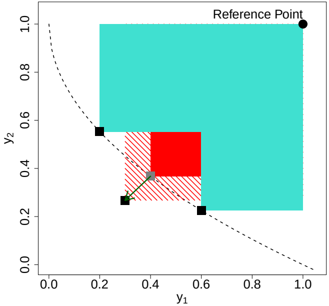
Optimization in Machine Learning

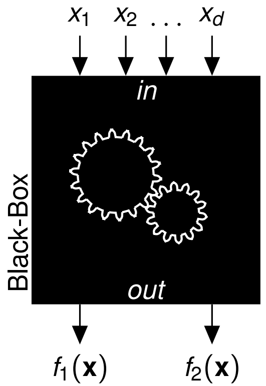
Bayesian Optimization
Multicriteria Bayesian Optimization



Learning goals

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

MULTICRITERIA BAYESIAN OPTIMIZATION



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

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

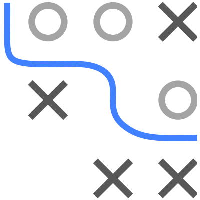
- A configuration $\mathbf{x}$ **dominates** ($\prec$) $\tilde{\mathbf{x}}$ if

$$\begin{aligned} \forall i \in \{1, ..., m\} : &\quad f_i(\mathbf{x}) \leq f_i(\tilde{\mathbf{x}}) \\ \text{and } \exists j \in \{1, ..., m\} : &\quad f_j(\mathbf{x}) < f_j(\tilde{\mathbf{x}}) \end{aligned}$$

- Set of non-dominated solutions:

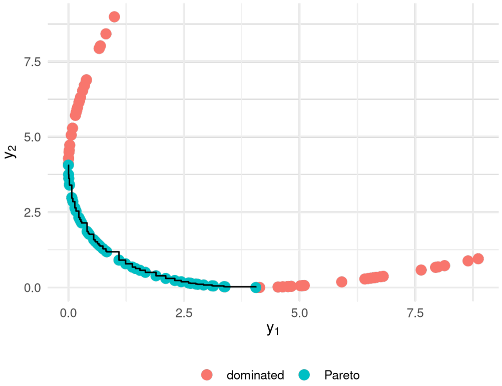
$$\mathcal{P} := \{\mathbf{x} \in \mathcal{S} | \nexists \tilde{\mathbf{x}} \in \mathcal{S} : \tilde{\mathbf{x}} \prec \mathbf{x}\}$$

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

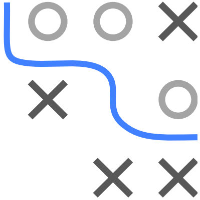


MULTICRITERIA BAYESIAN OPTIMIZATION / 2

Example Pareto front:



MULTICRITERIA BAYESIAN OPTIMIZATION / 3



The most popular quality indicator is the hypervolume indicator (also called dominated hypervolume or $\mathcal{S}$-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 $\text{domHC}_{\boldsymbol{r}}$ of all solution points $\mathbf{x} \in \hat{\mathcal{P}}$ regarding a reference point $\boldsymbol{r}$, i.e.,

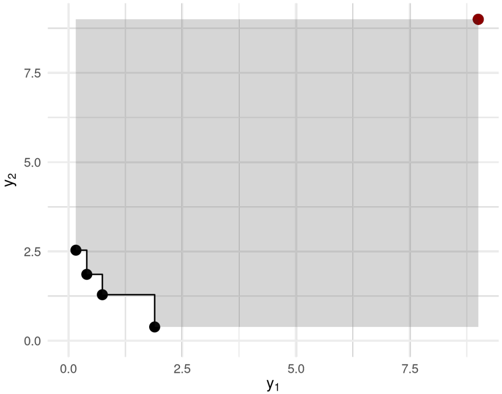
$$\text{HV}_{\boldsymbol{r}}(\hat{\mathcal{P}}) := \mu\left(\bigcup_{\mathbf{x}\in\hat{\mathcal{P}}} \text{domHC}_{\boldsymbol{r}}(\mathbf{x})\right)$$

where μ is the Lebesgue measure and the dominated hypercube is given as:

$$\text{domHC}_{\boldsymbol{r}}(\mathbf{x}) := \{\boldsymbol{u} \in \mathbb{R}^m \mid f_i(\mathbf{x}) \leq \boldsymbol{u}_i \leq \boldsymbol{r}_i \; \forall i \in \{1, \ldots, m\}\}$$

