Active Learning of (small) Quantile Sets

Romain Ait Abdelmalek-Lomenech[1]
joint work with Julien Bect[1], Vincent Chabridon[2] & Emmanuel Vazquez[1]

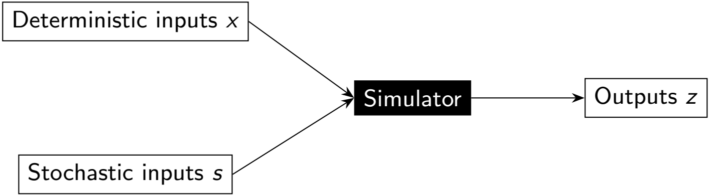
[1]Université Paris-Saclay, CNRS, CentraleSupélec, L2S
[2] EDF R&D, PRISME Team

ANR SAMOURAI final workshop,
December 11, Paris

Table of Contents

Consider an **expensive-to-evaluate** numerical simulator f, with inputs in a set $\mathbb{U} = \mathbb{X} \times \mathbb{S}$:

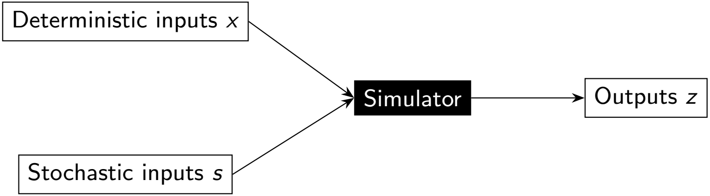
- $x \in \mathbb{X}$ (deterministic design choices).
- $s \in \mathbb{S}$ (stochastic factors).



For simplicity we assume a deterministic simulator $f : \mathbb{U} = \mathbb{X} \times \mathbb{S} \mapsto \mathbb{R}^q$.

Consider an **expensive-to-evaluate** numerical simulator f, with inputs in a set $\mathbb{U} = \mathbb{X} \times \mathbb{S}$:

- $x \in \mathbb{X}$ (deterministic design choices).
- $s \in \mathbb{S}$ (stochastic factors).



For simplicity we assume a deterministic simulator $f : \mathbb{U} = \mathbb{X} \times \mathbb{S} \mapsto \mathbb{R}^q$.

Given:

- $C \subset \mathbb{R}^q$ a subset of the outputs space $\mathbb{R}^q$.
- $\alpha \in (0, 1)$ a threshold.
- $\mathbb{P}_S$ a known distribution on $\mathbb{S}$.

We focus on the **quantile set inversion (QSI)** problem:

Estimate the set of all $x \in \mathbb{X}$ such that

$$\mathbb{P}\left(f(x, S) \in C\right) \leq \alpha, \qquad S \sim \mathbb{P}_S.$$

Given:

- $C \subset \mathbb{R}^q$ a subset of the outputs space $\mathbb{R}^q$.
- $\alpha \in (0, 1)$ a threshold.
- $\mathbb{P}_S$ a known distribution on $\mathbb{S}$.

We focus on the **quantile set inversion (QSI)** problem:

Estimate the set of all $x \in \mathbb{X}$ such that

$$\mathbb{P}\left(f(x, S) \in C\right) \leq \alpha, \qquad S \sim \mathbb{P}_S.$$

Given:

- $C \subset \mathbb{R}^q$ a subset of the outputs space $\mathbb{R}^q$.
- $\alpha \in (0, 1)$ a threshold.
- $\mathbb{P}_S$ a known distribution on $\mathbb{S}$.

We focus on the **quantile set inversion (QSI)** problem:

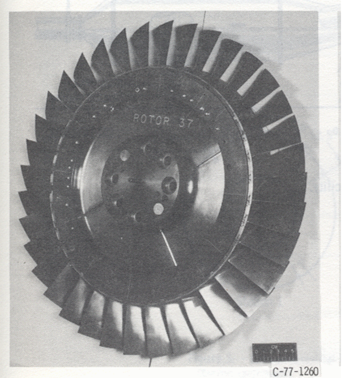
Estimate the set of all $x \in \mathbb{X}$ such that

$$\mathbb{P}\left(f(x, S) \in C\right) \leq \alpha, \qquad S \sim \mathbb{P}_S.$$

An example: the ROTOR37 compressor model

Function $f : \mathbb{X} \times \mathbb{S} \mapsto \mathbb{R}^3$ with two kind of inputs:

- $x \in \mathbb{X}$: design choice for the compressor
- $s \in \mathbb{S}$: manufacturing uncertainties, with $\mathbb{P}_S = \mathcal{U}(\mathbb{S})$.



(a) Rotor 37.

Simulator return three outputs:

- f_1: mass flow
- f_2: pressure ratio
- f_3: isentropic efficiency

We can set, for example, $\alpha = 5\%$ and

$$C = \left\{ z \in \mathbb{R}^3 : \frac{|z_1 - b_1|}{|b_1|} > 0.175 \quad \text{or} \quad \frac{|z_2 - b_2|}{|b_2|} > 0.175 \right\},$$

where b_1 and b_2 are baseline values for the mass flow and pressure ratio of the compressor.

For simplicity, we now assume $f : \mathbb{U} = \mathbb{X} \times \mathbb{S} \mapsto \mathbb{R}$, with $C = (-\infty, T]$.

The problem becomes

Estimate the quantile set:

$$\Gamma(f) = \{x \in \mathbb{X} : \mathbb{P}\left(f(x, S) \leq T\right) \leq \alpha\}.$$

Remark: With $C = (-\infty, T]$, the problem can be seen in term of quantile of $f(x, S)$. Indeed

$$x \in \Gamma(f) \quad \Longleftrightarrow \quad q_\alpha(f(x, S)) > T,$$

with $q_\alpha(f(x, S))$ the quantile of order α of $f(x, S)$ (with $S \sim \mathbb{P}_S$).

For simplicity, we now assume $f : \mathbb{U} = \mathbb{X} \times \mathbb{S} \mapsto \mathbb{R}$, with $C = (-\infty, T]$.

The problem becomes

Estimate the quantile set:

$$\Gamma(f) = \{x \in \mathbb{X} : \mathbb{P}\left(f(x, S) \leq T\right) \leq \alpha\}.$$

Remark: With $C = (-\infty, T]$, the problem can be seen in term of quantile of $f(x, S)$. Indeed

$$x \in \Gamma(f) \qquad \Longleftrightarrow \qquad q_\alpha(f(x, S)) > T,$$

with $q_\alpha(f(x, S))$ the quantile of order α of $f(x, S)$ (with $S \sim \mathbb{P}_S$).

Estimate the quantile set:

$$\Gamma(f) = \{x \in \mathbb{X} : \mathbb{P}(f(x,S) \leq T) \leq \alpha\},$$

Example of function and associated quantile set, with $T = 7.5$ and $\alpha = 5\%$.

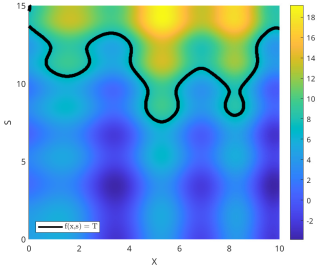


Figure: Representation of the function.

Estimate the quantile set:

$$\Gamma(f) = \{x \in \mathbb{X} : \mathbb{P}(f(x,S) \leq T) \leq \alpha\},$$

Example of function and associated quantile set, with $T = 7.5$ and $\alpha = 5\%$.

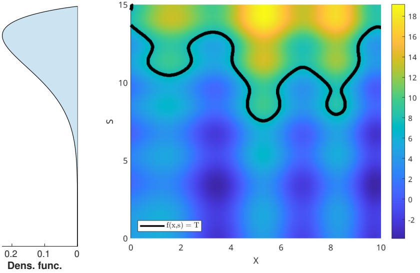


Figure: Representation of the function (right), the density of $\mathbb{P}_S$ (left)

Estimate the quantile set:

$$\Gamma(f) = \{x \in \mathbb{X} : \mathbb{P}(f(x,S) \leq T) \leq \alpha\},$$

Example of function and associated quantile set, with $T = 7.5$ and $\alpha = 5\%$.

