Statistical interpretation of an ensemble of surrogate models

.

How to create one stochastic surrogate model out of multiple deterministic surrogate models

Renaud Saltet
Director : Charles Audet, Co-director : Sébastien Le Digabel

Polytechnique Montréal

10 September 2020

Table of Contents

Table of Contents

Surrogate models in black-box optimization

True problem :

$$\begin{aligned} &\min_{x \in \mathbb{R}^n} f(x) \\ &s.t. \quad c_j(x) \leq 0, \quad \forall j \in \{1, 2, \ldots, m\} \end{aligned} \qquad (P)$$

Surrogate models in black-box optimization

True problem :

$$\min_{x \in \mathbb{R}^n} f(x) \qquad (P)$$
$$s.t. \quad c_j(x) \leq 0, \quad \forall j \in \{1, 2, \ldots, m\}$$

Surrogate problem :

$$\min_{x \in \mathbb{R}^n} \tilde{f}(x) \qquad (\tilde{P})$$
$$s.t. \quad \tilde{c}_j(x) \leq 0, \quad \forall j \in \{1, 2, \ldots, m\}$$

Surrogate models in black-box optimization

True problem :

$$\begin{aligned} &\min_{x \in \mathbb{R}^n} f(x) \\ &s.t. \quad c_j(x) \leq 0, \quad \forall j \in \{1, 2, \dots, m\} \end{aligned} \tag{P}$$

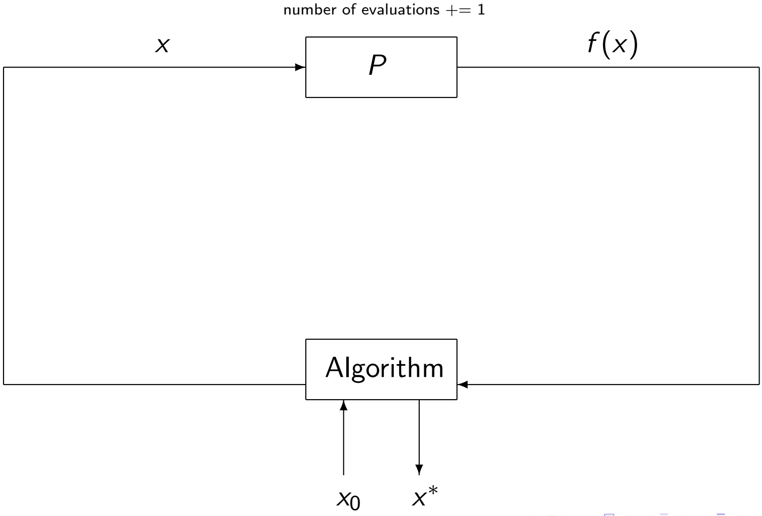
Surrogate problem :

$$\begin{aligned} &\min_{x \in \mathbb{R}^n} \tilde{f}(x) \\ &s.t. \quad \tilde{c}_j(x) \leq 0, \quad \forall j \in \{1, 2, \dots, m\} \end{aligned} \tag{$\tilde{P}$}$$

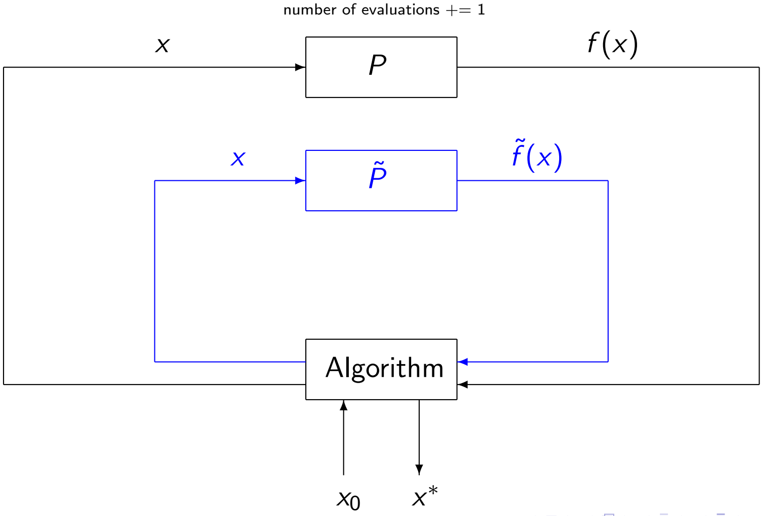
Sample points (cache) :

$$\begin{aligned} \mathbb{X} &= \{x^{(1)}, \dots, x^{(p)}\} \\ \mathbf{y} &= \{f(x^{(1)}), \dots, f(x^{(p)})\} \end{aligned}$$

