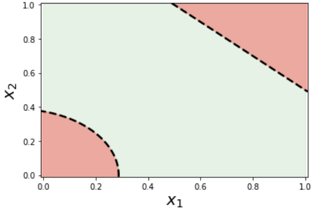
Learning hidden constraints using a Stepwise Uncertainty Reduction strategy with Gaussian Process Classifiers

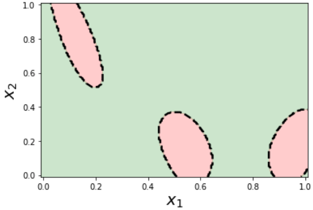
Morgane Menz, Miguel Munoz Zuniga, Delphine Sinoquet

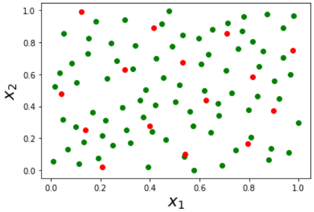
IFP Energies nouvelles
Department of Applied Mathematics - Digital Science and Technology Division
Paris, France

Problem statement: hidden constraints

Context: Expensive simulator outputs evaluation
➜ simulations crash: hidden constraints

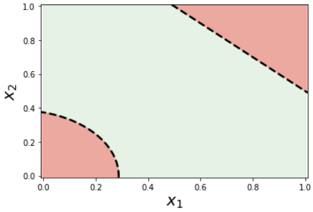


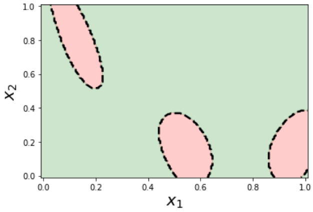


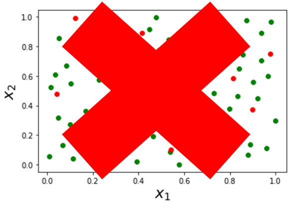


Problem statement: hidden constraints

Context: Expensive simulator outputs evaluation
➜ simulations crash: hidden constraints







Problem statement: hidden constraints

f: a computer code with inputs $x \in \Omega \subset \mathbb{R}^m$ with simulation failures on Ω.

Objective: determination of the feasible set:

$$\Gamma^* = \{x \in \Omega : f(x) \neq NAN\} = \{x \in \Omega : \mathbb{1}_{f(x) \neq NAN} = 1\}$$

Gaussian process based classification

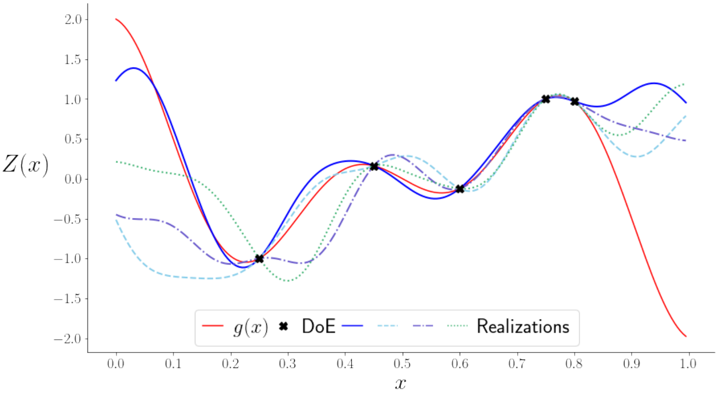
Learning hidden constraint is a binary classification problem:

- We have binary observations: $(\mathcal{X}, \mathcal{Y}) = (x_j, y_j)_{j=1,\dots,n}$, with $y_j = \mathbb{1}_{f(x_j) \neq NAN}$.
- Objective: predict the probability of belonging to the failure/non-failure class

➜ The formulation of the classification model is based on a Gaussian Process (GP) surrogate

Function approximation by Gaussian process model

Approximation of a function g by a GP $Z(x)$ conditioned on observations of g [Forrester et al., 2008, Rasmussen and Williams, 2006]

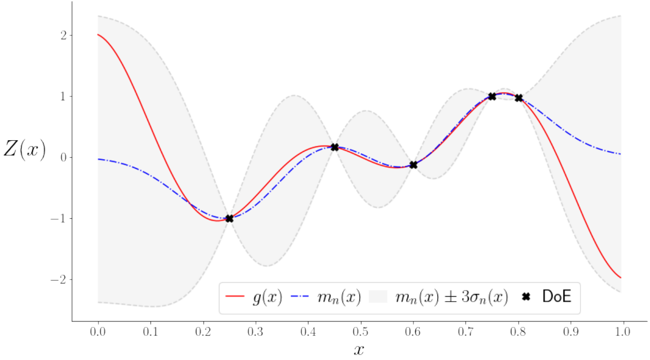


Function approximation by Gaussian process model

Approximation of a function g by a Gaussian process conditioned on observations of g defined by:

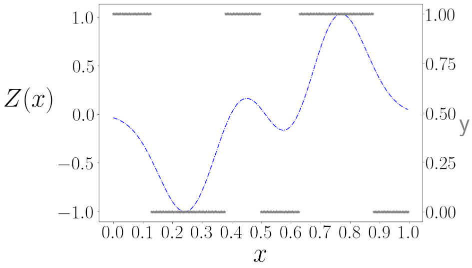
$$Z(x) \sim GP(m_n(\cdot), k_n(\cdot, \cdot))$$

$m_n(\cdot)$, $k_n(\cdot, \cdot)$ conditioned mean and kernel of $Z(x)$



Gaussian process classifier (GPC) formulation

A GPC is based on a latent GP Z conditioned on observations $\mathcal{X}, \mathcal{Y}$ (as $Z_n = (Z(x_1), ..., Z(x_n))$ is not available).