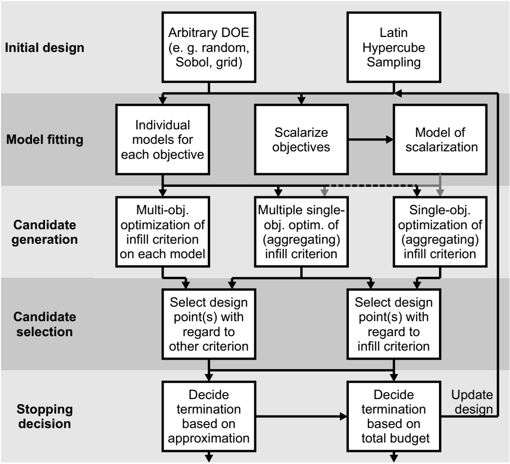
MULTICRITERIA BAYESIAN OPTIMIZATION / 4

Hypervolume example:



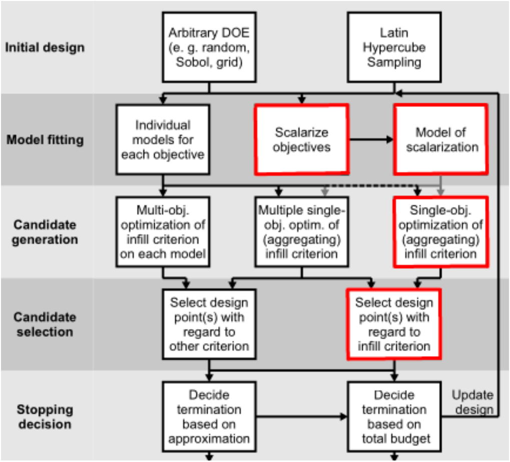
Reference point $\boldsymbol{r}$ in red, estimated Pareto front $\hat{\mathcal{F}}$ in black, corresponding $\mathrm{HV}_{\boldsymbol{r}}(\hat{\mathcal{P}})$ is given by the grey area

TAXONOMY



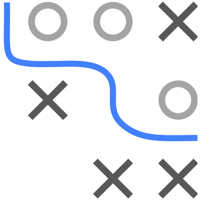
Horn, Wagner, Bischl et al. (2014).

PAREGO



Initial design
Model fitting
Candidate generation
Candidate selection
Stopping decision
Arbitrary DOE (e. g. random, Sobol, grid)
Latin Hypercube Sampling
Individual models for each objective
Scalarize objectives
Model of scalarization
Multi-obj. optimization of infill criterion on each model
Multiple single-obj. optim. of (aggregating) infill criterion
Single-obj. optimization of (aggregating) infill criterion
Select design point(s) with regard to other criterion
Select design point(s) with regard to infill criterion
Decide termination based on approximation
Decide termination based on total budget
Update design

PAREGO / 2



1. Scalarize standardized objectives using the augmented Tchebycheff norm

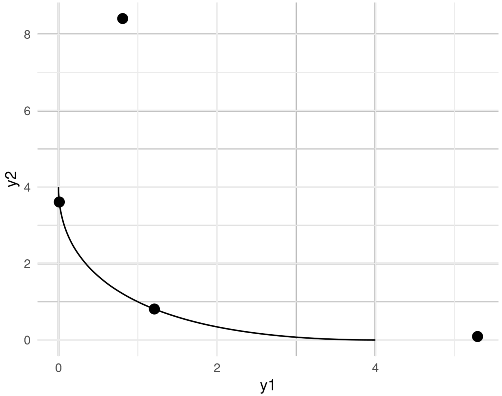
$$\max_{i \in \{1,\ldots,m\}} w_i f_i(\mathbf{x}) + \rho \sum_{i=1}^{m} w_i f_i(\mathbf{x})$$

   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
3. Proceed to use any standard single-objective acquisition function (EI, PI, LCB, ...)