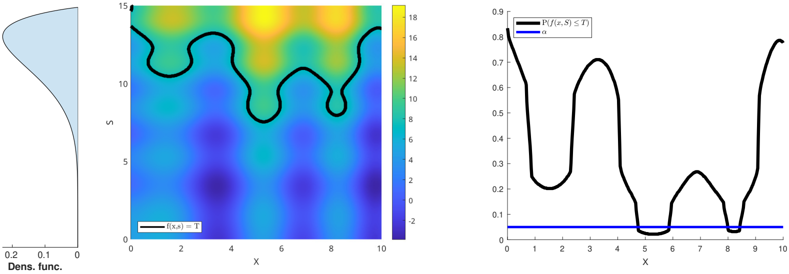


Figure: Representation of the function (middle), the density of $\mathbb{P}_S$ (left) and associated quantile set (right).

Estimate the quantile set:

$$\Gamma(f) = \{x \in \mathbb{X} : \mathbb{P}(f(x,S) \leq T) \leq \alpha\},$$

Example of function and associated quantile set, with $T = 7.5$ and $\alpha = 5\%$.

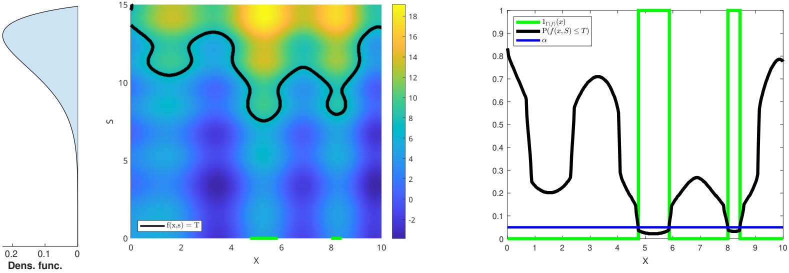


Figure: Representation of the function (middle), the density of $\mathbb{P}_S$ (left) and associated quantile set (right).

Given the expensive-to-evaluate nature of the underlying function, it is necessary to evaluate the function at points chosen with attention.

Active learning (or sequential design of experiments) approach:

Consider

- $\mathcal{I}_n = \{(u_1, f(u_1)), ..., (u_n, f(u_n))\}$ the current information,
- $a_n(u)$ a sampling criterion dependent on $\mathcal{I}_n$.

Until satisfied:

- Choose $u \in \mathbb{U}$ as the minimizer (or maximizer) of $a_n(u)$
- Evaluate f at u
- Update information $\mathcal{I}_n \mapsto \mathcal{I}_{n+1}$.

Question: How to construct such criterion for QSI?

Given the expensive-to-evaluate nature of the underlying function, it is necessary to evaluate the function at points chosen with attention.

Active learning (or sequential design of experiments) approach:

Consider

- $\mathcal{I}_n = \{(u_1, f(u_1)), ..., (u_n, f(u_n))\}$ the current information,
- $a_n(u)$ a sampling criterion dependent on $\mathcal{I}_n$.

Until satisfied:

- Choose $u \in \mathbb{U}$ as the minimizer (or maximizer) of $a_n(u)$
- Evaluate f at u
- Update information $\mathcal{I}_n \mapsto \mathcal{I}_{n+1}$.

Question: How to construct such criterion for QSI?

Given the expensive-to-evaluate nature of the underlying function, it is necessary to evaluate the function at points chosen with attention.

Active learning (or sequential design of experiments) approach:

Consider

- $\mathcal{I}_n = \{(u_1, f(u_1)), ..., (u_n, f(u_n))\}$ the current information,
- $a_n(u)$ a sampling criterion dependent on $\mathcal{I}_n$.

Until satisfied:

- Choose $u \in \mathbb{U}$ as the minimizer (or maximizer) of $a_n(u)$
- Evaluate f at u
- Update information $\mathcal{I}_n \mapsto \mathcal{I}_{n+1}$.

Question: How to construct such criterion for QSI?

Given the expensive-to-evaluate nature of the underlying function, it is necessary to evaluate the function at points chosen with attention.

Active learning (or sequential design of experiments) approach:

Consider

- $\mathcal{I}_n = \{(u_1, f(u_1)), ..., (u_n, f(u_n))\}$ the current information,
- $a_n(u)$ a sampling criterion dependent on $\mathcal{I}_n$.

Until satisfied:

- Choose $u \in \mathbb{U}$ as the minimizer (or maximizer) of $a_n(u)$
- Evaluate f at u
- Update information $\mathcal{I}_n \mapsto \mathcal{I}_{n+1}$.

Question: How to construct such criterion for QSI?

Table of Contents

In the following, we focus on Gaussian processes-based strategies.

Bayesian framework & notations:

$f \sim$ GP prior ξ on $\mathbb{U} = \mathbb{X} \times \mathbb{S}$, with constant mean μ and covariance k.

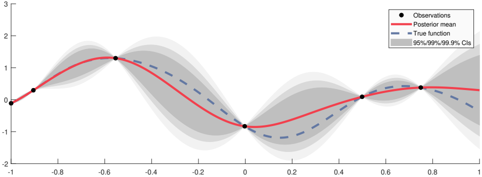


Figure: Illustration of a GP on an interval.

We denote:

- $\mathbb{P}_n$ and $\mathbb{E}_n$: conditional distribution and expectation given $\mathcal{I}_n$.
- μ_n, σ_n and k_n: posterior mean, st. deviation and covariance of ξ.
- $p_n(u) = \mathbb{P}_n(\xi(u) \leq T)$: posterior probability that $\{\xi(u) \leq T\}$

In the following, we focus on Gaussian processes-based strategies.

Bayesian framework & notations:

$f \sim$ GP prior ξ on $\mathbb{U} = \mathbb{X} \times \mathbb{S}$, with constant mean μ and covariance k.

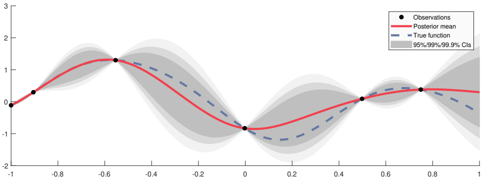


Figure: Illustration of a GP on an interval.

We denote:

- $\mathbb{P}_n$ and $\mathbb{E}_n$: conditional distribution and expectation given $\mathcal{I}_n$.
- μ_n, σ_n and k_n: posterior mean, st. deviation and covariance of ξ.
- $p_n(u) = \mathbb{P}_n(\xi(u) \leq T)$: posterior probability that $\{\xi(u) \leq T\}$

First approach: joint-space estimation

The QSI problem is related to the **estimation of the excursion set**

$$\Lambda(f) = \{u \in \mathbb{U} \; : \; f(u) \leq T\}$$

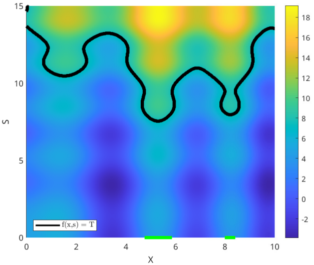


Figure: Example function. The black line delimits the set $\Lambda(f)$.

$$x \in \Gamma(f) \iff \mathbb{P}((x, S) \in \Lambda(f)) \leq \alpha,$$

Good approximation of $\Lambda(f) \implies$ good approximation of $\Gamma(f)$

First approach: joint-space estimation

The QSI problem is related to the **estimation of the excursion set**

$$\Lambda(f) = \{u \in \mathbb{U} \,:\, f(u) \leq T\}$$

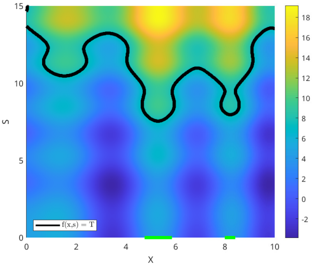


Figure: Example function. The black line delimits the set $\Lambda(f)$.