Surrogate models in black-box optimization



Surrogate models in black-box optimization

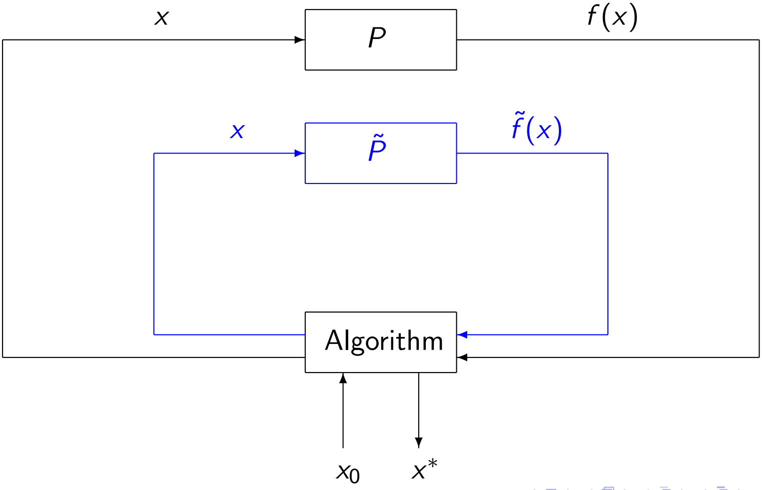


Surrogate models in black-box optimization

$$\mathbb{X} = \mathbb{X} \cup \{x\}$$

$$\mathbf{y} = \mathbf{y} \cup \{f(x)\}$$

number of evaluations += 1



Surrogate models in black-box optimization

Seminal article : Booker et al. (1999) [1]

Types of surrogate models :

- Radial Basis Functions [2, 3]
- Quadratic Models [4, 5]
- Kernel Smoothing [6, 7]
- Gaussian Processes [8, 9, 10]
- . . .

Table of Contents

Stochastic surrogate models

$$x \in \mathbb{R}^n \longrightarrow \tilde{f}(x)$$

Stochastic surrogate models

$$x \in \mathbb{R}^n \longrightarrow \tilde{f}(x),\ \tilde{\sigma}(x)$$

Stochastic surrogate models

$$x \in \mathbb{R}^n \longrightarrow \tilde{f}(x),\ \tilde{\sigma}(x)\ ;\ \tilde{c}_j(x),\ \tilde{\sigma}_j(x)$$

Stochastic surrogate models

$$x \in \mathbb{R}^n \longrightarrow \tilde{f}(x),\ \tilde{\sigma}(x)\ ;\ \tilde{c}_j(x),\ \tilde{\sigma}_j(x)$$

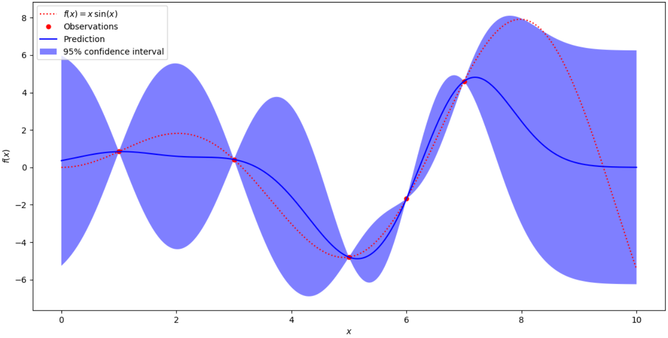


Figure: Gaussian process regression on $f : x \mapsto x \sin x$

Stochastic surrogate models

What is the use of a stochastic surrogate ?

Stochastic surrogate models

What is the use of a stochastic surrogate ?

- Expected Improvement [9] :

$$\begin{aligned} EI(x) &= \mathbb{E}[I(x)] \\ &= \mathbb{E}[\max\{fmin - f(x), 0\}] \end{aligned}$$

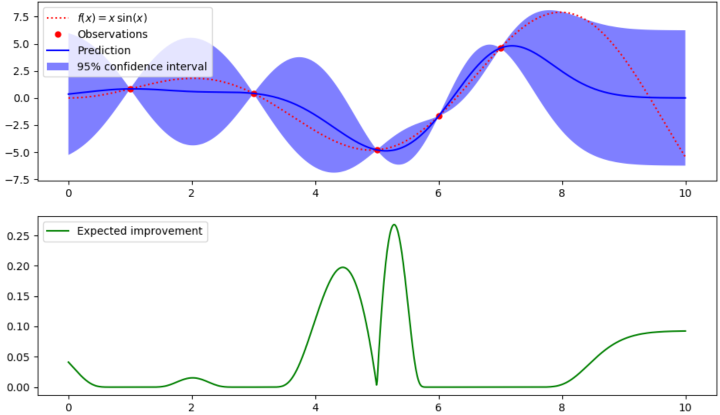
Stochastic surrogate models

What is the use of a stochastic surrogate ?

- Expected Improvement [9] :

$$\begin{aligned} EI(x) &= \mathbb{E}[I(x)] \\ &= \mathbb{E}[\max\{fmin - f(x), 0\}] \end{aligned}$$

Gaussian process regression and the resulting Expected Improvement



Stochastic surrogate models

What is the use of a stochastic surrogate ?

- Expected Improvement [9] :

$$EI(x) = \mathbb{E}[I(x)]$$
$$= \mathbb{E}[\max\{fmin - f(x), 0\}]$$

- Probability of feasibility [11] :

$$P(x) = \mathbb{P}[c_j(x) \leq 0, \ \forall j]$$

Stochastic surrogate models

What is the use of a stochastic surrogate ?

- Expected Improvement [9] :

$$EI(x) = \mathbb{E}[I(x)]$$
$$= \mathbb{E}[\max\{fmin - f(x), 0\}]$$

- Probability of feasibility [11] :

$$P(x) = \mathbb{P}[c_j(x) \leq 0, \ \forall j]$$

- Probability of Improvement [11] :

$$PI(x) = \mathbb{P}[I(x) > 0]$$
$$= \mathbb{P}[f_{min} > f(x)]$$

Stochastic surrogate models

What is the use of a stochastic surrogate ?

- Expected Feasible Improvement [11] :

$$EFI(x) = EI(x)P(x)$$

Stochastic surrogate models

What is the use of a stochastic surrogate ?

