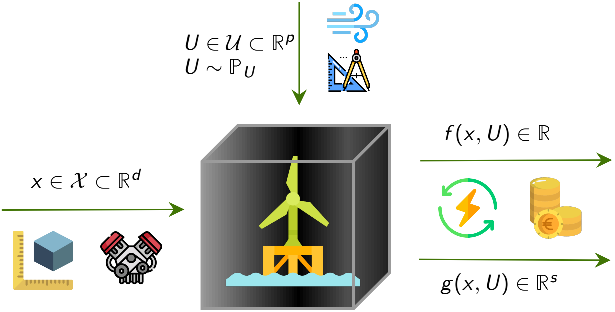
Sensitivity Analysis in Constrained Bayesian Optimization with Uncertainties

Noé Fellmann
Céline Helbert & Christophette Blanchet (ECL)
Adrien Spagnol & Delphine Sinoquet (IFPEN)

École Centrale de Lyon & IFP Énergie nouvelles

ETICS 2024, Septembre 2024

Black-box and robust optimization



Robust optimization problem

$$x^* = \arg\min_{\boldsymbol{x} \in \mathcal{K}} \mathbb{E}[f(\boldsymbol{x}, \boldsymbol{U})] \text{ where } \mathcal{K} = \{\boldsymbol{x} \in \mathcal{X} \text{ s.t. } \mathbb{P}[g(\boldsymbol{x}, \boldsymbol{U}) \leq 0] \geq 0.95\}$$

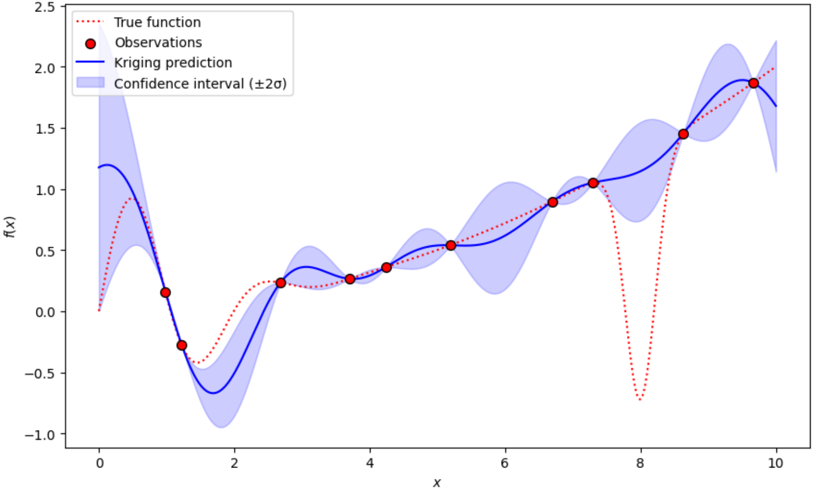
- f and g can be very expensive $\rightarrow$ Bayesian Optimization
- Robust Bayesian optimization : EFISUR [El-Amri et al. 2021]
- Input dimensions d and p can be very high $\rightarrow$ Dimension reduction

Variable selection by Sensitivity Analysis inside EFISUR

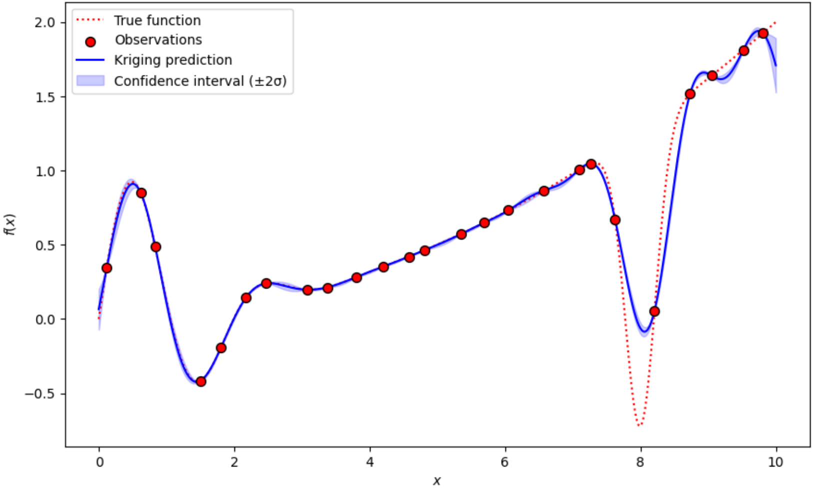
Table of Contents

Table of Contents

GP for optimization : LHS design



GP for optimization : LHS design



GP for optimization : sequential design = Bayesian optimization

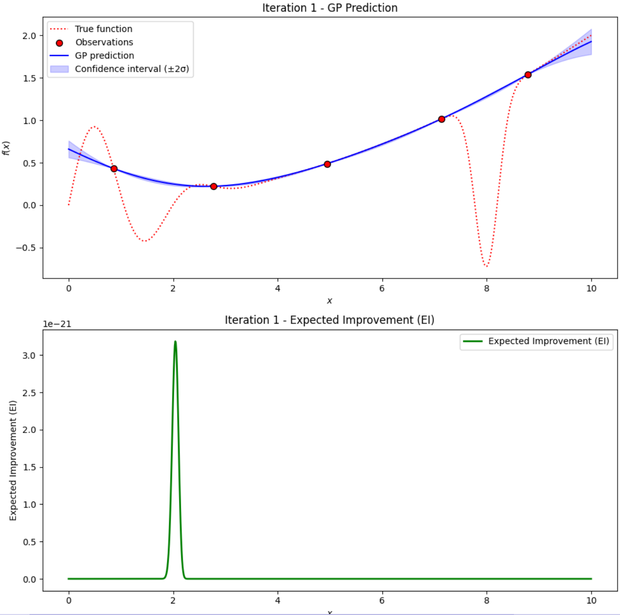
Algorithm Bayesian Optimization

Input : $f(x)$ to be minimized, prior GP F, acquisition function $a_F(x)$, initial design $\mathcal{D}^{(t)} = \{x^i\}$
Evaluate $y^i = f(x^i)$
while $t \leq$ *budget* **do**
    Fit GP $F^{(t)}$ of f with $\{(x^i, y^i)\}_{i=1}^{t}$
    Select the next point x^{t+1} by maximizing the acquisition function :