Gaussian process classifier (GPC) formulation

The GPC model allows to predict the probability of non failure of a simulation:

$$p_n(x) = \mathbb{P}[Y_n(x) = 1] = \mathbb{P}[Y(x) = 1 | \mathcal{X}, \mathcal{Y}]$$

This probability $p_n(x)$ is modeled on the basis of [Bachoc et al., 2020] by using the sign of the latent process Z:

$$p_n(x) = \mathbb{P}[\mathbb{1}_{Z(x)>0} = 1 | x, \mathcal{X}, \mathcal{Y}] = \int_{\mathbb{R}^n} \phi_{\mathcal{Y}}^{Z_n}(z_n) \bar{\Phi}\left(\frac{-m_n(x, z_n)}{\sqrt{k_n(x)}}\right) \mathrm{d}z_n$$

with $\phi_{\mathcal{Y}}^{Z_n}(z_n)$ the conditioned p.d.f of Z_n truncated to respect $sign(Z_n) = \mathcal{Y}$, and:

$$\bar{\Phi}\left(\frac{a}{b}\right) = \begin{cases} 1 - \Phi\left(\frac{a}{b}\right) & \text{si } b \neq 0 \\ \mathbb{1}_{-a>0} & \text{si } b = 0 \end{cases}$$

where Φ is the c.d.f. of the normal standard distribution.

Gaussian process classifier (GPC) formulation

Practical building of the GPC model $p_n(x)$ for any x:

- Optimization of the hyperparameters of the latent GP to maximize the likelihood: $\mathbb{P}[sign(Z_n) = \mathcal{Y}]$

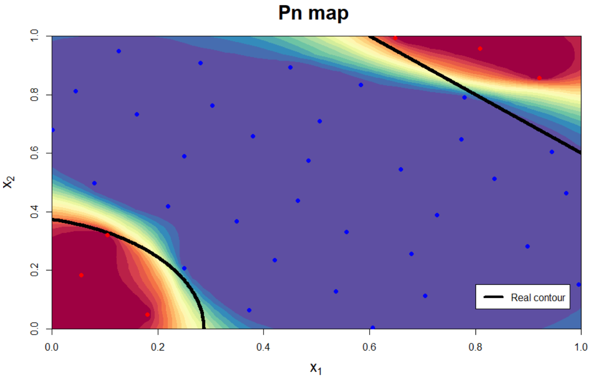
Gaussian process classifier (GPC) formulation

Practical building of the GPC model $p_n(x)$ for any x:

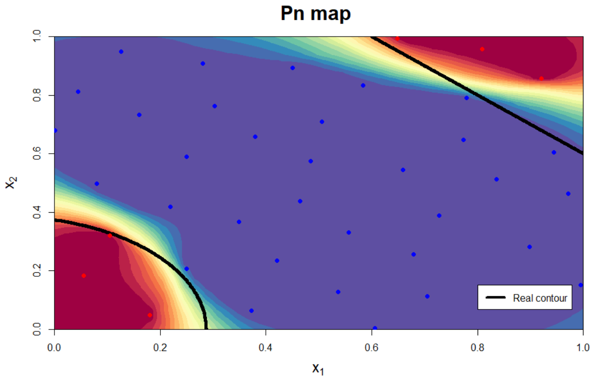
- Optimization of the hyperparameters of the latent GP to maximize the likelihood: $\mathbb{P}[sign(Z_n) = \mathcal{Y}]$
- Generation of realizations $z_n^{(i)}$ of $Z_n | sign(Z_n) = \mathcal{Y}$
  ➜ Approximation of $p_n(x)$:

$$\widehat{p_n}(x) = \frac{1}{N} \sum_{i=1}^{N} \bar{\Phi} \left(\frac{-m_n(x, z_n^{(i)})}{\sqrt{k_n(x)}} \right)$$

Example of a GPC for hidden constraint learning



Example of a GPC for hidden constraint learning



Characterisation of the feasible set:

$$Q_\alpha = \{x \in \Omega : p_n(x) \geq \alpha\}, \alpha \in (0, 1]$$

GPC based active learning

Principle: adaptively enrich the GPC using a learning criterion in order to obtain an accurate approximation of the set Γ^* contour

GPC based active learning

Principle: adaptively enrich the GPC using a learning criterion in order to obtain an accurate approximation of the set Γ^* contour

Idea: draw a methodology from existing criteria in Gaussian Process Regression active learning for feasible set estimation: stepwise uncertainty reduction strategy using the notion of random set [Bect et al., 2012, Molchanov, 2005]

Stepwise Uncertainty Reduction strategies

Let $\mathcal{U}_n$ be a measure of uncertainty about the excursion set knowing observations at points $\mathcal{X}$

Stepwise Uncertainty Reduction (SUR) strategies aim to minimize at each step:

$$J_n(x_{n+1}) = \mathbb{E}_n[\mathcal{U}_{n+1}(x_{n+1}, Z(x_{n+1}))] := \mathbb{E}_{Z(x_{n+1})}[\mathcal{U}_{n+1}(x_{n+1})|\mathcal{X}, Z(\mathcal{X})]$$

Stepwise Uncertainty Reduction strategy

The *Stepwise Uncertainty Reduction* strategy based on the uncertainty defined by the **vorob'ev deviation** $Var_n(\Gamma)$ [Chevalier, 2013, El Amri et al., 2021, Vorobyev and Lukyanova, 2013] is based on the following learning criterion:

$$J_n(x_{n+1}) = \mathbb{E}_n[Var_{n+1}(\Gamma)]$$

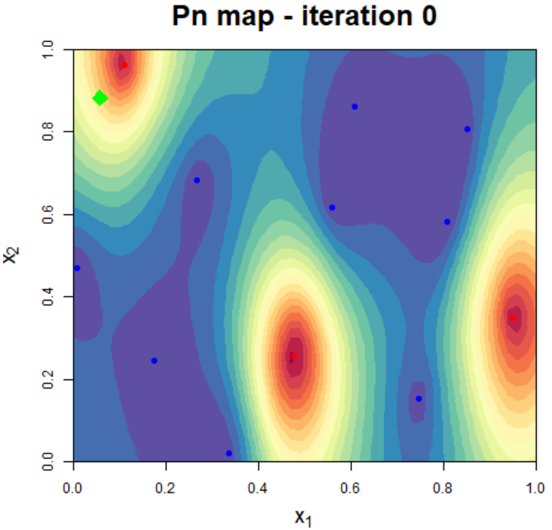
i.e.

$$J_n(x_{n+1}) = \mathbb{E}_{Z(x_{n+1})}\left[\int (1 - p_{n+1}(x))\mathbb{1}_{p_{n+1}(x) \geq \alpha^*}\mu(dx) + \int p_{n+1}(x)\mathbb{1}_{p_{n+1}(x) < \alpha^*}\mu(dx)\right]$$