- Expected Feasible Improvement [11] :

$$EFI(x) = EI(x)P(x)$$

- Uncertainty in the feasibility [11] :

$$\mu(x) = P(x)(1 - P(x))$$

Stochastic surrogate models

What is the use of a stochastic surrogate ?

- Expected Feasible Improvement [11] :

$$EFI(x) = EI(x)P(x)$$

- Uncertainty in the feasibility [11] :

$$\mu(x) = P(x)(1 - P(x))$$

- Probability of Feasible Improvement [11] :

$$PFI(x) = PI(x)P(x)$$

Stochastic surrogate models

B. Talgorn, S. Le Digabel, M. Kokkolaras
Statistical Surrogate Formulations for Simulation-Based Design Optimization, 2015 [11]

Stochastic surrogate models

B. Talgorn, S. Le Digabel, M. Kokkolaras
Statistical Surrogate Formulations for Simulation-Based Design Optimization, 2015 [11]

$$\min_{x\in\mathbb{R}^n} \tilde{f}(x)$$
$$s.t. \quad \tilde{c}_j(x) \leq 0, \quad \forall j \in \{1, 2, \ldots, m\} \qquad (\tilde{P})$$

Stochastic surrogate models

B. Talgorn, S. Le Digabel, M. Kokkolaras
Statistical Surrogate Formulations for Simulation-Based Design Optimization, 2015 [11]

$$\min_{x \in \mathbb{R}^n} \tilde{f}(x)$$
$$s.t. \quad \tilde{c}_j(x) \leq 0, \quad \forall j \in \{1, 2, \ldots, m\} \qquad (\tilde{P})$$

Stochastic surrogate models

B. Talgorn, S. Le Digabel, M. Kokkolaras
Statistical Surrogate Formulations for Simulation-Based Design Optimization, 2015 [11]

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \ & \tilde{f}(x) \\ s.t. \ & \tilde{c}_j(x) \leq 0, \quad \forall j \in \{1, 2, \ldots, m\} \end{aligned} \qquad (\tilde{P})$$

8 formulations of the surrogate problem $\tilde{P}$ integrating EI, P, PI, EFI, μ and PFI.

Table of Contents

Statistical interpretation of an ensemble of surrogate models

There is not one surrogate better than the others !

Statistical interpretation of an ensemble of surrogate models

There is not one surrogate better than the others !

We can use an ensemble of s surrogates $\mathcal{S} = \{\tilde{f}_1, \tilde{f}_2, \ldots, \tilde{f}_s\}$.

Statistical interpretation of an ensemble of surrogate models

There is not one surrogate better than the others !

We can use an ensemble of s surrogates $\mathcal{S} = \{\tilde{f}_1, \tilde{f}_2, \ldots, \tilde{f}_s\}$.

$$x \in \mathbb{R}^n \longrightarrow \tilde{f}(x) = \sum_{k=1}^{s} w_k \tilde{f}_k(x)$$

With $w_k \geq 0, \ \forall k = 1, \ldots, s$ and $\displaystyle\sum_{k=1}^{s} w_k = 1$.

Statistical interpretation of an ensemble of surrogate models

There is not one surrogate better than the others !

We can use an ensemble of s surrogates $\mathcal{S} = \{\tilde{f}_1, \tilde{f}_2, \ldots, \tilde{f}_s\}$.

$$x \in \mathbb{R}^n \;\longrightarrow\; \tilde{f}(x) = \sum_{k=1}^{s} w_k \tilde{f}_k(x)$$

With $w_k \geq 0,\ \forall k = 1, \ldots, s$ and $\sum_{k=1}^{s} w_k = 1$.

$$\tilde{\sigma}(x) \text{ ?}$$

Statistical interpretation of an ensemble of surrogate models

There is not one surrogate better than the others !

We can use an ensemble of s surrogates $\mathcal{S} = \{\tilde{f}_1, \tilde{f}_2, \ldots, \tilde{f}_s\}$.

$$x \in \mathbb{R}^n \;\longrightarrow\; \tilde{f}(x) = \sum_{k=1}^{s} w_k \tilde{f}_k(x)$$

With $w_k \geq 0, \;\forall k = 1, \ldots, s$ and $\displaystyle\sum_{k=1}^{s} w_k = 1$.

$$\tilde{\sigma}(x) \;?$$

$$\sout{EI,\; P,\; PI,\; EFI,\; \mu,\; PFI}$$

Statistical interpretation of an ensemble of surrogate models

There is not one surrogate better than the others !

We can use an ensemble of s surrogates $\mathcal{S} = \{\tilde{f}_1, \tilde{f}_2, \ldots, \tilde{f}_s\}$.

$$x \in \mathbb{R}^n \longrightarrow \tilde{f}(x) = \sum_{k=1}^{s} w_k \tilde{f}_k(x)$$