$$x \in \Gamma(f) \quad \Longleftrightarrow \quad \mathbb{P}((x, S) \in \Lambda(f)) \leq \alpha,$$

Good approximation of $\Lambda(f)$ $\Longrightarrow$ good approximation of $\Gamma(f)$

Several Bayesian methods focus on **estimating** $\Lambda(f)$. For example:

- **Maximal uncertainty sampling methods**:
  - Maximum misclassification probability [Bryan et al. (2005)]:

$$U_{n+1} \in \underset{u \in \mathbb{U}}{\operatorname{argmax}} \min(p_n(u), 1 - p_n(u))$$

  - [Ranjan et al. (2008); Echard et al. (2011), ...]
- **Stepwise uncertainty reduction (SUR)** methods:
  - For instance [Chevalier et al. (2014)] (Joint-SUR):

$$U_{n+1} \in \underset{u \in \mathbb{U}}{\operatorname{argmin}} \mathbb{E}_n(\mathcal{H}_{n+1} \mid U_{n+1} = u)$$

with $\mathcal{H}_n = \int_{\mathbb{U}} \min(p_n(u), 1 - p_n(u)) \, \mathrm{d}u$.

  - [Picheny et al. (2010); Marques et al. (2018), ...]

Several Bayesian methods focus on **estimating** $\Lambda(f)$. For example:

- **Maximal uncertainty sampling methods**:
  - Maximum misclassification probability [Bryan et al. (2005)]:

$$U_{n+1} \in \underset{u \in \mathbb{U}}{\operatorname{argmax}} \min(p_n(u), 1 - p_n(u))$$

  - [Ranjan et al. (2008); Echard et al. (2011), ...]
- **Stepwise uncertainty reduction (SUR)** methods:
  - For instance [Chevalier et al. (2014)] (Joint-SUR):

$$U_{n+1} \in \underset{u \in \mathbb{U}}{\operatorname{argmin}} \mathbb{E}_n(\mathcal{H}_{n+1} \mid U_{n+1} = u)$$

    with $\mathcal{H}_n = \int_{\mathbb{U}} \min(p_n(u), 1 - p_n(u)) \, \mathrm{d}u$.
  - [Picheny et al. (2010); Marques et al. (2018), ...]

Several Bayesian methods focus on **estimating $\Lambda(f)$**. For example:

- **Maximal uncertainty sampling methods**:
  - Maximum misclassification probability [Bryan et al. (2005)]:

$$U_{n+1} \in \operatorname*{argmax}_{u \in \mathbb{U}} \min(p_n(u), 1 - p_n(u))$$

  - [Ranjan et al. (2008); Echard et al. (2011), ...]
- **Stepwise uncertainty reduction (SUR)** methods:
  - For instance [Chevalier et al. (2014)] (Joint-SUR):

$$U_{n+1} \in \operatorname*{argmin}_{u \in \mathbb{U}} \mathbb{E}_n(\mathcal{H}_{n+1} \mid U_{n+1} = u)$$

with $\mathcal{H}_n = \int_{\mathbb{U}} \min(p_n(u), 1 - p_n(u)) \, \mathrm{d}u$.

  - [Picheny et al. (2010); Marques et al. (2018), ...]

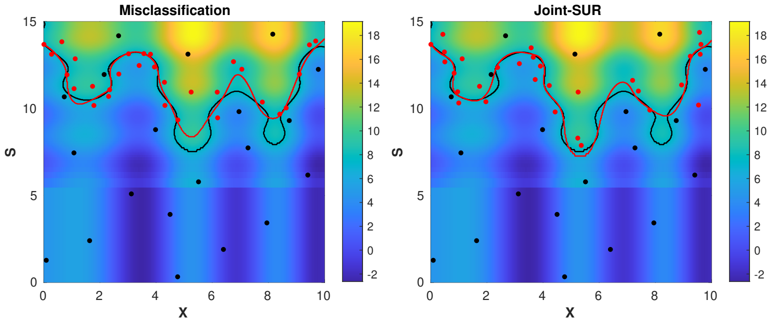


Figure: Examples of designs (red dots) obtained after $n = 30$ steps with the maximum misclassification and the 'joint-SUR' criteria.

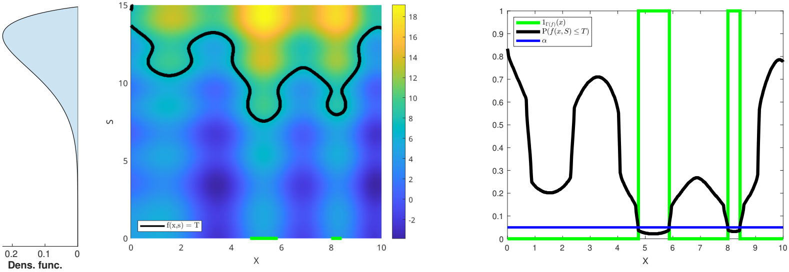


Figure: Representation of the function (middle), the density of $\mathbb{P}_S$ (left) and associated quantile set (right).

Second approach: Focusing directly on $\Gamma(f)$

To estimate $\Gamma(f)$, one only needs to focus on **'interesting parts'** of $\Lambda(f)$.

We denote:

- $\Gamma(\xi)$ the random quantile set associated to ξ.
- $\pi_n(x) = \mathbb{P}_n(x \in \Gamma(\xi))$, the posterior probability that x belongs to the (random) quantile set generated by ξ.
- $\mathcal{Q}_n = \int_{\mathbb{X}} \min(\pi_n(x), 1 - \pi_n(x)) \, \mathrm{d}x$.

QSI-SUR sampling criterion [Ait Abdelmalek-Lomenech et al. (2024)]:

$$(X_{n+1}, S_{n+1}) \in \underset{(x,s) \in \mathbb{X} \times \mathbb{S}}{\operatorname{argmin}} \mathbb{E}_n(\mathcal{Q}_{n+1} \mid (X_{n+1}, S_{n+1}) = (x, s)),$$

Second approach: Focusing directly on $\Gamma(f)$

To estimate $\Gamma(f)$, one only needs to focus on **'interesting parts'** of $\Lambda(f)$.

We denote:

- $\Gamma(\xi)$ the random quantile set associated to ξ.
- $\pi_n(x) = \mathbb{P}_n(x \in \Gamma(\xi))$, the posterior probability that x belongs to the (random) quantile set generated by ξ.
- $\mathcal{Q}_n = \int_{\mathbb{X}} \min(\pi_n(x), 1 - \pi_n(x))\,\mathrm{d}x$.

QSI-SUR sampling criterion [Ait Abdelmalek-Lomenech et al. (2024)]:

$$(X_{n+1}, S_{n+1}) \in \underset{(x,s)\in\mathbb{X}\times\mathbb{S}}{\operatorname{argmin}} \mathbb{E}_n(\mathcal{Q}_{n+1} \mid (X_{n+1}, S_{n+1}) = (x, s)),$$

The implementation proposed in [Ait Abdelmalek-Lomenech et al. (2024)] produces good results on moderately difficult examples.

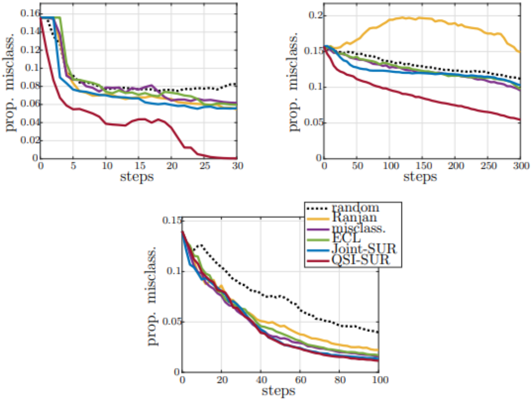


Figure: Median of the proportion of misclassified points vs. number of steps on several examples.