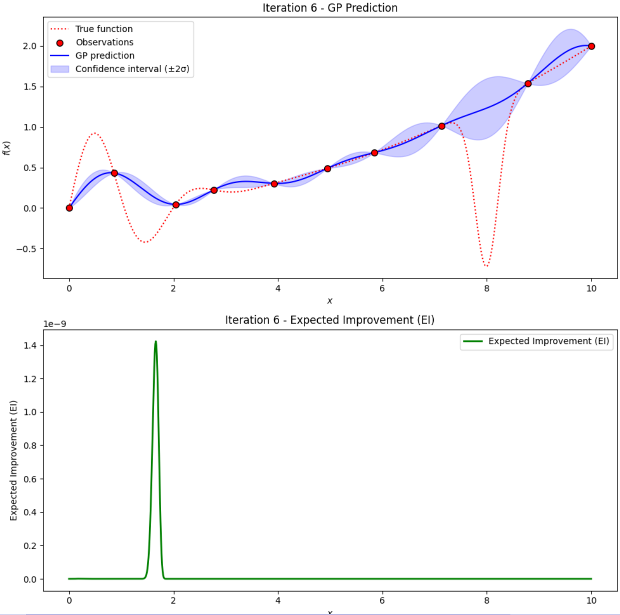
$$x^{t+1} = \arg\max_x a_{F^{(t)}}(x)$$

    Evaluate the objective function $y^{t+1} = f(x^{t+1})$
end while
Output : $x^* = \arg\min_i y^i$

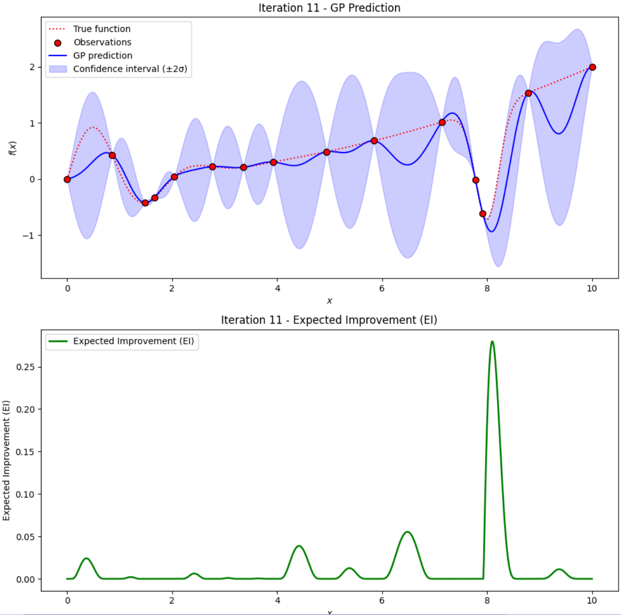
Efficient Global Optimization : Illustration



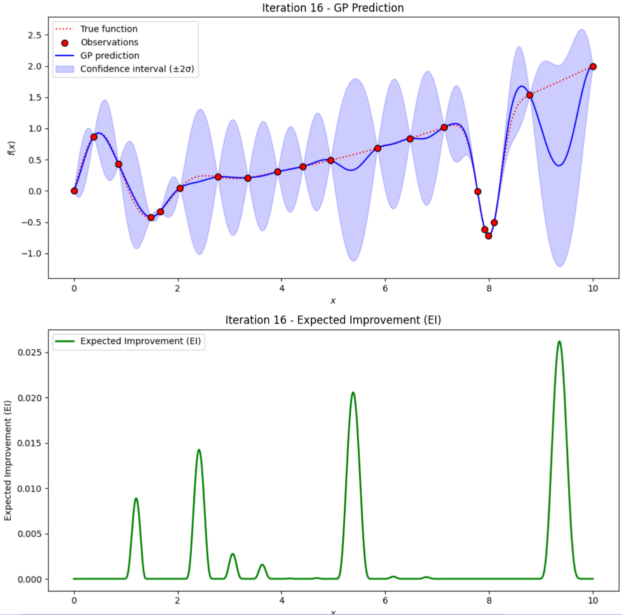
Efficient Global Optimization : Illustration



Efficient Global Optimization : Illustration



Efficient Global Optimization : Illustration



From EGO to EFISUR

From

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

to

$$x^* = \arg\min_{\boldsymbol{x} \in \mathcal{K}} \mathbb{E}[f(\boldsymbol{x}, \boldsymbol{U})] \text{ where } \mathcal{K} = \{\boldsymbol{x} \in \mathcal{X} \text{ s.t. } \mathbb{P}[g(\boldsymbol{x}, \boldsymbol{U}) \leq 0] \geq 0.95\}$$

- Constraints :
  - One GP G_i for each constraint g_i.
  - Expected Feasible Improvement (EFI)
- Uncertain inputs :
  - GPs F and G_i in the joint space $\mathcal{X} \times \mathcal{U}$
  - Expected (Gaussian) process $Z_x^{(t)} = \mathbb{E}_U[F_{(x,U)}^{(t)}]$
  - Constraint (not Gaussian) process $C_x^{(t)} = 0.95 - \mathbb{P}_U(G_{(x,U)}^{(t)} \leq 0)$
  - Criterion S to select u^{t+1} (Proxy of the one step-ahead feasible improvement variance)

Algorithm 1 : EFISUR [El-Amri et al. 2021]

Expected Feasible Improvement with Stepwise Uncertainty Reduction Sampling

---

Algorithm EFISUR

Design of Experiments (DoE) in the joint space $\mathcal{X} \times \mathcal{U}$:

$D^{(t)} = \{(\mathrm{x}_i, \mathrm{u}_i), i = 1, \ldots, t\}$

Calculate simulator responses : $f^{(t)} = (f(x_i, u_i))_{i=1}^{t}$ and $g^{(t)} = (g(x_i, u_i))_{i=1}^{t}$

while $t \leq$ maximum budget **do**

Fit GPs of f and g : $F^{(t)}_{(x,u)}$ and $G^{(t)}_{(x,u)}$

Calculate GP of the mean objective, and of the constraint :

- $Z_x^{(t)} = \mathbb{E}_U[F^{(t)}_{(x,U)}]$
- $C_x^{(t)} = 0.95 - \mathbb{P}_U(G^{(t)}_{(x,U)} \leq 0)$

