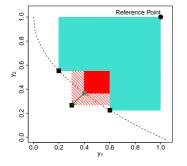
Optimization in Machine Learning

Bayesian Optimization Multicriteria Bayesian Optimization

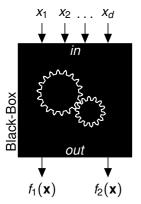




Learning goals

- Multicriteria Optimization
- Taxonomy
- ParEGO, SMS-EGO, EHI

MULTICRITERIA BAYESIAN OPTIMIZATION



$$f: \mathcal{S} \to \mathbb{R}^m$$

 $\min_{\mathbf{x} \in \mathcal{S}} f(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_m(\mathbf{x}))$

• A configuration **x** dominates (\prec) $\tilde{\mathbf{x}}$ if

$$\forall i \in \{1,...,m\}: f_i(\mathbf{x}) \leq f_i(\tilde{\mathbf{x}})$$

and $\exists j \in \{1,...,m\}: f_i(\mathbf{x}) < f_i(\tilde{\mathbf{x}})$

Set of non-dominated solutions:

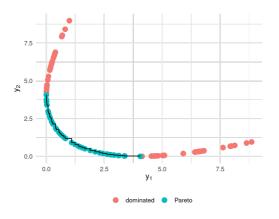
$$\mathcal{P} := \{ \mathbf{x} \in \mathcal{S} | \nexists \tilde{\mathbf{x}} \in \mathcal{S} : \tilde{\mathbf{x}} \prec \mathbf{x} \}$$

- Pareto set \mathcal{P} , Pareto front $\mathcal{F} = f(\mathcal{P})$
- Goal: Find \hat{P} of non-dominated points that estimates the true Pareto set P



MULTICRITERIA BAYESIAN OPTIMIZATION / 2

Example Pareto front:





MULTICRITERIA BAYESIAN OPTIMIZATION / 3

The most popular quality indicator is the hypervolume indicator (also called dominated hypervolume or $\mathcal{S}\text{-metric}$).

The hypervolume, HV, of an approximation of the Pareto front $\hat{\mathcal{F}} = f(\hat{\mathcal{P}})$ can be defined as the combined volume of the dominated hypercubes domHC_r of all solution points $\mathbf{x} \in \hat{\mathcal{P}}$ regarding a reference point \mathbf{r} , i.e.,

$$\mathsf{HV}_{\mathbf{r}}(\hat{\mathcal{P}}) := \mu \left(\bigcup_{\mathbf{x} \in \hat{\mathcal{P}}} \mathsf{dom}\mathsf{HC}_{\mathbf{r}}(\mathbf{x}) \right)$$

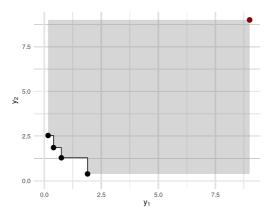
where $\boldsymbol{\mu}$ is the Lebesgue measure and the dominated hypercube is given as:

$$\mathsf{domHC}_{r}(\mathbf{x}) := \{ \mathbf{u} \in \mathbb{R}^{m} \mid f_{i}(\mathbf{x}) \leq \mathbf{u}_{i} \leq \mathbf{r}_{i} \ \forall i \in \{1, \dots, m\} \}$$



MULTICRITERIA BAYESIAN OPTIMIZATION / 4

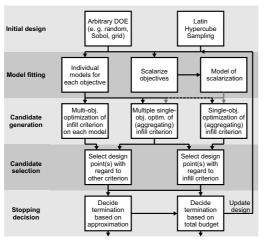
Hypervolume example:





Reference point r in red, estimated Pareto front $\hat{\mathcal{F}}$ in black, corresponding $\mathsf{HV}_r(\hat{\mathcal{P}})$ is given by the grey area

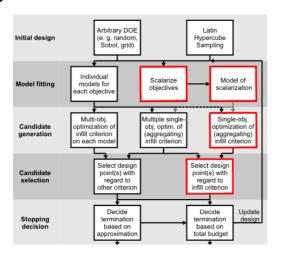
TAXONOMY



Horn, Wagner, Bischl et al. (2014).



PAREGO





 Scalarize standardized objectives using the augmented Tchebycheff norm

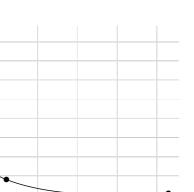
$$\max_{i \in \{1,\dots,m\}} w_i f_i(\mathbf{x}) + \rho \sum_{i=1}^m w_i f_i(\mathbf{x})$$

with weight vector \mathbf{w} drawn uniformly from the set of evenly distributed weight vectors \mathcal{W}

- 2 Fit SM on the scalarized objective function
- Proceed to use any standard single-objective acquisition function (EI, PI, LCB, ...)

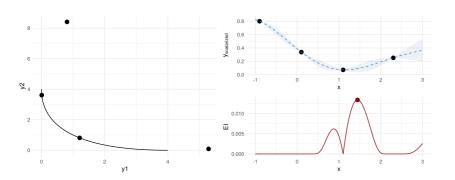


ParEGO Example, initial design and true Pareto front in black ...



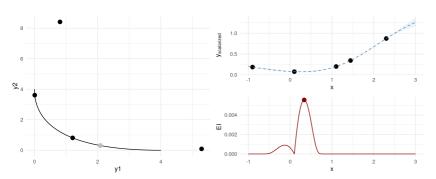


... standardize objectives, obtain scalarized objective via augmented Tchebycheff norm, fit SM and optimize EI ...



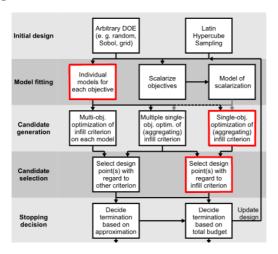


... note that the specific scalarization is different at each iteration!





The grey point visualizes the candidate we choose to evaluate in the previous iteration

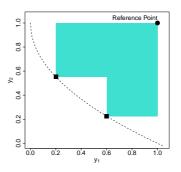




Individual models for each objective f_i

Single-objective optimization of aggregating acquisition function: Calculate contribution of the confidence bound of candidate to the current front approximation

- Calculate LCB for each objective
- Measure contribution with regard to the hypervolume improvement
- For ε -dominated (\prec_{ε}) solutions, a penalty is added

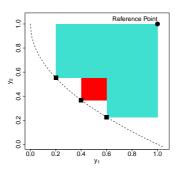




Individual models for each objective f_i

Single-objective optimization of aggregating acquisition function: Calculate contribution of the confidence bound of candidate to the current front approximation

- Calculate LCB for each objective
- Measure contribution with regard to the hypervolume improvement
- For ε -dominated (\prec_{ε}) solutions, a penalty is added

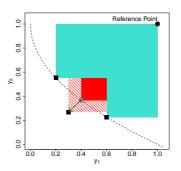




Individual models for each objective f_i

Single-objective optimization of aggregating acquisition function: Calculate contribution of the confidence bound of candidate to the current front approximation

- Calculate LCB for each objective
- Measure contribution with regard to the hypervolume improvement
- For ε -dominated (\prec_{ε}) solutions, a penalty is added





OUTLOOK

Many more options exist:

- Expected Hypervolume improvement
- Multi-Ego
- Entropy based: PESMO, MESMO
- ...

