Sampling criteria for constrained Bayesian optimization under uncertainty

Julien Pelamatti [1] R. El-Amri [2] R. Le Riche [3], C. Helbert [4], C. Blanchet-Scalliet [4]

[1] EDF R&D

[2] IRT Saint-Exupéry

[3] CNRS at LIMOS

[4] Ecole Centrale de Lyon

SIAM UQ, Atlanta, 12-15 April 2022

Overview

This presentation summarizes the research done by :

- Reda El-Amri [El-Amri et al. 2021]
- Julien Pelamatti [Pelamatti et al. 2022]

during two consequent post-doctoral projets in collaboration with :

- Christophette Blanchet-Scalliet
- Céline Helbert
- Rodolphe Le Riche

This work was partly supported by the OQUAIDO research Chair in Applied Mathematics.

Context

Design optimization of complex engineering systems:

- Computationally costly system performance simulations
  - Computational Fluid-Dynamics
  - Finite Element Models
  - Internal optimizations
- Presence of uncertain parameters
  - Manufacturing errors
  - Meteorological conditions
  - Physical model limitations

Necessity to determine a robust optimal design in terms of objective function value and feasibility with a finite amount of simulations

Problem formulation (1)

We consider 2 types of variables [Valdebenito et al. 2010]

$$\mathbf{x} \in \mathcal{S}_x \subset \mathbb{R}^d \qquad \text{Design variables}$$

$$\mathbf{u} \in \mathcal{S}_u \subset \mathbb{R}^m \qquad \text{Uncertain variables}$$

$\mathbf{u} \sim \mathbf{U}$ with known density function $\rho_{\mathbf{U}}$.

The system performance can be modeled as an objective function:

$$f(\mathbf{x}, \mathbf{U})$$

subject to multiple constraints:

$$g_i(\mathbf{x}, \mathbf{U}) \leq 0, \; i = 1, \ldots, n_g$$

Problem formulation (2)

In the presence of uncertain variables, an optimal robust and feasible design can be defined as:

$$\begin{aligned} \min_{\mathbf{x}\in\mathcal{S}_x} \quad & \mathbb{E}_{\mathbf{U}}[f(\mathbf{x},\mathbf{U})] \\ \text{s.t.} \quad & \mathbb{P}(g_i(\mathbf{x},\mathbf{U}) \leq 0, i = 1,\ldots,n_g) \geq 1-\alpha, \qquad 0 < \alpha < 1 \end{aligned}$$

Alternatively:

$$\mathbb{P}(g_i(\mathbf{x},\mathbf{U}) \leq 0, i = 1,\ldots,n_g) \geq 1-\alpha \quad \longleftrightarrow \quad c(\mathbf{x}) \leq 0$$

Bayesian optimization

Necessity to solve the optimization problem with a limited amount of costly simulations $\rightarrow$ Bayesian optimization

- Surrogate Model Based Design Optimization (SMBDO) relying on Gaussian Process (GP) modeling
- GPs are used to define an **acquisition function**, characterizing how promising an unmapped location in the search space is
- At each iteration, the costly system performance is only evaluated at **one or few** promising locations

A 2-step Bayesian optimization algorithm for robust optimization under uncertainties is therefore proposed

Robust Bayesian optimization (1)

Generic approach : two-step acquisition function

Let $P(\mathbf{x})$ be a **random** progress measure at $\mathbf{x}$ computed w.r.t. the GPs of f and $g_1, \dots, g_{n_g}$

- Step 1: Define a desirable $\mathbf{x}_{target}$ by maximizing the expectation of the progress measure

$$\mathbf{x}_{target} = \arg\max_{\mathbf{x}} \mathbb{E}(P^{(t)}(\mathbf{x}))$$

- Step 2: Define the complete next iterate by minimizing the one step-ahead variance of the progress measure at $\mathbf{x}_{target}$:

$$(\mathbf{x}_{t+1}, \mathbf{u}_{t+1}) = \arg\min_{\mathbf{x},\mathbf{u}} Var(P^{(t+1)}(\mathbf{x}_{target}))$$

What points should be added to the training set in order to minimize the progress variance at $\mathbf{x}_{target}$?