Maximize $\mathrm{EFI}^{(t)}(\mathrm{x}) = \mathbb{E}(\mathrm{FI}^{(t)}) = \mathrm{EI}^{(t)}(\mathrm{x})\mathbb{P}\left(C^{(t)}(\mathrm{x}) \leq 0\right) \rightarrow \mathrm{x}_{\mathbf{targ}}$

Set $\mathrm{x}_{t+1} = \mathrm{x}_{\mathbf{targ}}$

Sample the next uncertain point by minimizing (a proxy of) the one-step ahead variance of the Feasible Improvement :

$\mathrm{u}_{t+1} = \arg\min_{\bar{u} \in \mathcal{S}_{\mathcal{U}}} S(\mathrm{x}_{\mathbf{targ}}, \bar{u})$

Calculate simulator responses at the next point $(\mathrm{x}_{t+1}, \mathrm{u}_{t+1})$

Update the DoE

end while

Progress Measure : $FI^{(t)}(\mathrm{x}) = I^{(t)}(\mathrm{x})\mathbb{1}_{\{C^{(t)}(\mathrm{x}) \leq 0\}}$ where $I^{(t)}(x) = \left(z_{\min}^{\mathrm{feas}} - Z^{(t)}(\mathrm{x})\right)^{+}$

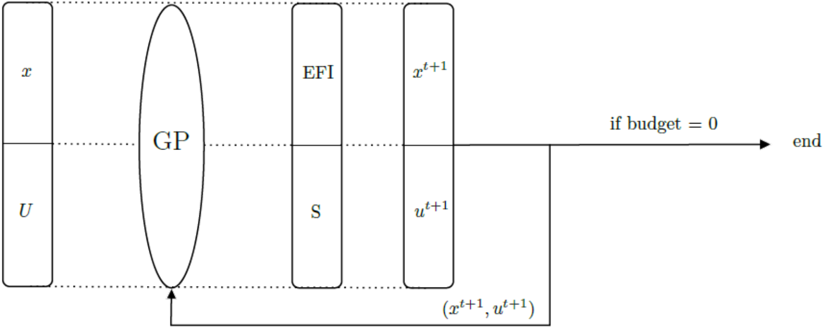


Figure – EFISUR

Issues of Bayesian optimization in high dimension

High input dimension issues in Bayesian optimization [Binois et Wycoff 2022] :

- Fit GPs in high dimension
- Optimize acquisition criterion in high dimension

Examples of solutions :

- Adapt the way to fit the GP : sum of low dimensional GPs
- Adapted kernels
- Adaptations of the acquisition criterion
- **Dimension reduction**

Dimension reduction in Bayesian optimization :

- Offline approaches : Dimension reduction before fitting the GPs $\rightarrow$ GPs in low dimensions
- Onlines approaches : Dimension reduction of the acquisition criterion optimization

How to reduce the dimension ? $\longrightarrow$ Sensitivity Analysis

Table of Contents

Goal oriented Sensitivity analysis

Sensitivity analysis (SA)

$$(X_1, ..., X_d) \overset{f}{\mapsto} Y = f(X_1, ..., X_d)$$

How can the uncertainty of Y be divided and allocated to the uncertainty of the inputs X_i ?

- Sobol indices : $S_i = \frac{\text{Var}\,\mathbb{E}(Y|X_i)}{\text{Var}\,Y}$
- Dependence measures : $S_i = ||\mathbb{P}_{(X_i,Y)} - \mathbb{P}_{X_i} \otimes \mathbb{P}_Y|| \rightarrow \text{HSIC}(X, Y)$

Screening : $X_1, ..., X_k$ are influential and $X_{k+1}, ...X_d$ are not influential
Ranking : $X_1 \prec ... \prec X_d$

SA for optimization

For deterministic optimization [Spagnol 2020] :

$$x^* = \underset{x \in \mathcal{X}}{\arg\min}\, f(x) \text{ s.t. } g(x) \leq 0$$

SA on

$$X = (X_1, ..., X_d) \rightarrow \mathbb{1}_{f(X) \leq q, g(X) \leq 0}$$

to identify the important input for the optimization.
$\rightarrow$ smoothed transformations also possible (see [Marrel et Chabridon 2021])

Our way to do Goal-oriented SA for robust optimization

$$x^* = \arg\min_{\boldsymbol{x} \in \mathcal{K}} \mathbb{E}[f(\boldsymbol{x}, \boldsymbol{U})] \text{ where } \mathcal{K} = \{\boldsymbol{x} \in \mathcal{X} \text{ s.t. } \mathbb{P}[g(\boldsymbol{x}, \boldsymbol{U}) \leq 0] \geq 0.95\}$$

GOSA on the x

$$HSIC(X_i, \mathbb{1}_{E \leq q, P \geq \alpha})$$

where $E = \mathbb{E}_U[f(X, U)]$ and $P = \mathbb{P}_U(g(X, U) \leq 0)$

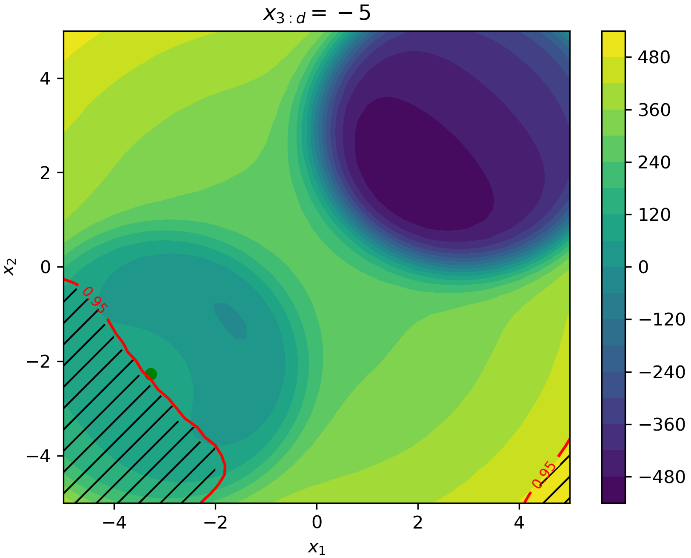
GOSA on the U

Influence of the U on excursion sets :