The QSI-SUR criterion focus on part of $\Lambda(f)$ that gives relevant information on $\Gamma(f)$.

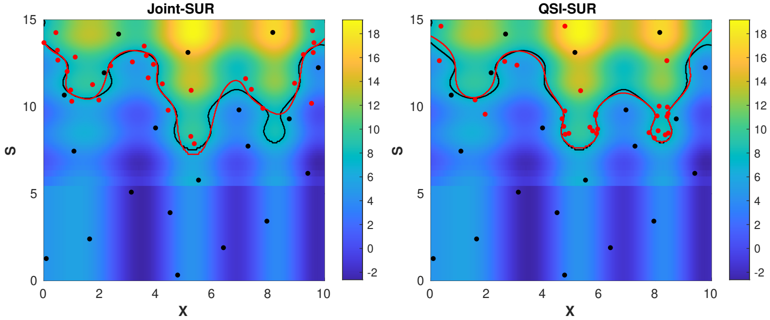


Figure: Example of design obtained with the QSI-SUR criterion and the Joint-SUR criterion.

The QSI-SUR criterion focus on part of $\Lambda(f)$ that gives relevant information on $\Gamma(f)$.

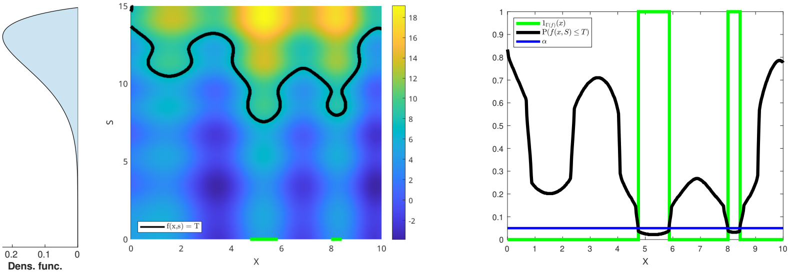


Figure: Representation of the function (middle), the density of $\mathbb{P}_S$ (left) and associated quantile set (right).

The QSI-SUR criterion is based on

$$\mathcal{Q}_n = \int_{\mathbb{X}} \min(\pi_n(x), 1 - \pi_n(x))\, \mathrm{d}x.$$

Two main issues in the implementation:

- **First issue: Computational complexity.**
  - π_n approximated using conditional sample paths of $\xi(x, \cdot)$.
  - Complexity is $O(m^3)$, where m is the number of points used for the approximation. Due to the Cholesky factorization of the covariance matrix.
  - Criterion too expensive for continouous optimization and batch design.
- **Second issue: Not adapted to "small" $\Gamma(f)$.**
  - Integral over $\mathbb{X}$ in $\mathcal{Q}_n$ is discretized.
  - Necessity of points (in $\mathbb{X}$) close to the boundary of $\Gamma(f)$
  - When $\Gamma(f)$ is small, importance sampling can prove insufficient.

The QSI-SUR criterion is based on

$$\mathcal{Q}_n = \int_{\mathbb{X}} \min(\pi_n(x), 1 - \pi_n(x)) \, \mathrm{d}x.$$

Two main issues in the implementation:

- **First issue: Computational complexity.**
  - π_n approximated using conditional sample paths of $\xi(x, \cdot)$.
  - Complexity is $O(m^3)$, where m is the number of points used for the approximation. Due to the Cholesky factorization of the covariance matrix.
  - Criterion too expensive for continouous optimization and batch design.
- **Second issue: Not adapted to "small" $\Gamma(f)$.**
  - Integral over $\mathbb{X}$ in $\mathcal{Q}_n$ is discretized.
  - Necessity of points (in $\mathbb{X}$) close to the boundary of $\Gamma(f)$
  - When $\Gamma(f)$ is small, importance sampling can prove insufficient.

Table of Contents

To resolve the issues listed previously, we use a two part solution:

- **First issue: Computational complexity**
  - We introduce a new type of method called "Maximum Expected Estimator Modification" (MEEM).
  - We derive a MEEM criterion with complexity $O(m^2)$.
  - This criterion allows continuous optimization and batch design of experiments.
- **Second issue: Not adapted to "small" $\Gamma(f)$.**
  - We introduce a sequential Monte Carlo (SMC) framework.
  - We estimate a sequence of decreasing quantile sets, converging towards the set of interest.

Maximum Expected Estimator Modification (MEEM)

Let us consider a sequence of estimators $(\widehat{\Gamma}_n)_n$ such that

$$\widehat{\Gamma}_n : \mathcal{I}_n \mapsto \mathcal{P}(\mathbb{X}),$$

and a "distance"

$$d : \mathcal{P}(\mathbb{X})^2 \mapsto \mathbb{R}^+.$$

MEEM principle: Choose the point that maximize the expected change in the estimation, i.e.

$$U_{n+1} \in \arg\max_{u \in \mathbb{U}} \mathbb{E}_n(d(\widehat{\Gamma}_{n+1}, \widehat{\Gamma}_n) \,|\, U_{n+1} = u)$$

NB: For convenience, we assume a batch size of 1.

Maximum Expected Estimator Modification (MEEM)

Let us consider a sequence of estimators $(\widehat{\Gamma}_n)_n$ such that

$$\widehat{\Gamma}_n : \mathcal{I}_n \mapsto \mathcal{P}(\mathbb{X}),$$

and a "distance"

$$d : \mathcal{P}(\mathbb{X})^2 \mapsto \mathbb{R}^+.$$

MEEM principle: Choose the point that maximize the expected change in the estimation, i.e.

$$U_{n+1} \in \underset{u \in \mathbb{U}}{\arg\max}\, \mathbb{E}_n(d(\widehat{\Gamma}_{n+1}, \widehat{\Gamma}_n) \,|\, U_{n+1} = u)$$

NB: For convenience, we assume a batch size of 1.

"Duality" SUR / MEEM

Several SUR strategies in the litterature are equivalent to MEEM strategies.

For example:

- Optimization:
  - Expected Improvement
  - MEEM method with $d(a, b) = |a - b|$ and estimator $f_n^* = \min\{U_1, ..., U_n\}$
- Function approximation:
  - SUR method : $U_{n+1} \in \arg\min \mathbb{E}_n \left(\int_{\mathbb{U}} \sigma_n^2(u) \, \mathrm{d}u \,|\, U_{n+1} = u \right)$.
  - MEEM method with $d(h, g) = \int_{\mathbb{U}} (h(u) - g(u))^2 \, \mathrm{d}u$ and estimator $f_n = \mu_n$

"Duality" SUR / MEEM

Several SUR strategies in the litterature are equivalent to MEEM strategies.

