Bayesian multi-objective optimization with noisy evaluations
Bruno Barracosa, Julien Bect, Héloïse Dutrieux Baraffe, Juliette Morin,
Josselin Fournel, Emmanuel Vazquez

To cite this version:
Bruno Barracosa, Julien Bect, Héloïse Dutrieux Baraffe, Juliette Morin, Josselin Fournel, et al.. Bayesian multi-objective optimization with noisy evaluations. MASCOT PhD student 2020 Meeting, Sep 2020, Grenoble, France. hal-03022267
BAYESIAN MULTI-OBJECTIVE OPTIMIZATION WITH NOISY EVALUATIONS

1. CONTEXT

- Multi-objective optimization of the parameters of a planning strategy for the multi-year planning of the electricity distribution grid [DUT15]
- A stochastic black box provides noisy evaluation results of the objective functions $f_1, \ldots, f_d$ defined on a discrete search domain $X \subset \mathbb{R}^d$
- Previous $n$ evaluations at $x = (x_1, \ldots, x_n)$ assumed:
  - $Z_{i,j} = f_1(x_i) + \xi_{i,j}, \ldots, Z_{i,q} = f_q(x_i) + \xi_{i,q}$
  - where $\xi_{i,j}$ are zero-mean random variables

2. OBJECTIVE

Estimate the Pareto-optimal solutions (or Pareto set $\Gamma$) of the problem:

$$x^* = \arg\max_{x \in X} f(x)$$

Defined as:

$$\Gamma = \{ x \in X : \exists x' \in X \text{ such that } f(x') < f(x) \}$$

where $\prec$ stands for the Pareto domination rule:

$$y = (y_1, \ldots, y_q) \prec y' = (y'_1, \ldots, y'_q) \iff \forall i \leq q, y_i \leq y'_i \quad \text{and} \quad \exists j \leq q, y_j < y'_j$$

Illustration of the Pareto domination rule:

- $y_1, y_2, y_3$ are non-dominated points
- $y_4$ is only dominated by $y_1$
- $y_5$ is only dominated by $y_2$
- $y_6$ is dominated by all other points

3. BAYESIAN OPTIMIZATION

- Define a probabilistic model for each $f$ conditional on previous observations
- Use a sampling criterion to select new evaluation points

Provides estimate of $f$ and a measure of uncertainty of the estimation

**PARETO-OPTIMAL ESTIMATES**
Built from the estimates of $f$

**PROPOSED APPROACH**

- Replace the multi-objective problem by the minimization of a single augmented Tchebycheff function [KNO06]:
  $$f(x) = \max_{x \in X} \omega f_j(x) + \rho \sum_{j=1}^{d} \omega_j f_j(x), \quad \sum_{j=1}^{d} \omega_j = 1, \rho > 0$$
- At each iteration, generate random weights $\omega_j$ and apply this function to the $n$ previous observations:
  $$Z_{i,j} = \max_{x \in X} \omega f_j(x) + \rho \sum_{i=1}^{m} \omega_j Z_{i,j}, \quad i = 1, \ldots, n$$
- Assume a homoscedastic Gaussian noise model and fit to $Z = (Z_1, \ldots, Z_n)$ a Gaussian Process model $\xi$ with parameters estimated by maximum likelihood
- We use the Knowledge Gradient (KG) criterion [FRA09] to select new point $x_{n+1}$ based on previous observations. The idea is to identify a point that is expected to reduce the minimum of the posterior mean of $\xi$:
  $$x_{n+1} = \arg\max_{x \in X} KG(x)$$

4. NUMERICAL EXPERIMENTS

- Compare the proposed approach (PA) with random selection of points in a bi-dimensional bi-objective problem
- Compare use of batches of 1, 4 or 10 evaluations

5. OPEN QUESTIONS

- ‘Ideal’ batch of evaluations?
- Performance comparison of the proposed approach to other methods in the literature?

REFERENCES


Contacts:
Bruno Tebbal-Barracosa (EDF R&D, L2S, CentraleSupélec), Julien Bect (L2S, CentraleSupélec), Hélène Dutrieux Baraffé (EDF R&D), Juliette Morin (EDF R&D), Joasselin Fournel (EDF R&D), Emmanuel Vazquez (L2S, CentraleSupélec)
*bruno.tebbal-barracosa@edf.fr