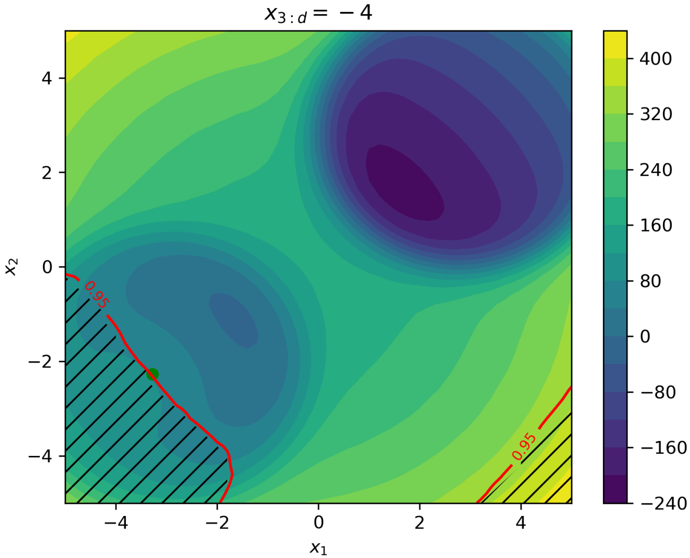
$$HSIC(U, \Gamma)$$

where $\Gamma = \{x, f(x, U) \leq q, g(x, U) \leq 0\}$
→ See [Fellmann et al. 2024] for HSIC of set-valued outputs

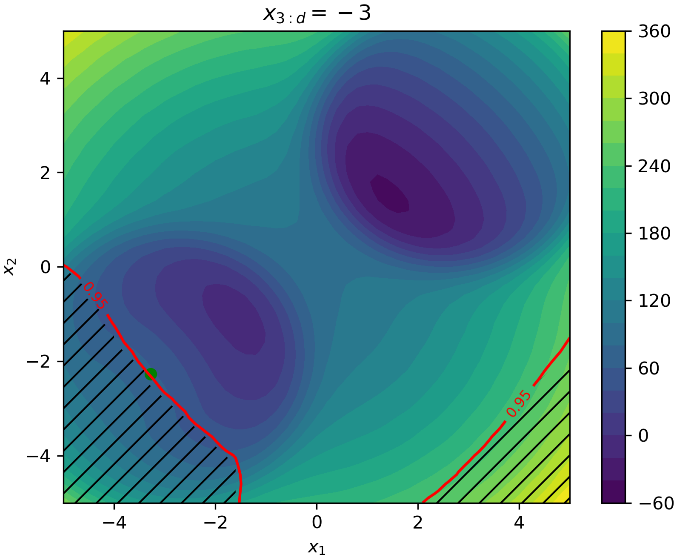
Toy function in higher dimension : PLOTS



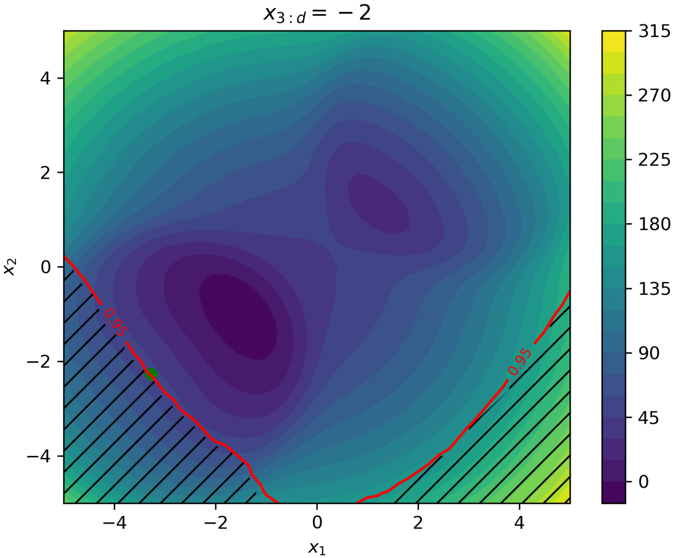
Toy function in higher dimension : PLOTS



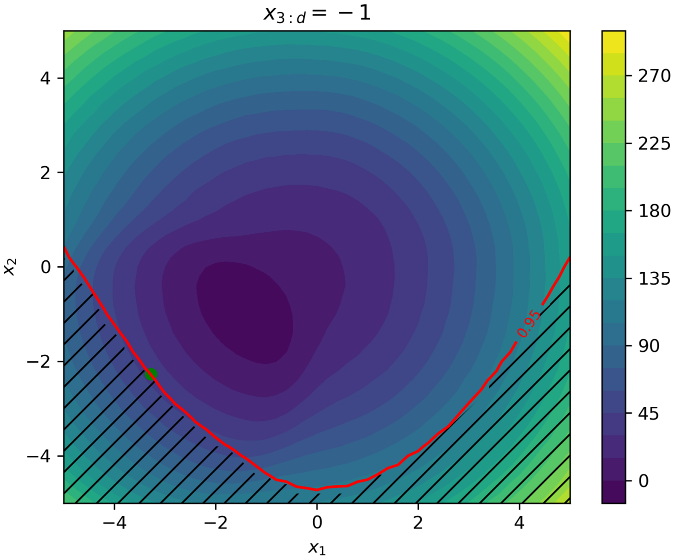
Toy function in higher dimension : PLOTS



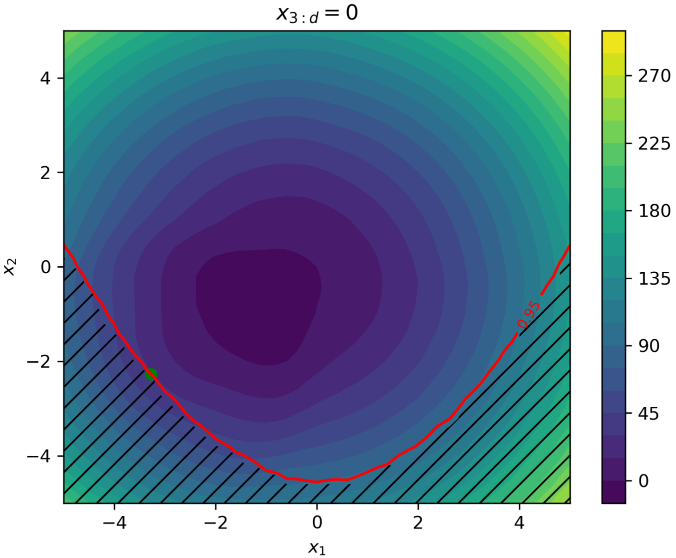
Toy function in higher dimension : PLOTS