For example:

- Optimization:
  - Expected Improvement
  - MEEM method with $d(a, b) = |a - b|$ and estimator $f_n^* = \min\{U_1, ..., U_n\}$
- Function approximation:
  - SUR method : $U_{n+1} \in \arg\min \mathbb{E}_n \left(\int_{\mathbb{U}} \sigma_n^2(u) \, \mathrm{d}u \mid U_{n+1} = u \right)$.
  - MEEM method with $d(h, g) = \int_{\mathbb{U}} (h(u) - g(u))^2 \, \mathrm{d}u$ and estimator $f_n = \mu_n$

MEEM method for QSI

We choose the divergence $d(\widehat{\Gamma}_{n+1}, \widehat{\Gamma}_n) = \lambda(\widehat{\Gamma}_{n+1} \Delta \widehat{\Gamma}_n)$ we have

$$d(\widehat{\Gamma}_{n+1}, \widehat{\Gamma}_n) = \int_{\mathbb{X}} |\mathbb{1}_{\widehat{\Gamma}_{n+1}}(x) - \mathbb{1}_{\widehat{\Gamma}_n}(x)| \, \mathrm{d}x,$$

coupled with the sequence of plug-in estimators

$$\widehat{\Gamma}_n = \{x \in \mathbb{X} : \mathbb{P}(\mu_n(x, S) \leq T) \leq \alpha\},$$

We obtain the **QSI-MEEM** strategy:

$$U_{n+1} \in \underset{u \in \mathbb{U}}{\arg\max} \int_{\mathbb{X}} \mathbb{E}_n \left(|\mathbb{1}_{\widehat{\Gamma}_{n+1}}(x) - \mathbb{1}_{\widehat{\Gamma}_n}(x)| \,\middle|\, U_{n+1} = u \right) \mathrm{d}x.$$

MEEM method for QSI

We choose the divergence $d(\widehat{\Gamma}_{n+1}, \widehat{\Gamma}_n) = \lambda(\widehat{\Gamma}_{n+1} \Delta \widehat{\Gamma}_n)$ we have

$$d(\widehat{\Gamma}_{n+1}, \widehat{\Gamma}_n) = \int_{\mathbb{X}} |\mathbb{1}_{\widehat{\Gamma}_{n+1}}(x) - \mathbb{1}_{\widehat{\Gamma}_n}(x)| \, \mathrm{d}x,$$

coupled with the sequence of plug-in estimators

$$\widehat{\Gamma}_n = \{x \in \mathbb{X} \, : \, \mathbb{P}(\mu_n(x, S) \leq T) \leq \alpha\},$$

We obtain the **QSI-MEEM** strategy:

$$U_{n+1} \in \underset{u \in \mathbb{U}}{\arg\max} \int_{\mathbb{X}} \mathbb{E}_n \left(|\mathbb{1}_{\widehat{\Gamma}_{n+1}}(x) - \mathbb{1}_{\widehat{\Gamma}_n}(x)| \, \middle| \, U_{n+1} = u \right) \mathrm{d}x.$$

Criterion does not need to be approximated using conditional sample paths of ξ.

As a consequence of the kriging update formula, we have:

Proposition

Given $\mathcal{I}_n$ and U_{n+1}, $\widehat{\Gamma}_{n+1}$ is a function of a standard Gaussian variable Z:

$$\widehat{\Gamma}_{n+1}(z) = \{x \in \mathbb{X} : P\left(\mu_n(x, S) + \kappa_n(x, S) z \leq T\right) \leq \alpha\},$$

with $\kappa_n(x, s) = k_n(U_{n+1}, (x, s))/\sigma_n(U_{n+1})$.

$\Longrightarrow$ Computational complexity of the criterion: $O(m^2)$, with m the number of points (in $\mathbb{X} \times \mathbb{S}$) used for the approximation.

Criterion does not need to be approximated using conditional sample paths of ξ.

As a consequence of the kriging update formula, we have:

Proposition

Given $\mathcal{I}_n$ and U_{n+1}, $\widehat{\Gamma}_{n+1}$ is a function of a standard Gaussian variable Z:

$$\widehat{\Gamma}_{n+1}(z) = \{x \in \mathbb{X} : P\left(\mu_n(x, S) + \kappa_n(x, S)z \leq T\right) \leq \alpha\},$$

with $\kappa_n(x, s) = k_n(U_{n+1}, (x, s))/\sigma_n(U_{n+1})$.

$\Longrightarrow$ Computational complexity of the criterion: $O(m^2)$, with m the number of points (in $\mathbb{X} \times \mathbb{S}$) used for the approximation.

Criterion does not need to be approximated using conditional sample paths of ξ.

As a consequence of the kriging update formula, we have:

Proposition

Given $\mathcal{I}_n$ and U_{n+1}, $\widehat{\Gamma}_{n+1}$ is a function of a standard Gaussian variable Z:

$$\widehat{\Gamma}_{n+1}(z) = \{x \in \mathbb{X} : P\left(\mu_n(x, S) + \kappa_n(x, S) z \leq T\right) \leq \alpha\},$$

with $\kappa_n(x, s) = k_n(U_{n+1}, (x, s))/\sigma_n(U_{n+1})$.

$\implies$ Computational complexity of the criterion: $O(m^2)$, with m the number of points (in $\mathbb{X} \times \mathbb{S}$) used for the approximation.

Sanity check

We observe that the QSI-MEEM method produces results similar (or better) than the QSI-SUR strategy on case with relatively large quantile sets.

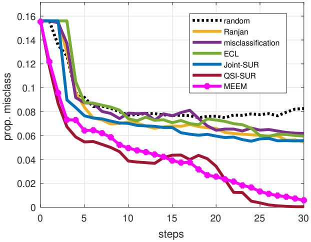


Figure: Median (left) and quantile of order 0.9 of the proportion of misclassified points vs. number of steps, for 100 repetitions of the algorithms on the introductory example.

Second issue: Estimation of small quantile sets

Idea: Multilevel splitting/**subset simulation** [Kahn and Harris (1951); Au and Beck (2001)] to efficiently sample points in $\mathbb{X}$.

- Sequentially estimate a sequence of **decreasing quantile sets**

$$\Gamma^0(f) \supset \Gamma^1(f) \supset \ ... \ \supset \Gamma^K(f) = \Gamma(f),$$

using the MEEM strategy described previously.

- Such sets can be defined by setting

$$\Gamma^k(f) = \{x \in \mathbb{X} \, : \, \mathbb{P}(f(x, S) \leq T_k) \leq \alpha\},$$

with $T_k \leq T_{k+1}$.

Second issue: Estimation of small quantile sets

Idea: Multilevel splitting/**subset simulation** [Kahn and Harris (1951); Au and Beck (2001)] to efficiently sample points in $\mathbb{X}$.

- Sequentially estimate a sequence of **decreasing quantile sets**

$$\Gamma^0(f) \supset \Gamma^1(f) \supset \ldots \supset \Gamma^K(f) = \Gamma(f),$$

using the MEEM strategy described previously.

- Such sets can be defined by setting

$$\Gamma^k(f) = \{x \in \mathbb{X} : \mathbb{P}(f(x, S) \leq T_k) \leq \alpha\},$$

with $T_k \leq T_{k+1}$.

We propose a **SMC-based** algorithm inspired by **BSS** [Li (2012); Bect et al. (2017)]

It alternates two distinct phases:

- **Estimation phase**
  - Define a new intermediary quantile set to estimate.
  - Sample points $U_n, \ldots, U_{n+r}$ using the MEEM criterion.
- **Move phase**
  - Concentrate the particles towards the previously estimated set.

For simplicity, we still assume $C = (-\infty, T]$ and a batch size of 1.

We propose a **SMC-based** algorithm inspired by **BSS** [Li (2012); Bect et al. (2017)]

It alternates two distinct phases:

- **Estimation phase**
  - Define a new intermediary quantile set to estimate.
  - Sample points $U_n, ..., U_{n+r}$ using the MEEM criterion.
- **Move phase**
  - Concentrate the particles towards the previously estimated set.

For simplicity, we still assume $C = (-\infty, T]$ and a batch size of 1.

Let $q_{n,k}$ a density targeting $\Gamma^k(f) = \{x \in \mathbb{X} : \mathbb{P}(f(x,S) \in C_k) \leq \alpha\}$ at step n.

Estimation phase:

- Set T_{k+1} such that

$$\mathrm{ESS}\left(\frac{\mathbb{1}_{\widehat{\Gamma}_n^{k+1}}(x)}{\mathbb{1}_{\widehat{\Gamma}_n^{k}}(x)}\right) \approx 30\%.$$

- Sample point

$$U_{n+1} \in \operatorname{argmax} J_n(u),$$

with J_n the MEEM criterion targeting $\Gamma^{k+1}(f)$.