Robust Bayesian optimization (2)

f and g modeled in the **joint search space** $\mathbb{S}_x \times \mathbb{S}_u$ as [Janusevskis et al. 2013] :

$$F(\mathbf{x}, \mathbf{u}) \sim \mathcal{GP}(m_F(\mathbf{x}, \mathbf{u}), k_F(\mathbf{x}, \mathbf{u}; \mathbf{x}', \mathbf{u}'))$$

$$G_i(\mathbf{x}, \mathbf{u}) \sim \mathcal{GP}(m_{G_i}(\mathbf{x}, \mathbf{u}), k_{G_i}(\mathbf{x}, \mathbf{u}; \mathbf{x}', \mathbf{u}')), \quad \text{for } i = 1, \dots, n_g$$

$F^{(t)}$, $G_i^{(t)}$ represent the GPs conditioned on t observations. e.g., $\{f(\mathbf{x}_1, \mathbf{u}_1), \dots, f(\mathbf{x}_t, \mathbf{u}_t)\}$

Projected process:

$$Z(\mathbf{x}) = \mathbb{E}_U[F(\mathbf{x}, \mathbf{u})] \sim \mathcal{GP}(m_Z(\mathbf{x}), k_Z(\mathbf{x}, \mathbf{x}'))$$

Robust Bayesian optimization (3)

We choose the Feasible Improvement (FI) as the measure of progress

Feasible improvement (projected space)

$$P(\mathbf{x}) = FI^{(t)}(\mathbf{x}) = I^{(t)}(\mathbf{x})\mathbb{1}_{\{C^{(t)}(\mathbf{x})\leq 0\}}$$
$$= (z_{min}^{feas} - Z^{(t)}(\mathbf{x}))^{+}\mathbb{1}_{\{C^{(t)}(\mathbf{x})\leq 0\}}$$

where

$$C^{(t)}(\mathbf{x}) = 1 - \alpha - \mathbb{E}_{\mathbf{U}}[\mathbb{1}_{\cap_{i=1}^{n_g}\{G_i^{(t)}(\mathbf{x},\mathbf{U})\leq 0, i=1,\dots,n_g\}}]$$

$z_{\text{min}}^{\text{feas}}$ is not known and must be estimated through an optimization routine

Robust Bayesian optimization (4)

The most promising set of design variables is computed as:

Expected Feasible improvement (projected space)

$$EFI^{(t)}(\mathbf{x}) = \mathbb{E}[(z_{min}^{feas} - Z^{(t)}(\mathbf{x}))^{+}\mathbb{1}_{\{C^{(t)}(\mathbf{x})\leq 0\}}]$$
$$= EI^{(t)}(\mathbf{x})\mathbb{P}(C^{(t)}(\mathbf{x}) \leq 0)$$

$$\mathbf{x}_{target} = \arg\max_{\mathbf{x}\in\mathbb{S}_x} EFI^{(t)}(\mathbf{x})$$

$C^{(t)}(\mathbf{x})$ is not Gaussian $\rightarrow$ no closed form expression for $\mathbb{P}(C^{(t)}(\mathbf{x}) \leq 0)$

Robust Bayesian optimization (5)

$\mathbf{x}_{t+1}, \mathbf{u}_{t+1}$ selected in order to minimize the one-steap-ahead variance of the FI at $\mathbf{x}_{target}$ we would obtain if we added $\{\mathbf{x}_{t+1}, \mathbf{u}_{t+1}\}$ to the training set:

One-step-ahead FI variance

$$Var\left(I^{(t+1)}(\mathbf{x}_{target})\mathbb{1}_{\{C^{(t+1)}(\mathbf{x})\leq 0\}}\right)$$

$$\text{where } F^{(t+1)} \mid \{f(\mathbf{x}_1, \mathbf{u}_1), \dots, f(\mathbf{x}_t, \mathbf{u}_t)\} \cup f(\mathbf{x}_{t+1}, \mathbf{u}_{t+1})$$

$$G_i^{(t+1)} \mid \{g_i(\mathbf{x}_1, \mathbf{u}_1), \dots, g_i(\mathbf{x}_t, \mathbf{u}_t)\} \cup g_i(\mathbf{x}_{t+1}, \mathbf{u}_{t+1})$$

Which new evaluation would provide the largest amount of information on the robust optimal design?