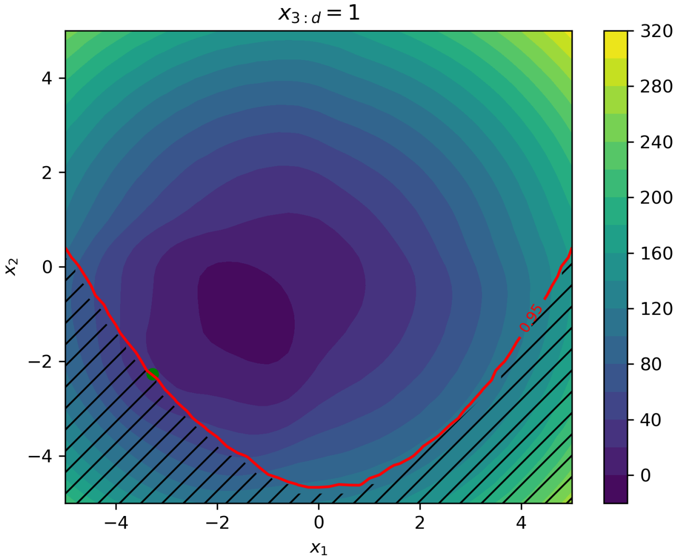
Toy function in higher dimension : PLOTS



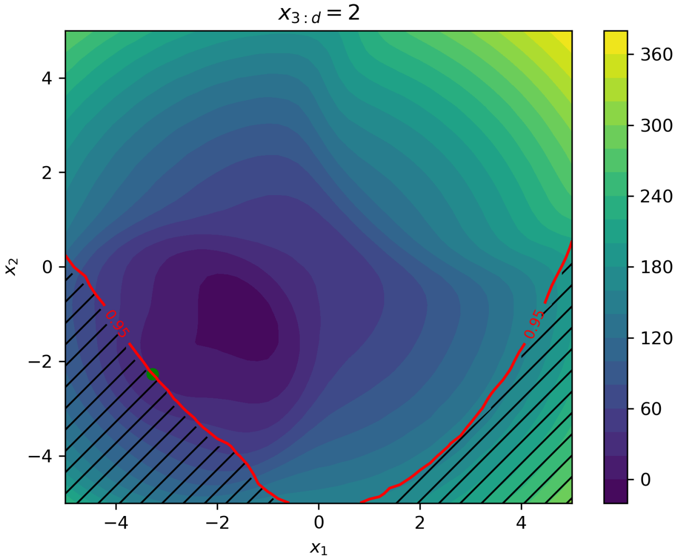
Toy function in higher dimension : PLOTS



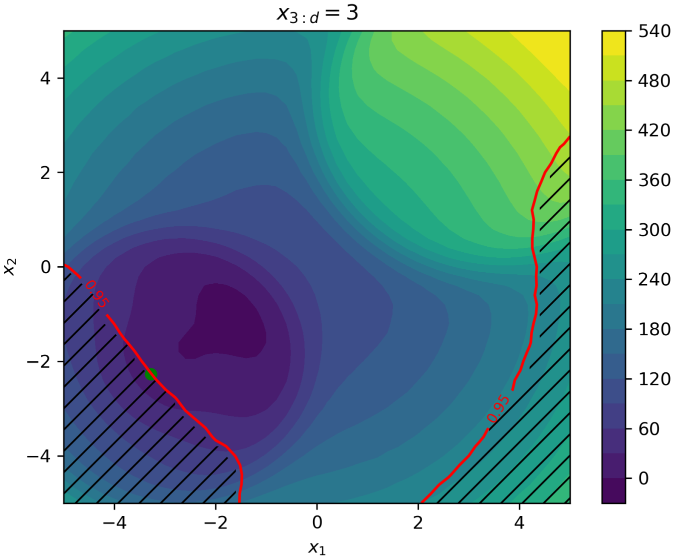
Toy function in higher dimension : PLOTS



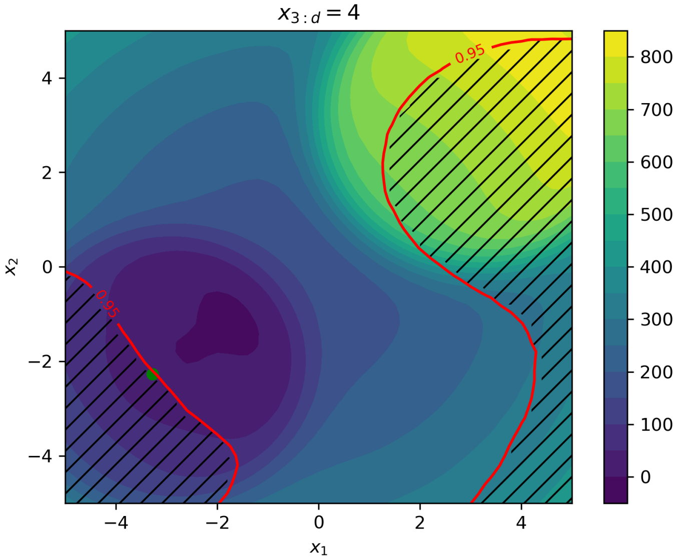
Toy function in higher dimension : PLOTS



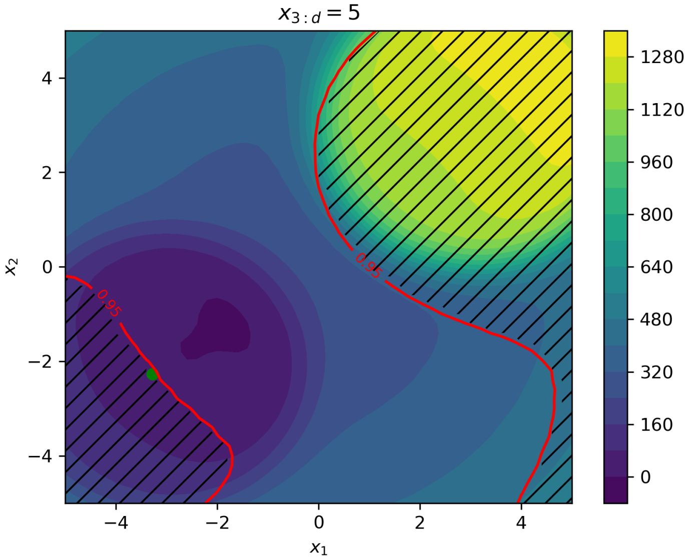
Toy function in higher dimension : PLOTS