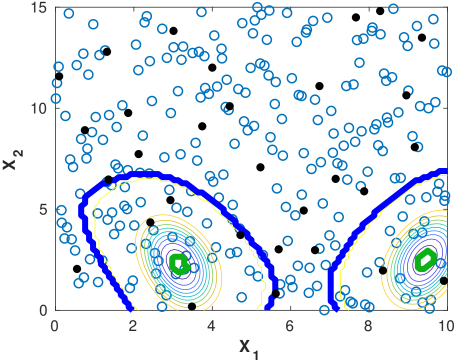


Figure: Temporary quantile set (blue line), final quantile set (green line), particles (blue dots). - $n = 0$.

Let $q_{n,k}$ a density targeting $\Gamma^k(f) = \{x \in \mathbb{X} : \mathbb{P}(f(x,S) \in C_k) \leq \alpha\}$ at step n.

Estimation phase:

- Set T_{k+1} such that

$$\mathrm{ESS}\left(\frac{\mathbb{1}_{\widehat{\Gamma}_n^{k+1}}(x)}{\mathbb{1}_{\widehat{\Gamma}_n^{k}}(x)}\right) \approx 30\%.$$

- Sample point

$$U_{n+1} \in \operatorname{argmax} J_n(u),$$

with J_n the MEEM criterion targeting $\Gamma^{k+1}(f)$.

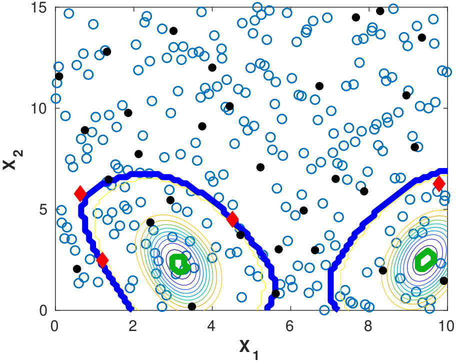


Figure: Temporary quantile set (blue line), final quantile set (green line), particles (blue dots) and projection of the sequential design (red dots). - $n = 4$.

Let $q_{n,k}$ a density targeting $\Gamma^k(f) = \{x \in \mathbb{X} : \mathbb{P}(f(x,S) \leq T_k) \leq \alpha\}$ at step n.

Move phase:

When stopping condition is met:

- Residual resampling.
- Move particles in $\Gamma^{k+1}(f)$ using MHRW with target density $q_{n,k+1}$.
- Adapt walk's variance to target acceptation rate 25%.

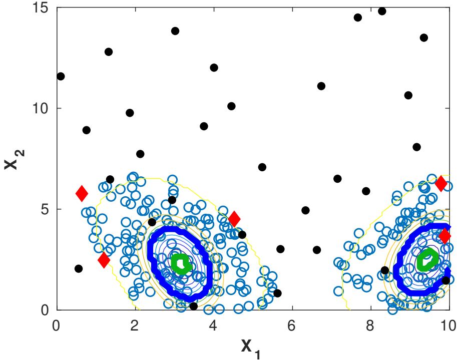


Figure: Temporary quantile set (blue line), final quantile set (green line), particles (blue dots) and projection of the sequential design (red dots). - $n = 5$.

Table of Contents

To evaluate the performances of the proposed strategy, we focus on the relative error

$$\frac{\lambda_{\mathbb{X}}(\Gamma(f)\Delta\widehat{\Gamma}_n)}{\lambda_{\mathbb{X}}(\Gamma(f))}$$

obtained at the end of the strategy, and the number of steps required to obtain these results.

The functions considered are modeled by a GP with constant mean and Matern covariance kernel (with $\nu \in \{1/2, 3/2, 5/2, +\infty\}$).

Covariance parameters are estimated at each step using ReML.

Each strategy is repeated 50 times, with different maximin LHS initial designs.

To evaluate the performances of the proposed strategy, we focus on the relative error

$$\frac{\lambda_{\mathbb{X}}(\Gamma(f)\Delta\widehat{\Gamma}_n)}{\lambda_{\mathbb{X}}(\Gamma(f))}$$

obtained at the end of the strategy, and the number of steps required to obtain these results.

The functions considered are modeled by a GP with constant mean and Matern covariance kernel (with $\nu \in \{1/2, 3/2, 5/2, +\infty\}$).

Covariance parameters are estimated at each step using ReML.

Each strategy is repeated 50 times, with different maximin LHS initial designs.

1st example: 5-Trid function [Adorio and U.P. (2005)]

- $\mathbb{U} = \mathbb{X} \times \mathbb{S}$ with $\mathbb{X} = [-25, 25]^3$ and $\mathbb{S} = [-25, 25]^2$,
- $\mathbb{P}_S = \mathcal{U}(\mathbb{S})$
- $\alpha = 5\%$ and $C = [1098.5, +\infty)$
- Relative size of $\Gamma(f)$: $\lambda_{\mathbb{X}}(\Gamma(f)) \sim 10^{-6}$.

$$f(u) = \sum_{i=1}^{5} (u_i - 1)^2 - \sum_{i=2}^{5} u_i u_{i-1}$$

Size of the initial design: 50.

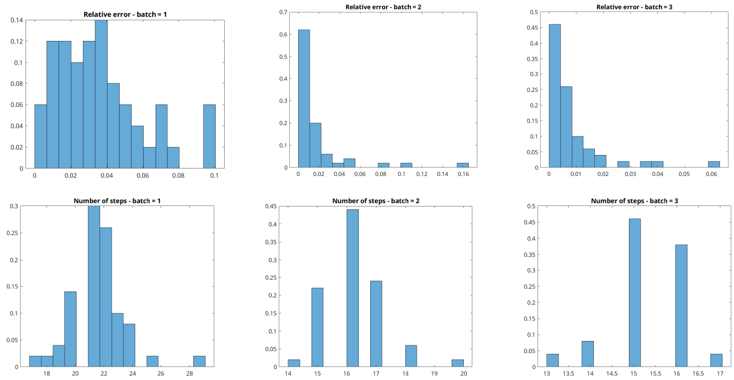


Figure: Distribution of the relative error (top) and the number of steps (bottom), for different batch size (left to right: 1, 2, 3). - (50 runs)

2nd example: OTL circuit function [E. N. and D. M. (2007)]

- $\mathbb{U} = \mathbb{X} \times \mathbb{S}$
- $\mathbb{X} = [-50, 150] \times [25, 70] \times [0.5, 3] \times [1.2, 2.5] \times [0.25, 1.2]$
- $\mathbb{S} = [-50, 300]$,
- $\mathbb{P}_S = trunc\mathcal{N}(175, 50)$
- $\alpha = 5\%$ and $C = [2.65, +\infty)$
- Relative size of $\Gamma(f)$: $\lambda_{\mathbb{X}}(\Gamma(f)) \sim 10^{-7}$.

$f(x, s)$ represents the midpoint voltage of a circuit given the choice of resistances designs x and the current gain s.

Size of the initial design: 60.

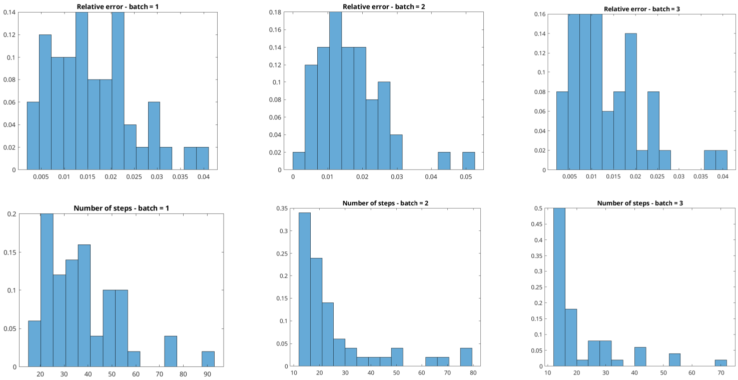


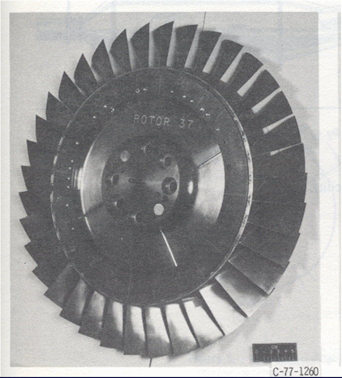
Figure: Distribution of the relative error (top) and the number of steps (bottom), for different batch size (left to right: 1, 2, 3). - (50 runs).