Robust Bayesian optimization (6)

We consider $\mathbf{x}_{t+1} = \mathbf{x}_{target}$ and a FI variance **proxy**:

$$\mathbf{u}_{t+1} = \arg\min_{\mathbf{u}_{t+1}} Var\left(I^{(t+1)}(\mathbf{x}_{target})\right) \int_{\mathbb{R}^m} Var\left(\mathbb{1}_{\cap_{i=1}^{n_g}\{G_i^{(t+1)}(\mathbf{x},\mathbf{U})\leq 0, i=1,\ldots,n_g\}}\right) \rho_U d\mathbf{u}$$

- Product of variances instead of variance of product

The conditioning values $f(\mathbf{x}_{t+1}, \mathbf{u}_{t+1})$ and $g_i(\mathbf{x}_{t+1}, \mathbf{u}_{t+1})$ are **unknown**

The FI variance at t+1 is integrated w.r.t. the distributions of $f(\mathbf{x}_{t+1}, \mathbf{u}_{t+1})$ and $g_i(\mathbf{x}_{t+1}, \mathbf{u}_{t+1})$

Results, part 1

3 compared algorithms

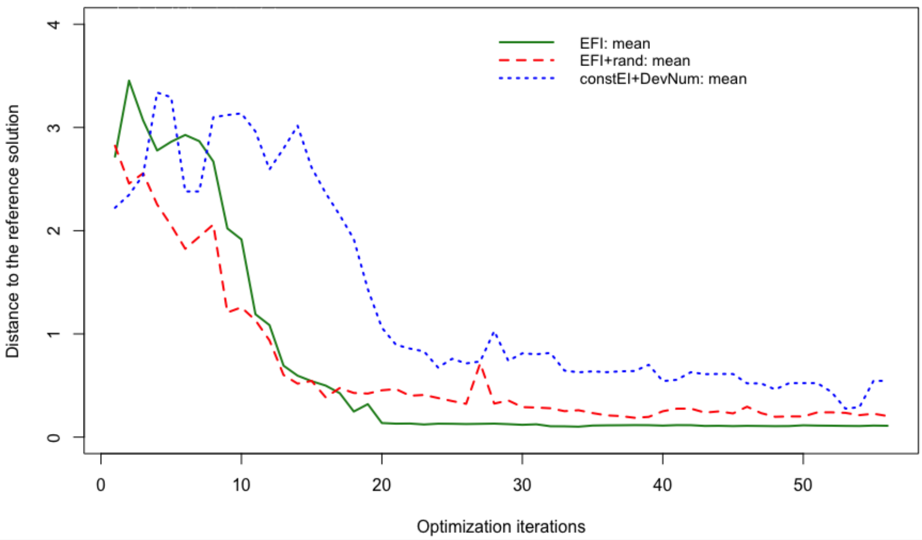
- EFI : presented RBO algorithm
- EFI + random : $\mathbf{x}_{t+1}$ determined through EFI maximisation, $\mathbf{u}_{t+1}$ sampled according to its distribution
- EFI + random : $\mathbf{x}_{t+1}$ determined through constrained EI maximisation, $\mathbf{u}_{t+1}$ is computed by minimizing the deviation number:
  $D(\mathbf{u}) = \min_i \frac{|m_{G_i^{(t)}}(\mathbf{x}_{t+1},\mathbf{u})|}{\sigma_{G_i}^{(t)}(\mathbf{x}_{t+1},\mathbf{u})}$,

4-d test-case

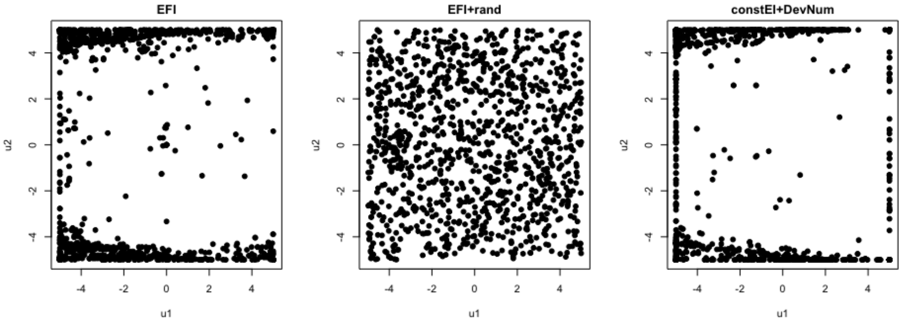
$$
\begin{aligned}
\min_{\mathbf{x}} E_U[f(\mathbf{x},\mathbf{U})] \quad & s.t. \quad P(g(\mathbf{x},\mathbf{U}) \leq 0) \geq 0.95 \\
f(x_1,x_2,u_1,u_2) \quad & = \quad 5(x_1^2+x_2^2) - (u_1^2+u_2^2) + x_1(u_2-u_1+5) \\
& \quad\quad +x_2(u_1-u_2+3) \\
g(x_1,x_2,u_1,u_2) \quad & = \quad -x_1^2 + 5x_2 - u_1 + u_2^2 - 1 \\
\text{with} \quad \mathbf{x} \quad & \in [-5,5]^2 \text{ and } \mathbf{U} \sim \mathcal{U}([-5,5]^2)
\end{aligned}
$$