Toy function in higher dimension : PLOTS



Toy function in higher dimension : PLOTS



SA and GOSA of the toy function

HSIC	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
Classic	1.00	1.00	1.00	0.85	0.15	0.00	0.00	0.10	0.00	0.05
TSA	1.00	1.00	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.10
CSA	0.10	0.70	0.00	0.10	0.10	0.00	0.10	0.15	0.15	0.125

Table – Rejection rates on 20 repetitions

HSIC	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
Classic	1.00	1.00	0.25	0.00	0.05	0.15	0.05	0.10	0.05	0.40
TSA	1.00	1.00	0.05	0.05	0.00	0.05	0.00	0.10	0.05	1.00
CSA	0.70	0.55	0.10	0.00	0.10	0.15	0.05	0.05	0.00	1.00

Table – Rejection rates zoomed in the south-west quadrant ($x_1 \leq 0$ and $x_2 \leq 0$).

HSIC of sets	U_1	U_2	U_3	U_4	U_5	U_6	U_7	U_8	U_9	U_{10}
SA	1.00	1.00	0.85	0.25	0.00	0.00	0.15	0.05	0.05	0.00
GOSA	1.00	1.00	0.10	0.10	0.15	0.10	0.10	0.10	0.15	1.00

Table – Rejection rates on 20 repetitions

Table of Contents

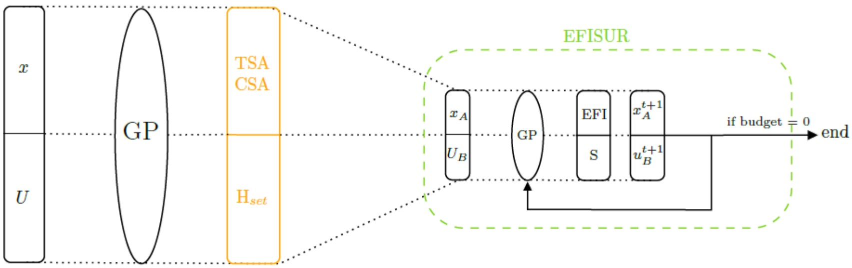


Figure – Off-SA-EFISUR

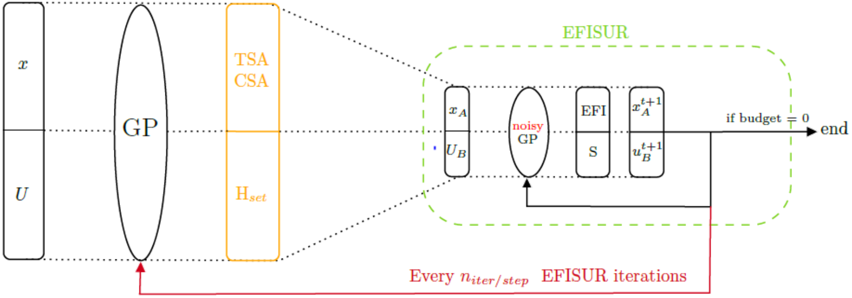


Figure – Off-SA-seq-EFISUR

Gaussian Process modelling in Off-seq-SA

SA steps are too coslty to be done directly on f and g evaluations

- SA on the GP means μ_F, μ_G, μ_Z and μ_C.

As the dimensions of interest evolve during the algorithm, model evaluations must be done without fixing variables

- Noisy GPs are used, where the (heteroscedastic) noise corresponds to the effect of current negligible inputs.

How to fit the noise ?

- We use the Goal-oriented transformation to fit the heteroscedastic noise, fixing its variance proportional to $1/h$.

Table of Contents

Algorithm 1 : EFISUR (El-Amri et al. 2021)

Expected Feasible Improvement with Stepwise Uncertainty Reduction Sampling

Algorithm EFISUR

Design of Experiments (DoE) in the joint space $\mathcal{X} \times \mathcal{U}$:

$D^{(t)} = \{(\mathrm{x}_i, \mathrm{u}_i), i = 1, \ldots, t\}$

Calculate simulator responses : $f^{(t)} = (f(x_i, u_i))_{i=1}^t$ and $g^{(t)} = (g(x_i, u_i))_{i=1}^t$

while $t \leq$ maximum budget **do**

Fit GPs of f and g : $F^{(t)}_{(x,u)}$ and $G^{(t)}_{(x,u)}$

Calculate GP of the mean objective, and of the constraint :

- $Z_x^{(t)} = \mathbb{E}_U[F^{(t)}_{(x,U)}]$
- $C_x^{(t)} = 0.95 - \mathbb{P}_U(G^{(t)}_{(x,U)} \leq 0)$

Maximize $\mathrm{EFI}^{(t)}(\mathrm{x}) = \mathbb{E}(\mathrm{FI}^{(t)}) = \mathrm{EI}^{(t)}(\mathrm{x})\mathbb{P}\left(C^{(t)}(\mathrm{x}) \leq 0\right) \to \mathrm{x}_{\mathbf{targ}}$

Set $\mathrm{x}_{t+1} = \mathrm{x}_{\mathbf{targ}}$

Sample the next uncertain point by minimizing (a proxy of) the one-step ahead variance of the Feasible Improvement :

$\mathrm{u}_{t+1} = \arg\min_{\bar{u} \in \mathcal{S}_{\mathcal{U}}} S(\mathrm{x}_{\mathbf{targ}}, \bar{u})$

Calculate simulator responses at the next point $(\mathrm{x}_{t+1}, \mathrm{u}_{t+1})$

Update the DoE

end while

Algorithm 1 : EFISUR (El-Amri et al. 2021)

Expected Feasible Improvement with Stepwise Uncertainty Reduction Sampling

Algorithm EFISUR

Design of Experiments (DoE) in the joint space $\mathcal{X} \times \mathcal{U}$:

$D^{(t)} = \{(\mathrm{x}_i, \mathrm{u}_i), i = 1, \dots, t\}$

Calculate simulator responses : $f^{(t)} = (f(x_i, u_i))_{i=1}^{t}$ and $g^{(t)} = (g(x_i, u_i))_{i=1}^{t}$

while $t \leq$ maximum budget **do**

Fit GPs of f and g : $F^{(t)}_{(x,u)}$ and $G^{(t)}_{(x,u)}$

Calculate GP of the mean objective, and of the constraint :

- $Z_x^{(t)} = \mathbb{E}_U[F^{(t)}_{(x,U)}]$
- $C_x^{(t)} = 0.95 - \mathbb{P}_U(G^{(t)}_{(x,U)} \leq 0)$

Maximize $\mathrm{EFI}^{(t)}(\mathrm{x}) = \mathbb{E}(\mathrm{FI}^{(t)}) = \mathrm{EI}^{(t)}(\mathrm{x})\mathbb{P}\left(C^{(t)}(\mathrm{x}) \leq 0\right) \to \mathrm{x}_{\mathrm{targ}}$

Set $\mathrm{x}_{t+1} = \mathrm{x}_{\mathrm{targ}}$

Sample the next uncertain point by minimizing (a proxy of) the one-step ahead variance of the Feasible Improvement :

$\mathrm{u}_{t+1} = \arg\min_{\bar{u} \in \mathcal{S}_\mathcal{U}} S(\mathrm{x}_{\mathrm{targ}}, \bar{u}))$

Calculate simulator responses at the next point $(\mathrm{x}_{t+1}, \mathrm{u}_{t+1})$

Update the DoE

end while

Algorithm 1 : EFISUR (El-Amri et al. 2021)

Expected Feasible Improvement with Stepwise Uncertainty Reduction Sampling

Algorithm EFISUR

Design of Experiments (DoE) in the joint space $\mathcal{X} \times \mathcal{U}$:

$D^{(t)} = \{(\mathrm{x}_i, \mathrm{u}_i), i = 1, \ldots, t\}$

Calculate simulator responses : $f^{(t)} = (f(x_i, u_i))_{i=1}^{t}$ and $g^{(t)} = (g(x_i, u_i))_{i=1}^{t}$

while $t \leq$ maximum budget **do**

Fit GPs of f and g : $F^{(t)}_{(x,u)}$ and $G^{(t)}_{(x,u)}$

Calculate GP of the mean objective, and of the constraint :

- $Z^{(t)}_x = \mathbb{E}_U[F^{(t)}_{(x,U)}]$
- $C^{(t)}_x = 0.95 - \mathbb{P}_U(G^{(t)}_{(x,U)} \leq 0)$

Dimension reduction with GOSA on the x's : HSIC(X, $EFI(X)$)

Maximize $\mathrm{EFI}^{(t)}(\mathrm{x}) = \mathbb{E}(\mathrm{FI}^{(t)}) = \mathrm{EI}^{(t)}(\mathrm{x})\mathbb{P}\left(C^{(t)}(\mathrm{x}) \leq 0\right) \rightarrow \mathrm{x}_{\mathrm{targ}}$

Set $\mathrm{x}_{t+1} = \mathrm{x}_{\mathrm{targ}}$

Sample the next uncertain point by minimizing (a proxy of) the one-step ahead variance of the Feasible Improvement :

$\mathrm{u}_{t+1} = \arg\min_{\bar{u} \in \mathcal{S}_{\mathcal{U}}} S(\mathrm{x}_{\mathrm{targ}}, \bar{u}))$

Calculate simulator responses at the next point $(\mathrm{x}_{t+1}, \mathrm{u}_{t+1})$

Update the DoE

end while

Algorithm 1 : EFISUR (El-Amri et al. 2021)

Expected Feasible Improvement with Stepwise Uncertainty Reduction Sampling

Algorithm EFISUR

Design of Experiments (DoE) in the joint space $\mathcal{X} \times \mathcal{U}$:
    $D^{(t)} = \{(\mathrm{x}_i, \mathrm{u}_i), i = 1, \ldots, t\}$
Calculate simulator responses : $f^{(t)} = (f(x_i, u_i))_{i=1}^t$ and $g^{(t)} = (g(x_i, u_i))_{i=1}^t$
while $t \leq$ maximum budget **do**
    Fit GPs of f and g : $F^{(t)}_{(x,u)}$ and $G^{(t)}_{(x,u)}$
    Calculate GP of the mean objective, and of the constraint :

- $Z_x^{(t)} = \mathbb{E}_U[F^{(t)}_{(x,U)}]$
- $C_x^{(t)} = 0.95 - \mathbb{P}_U(G^{(t)}_{(x,U)} \leq 0)$

    Dimension reduction with GOSA on the x's : HSIC$(X, EFI(X))$
    Maximize $\mathrm{EFI}^{(t)}(\mathrm{x}) = \mathbb{E}(\mathrm{FI}^{(t)}) = \mathrm{EI}^{(t)}(\mathrm{x})\mathbb{P}\left(C^{(t)}(\mathrm{x}) \leq 0\right) \rightarrow \mathrm{x}_{\mathrm{targ}}$
    Set $\mathrm{x}_{t+1} = \mathrm{x}_{\mathrm{targ}}$
    Dimension reduction with GOSA on the U's : HSIC$(U, S(x_{targ}, U)$
    Sample the next uncertain point by minimizing (a proxy of) the one-step ahead variance of the Feasible Improvement :
        $\mathrm{u}_{t+1} = \arg\min_{\bar{u} \in \mathcal{S}_{\mathcal{U}}} S(\mathrm{x}_{\mathrm{targ}}, \bar{u}))$
    Calculate simulator responses at the next point $(\mathrm{x}_{t+1}, \mathrm{u}_{t+1})$
    Update the DoE
end while