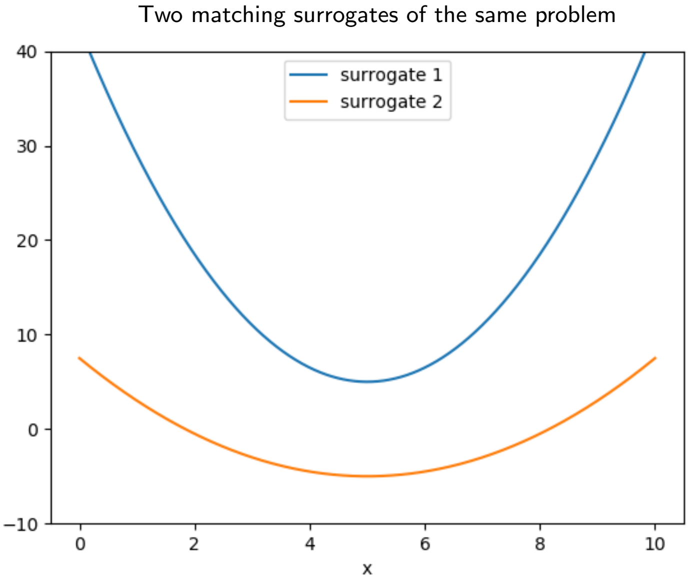
With $w_k \geq 0, \ \forall k = 1, \ldots, s$ and $\sum_{k=1}^{s} w_k = 1$.

$$\tilde{\sigma}(x) \ ?$$

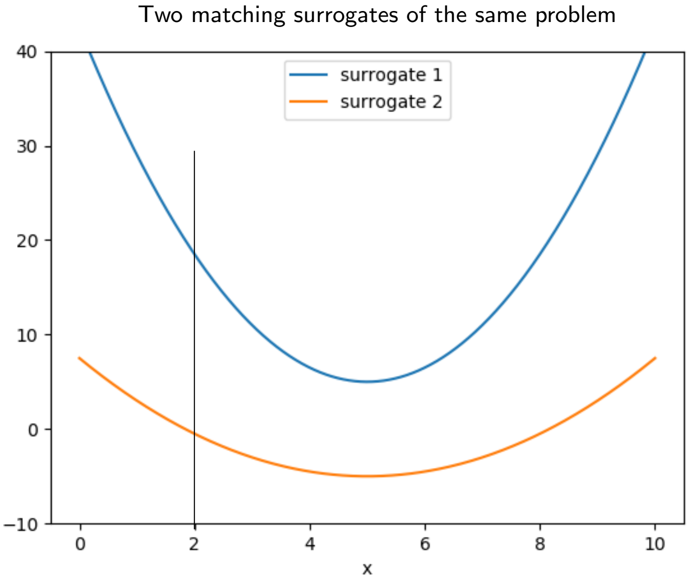
~~$EI, \ P, \ PI, \ EFI, \ \mu, \ PFI$~~

How to create one stochastic surrogate model out of multiple deterministic surrogate models ?

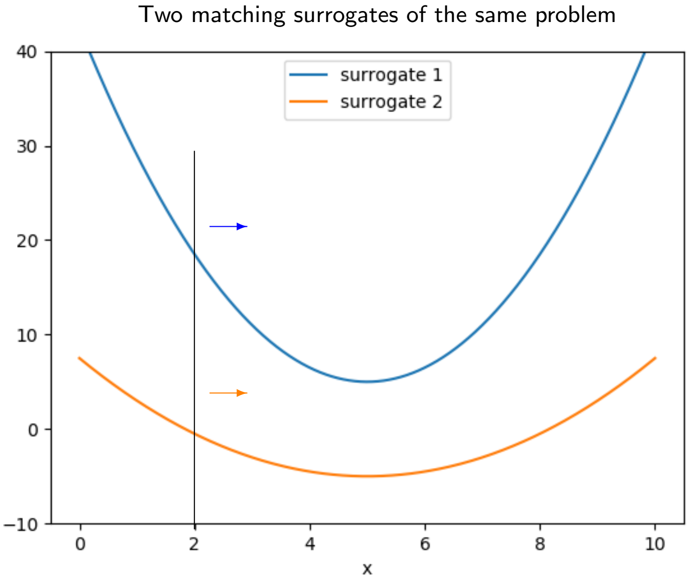
Statistical interpretation of an ensemble of surrogate models



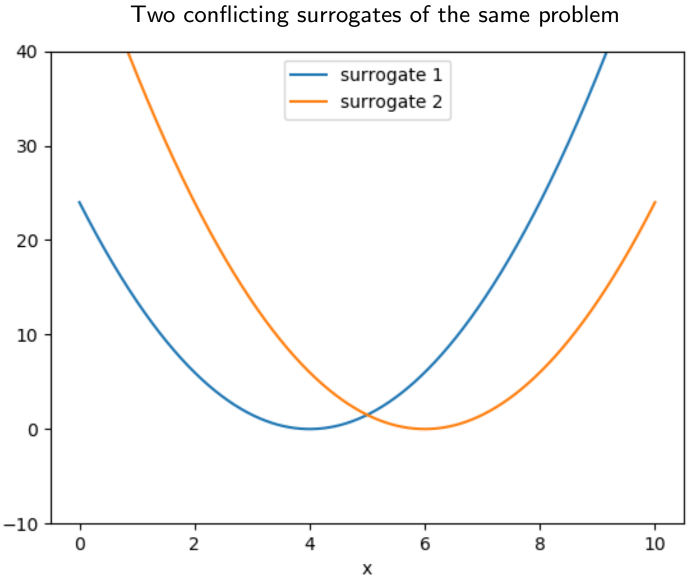
Statistical interpretation of an ensemble of surrogate models