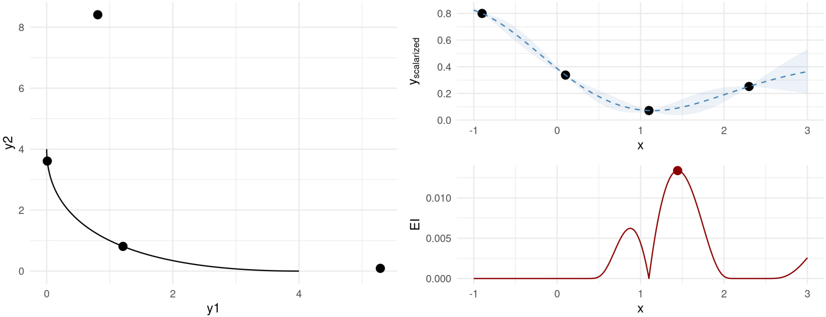
PAREGO / 3

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



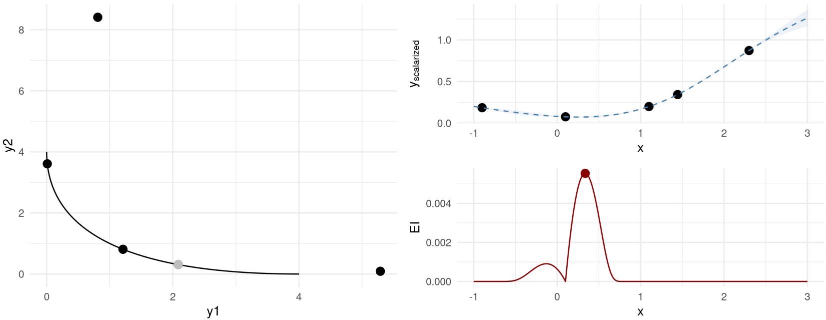
PAREGO / 4

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



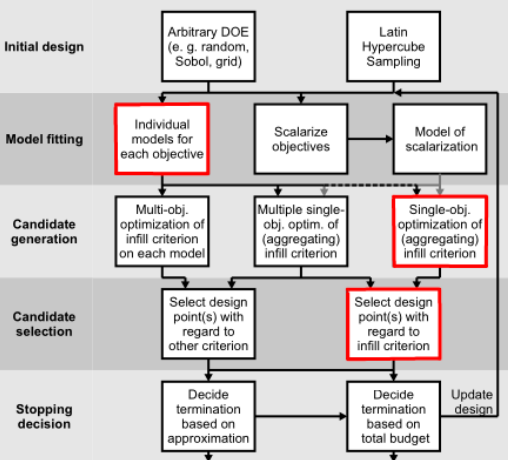
PAREGO / 5

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



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

SMS-EGO



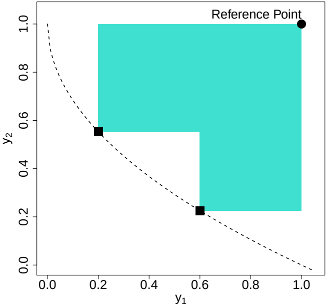
Initial design
Arbitrary DOE (e. g. random, Sobol, grid)
Latin Hypercube Sampling
Model fitting
Individual models for each objective
Scalarize objectives
Model of scalarization
Candidate generation
Multi-obj. optimization of infill criterion on each model
Multiple single-obj. optim. of (aggregating) infill criterion
Single-obj. optimization of (aggregating) infill criterion
Candidate selection
Select design point(s) with regard to other criterion
Select design point(s) with regard to infill criterion
Stopping decision
Decide termination based on approximation
Decide termination based on total budget
Update design

SMS-EGO

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

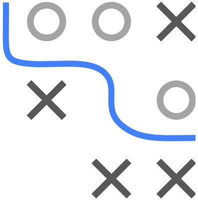
- Calculate LCB for each objective
- Measure contribution with regard to the hypervolume improvement
- For ε-dominated ($\prec_\varepsilon$) solutions, a penalty is added



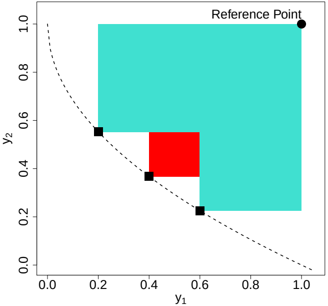
SMS-EGO

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 ($\prec_\varepsilon$) solutions, a penalty is added

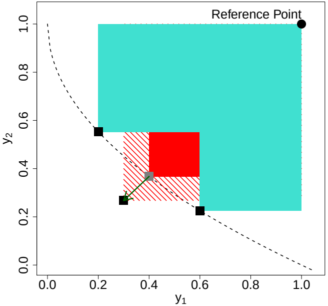


SMS-EGO

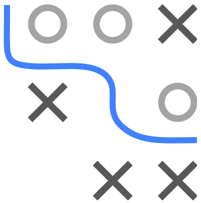
Individual models for each objective f_i

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 ($\prec_\varepsilon$) solutions, a penalty is added



OUTLOOK



Many more options exist:

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