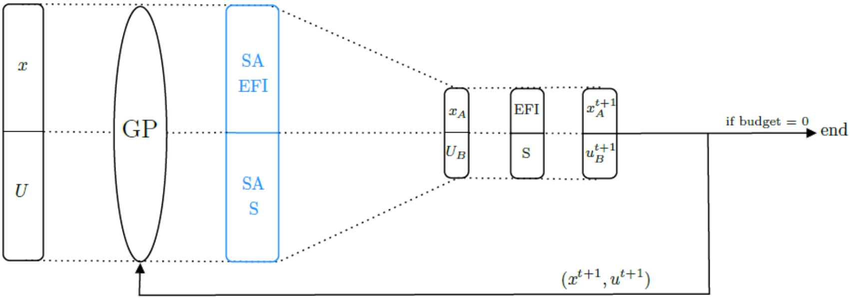


Figure – On-SA-EFISUR

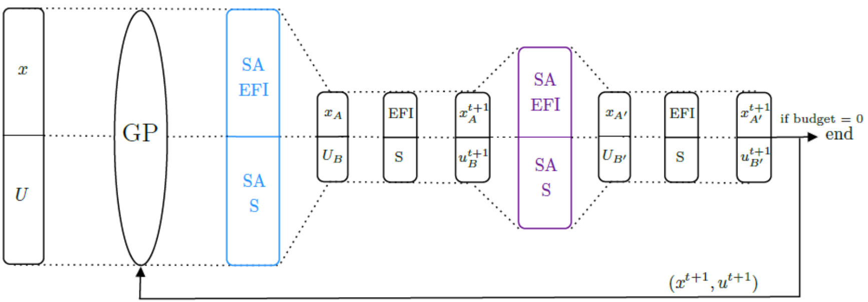
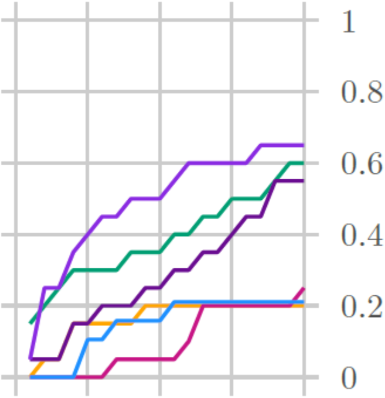


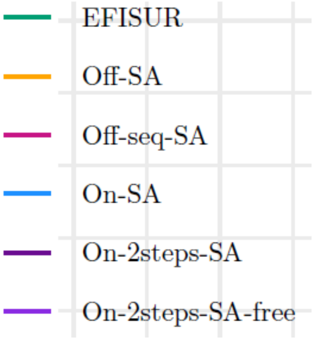
Figure – On-2-SA-seq-EFISUR

Table of Contents

Method comparison on the toy function

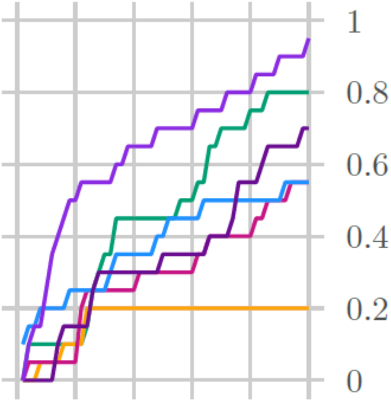


(a) 180 (initial design) + 20 iterations

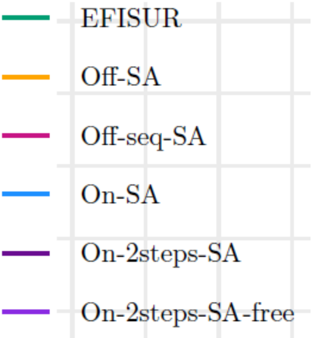


(b) Algorithms

Method comparison on the toy function

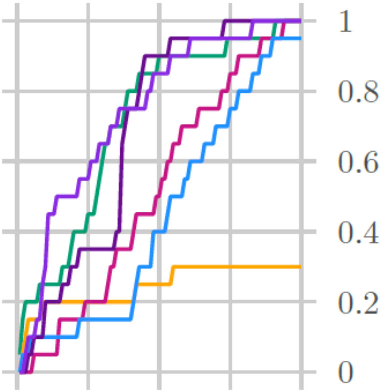


(c) 150 (initial design)+ 50 iterations

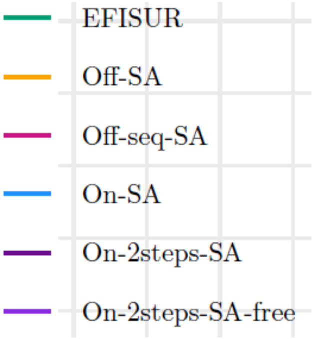


(d) Algorithms

Method comparison on the toy function

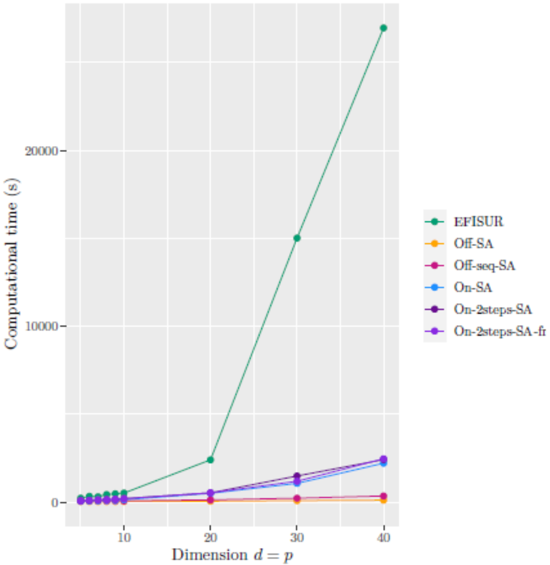


(e) 100 (initial design) + 100 iterations



(f) Algorithms

Comparison in time between the methods



El-Amri, Reda et al. (2021). *A sampling criterion for constrained Bayesian optimization with uncertainties*. arXiv : 2103.05706 [stat.ML].

Binois, Mickael et Nathan Wycoff (2022). "A survey on high-dimensional Gaussian process modeling with application to Bayesian optimization". In : *ACM Transactions on Evolutionary Learning and Optimization* 2.2, p. 1-26.

Fellmann, Noé et al. (2024). "Kernel-based sensitivity analysis for (excursion) sets". In : *Technometrics*, p. 1-21.

Marrel, Amandine et Vincent Chabridon (2021). "Statistical developments for target and conditional sensitivity analysis : Application on safety studies for nuclear reactor". In : *Reliability Engineering and System Safety* 214, p. 107711. issn : 0951-8320. doi : https://doi.org/10.1016/j.ress.2021.107711. url : https://www.sciencedirect.com/science/article/pii/S0951832021002465.

Spagnol, Adrien (juill. 2020). "Kernel-based sensitivity indices for high-dimensional optimization problems". Theses. Université de Lyon. url : https://tel.archives-ouvertes.fr/tel-03173192.