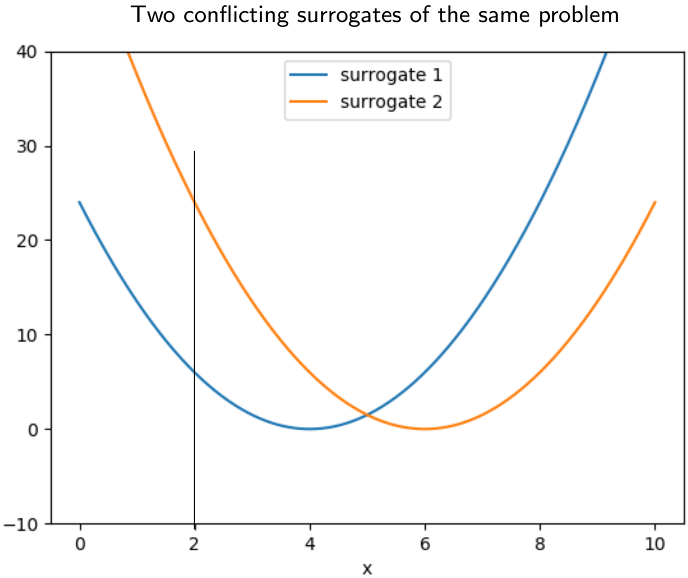
Statistical interpretation of an ensemble of surrogate models



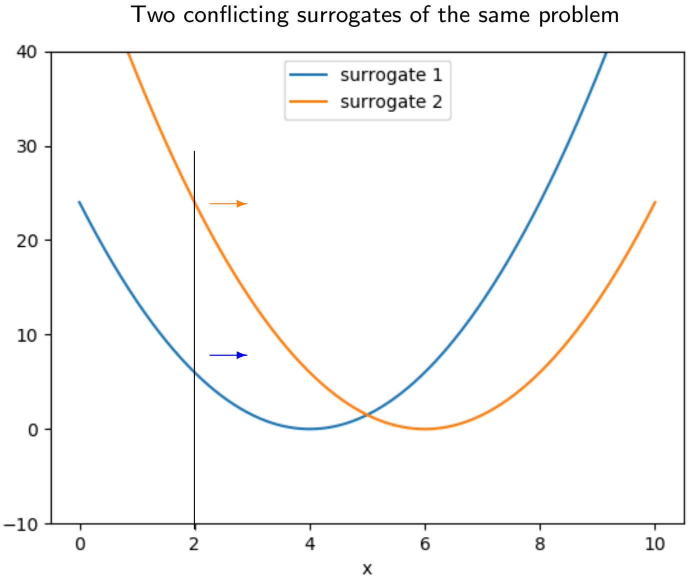
Statistical interpretation of an ensemble of surrogate models



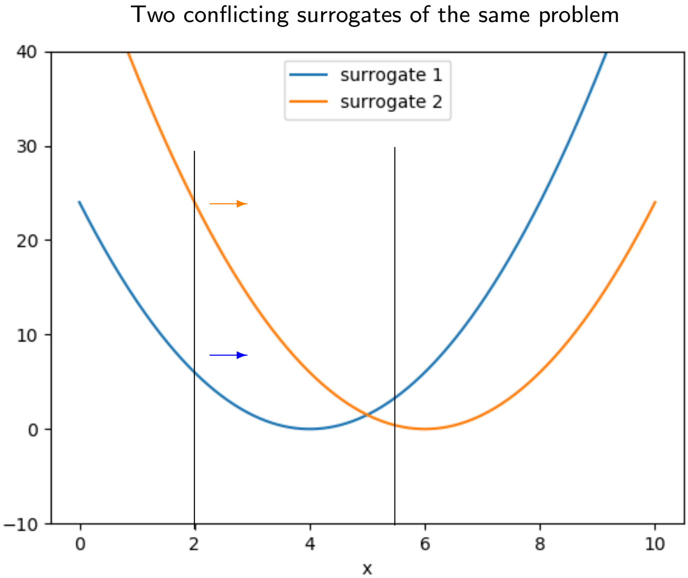
Statistical interpretation of an ensemble of surrogate models



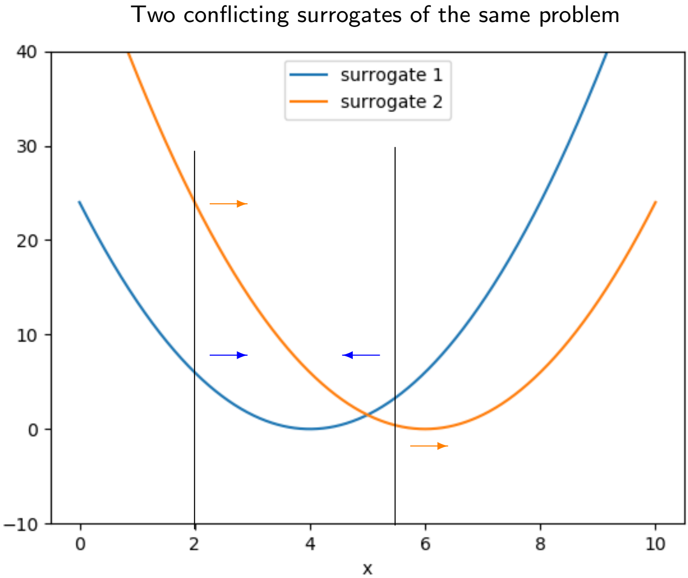
Statistical interpretation of an ensemble of surrogate models