This expression can be developed using the expression of $p_n(x)$ given for the GPC model [Bachoc et al., 2020] and GP update formulae provided in [Chevalier, 2013]

Example - Active learning based on Vorob'ev criterion

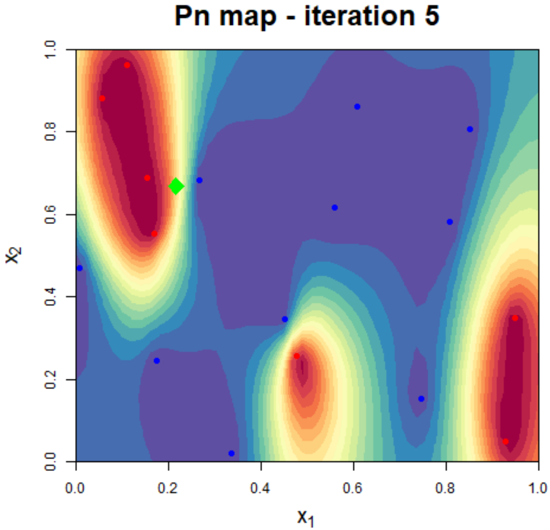
Classification problem based on the Branin function



 |

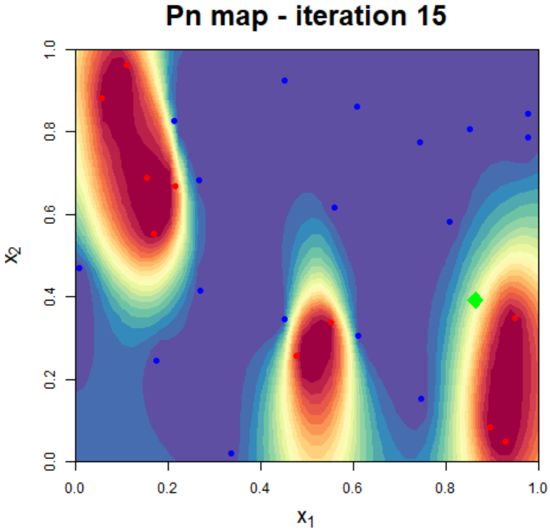
Example - Active learning based on Vorob'ev criterion

Classification problem based on the Branin function



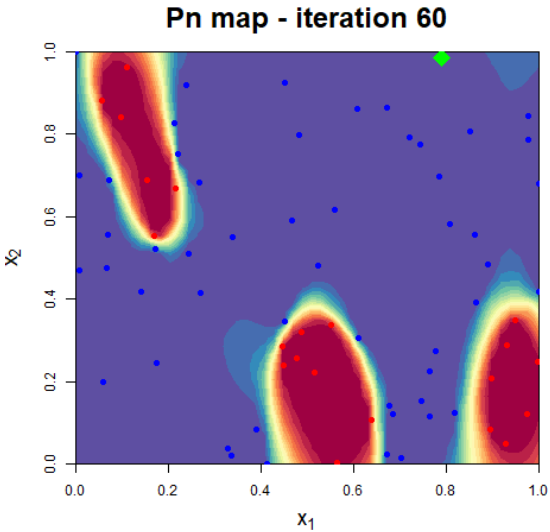
Example - Active learning based on Vorob'ev criterion

Classification problem based on the Branin function



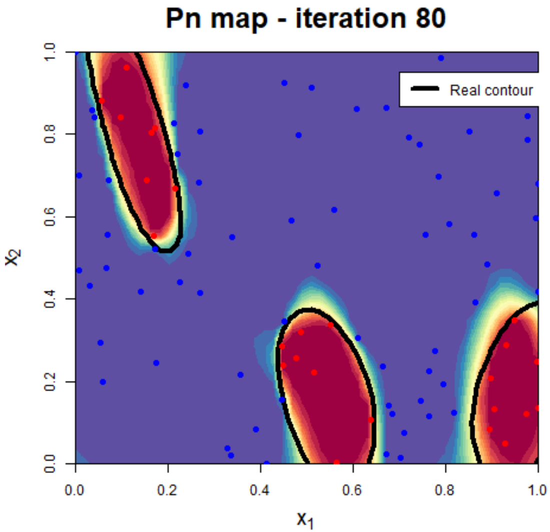
Example - Active learning based on Vorob'ev criterion

Classification problem based on the Branin function



Example - Active learning based on Vorob'ev criterion

Classification problem based on the Branin function



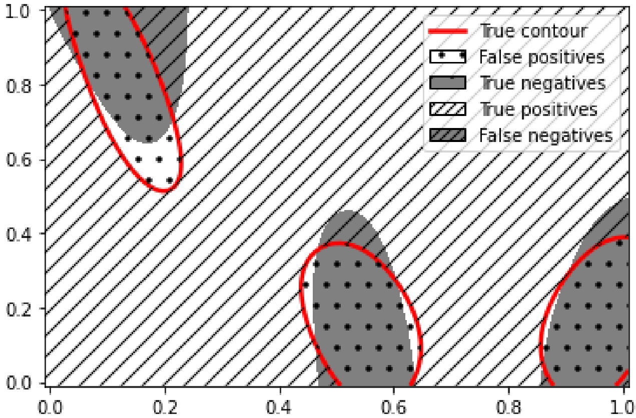
Results: comparison of different enrichment criteria

Compared strategies:

- SUR criterion for classification
- Mixed enrichment criterion: add of the point corresponding to the maximum of the GP variance (exploration) and the one where $p_n(x)$ value is the closest to $\frac{1}{2}$ (exploitation) simultanously
- SMOCU enrichment measure: Soft-MOCU (Mean Objective Cost of Uncertainty) method [Zhao et al., 2021]

Results: comparison criteria

1 - Number of true positives ($Q_\alpha \cap \Gamma^*$) and true negatives ($\Omega \setminus (Q_\alpha \cap \Gamma^*)$)

