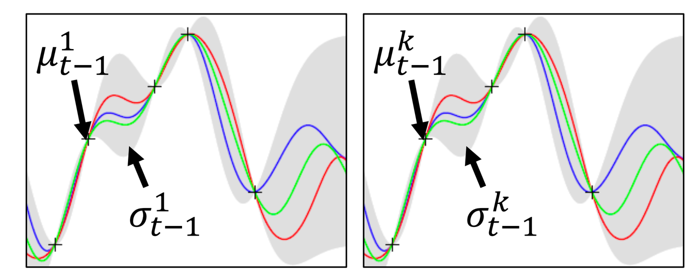
- Equivalent to expected reduction in entropy over the Pareto front $\boldsymbol{Y}^*$
- Relies on a computationally cheap and low-dimensional $m.k \ll m.d$ distribution, where k is the number of objectives

❑ **Key advantages of MESMOC**

- Robust to the number of samples for AF computation
- Scalable for high-dimensions via output space entropy search
- Tight approximation with closed-form expression
- Two real world applications to show the effectiveness of our algorithm

MESMOC Algorithm

Posteriors estimation



→ Sample a set of optimal pareto fronts Y^*using functions and constraints sampled from models

→ Define the acquisition function $\alpha_t(x) = I(\{x, y\}: Y^* \mid D_t)$

→ Evaluate the functions $f_1, \ldots, f_k, C_1, \ldots, C_L$ at $x_t = \arg\max \alpha_t(x)$ St. $\mu_{c_1}, \ldots, \mu_{c_L}$

→ Posteriors estimation

❑ Input space entropy-based acquisition function

$$\begin{aligned}\alpha(\mathbf{x}) &= I(\{\mathbf{x},\mathbf{y}\}, \boxed{\mathcal{X}^*} \mid D) \\ &= H(\mathcal{X}^* \mid D) - \mathbb{E}_y[H(\mathcal{X}^* \mid D \cup \{\mathbf{x},\mathbf{y}\})] \\ &= H(\mathbf{y} \mid D, \mathbf{x}) - \boxed{\mathbb{E}_{\mathcal{X}^*}[H(\mathbf{y} \mid D, \mathbf{x}, \mathcal{X}^*)]}\end{aligned}$$

Input dimension d

Requires approximation

❑ Output space entropy-based acquisition function

$$\begin{aligned}\alpha(\mathbf{x}) &= I(\{\mathbf{x},\mathbf{y}\}, \boxed{\mathcal{Y}^*} \mid D) \\ &= H(\mathcal{Y}^* \mid D) - \mathbb{E}_y[H(\mathcal{Y}^* \mid D \cup \{\mathbf{x},\mathbf{y}\})] \\ &= H(\mathbf{y} \mid D, \mathbf{x}) - \boxed{\mathbb{E}_{\mathcal{Y}^*}[H(\mathbf{y} \mid D, \mathbf{x}, \mathcal{Y}^*)]}\end{aligned}$$

Output dimension $k << d$

Sum of truncated Gaussians

❑ How to sample $\boldsymbol{Y}^*$?

- Sample functions from posterior GPs based on random Fourier features sampling procedure. We approximate each GP prior as $\tilde{f}_i = \phi(x)^T\theta$ and $\tilde{C}_i = \phi(x)^T\theta$ where $\theta \sim N(0, I)$.
- Solve a cheap constrained multi-objective optimization problem over the sampled functions and constraints $\tilde{f}_1 \ldots \tilde{f}_k, \tilde{C}_1 \ldots \tilde{C}_L$to and compute sample Pareto front

❑ For each function and constraint, select the maximum-value in the cheap Pareto front as an upper bound for the truncated Gaussian

❑ MESMOC's Acquisition Function

$$\alpha(\mathbf{x}) \simeq \frac{1}{S}\sum_{s=1}^{S}\left[\sum_{j=1}^{K}\frac{\gamma_s^{f_j}(\mathbf{x})\phi(\gamma_s^{f_j}(\mathbf{x}))}{2\Phi(\gamma_s^{f_j}(\mathbf{x}))} - \ln\Phi(\gamma_s^{f_j}(\mathbf{x})) + \sum_{j=1}^{L}\frac{\gamma_s^{c_j}(\mathbf{x})\phi(\gamma_s^{c_j}(\mathbf{x}))}{2\Phi(\gamma_s^{c_j}(\mathbf{x}))} - \ln\Phi(\gamma_s^{c_j}(\mathbf{x}))\right]$$

Experiments and Results

❑ Analog circuit optimization :

- Optimizing the design of a multi-output switched-capacitor voltage regulator via Cadence circuit simulator
- Each candidate circuit design is defined by 33 input variables (d=32)
- Our problem has a total of nine functions and nine constraints

❑ Electrified aviation power system design:

- Optimizing the design of electrified aviation power system of unmanned aerial vehicle (UAV) via a time-based static simulation
- Each candidate design is defined by 5 input variables (d=5)
- Our problem has a total of two functions and five constraints
- Only 9% of all designs satisfy all constraints

❑ Evaluation metrics

- *The Pareto hypervolume* : The volume between $\tilde{Y}_t$ and a reference point

❑ MESMOC vs. State-of-the-art

- MESMOC performs better than all baselines and converges faster
- For UAV experiment, despite the hardness of the problem, 50% of the designs selected by MESMOC satisfy all the constraints while for PESMOC, MOEAD, and NSGA-II this was 1.5%, 9.5%, and 7.5% respectively
- For circuit experiment: MESMOC can achieve the highest conversion efficiency of 88.81% (12.61% improvement when compared with PESMOC and 17.86% improvement when compared with NSGA-II) with similar output ripples.