Results over 20 repetitions

4-d test-case


EFI: mean
EFI+rand: mean
constEI+DevNum: mean
Distance to the reference solution
0
1
2
3
4
0
10
20
30
40
50
Optimization iterations

4-d test-case



Algorithm variants (1)

- In general, the constraints are considered independent and are modeled separately : $G_i(\mathbf{x}, \mathbf{u})$
- The objective function and all of the constraints are computed at every iteration and at the same location $\mathbf{x}_{t+1}, \mathbf{u}_{t+1}$

Some additional notes can be made:

- Constraints often depend on correlated physical phenomena, e.g.,
  - Displacement & Von Mises stress
  - Aerodynamic and structural responses of a wing
- It is often possible to separately simulate a single given constraint
- Constraints often drive the computational cost of the optimization

Algorithm variants (2)

The proposed extensions stem from modeling the n_g constraints as a multi-output GP [Alvarez et al. 2011], [Evgeniou et al. 2006]

A single GP simultaneously models all of the constraints characterizing the problem

This allows for:

- Possibly more accurate GP modeling of constraints
- Possibility of selecting different values of $\mathbf{u}$ for f and for each g_i
- Possibility of selecting the constraints which provide the most information w.r.t. the robust optimum location

$\rightarrow$ **Overall computational cost reduction**

Multi-output GP modeling (1)

The **output-as-input** approach is considered:

$$G(\mathbf{x}, \mathbf{u}, p) \sim \mathcal{GP}(m_g(\mathbf{x}, \mathbf{u}, p), k_g(\mathbf{x}, \mathbf{u}, p, \mathbf{x}', \mathbf{u}', p'))$$

$$k_g(\mathbf{x}, \mathbf{u}, p, \mathbf{x}', \mathbf{u}', p') = k_{g_c}(\mathbf{x}, \mathbf{u}, \mathbf{x}', \mathbf{u}') \times k_{g_d}(p, p')$$

- $p \in \{1, \dots, n_g\}$ indicates the selected output
- $k_{g_d}(p, p')$ is a kernel characterizing the covariance between discrete levels [Roustant et al. 2018]
- The design space is augmented: the training sets for all constraints are considered simultaneously

Multi-output GP modeling (2)

$$\mathbf{G}(\mathbf{x},\mathbf{u}) = \{G_1(\mathbf{x},\mathbf{u}),\dots,G_{n_g}(\mathbf{x},\mathbf{u})\}$$
$$\mathbf{G}(\mathbf{x},\mathbf{u}) \sim \mathcal{GP}(\mathbf{m}_g(\mathbf{x},\mathbf{u}),\mathbf{K}_g(\mathbf{x},\mathbf{u},\mathbf{x}',\mathbf{u}'))$$

$$C^{(t)}(\mathbf{x}) = 1-\alpha-\mathbb{E}_{\mathbf{U}}[\mathbb{1}_{\mathbf{G}^{(t)}(\mathbf{x},\mathbf{U})\leq\mathbf{0}}]$$

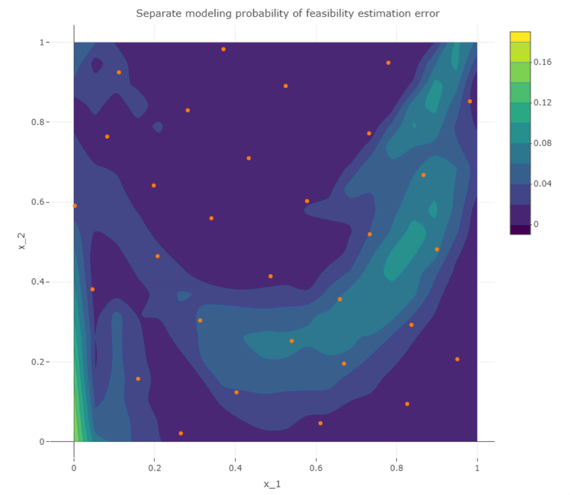
where:

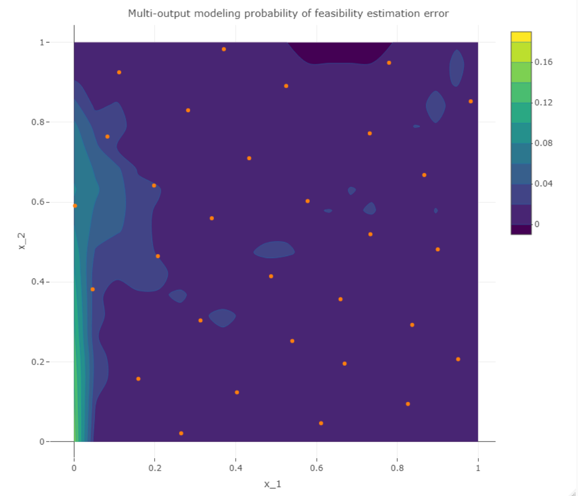
- $\mathbf{m}_g(\mathbf{x},\mathbf{u})$ is an $n_g \times 1$ vector
- $\mathbf{K}_g$ is a $n_g \times n_g$ matrix

The prediction of $\mathbf{G}(\mathbf{x},\mathbf{u})$ at un unmapped location follows a multivariate normal distribution

Multi-output GP modeling (3)

Example





RBO Variant (1)

It can be shown that the FI equations remain valid if multi-output modeling is considered

A first straightforward extension is obtained as:

$$\mathbf{x}_{target} = \arg\max_{\mathbf{x}\in\mathbb{S}_x} EFI(\mathbf{x})$$

$$\mathbf{u}_{t+1} = \arg\min_{\mathbf{u}_{t+1}} Var\left(I^{(t+1)}(\mathbf{x}_{target})\right)\int_{\mathbb{R}^m} Var\left(\mathbb{1}_{\{\mathbf{G}(\mathbf{x},\mathbf{u})\leq\mathbf{0}\}}\right)\rho_U d\mathbf{u}$$

NB: the objective function is considered independent and uncorrelated from the constraints

RBO Variant (2)

Notice: the feasible improvement variance **proxy** at t+1 can be decomposed into two separate terms:

$$
\begin{aligned}
S(\mathbf{x}_{target}, \mathbf{u}) &= Var\left(I^{(t+1)}(\mathbf{x}_{target})\right) \int_{\mathbb{R}^m} Var\left(\mathbb{1}_{\{\mathbf{G}(\mathbf{x},\mathbf{u})\leq \mathbf{0}\}}\right) \rho_U d\mathbf{u} \\
&= S_{T1}(\mathbf{x}_{target}, \mathbf{u}) S_{T2}(\mathbf{x}_{target}, \mathbf{u})
\end{aligned}
$$

S_{T1} only relates to the objective function GP, while S_{T2} only relates to the constraints GP

What if we refine objective and constraint functions with different values of $\mathbf{u}$?

RBO Variant (3)

Independent refining of each constraint

The search space of the one-step-ahead FI variance can be augmented and decomposed:

One-step-ahead FI variance proxy variant

$$S_{T1}(\mathbf{x}, \mathbf{u}_f) = Var((z_{min}^{feas} - Z(\mathbf{x}_{target}))^+ | \mathbf{f}^t, f(\mathbf{x}, \mathbf{u}_f))$$

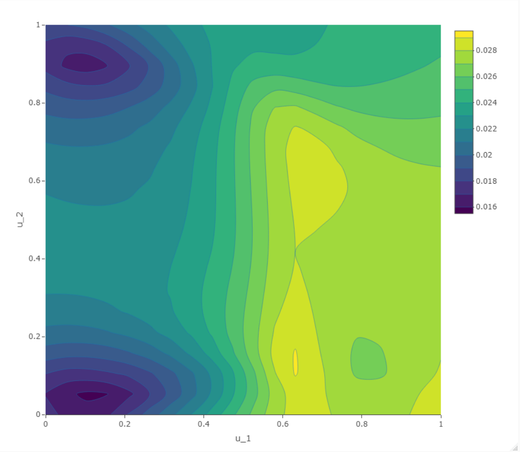
$$S_{T2}(\mathbf{x}, \mathbf{u}_1, \dots, \mathbf{u}_{n_g}) = \int_{\mathbb{R}^m} Var\left(\mathbb{1}_{\{\mathbf{G}(\mathbf{x}_{target}, \mathbf{u})^{t+1} \leq \mathbf{0}\} | \mathbf{g}_1^t, \dots, \mathbf{g}_{n_g}^t, g_1(\mathbf{x}, \mathbf{u}_1), \dots, g_{n_g}(\mathbf{x}, \mathbf{u}_{n_g})}\right) \rho_U(\mathbf{u}) d\mathbf{u}$$