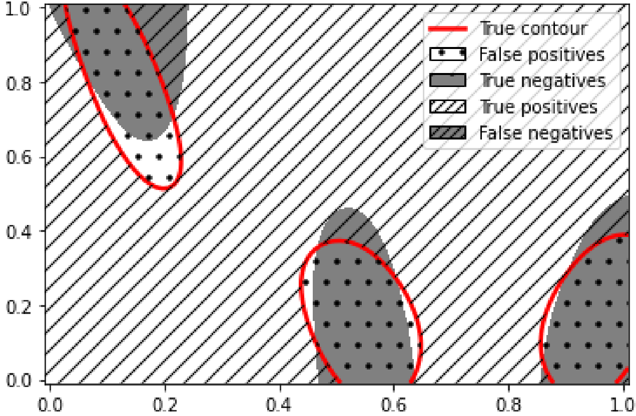


$$crit_P = \frac{TP}{P}$$

$$crit_N = \frac{TN}{N}$$

Results: comparison criteria

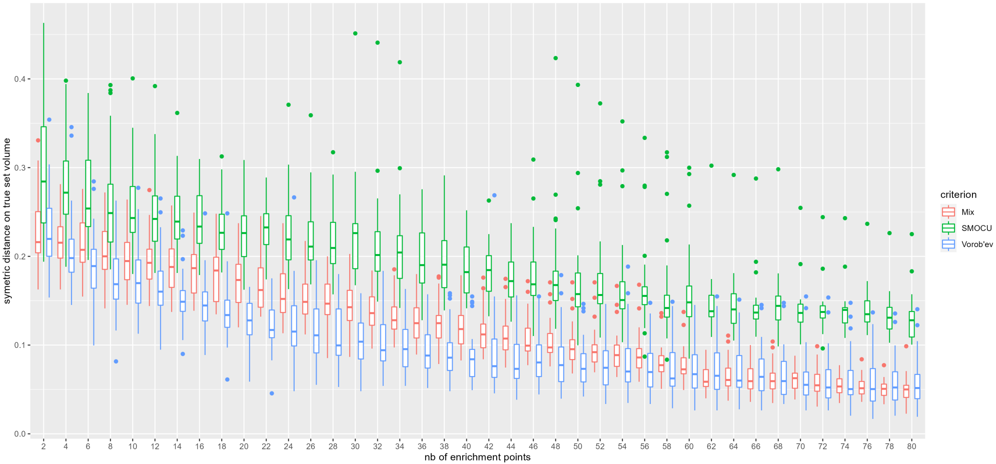
2 - Relative error of the feasible set estimation



$$crit_F = \frac{FN + FP}{P}$$

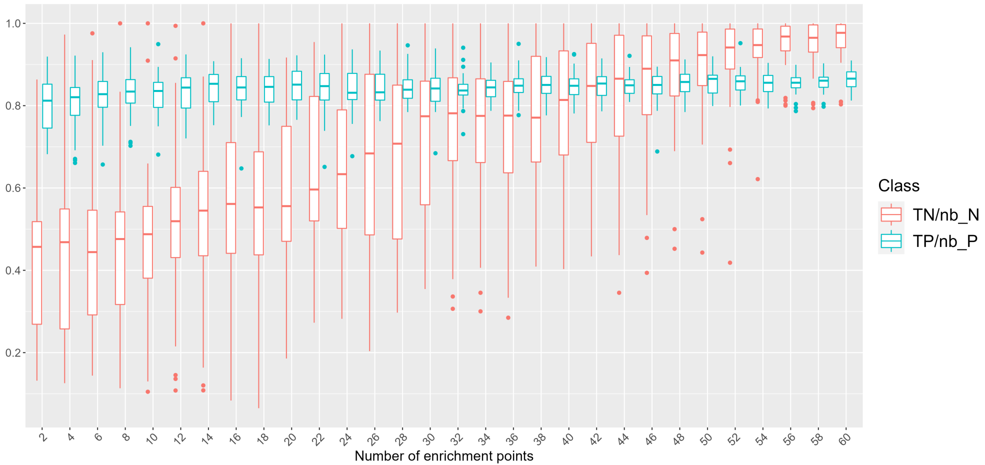
Results on the classification problem based on Branin function

Evolution of $crit_F$ for mixed, SMOCU and SUR criteria



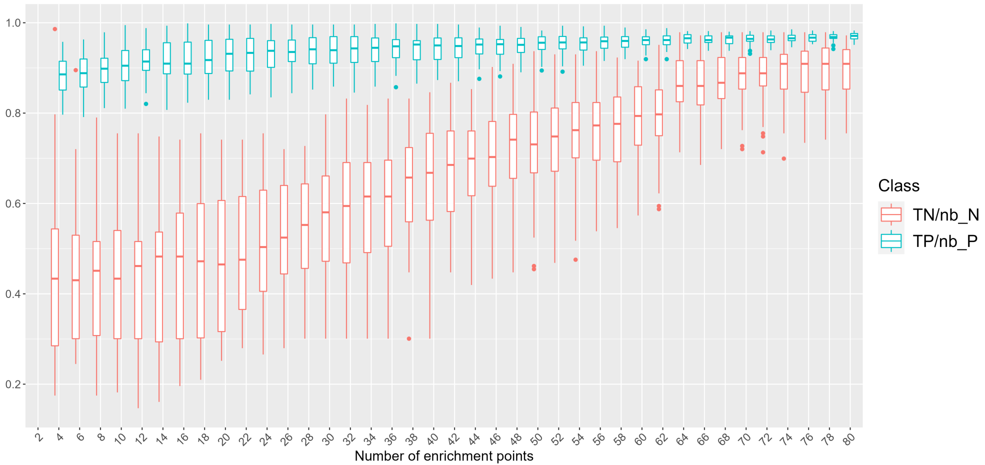
Results on the classification problem based on Branin function