3rd example: ROTOR37 model [Reid and Moore (1978)]

Gaussian metamodel (provided by S. Da Veiga and SafranTech) of the ROTOR37 compressor model.

The function $f : \mathbb{X} \times \mathbb{S} \mapsto \mathbb{R}^3$ takes two kind of inputs:

- $x \in \mathbb{X} = [0, 1]^{13}$: design choice for the compressor
- $s \in \mathbb{S} = [0, 1]^5$: manufacturing uncertainties ($\mathbb{P}_S = \mathcal{U}(\mathbb{S})$)



Simulator returns three outputs:

- f_1: the mass flow
- f_2: the pressure ratio
- f_3: the isentropic efficiency

Goal: finding the set of determinisitic design choice leading to values of the mass flow and pressure ratio being close to baselines values (b_1, b_2) with sufficiently high probability.

We consider:

- $C = \left\{ z \in \mathbb{R}^3 : \frac{|z_1 - b_1|}{|b_1|} > 0.175 \quad \text{or} \quad \frac{|z_2 - b_2|}{|b_2|} > 0.175 \right\}$
- $\alpha = 5\%$

Relative size of $\Gamma(f)$: $\lambda_{\mathbb{X}}(\Gamma(f)) \sim 10^{-8}$

We model the different outputs using independant GPs.

Simulator returns three outputs:

- f_1: the mass flow
- f_2: the pressure ratio
- f_3: the isentropic efficiency

Goal: finding the set of determinisitic design choice leading to values of the mass flow and pressure ratio being close to baselines values (b_1, b_2) with sufficiently high probability.

We consider:

- $C = \left\{ z \in \mathbb{R}^3 : \frac{|z_1 - b_1|}{|b_1|} > 0.175 \quad \text{or} \quad \frac{|z_2 - b_2|}{|b_2|} > 0.175 \right\}$
- $\alpha = 5\%$

Relative size of $\Gamma(f)$: $\lambda_{\mathbb{X}}(\Gamma(f)) \sim 10^{-8}$

We model the different outputs using independant GPs.

Simulator returns three outputs:

- f_1: the mass flow
- f_2: the pressure ratio
- f_3: the isentropic efficiency

Goal: finding the set of determinisitic design choice leading to values of the mass flow and pressure ratio being close to baselines values (b_1, b_2) with sufficiently high probability.

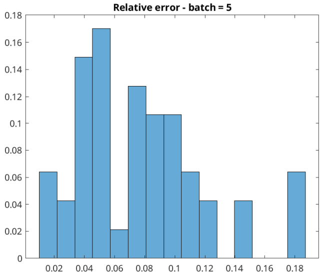
We consider:

- $C = \left\{ z \in \mathbb{R}^3 : \frac{|z_1 - b_1|}{|b_1|} > 0.175 \quad \text{or} \quad \frac{|z_2 - b_2|}{|b_2|} > 0.175 \right\}$
- $\alpha = 5\%$

Relative size of $\Gamma(f)$: $\lambda_{\mathbb{X}}(\Gamma(f)) \sim 10^{-8}$

We model the different outputs using independant GPs.

We start the strategy from an initial design of size 90. A batch size of 5 is used.



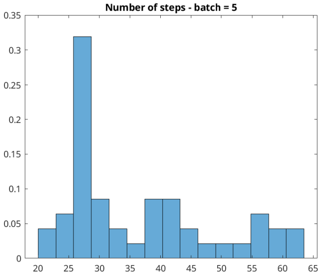


Figure: Distribution of the relative error (left) and the number of steps (right), for a batch size of 5. - (50 runs).

Conclusion:

- Introducing the concept of MEEM allows to reproduce (or improve) the results obtained by using the QSI-SUR criterion on moderately difficult examples.
- The MEEM criterion permits a welcomed gain regarding the computational complexity, due to the absence of conditional Gaussian sample paths
- Coupled with a SMC framework, the criterion allows to accurately estimate small quantile sets (size of order $10^{-6} - 10^{-8}$).

Future research might be dedicated to treating cases where $\alpha \sim 0$.

Conclusion:

- Introducing the concept of MEEM allows to reproduce (or improve) the results obtained by using the QSI-SUR criterion on moderately difficult examples.
- The MEEM criterion permits a welcomed gain regarding the computational complexity, due to the absence of conditional Gaussian sample paths
- Coupled with a SMC framework, the criterion allows to accurately estimate small quantile sets (size of order $10^{-6} - 10^{-8}$).

Future research might be dedicated to treating cases where $\alpha \sim 0$.

Thank you for your attention!

First part (QSI-SUR criterion) is based on:

R. Ait Abdelmalek-Lomenech, J. Bect, V. Chabridon and E. Vazquez. *Bayesian Sequential Design of Experiments for Quantile Set Inversion*. 2024. Technometrics (to appear).

The second part (QSI-MEEM criterion) is based on:

R. Ait Abdelmalek-Lomenech, J. Bect and E. Vazquez. *Active Learning of (small) Quantile Sets through Expected Estimator Modification.* 2025. ArXiv preprint (to appear).

This work has been funded by the French National Research Agency (ANR) in the context of the project SAMOURAI (ANR-20-CE46-0013).
The authors are thankfull to S. Da Veiga and Safran Tech for providing the ROTOR37 metamodel used as application.

Thank you for your attention!

First part (QSI-SUR criterion) is based on:

R. Ait Abdelmalek-Lomenech, J. Bect, V. Chabridon and E. Vazquez. *Bayesian Sequential Design of Experiments for Quantile Set Inversion*. 2024. Technometrics (to appear).

The second part (QSI-MEEM criterion) is based on:

R. Ait Abdelmalek-Lomenech, J. Bect and E. Vazquez. *Active Learning of (small) Quantile Sets through Expected Estimator Modification.* 2025. ArXiv preprint (to appear).

This work has been funded by the French National Research Agency (ANR) in the context of the project SAMOURAI (ANR-20-CE46-0013).
The authors are thankfull to S. Da Veiga and Safran Tech for providing the ROTOR37 metamodel used as application.

Thank you for your attention!

First part (QSI-SUR criterion) is based on:

R. Ait Abdelmalek-Lomenech, J. Bect, V. Chabridon and E. Vazquez. *Bayesian Sequential Design of Experiments for Quantile Set Inversion*. 2024. Technometrics (to appear).

The second part (QSI-MEEM criterion) is based on:

R. Ait Abdelmalek-Lomenech, J. Bect and E. Vazquez. *Active Learning of (small) Quantile Sets through Expected Estimator Modification.* 2025. ArXiv preprint (to appear).

This work has been funded by the French National Research Agency (ANR) in the context of the project SAMOURAI (ANR-20-CE46-0013).
The authors are thankfull to S. Da Veiga and Safran Tech for providing the ROTOR37 metamodel used as application.

References

Adorio, E. P. and U.P., D. (2005). Mvf - multivariate test functions library in c for unconstrained global optimization.

Ait Abdelmalek-Lomenech, R., Bect, J., Chabridon, V., and Vazquez, E. (2024). Bayesian sequential design of computer experiments for quantile set inversion. *Technometrics*, 0(0):1–10.

Au, S. and Beck, J. L. (2001). Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, 16:263–277.