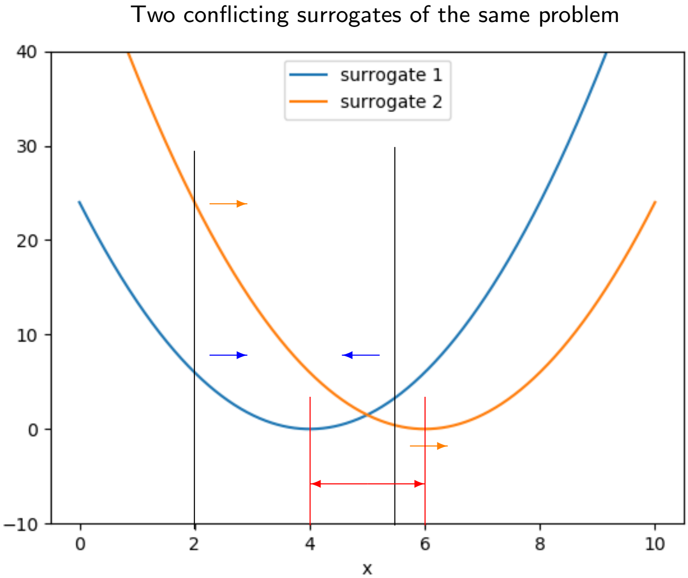
Statistical interpretation of an ensemble of surrogate models



Statistical interpretation of an ensemble of surrogate models



Statistical interpretation of an ensemble of surrogate models



Statistical interpretation of an ensemble of surrogate models

- $\mathcal{S} = \{\tilde{f}_1, \tilde{f}_2, \ldots, \tilde{f}_s\}$

Statistical interpretation of an ensemble of surrogate models

- $\mathcal{S} = \{\tilde{f}_1, \tilde{f}_2, \ldots, \tilde{f}_s\}$
- $\mathcal{D}$ a positive spanning set of $\mathbb{R}^n$

Statistical interpretation of an ensemble of surrogate models

- $\mathcal{S} = \{\tilde{f}_1, \tilde{f}_2, \ldots, \tilde{f}_s\}$
- $\mathcal{D}$ a positive spanning set of $\mathbb{R}^n$
- t a "finite difference" size parameter

Statistical interpretation of an ensemble of surrogate models

- $\mathcal{S} = \{\tilde{f}_1, \tilde{f}_2, \ldots, \tilde{f}_s\}$
- $\mathcal{D}$ a positive spanning set of $\mathbb{R}^n$
- t a "finite difference" size parameter

For a given point $x \in \mathbb{R}^n$, we define :

- The number of surrogates that predict a decrease in the direction $d \in \mathcal{D}$:

$$n_{d<x} = \left|\left\{\tilde{f} \in \mathcal{S} : \; \tilde{f}(x + td) < \tilde{f}(x)\right\}\right|$$

Statistical interpretation of an ensemble of surrogate models

- The number of surrogates that predict a decrease in the direction $d \in \mathcal{D}$:

$$n_{d<x} = \left|\left\{\tilde{f} \in \mathcal{S} : \ \tilde{f}(x+td) < \tilde{f}(x)\right\}\right|$$

Statistical interpretation of an ensemble of surrogate models

- The number of surrogates that predict a decrease in the direction $d \in \mathcal{D}$:

$$n_{d<x} = \left| \left\{ \tilde{f} \in \mathcal{S} : \; \tilde{f}(x + td) < \tilde{f}(x) \right\} \right|$$

- The *order-based uncertainty* (inspired by the article [12]) :

$$\mathcal{U}^{f}_{OB}(x) = \frac{1}{|\mathcal{D}|} \sum_{d \in \mathcal{D}} \frac{n_{d<x}(s - n_{d<x})}{(s/2)^2}$$

Statistical interpretation of an ensemble of surrogate models

- The number of surrogates that predict a decrease in the direction $d \in \mathcal{D}$:

$$n_{d<x} = \left|\left\{\tilde{f} \in \mathcal{S} : \ \tilde{f}(x+td) < \tilde{f}(x)\right\}\right|$$

- The *order-based uncertainty* (inspired by the article [12]) :

$$\mathcal{U}^{f}_{OB}(x) = \frac{1}{|\mathcal{D}|} \sum_{d \in \mathcal{D}} \frac{n_{d<x}(s - n_{d<x})}{(s/2)^2}$$

For $d \in \mathcal{D}$:

$$(n_{d<x} = 0 \text{ or } n_{d<x} = s) \ \Rightarrow \ \frac{n_{d<x}(s - n_{d<x})}{(s/2)^2} = 0$$

Statistical interpretation of an ensemble of surrogate models

- The number of surrogates that predict a decrease in the direction $d \in \mathcal{D}$:

$$n_{d<x} = \left|\left\{\tilde{f} \in \mathcal{S} :\ \tilde{f}(x + td) < \tilde{f}(x)\right\}\right|$$

- The *order-based uncertainty* (inspired by the article [12]) :