Evolution of the number of true positives/negatives for SMOCU method



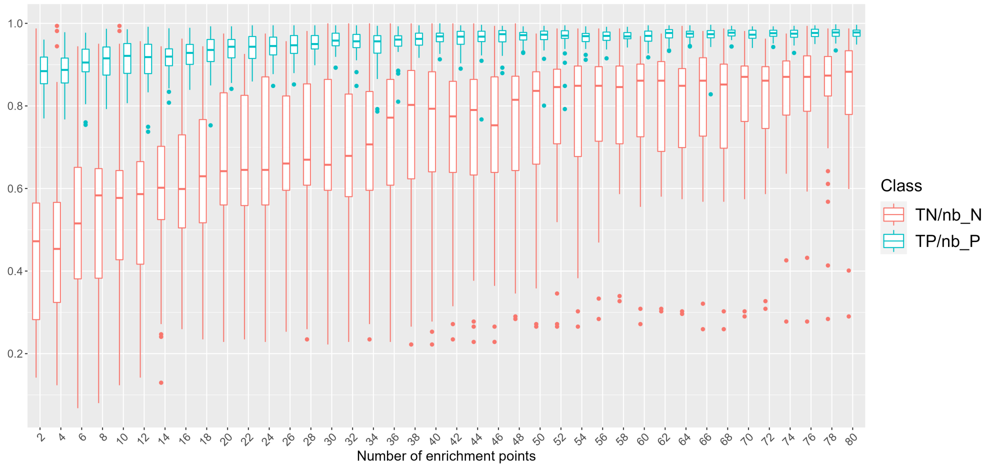
Results on the classification problem based on Branin function

Evolution of the number of true positives/negatives for mixed criterion



Results on the classification problem based on Branin function

Evolution of the number of true positives/negatives for SUR criterion



Example in 4 dimensions

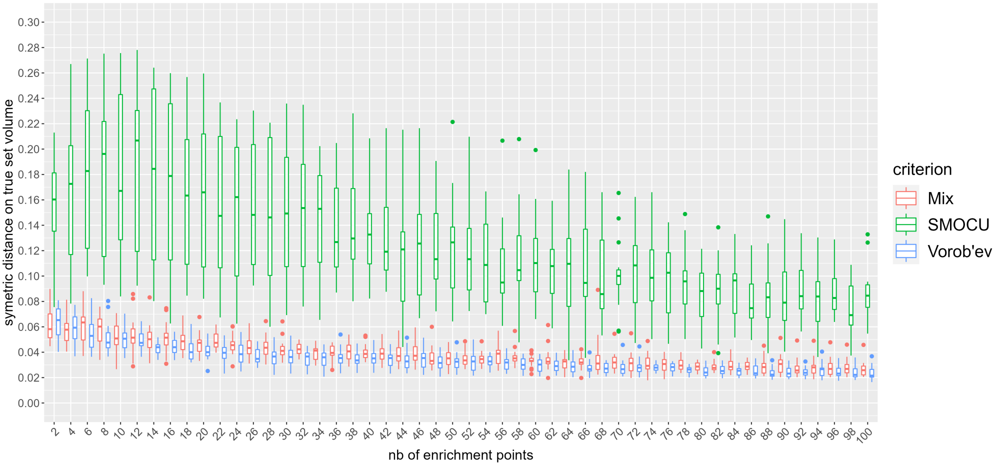
Classification problem with two constraints defined in $[0,1]^4$:

$$f_1(x_1, x_2, x_3, x_4) = x_2 + x_1 - 1.6 + 0.1 * x_3$$

$$f_2(x_1, x_2, x_3, x_4) = 0.15 - (x_2^2 + (x_1 + 0.1)^2)$$

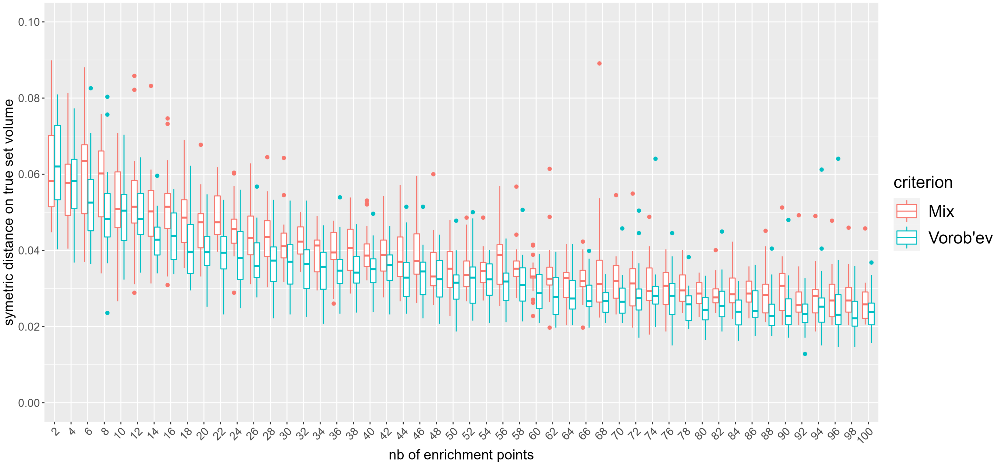
Results on the classification problem in 4 dimensions

Evolution of $crit_F$ for mixed, SMOCU and SUR criteria



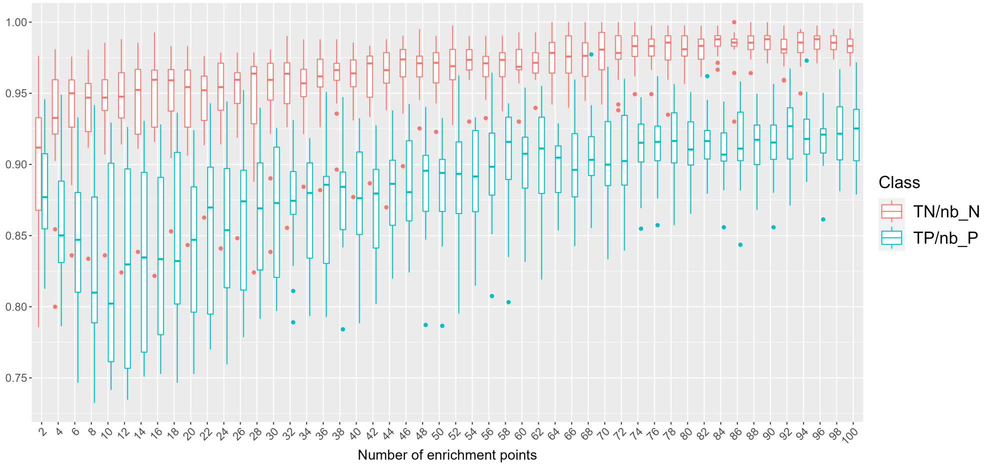
Results on the classification problem in 4 dimensions