- $\{\mathbf{x}_{target}, \mathbf{u}_f\}$ represents the objective function infilled sample
- $\{\mathbf{x}_{target}, \mathbf{u}_i\}$ with $i = 1, \dots, n_g$ represents the sample infilled associated to the i-th constraint

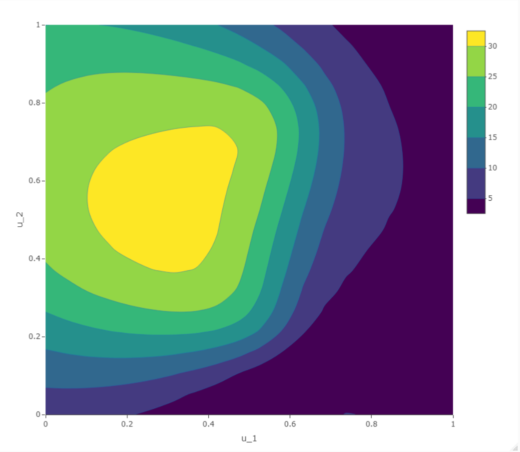
RBO Variant (4)

Example

Feasibility variance at t+1



Improvement variance at t+1



RBO Variant (5)

What if we do not evaluate all of the constraints at every iteration?

Selection of the most informative constraint

One-step-ahead FI variance proxy variant

$$S_{T1}(\mathbf{x}, \mathbf{u}_f) = Var((z_{min}^{feas} - Z(\mathbf{x}_{target}))^+ | \mathbf{f}^t, f(\mathbf{x}, \mathbf{u}_f))$$

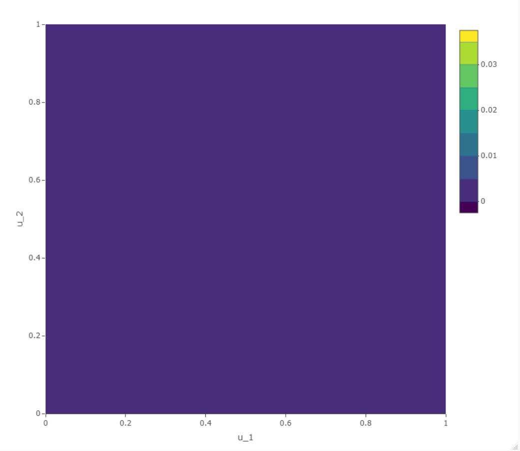
$$S_{T2}(\mathbf{x}, \mathbf{u}_g, p) = \int_{\mathbb{R}^m} Var\left(\mathbb{1}_{\{\mathbf{G}(\mathbf{x}_{target}, \mathbf{u})^{t+1} \leq \mathbf{0}\} | \mathbf{g}_1^t, \ldots, \mathbf{g}_{n_g}^t, g_p(\mathbf{x}, \mathbf{u}_g)}\right) \rho_U(\mathbf{u}) d\mathbf{u}$$

- $\{\mathbf{x}_{target}, \mathbf{u}_f\}$ represents the objective function infilled sample
- $\{\mathbf{x}_{target}, \mathbf{u}_g, p\}$ represents the sample infilled to the constraint associated to the indicator p

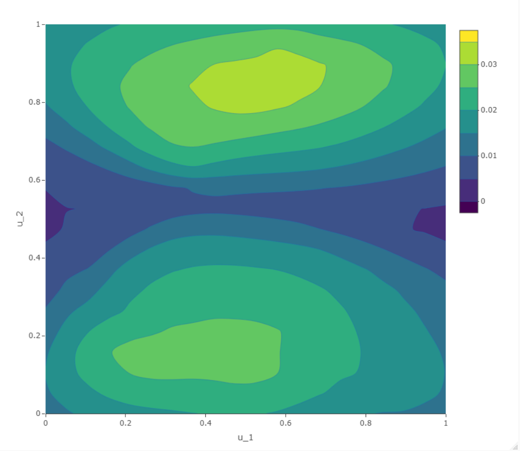
RBO Variant (6)