Bect, J., Li, L., and Vazquez, E. (2017). Bayesian Subset Simulations. *SIAM/ASA Journal on Uncertainty Quantification*, 5:762–786.

Bryan, B., Nichol, R. C., Genovese, C. R., Schneider, J., Miller, C. J., and Wasserman, L. (2005). Active learning for identifying function threshold boundaries. In Weiss, Y., Schölkopf, B., and Platt, J., editors, *Advances in Neural Information Processing Systems*, volume 18. MIT Press.

Chevalier, C., Bect, J., Ginsbourger, D., Vazquez, E., Picheny, V., and Richet, Y. (2014). Fast parallel kriging-based stepwise uncertainty reduction with application to the identification of an excursion set. *Technometrics*, 56(4):455–465.

Dubourg, V., Sudret, B., and Deheeger, F. (2013). Metamodel-based importance sampling for structural reliability analysis. *Probabilistic Engineering Mechanics*, 33:47–57.

E. N., B.-A. and D. M., S. (2007). Modeling data from computer experiments: an empirical comparison of kriging with mars and projection pursuit regression. *Quality Engineering*, 19(4):327–338.

Echard, B., Gayton, N., and Lemaire, M. (2011). AK-MCS: An active learning reliability method combining Kriging and Monte Carlo Simulation. *Structural Safety*, 33(2):145–154.

References (cont.)

Kahn, H. and Harris, T. E. (1951). Estimation of particle transmission by random sampling. *National Bureau of Standards applied mathematics series*, 12:27–30.

Li, L. (2012). *Sequential Design of Experiments to Estimate a Probability of Failure.* PhD thesis.

Marques, A., Lam, R., and Willcox, K. (2018). Contour location via entropy reduction leveraging multiple information sources. In *Advances in Neural Information Processing Systems 31 (NeurIPS 2018)*, pages 1–11.

Picheny, V., Ginsbourger, D., Roustant, O., Haftka, R. T., and Kim, N.-H. (2010). Adaptive designs of experiments for accurate approximation of a target region. *Journal of Mechanical Design*, 132(7):071008.

Ranjan, P., Bingham, D., and Michailidis, G. (2008). Sequential experiment design for contour estimation from complex computer codes. *Technometrics*, 50(4):527–541.

Reid, L. and Moore, R. D. (1978). Design and overall performance of four highly loaded, high speed inlet stages for an advanced high-pressure-ratio core compressor. Technical report, NASA Lewis Research Center Cleveland, OH, United States.

Choice of the target densities:

Natural idea (in the spirit of [Dubourg et al. (2013); Bect et al. (2017)]):

$$q_{n,k}(x) \propto \pi_n^k(x) = \mathbb{P}_n(x \in \Gamma^k(\xi))$$

- Does not admit a closed-form expression.
- Expensive to estimate.

Idea: Replace $\pi_n^k(x)$ by $\mathbb{1}(x \in \Gamma_{n,k}^+)$. How to define $\Gamma_{n,k}^+$?

Given $x_0 \in \mathbb{X}$, μ_n and σ_n the posterior mean and standard deviation of ξ and $\beta \in (1/2, 1)$, consider the **quantile function**:

$$\xi_n^+(x_0, \cdot) = \mu_n(x_0, \cdot) + \Phi^{-1}(\beta)\sigma_n(x_0, \cdot),$$

Choice of the target densities:

Natural idea (in the spirit of [Dubourg et al. (2013); Bect et al. (2017)]):

$$q_{n,k}(x) \propto \pi_n^k(x) = \mathbb{P}_n(x \in \Gamma^k(\xi))$$

- Does not admit a closed-form expression.
- Expensive to estimate.

Idea: Replace $\pi_n^k(x)$ by $\mathbb{1}(x \in \Gamma_{n,k}^+)$. How to define $\Gamma_{n,k}^+$?

Given $x_0 \in \mathbb{X}$, μ_n and σ_n the posterior mean and standard deviation of ξ and $\beta \in (1/2, 1)$, consider the **quantile function**:

$$\xi_n^+(x_0, \cdot) = \mu_n(x_0, \cdot) + \Phi^{-1}(\beta)\sigma_n(x_0, \cdot),$$

$C = (-\infty, T]$ and $\xi(x_0, \cdot)$ is a **high quantile**

- $\mathbb{P}(\xi_n^+(x_0, S) \in C_k)$ is an **optimistic estimation of the probability of failure** at point x_0.

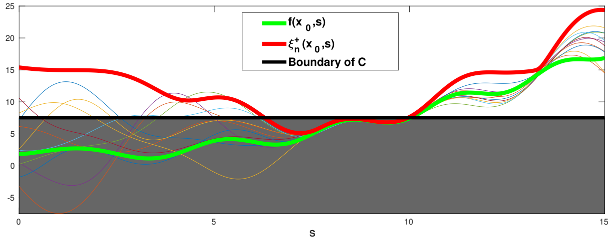


Figure: Example of quantile function $\xi_n^+(x_0, \cdot)$, with a fixed x_0.

Setting $\Gamma_{n,k}^+ = \Gamma_n^k(\xi_n^+)$ eliminates x_0 if $\{x_0 \in \Gamma^k(\xi)\}$ is **very improbable**.

We define the target densities as

$$q_{n,k}(x) \propto \mathbb{1}(x \in \Gamma_n^k(\xi_n^+))$$

NB: The MHRW step becomes a constrained random walk.

$C = (-\infty, T]$ and $\xi(x_0, \cdot)$ is a **high quantile**

- $\mathbb{P}(\xi_n^+(x_0, S) \in C_k)$ is an **optimistic estimation of the probability of failure** at point x_0.

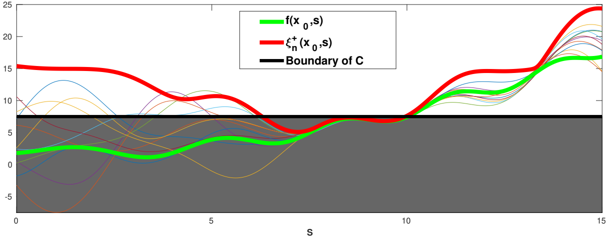


Figure: Example of quantile function $\xi_n^+(x_0, \cdot)$, with a fixed x_0.

Setting $\Gamma_{n,k}^+ = \Gamma_n^k(\xi_n^+)$ eliminates x_0 if $\{x_0 \in \Gamma^k(\xi)\}$ is **very improbable**.

We define the target densities as

$$q_{n,k}(x) \propto \mathbb{1}(x \in \Gamma_n^k(\xi_n^+))$$

NB: The MHRW step becomes a constrained random walk.

$C = (-\infty, T]$ and $\xi(x_0, \cdot)$ is a **high quantile**

- $\mathbb{P}(\xi_n^+(x_0, S) \in C_k)$ is an **optimistic estimation of the probability of failure** at point x_0.

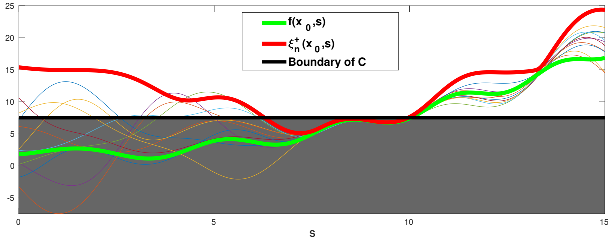


Figure: Example of quantile function $\xi_n^+(x_0, \cdot)$, with a fixed x_0.

Setting $\Gamma_{n,k}^+ = \Gamma_n^k(\xi_n^+)$ eliminates x_0 if $\{x_0 \in \Gamma^k(\xi)\}$ is **very improbable**.

We define the target densities as

$$q_{n,k}(x) \propto \mathbb{1}(x \in \Gamma_n^k(\xi_n^+))$$

NB: The MHRW step becomes a constrained random walk.