Evolution of $crit_F$ for mixed, SMOCU and SUR criteria



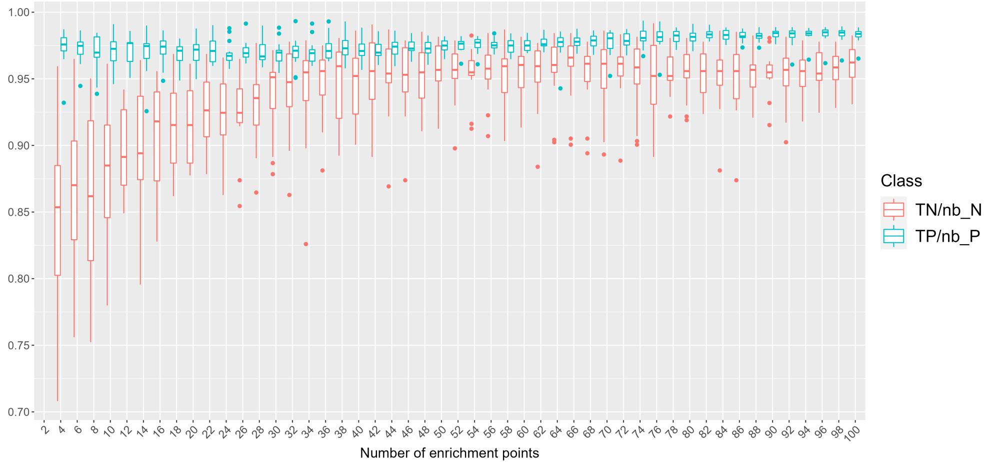
Results on the classification problem in 4 dimensions

Evolution of the number of true positives/negatives for SMOCU method

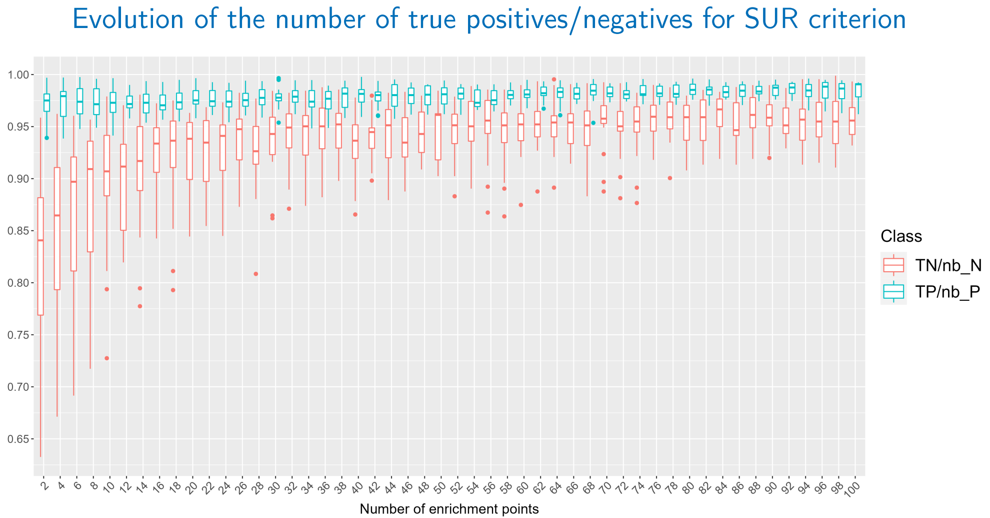


Results on the classification problem in 4 dimensions

Evolution of the number of true positives/negatives for mixed criterion



Results on the classification problem in 4 dimensions



Conclusions and outlook

Conclusions:

- great potential to learn feasible sets in terms of crash constraints
- drawbacks: enrichement criterion computation (integration) and optimization time in high dimension

Perspectives:

- Coupling with an optimization ➜ e.g. NOMAD, see session MB8: Stéphane Jacquet's presentation (at 15h30)
- Take into account the simulation robustness or convergence level (more than two classes)

References

François Bachoc, Céline Helbert, and Victor Picheny. Gaussian process optimization with failures: classification and convergence proof. *Journal of Global Optimization*, 78(3):483–506, November 2020. ISSN 0925-5001, 1573-2916. doi: 10.1007/s10898-020-00920-0. URL https://link.springer.com/10.1007/s10898-020-00920-0.

Julien Bect, David Ginsbourger, Ling Li, Victor Picheny, and Emmanuel Vazquez. Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, 22(3):773–793, May 2012. ISSN 0960-3174, 1573-1375. doi: 10.1007/s11222-011-9241-4. URL https://link.springer.com/article/10.1007/s11222-011-9241-4.

Clément Chevalier. *Fast uncertainty reduction strategies relying on Gaussian process models*. PhD Thesis, 2013.

Reda El Amri, Céline Helbert, Miguel Munoz Zuniga, Clémentine Prieur, and Delphine Sinoquet. Set inversion under functional uncertainties with gaussian process regression defined in the joint space of control and uncertain. 2021.

Alexander I. J. Forrester, Andrs Sbester, and Andy J. Keane. *Engineering Design via Surrogate Modelling*. John Wiley & Sons, Ltd, Chichester, UK, July 2008. ISBN 978-0-470-77080-1 978-0-470-06068-1. doi: 10.1002/9780470770801. URL http://doi.wiley.com/10.1002/9780470770801.

Ilya Molchanov. *Theory of random sets*, volume 19. Springer, 2005.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning*. Adaptive computation and machine learning. MIT Press, Cambridge, Mass, 2006. ISBN 978-0-262-18253-9.

Oleg Yu. Vorobyev and Natalia A. Lukyanova. A Mean Probability Event for a Set of Events. *Journal of Siberian Federal University. Mathematics and Physics.*, pages 128—-136, 2013.

Guang Zhao, Edward R Dougherty, Byung-Jun Yoon, Francis J Alexander, and Xiaoning Qian. Efficient Active Learning for Gaussian Process Classification by Error Reduction. page 13, 2021.