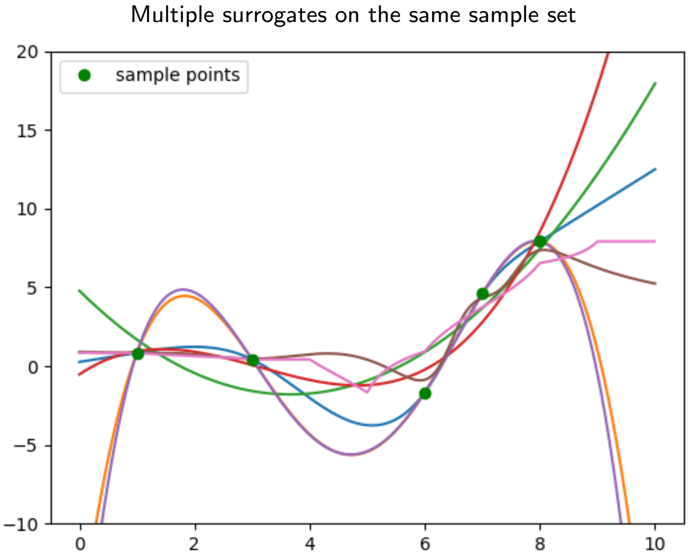
$$\mathcal{U}^f_{OB}(x) = \frac{1}{|\mathcal{D}|} \sum_{d \in \mathcal{D}} \frac{n_{d<x}(s - n_{d<x})}{(s/2)^2}$$

For $d \in \mathcal{D}$:

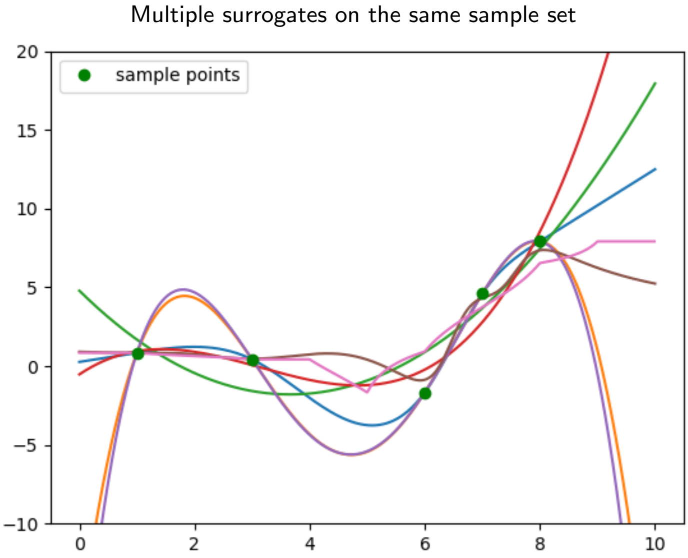
$$(n_{d<x} = 0 \text{ or } n_{d<x} = s) \ \Rightarrow\ \frac{n_{d<x}(s - n_{d<x})}{(s/2)^2} = 0$$

$$n_{d<x} = s/2 \ \Rightarrow\ \frac{n_{d<x}(s - n_{d<x})}{(s/2)^2} = 1$$

Statistical interpretation of an ensemble of surrogate models



Statistical interpretation of an ensemble of surrogate models

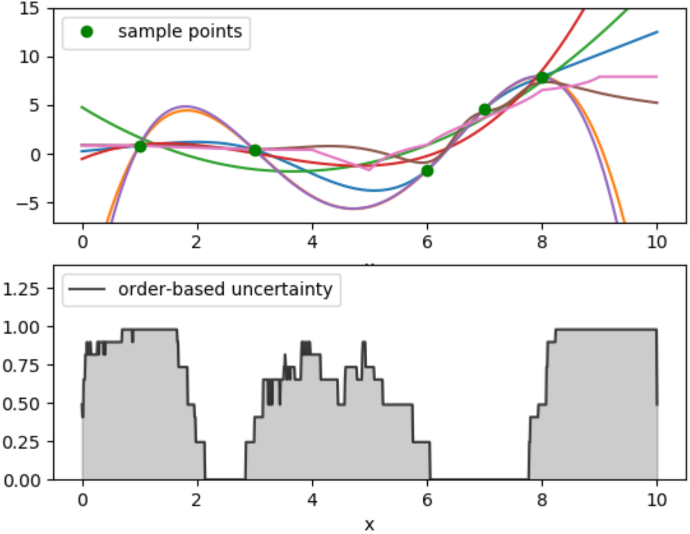


$\mathcal{D} = \{-1, +1\}$

$t = 0.3$

Statistical interpretation of an ensemble of surrogate models

Multiple surrogates on the same sample set
and order-based uncertainty



$\mathcal{D} = \{-1, +1\}$

$t = 0.3$

Statistical interpretation of an ensemble of surrogate models

Adaptation to the constraint j : $\mathcal{S} = \{\tilde{c}_j^1, \tilde{c}_j^2, \ldots, \tilde{c}_j^s\}$

- The number of surrogates that predict feasibility :

$$n_{x<0} = \left|\left\{\tilde{c}_j \in \mathcal{S} : \ \tilde{c}_j(x) < 0\right\}\right|$$

Statistical interpretation of an ensemble of surrogate models

Adaptation to the constraint j : $\mathcal{S} = \{\tilde{c}_j^1, \tilde{c}_j^2, \ldots, \tilde{c}_j^s\}$

- The number of surrogates that predict feasibility :

$$n_{x<0} = \left|\left\{\tilde{c}_j \in \mathcal{S} : \; \tilde{c}_j(x) < 0\right\}\right|$$

- The *order-based uncertainty* on the constraint j :

$$\mathcal{U}_{OB}^j(x) = \frac{n_{x<0}(s - n_{x<0})}{(s/2)^2}$$

Statistical interpretation of an ensemble of surrogate models

Adaptation to the constraint j : $\mathcal{S} = \{\tilde{c}_j^1, \tilde{c}_j^2, \ldots, \tilde{c}_j^s\}$