Example

Feasibility variance reduction g_1



Feasibility variance reduction g_2



RBO Variant (7)

One-step-ahead FI variance proxy variant

$$\mathbf{u}_f^{t+1} = \arg\min_{\mathbf{u}_f} S_{T1}(\mathbf{x}_{target}, \mathbf{u}_1)$$

$$\mathbf{u}_g^{t+1}, p^{t+1} = \arg\min_{\mathbf{u}_g, p} S_{T2}(\mathbf{x}_{target}, \mathbf{u}_g, p)$$

- Different 'random scenarios' are considered for the objective and the constraint function
- At each iteration, only the single constraint which yields the largest FI variance reduction is evaluated
- Reduction of costly function evaluations
- Possibility of separately optimizing S_{T1} and S_{T2}

Results, part 2

- 4 compared methods:
  - **REF**: Reference method, separate Modeling and common value of U
  - **SMCS**: Separate Modeling and Constraints selection
  - **MMCU**: Multi-output Modeling of constraints and Common value of U
  - **MMCS**: Multi-output Modeling and Constraint Selection

4-d test-case

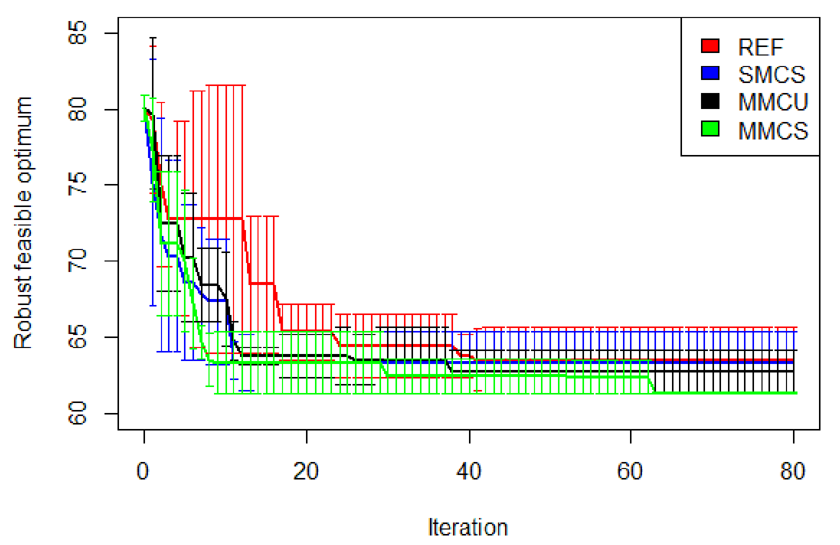
$$\min_{\mathbf{x}} E_U[f(\mathbf{x},\mathbf{U})] \quad s.t. \quad P(g_i(\mathbf{x},\mathbf{U}) \leq 0, i=1,2) \geq 0.95$$

$$\begin{aligned} f(x_1,x_2,u_1,u_2) &= 5(x_1^2+x_2^2)-(u_1^2+u_2^2)+x_1(u_2-u_1+5) \\ &\quad +x_2(u_1-u_2+3) \\ g_1(x_1,x_2,u_1,u_2) &= -x_1^2+5x_2-u_1+u_2^2-1 \\ g_2(x_1,x_2,u_1,u_2) &= (-x_1^2+5x_2-u_1+u_2^2-1)(x_1+5)/5-u_1-1 \end{aligned}$$

with $\mathbf{x} \in [-5,5]^2$ and $\mathbf{U} \sim \mathcal{U}([-5,5]^2)$.