Expressions of m_n and k_n for the GPC model

$$m_n(x, z_n^{(i)}) = \mu_Z + k_\theta^Z(x, \mathcal{X})(k_\theta^Z(\mathcal{X}, \mathcal{X}))^{-1}(z_n^{(i)} - \mu_Z)$$

and

$$k_n(x, x') = k_\theta^Z(x, x') + k_\theta^Z(x, \mathcal{X})(k_\theta^Z(\mathcal{X}, \mathcal{X}))^{-1}k_\theta^Z(\mathcal{X}, x')$$

GP update formulae

Updated GP mean for r new points $x^{(r)} = (x_{n+1}, \ldots, x_{n+r})$:

$$m_{n+r}(x) = m_n(x) + \lambda_{new}(x)^T Z_c(x^{(r)})$$

with:

- $Z_c(x^{(r)}) = Z(x^{(r)}) - m_n(x^{(r)})$ le vecteur des réponses centrées de distribution $\mathcal{N}(0, \Sigma)$, où Σ est la matrice de covariance de $(Z_n(x_{n+1}), \ldots, Z(x_{n+r}))$
- $\lambda_{new}(x) = K_{new}^{-T} k_n(x, x^{(r)})$ et $K_{new} = k_n(x^{(r)}, x^{(r)})$

Updated covariance function of GP Z:

$$k_{n+r}(x, x') = k_n(x, x') - k_n(x, x^{(r)})^T K_{new}^{-1} k_n(x', x^{(r)})$$

Vorob'ev expectation and deviation

Vorob'ev deviation can be expressed as follows:

$$
\begin{aligned}
var_n(\Gamma) &= \mathbb{E}[\mu(Q_\alpha \Delta \Gamma)|z] \\
&= \int_{Q_\alpha} (1 - p_n(x))\mu(dx) + \int_{Q_\alpha^C} p_n(x)\mu(dx) \\
&= \int (1 - p_n(x)) \mathbb{1}_{p_n(x) \geq \alpha^*} \mu(dx) + \int p_n(x) \mathbb{1}_{p_n(x) < \alpha^*} \mu(dx)
\end{aligned}
$$

with Δ such that $A \Delta B = (A \cup B) \setminus (A \cap B)$

GPC and *Jn* criterion

MOBILITÉ DURABLE

Expression of the first term of J_n criterion in the integral on x:

$$\mathbb{E}_n\left[p_{n+1}(x)\mathbb{1}_{p_{n+1}(x)<\alpha^*}\right]$$

$$= \mathbb{E}_n\left[\frac{1}{N}\sum_{i=1}^{N}\bar{\Phi}\left(\frac{-m_{n+1}(x,z_n^{(i)})}{\sqrt{k_{n+1}(x)}}\right)\mathbb{1}_{\frac{1}{N}\sum_{j=1}^{N}\bar{\Phi}\left(\frac{-m_{n+1}(x,z_n^{(j)})}{\sqrt{k_{n+1}(x)}}\right)<\alpha^*}\right]$$

$$= \mathbb{E}_n\left[\frac{1}{N}\sum_{i}\bar{\Phi}\left(\frac{-(m_n(x,z_n^{(i)})+\lambda_{new}(x)^T\mathrm{U})}{\sqrt{k_{n+1}(x)}}\right)\mathbb{1}_{\frac{1}{N}\sum_{j}\bar{\Phi}\left(\frac{-(m_n(x,z_n^{(j)})+\lambda_{new}(x)^T\mathrm{U})}{\sqrt{k_{n+1}(x)}}\right)<\alpha^*}\right]$$

avec $\mathrm{U} \sim \mathcal{N}(0,\Sigma)$, où Σ est la matrice de covariance de $(Z_n(x_{n+1}),\ldots,Z(x_{n+r}))$.

Remark: The realizations $z_n^{(i)}$ are not updated because $sign(Z(x_{n+1}))$ do not provide any information on Z_n.

GPC and Vorob'ev based *Jn* criterion

The third can be expressed as follows:

$$\mathbb{E}_n\left[\mathbb{1}_{p_{n+q}(x)\geq\alpha^*}\right]$$

$$= \mathbb{E}_n\left[\mathbb{1}_{\frac{1}{N}\sum_j \bar{\Phi}\left(\frac{-(m_n(x,z_n^{(i)})+\lambda_{new}(x)^T\mathbf{U})}{\sqrt{k_{n+q}(x)}}\right)<\alpha^*}\right]$$

Update of the realizations $z_n^{(i)}$

MOBILITÉ DURABLE

$$z_{n+q}^{(i)} = z_n^{(i)} + \lambda_{new}^Z(\mathcal{X})^T (Z_n(x_{new})_{real} - Z_n(x_{new})_{simu})$$

$$\begin{pmatrix} Z(x_1)^{(i)} \\ \vdots \\ Z(x_n)^{(i)} \end{pmatrix}_{n+q} = \begin{pmatrix} Z(x_1)^{(i)} \\ \vdots \\ Z(x_n)^{(i)} \end{pmatrix}_n + \lambda_{new}^Z(\mathcal{X})^T Z_{n,c}(x_{new})$$

with $Z_{n,c}(x_{new}) \sim \mathcal{N}(0, k^Z(x_{new}, x_{new}))$ and $\lambda_{new}^Z(\mathcal{X}) = k^Z(x_{new}, x_{new})^{-T} k^Z(x_{new}, \mathcal{X})$.

Yet:

- we do not guarantee that the signs of $z_{n+q}^{(i)}$ correspond to the observed values $y_1, \ldots, y_n$ anymore,
- it can be verified that as for Z_n, Z_{n+q} can be modeled by a GP with constant mean μ_Z and a stationary covariance function k_θ^Z

INSTITUT CARNOT IFPEN Transports Energie

Energies nouvelles

Vorob'ev expectation and deviation

MOBILITÉ DURABLE

Let us consider the random set:

$$\Gamma = \{x \in \Omega : Y_n(x) = 1\}$$

The function

$$\begin{aligned} P_\Gamma(x) &= \mathbb{P}[x \in \Gamma] \\ &= p_n(x) \end{aligned}$$