- The number of surrogates that predict feasibility :

$$n_{x<0} = \left|\left\{\tilde{c}_j \in \mathcal{S} : \; \tilde{c}_j(x) < 0\right\}\right|$$

- The *order-based uncertainty* on the constraint j :

$$\mathcal{U}_{OB}^j(x) = \frac{n_{x<0}(s - n_{x<0})}{(s/2)^2}$$

$$(n_{x<0} = 0 \text{ or } n_{d<x} = s) \;\Rightarrow\; \frac{n_{x<0}(s - n_{x<0})}{(s/2)^2} = 0$$

Statistical interpretation of an ensemble of surrogate models

Adaptation to the constraint j : $\mathcal{S} = \{\tilde{c}_j^1, \tilde{c}_j^2, \ldots, \tilde{c}_j^s\}$

- The number of surrogates that predict feasibility :

$$n_{x<0} = \left|\left\{\tilde{c}_j \in \mathcal{S} : \; \tilde{c}_j(x) < 0\right\}\right|$$

- The *order-based uncertainty* on the constraint j :

$$\mathcal{U}_{OB}^j(x) = \frac{n_{x<0}(s - n_{x<0})}{(s/2)^2}$$

$$(n_{x<0} = 0 \text{ or } n_{d<x} = s) \;\Rightarrow\; \frac{n_{x<0}(s - n_{x<0})}{(s/2)^2} = 0$$

$$n_{x<0} = s/2 \;\Rightarrow\; \frac{n_{x<0}(s - n_{x<0})}{(s/2)^2} = 1$$

Table of Contents

Incorporation in a black-box optimization algorithm

Initial objective : recover $\tilde{\sigma}(x)$, $\tilde{\sigma}_j(x)$, $EI(x)$, $P(x)$ and $PI(x)$.

Incorporation in a black-box optimization algorithm

Initial objective : recover $\tilde{\sigma}(x)$, $\tilde{\sigma}_j(x)$, $EI(x)$, $P(x)$ and $PI(x)$.

$$\tilde{\sigma}(x) \longrightarrow \mathcal{U}^f_{OB}$$
$$\tilde{\sigma}_j(x) \longrightarrow \mathcal{U}^j_{OB}$$

Incorporation in a black-box optimization algorithm

Initial objective : recover $\tilde{\sigma}(x)$, $\tilde{\sigma}_j(x)$, $EI(x)$, $P(x)$ and $PI(x)$.

$$\begin{aligned} \tilde{\sigma}(x) &\longrightarrow \mathcal{U}^f_{OB} \\ \tilde{\sigma}_j(x) &\longrightarrow \mathcal{U}^j_{OB} \end{aligned}$$

$$\begin{aligned} EI(x) &\longrightarrow ? \\ P(x) &\longrightarrow ? \\ PI(x) &\longrightarrow ? \end{aligned}$$

Incorporation in a black-box optimization algorithm

Future work :

Incorporation in a black-box optimization algorithm

Future work :

- Adapt EI, P and PI

Incorporation in a black-box optimization algorithm

Future work :

- Adapt EI, P and PI
- Implement the method

Incorporation in a black-box optimization algorithm

Future work :

- Adapt EI, P and PI
- Implement the method
- Adjust the "finite difference" parameter t

References I

A.J. Booker, J.E. Dennis, Jr., P.D. Frank, D.B. Serafini, V. Torczon, and M.W. Trosset.
A Rigorous Framework for Optimization of Expensive Functions by Surrogates.
Structural and Multidisciplinary Optimization, 17(1):1–13, 1999.

M.D. Buhmann.
Radial Basis Functions: Theory and Implementations.
Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2003.

M.J.D. Powell.
The theory of radial basis function approximation in 1990.
In W.A. Light, editor, *Advances in Numerical Analysis, Vol. II: Wavelets, Subdivision Algorithms and Radial Basis Functions*, pages 105–210. Oxford University Press, Cambridge, 1992.

A.L. Custódio, H. Rocha, and L.N. Vicente.
Incorporating minimum Frobenius norm models in direct search.
Computational Optimization and Applications, 46(2):265–278, 2010.

A.R. Conn and S. Le Digabel.
Use of quadratic models with mesh-adaptive direct search for constrained black box optimization.
Optimization Methods and Software, 28(1):139–158, 2013.

E. Acar and M. Rais-Rohani.
Ensemble of metamodels with optimized weight factors.
Structural and Multidisciplinary Optimization, 37(3):279–294, 2009.

T. Hastie, R. Tibshirani, and J. Friedman.
The Elements of Statistical Learning.
Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2001.

References II

C.E. Rasmussen and C.K.I. Williams.
Gaussian Processes for Machine Learning.
The MIT Press, 2006.

D.R Jones, M. Schonlau, and W.J. Welch.
Efficient Global Optimization of Expensive Black Box Functions.
Journal of Global Optimization, 13(4):455–492, 1998.

D.R. Jones.
A Taxonomy of Global Optimization Methods Based on Response Surfaces.
Journal of Global Optimization, 21:345–383, 2001.

B. Talgorn, S. Le Digabel, and M. Kokkolaras.
Statistical Surrogate Formulations for Simulation-Based Design Optimization.
Journal of Mechanical Design, 137(2):021405–1–021405–18, 2015.

C. Audet, M. Kokkolaras, S. Le Digabel, and B. Talgorn.
Order-based error for managing ensembles of surrogates in mesh adaptive direct search.
Journal of Global Optimization, 70(3):645–675, 2018.