- Initial data-set : 30 samples (15 per constraint)
- Optimization performed with 80 constraint evaluations

4-d test-case

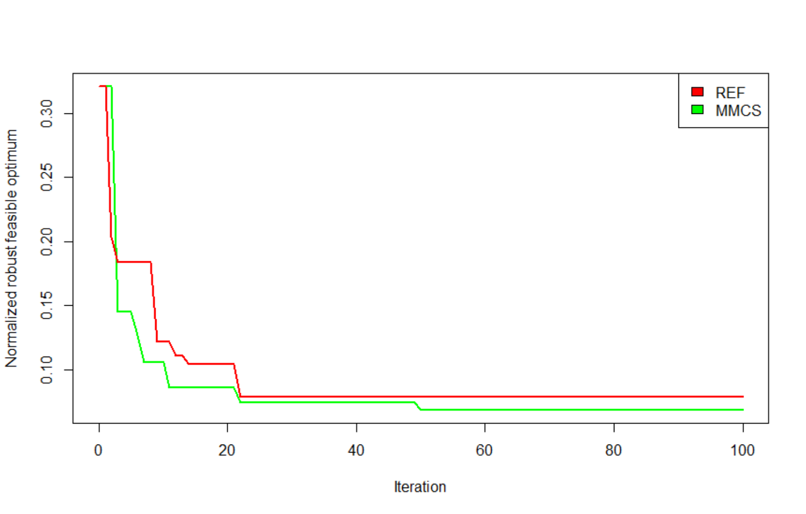

REF
SMCS
MMCU
MMCS
Robust feasible optimum
60
65
70
75
80
85
0
20
40
60
80
Iteration

Industrial test case: compressor rotor design

Design optimization problem for the NASA rotor 37. The objective is to maximize the compressor polytropic efficiency.

- 20 geometric design variables
- 7 random variables
- 5 Constraints
- Initial data-set : 100 samples (20 per constraint)
- Single repetition due to computational costs
- Only 2 methods are compared: REF and MMCS

Industrial test case: compressor rotor design



Industrial test case: compressor rotor design

Constraint	g_1	g_2	g_3	g_4	g_5
N° evaluations REF	100	100	100	100	100
N° evaluations MMCS	103	202	17	98	80

Conclusions

- Novel two-step robust Bayesian optimization algorithm, alloving to solve chance constrained problems for which both objective function and constraints are influenced by random variables
- Different variants of the algorithm are proposed and compared
- Possibility of refining the surrogate models more efficiently
    - Different random variable values for each constraint
    - Selection of most informational constraints
- The variants performance depends on the considered optimization problem nature

Perspectives

- Possibility of selecting the most relevant batch of constraints (and associated value of $\mathbf{u}$) rather than a single one
- Possibility of relying on multi-output GPs only for constraints for which the presence of correlation is known
- Possibility of taking into account the correlation between the objective function and the constraints
- Improvement of the internal optimization routines
- Extension of the method to inversion problems


"That's all Folks!"

References I

[1] Mauricio A Alvarez, Lorenzo Rosasco, and Neil D Lawrence.
Kernels for vector-valued functions: A review.
arXiv preprint arXiv:1106.6251, 2011.

[2] Reda El Amri, Rodolphe Le Riche, Céline Helbert, Christophette Blanchet-Scalliet, and Sébastien Da Veiga.
A sampling criterion for constrained bayesian optimization with uncertainties.
arXiv preprint arXiv:2103.05706, 2021.

[3] Clément Chevalier, David Ginsbourger, and Xavier Emery.
Corrected kriging update formulae for batch-sequential data assimilation.
In *Mathematics of Planet Earth*, pages 119–122. Springer, 2014.

References II

[4] Theodoros Evgeniou, Charles A Micchelli, and Massimiliano Pontil.
Learning multiple tasks with kernel methods.
Journal of machine learning research, 6(Apr):615–637, 2005.

[5] Janis Janusevskis and Rodolphe Le Riche.
Simultaneous kriging-based estimation and optimization of mean response.
Journal of Global Optimization, 55(2):313–336, 2013.

[6] Julien Pelamatti, Rodolphe Le Riche, Céline Helbert, Christophette Blanchet-Scalliet, and Sébastien Da Veiga.
A coupling and selecting constraints in bayesian optimization under uncertainties sampling criterion for constrained bayesian optimization with uncertainties.
arXiv preprint arXiv:2204.00527, 2022.

References III

[7] Olivier Roustant, Esperan Padonou, Yves Deville, Aloïs Clément, Guillaume Perrin, Jean Giorla, and Henry Wynn.
Group kernels for gaussian process metamodels with categorical inputs.
arXiv preprint arXiv:1802.02368, 2018.

[8] Marcos A Valdebenito and Gerhart I Schuëller.
A survey on approaches for reliability-based optimization.
Structural and Multidisciplinary Optimization, 42(5):645–663, 2010.