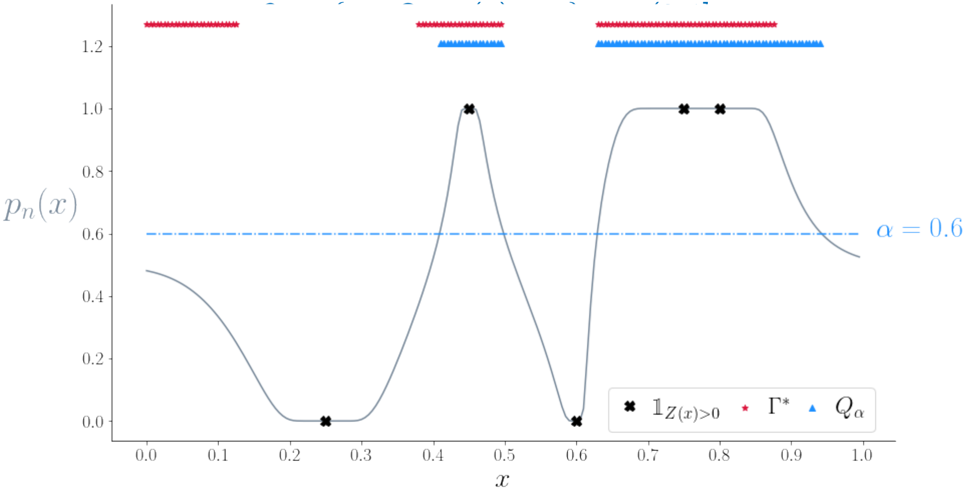
is the cover function of Γ and its level sets are the α-percentiles of Γ:

$$Q_\alpha = \{x \in \Omega : p_n(x) \geq \alpha\}, \alpha \in (0, 1]$$

INSTITUT CARNOT IFPEN Transports Energie

Energies nouvelles

Vorob'ev expectation and deviation

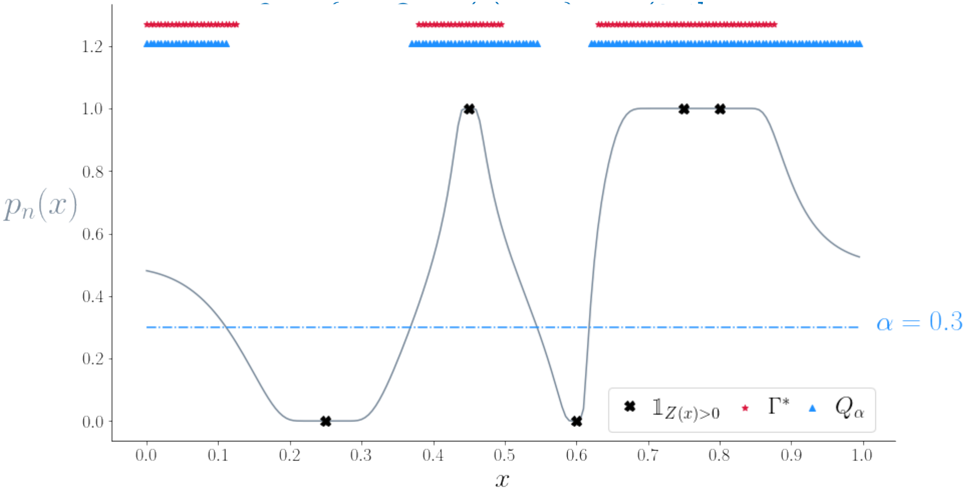


INSTITUT CARNOT IFPEN Transports Energie

IFP Energies nouvelles

Vorob'ev expectation and deviation

MOBILITÉ DURABLE



Vorob'ev expectation and deviation

Definition

Vorob'ev expectation is defined as the α^*-percentile of Γ, where α^* is determined by:

$$\mathbb{E}[\mu(\Gamma)] = \mu(Q_{\alpha^*})$$

with μ the Lebesgue measure.

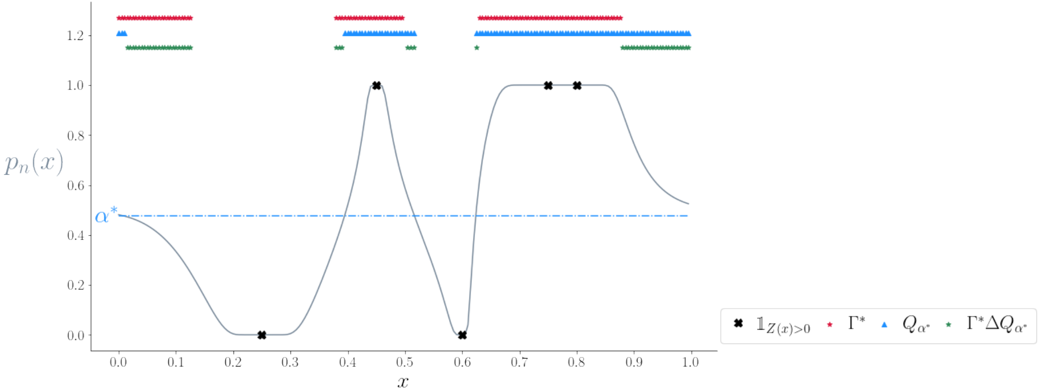
Vorob'ev expectation is a global minimiser of the vorob'ev deviation $Var_n(\Gamma)$ among closed sets of volume equal to the mean volume of Γ:

$$Var_n(\Gamma) = \mathbb{E}[\mu(Q_\alpha \Delta \Gamma) | \mathcal{X}, \mathcal{Y}]$$

with $\Gamma \Delta Q_\alpha = (\Gamma \setminus Q_\alpha) \cup (Q_\alpha \setminus \Gamma)$

Vorob'ev expectation and deviation

$$Q_\alpha = \{x \in \Omega : p_n(x) \geq \alpha\}, \alpha \in (0, 1]$$



MOCU/SMOCU criteria

$$U^{MOCU}(x) = \mathbb{E}_{x_s}[\mathbb{E}_n[max(p_{n+1}(x_s), 1 - p_{n+1}(x_s))] - max(p_n(x_s), 1 - p_n(x_s))]$$

i.e.

$$U^{MOCU}(x) = \mathbb{E}_{x_s}[1 - max(p_n(x_s), 1 - p_n(x_s))] - \mathbb{E}_n[\mathbb{E}_{x_s}[1 - max(p_{n+1}(x_s), 1 - p_{n+1}(x_s))]$$

et

$$\begin{aligned} U^{SMOCU}(x) &= \mathbb{E}_{x_s}[\mathbb{E}_n[\tfrac{1}{k} ln(exp(k * p_{n+1}(x_s)) + exp(k(1 - p_{n+1}(x_s))))] \\ &- \tfrac{1}{k} ln(exp(k * p_n(x_s)) + exp(k(1 - p_n(x_s))))] \end{aligned}$$

