Interprétation statistique d'un ensemble de substituts

Renaud Saltet

Directeur : Charles Audet ; Codirecteur : Sébastien Le Digabel

Polytechnique Montréal

29 mars 2021

Table des matières

Table des matières

Utilisation de substituts en optimisation de boîtes noires

Vrai problème :

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

Utilisation de substituts en optimisation de boîtes noires

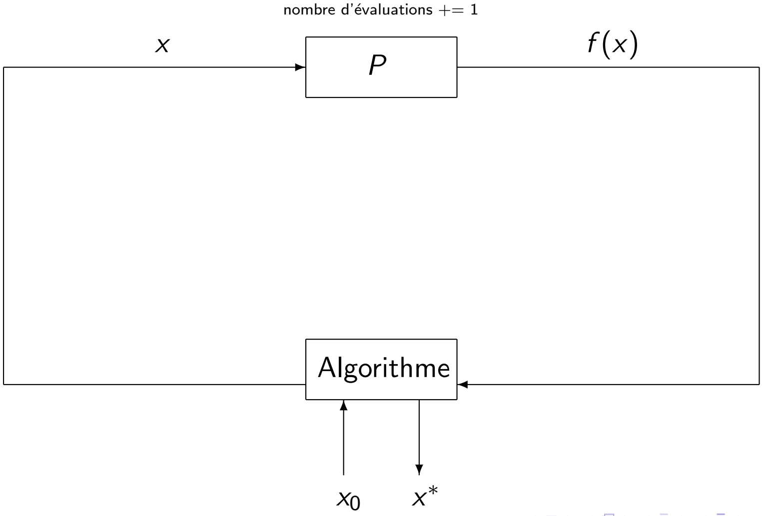
Vrai problème :

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

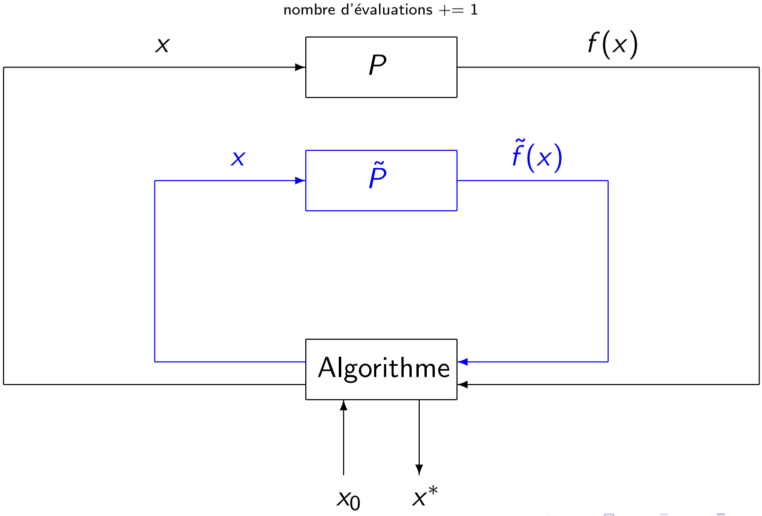
Problème substitut :

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

Utilisation de substituts en optimisation de boîtes noires



Utilisation de substituts en optimisation de boîtes noires



Utilisation de substituts en optimisation de boîtes noires

Article fondateur : Booker et al. (1999) [1]

Types de modèles :

- Fonctions à base radiale [2, 3]
- Modèles quadratiques [4, 5]
- Lissage par noyau [6, 7]
- Processus gaussiens [8, 9, 10]
- ...

Table des matières

Ensemble de modèles

Il n'y a pas *un* modèle meilleur que les autres dans toutes les situations.

Ensemble de modèles

Il n'y a pas *un* modèle meilleur que les autres dans toutes les situations.

Modèle aggrégé :

$$\hat{f}(x) = \sum_{k=1}^{s} w^k \tilde{f}^k(x)$$

- $\{\tilde{f}^1, \tilde{f}^2, \ldots, \tilde{f}^s\}$ est un ensemble de modèles de l'objectif f
- $\{w^1, w^2, \ldots, w^s\}$ est une collection de poids non négatifs dont la somme vaut 1

Ensemble de modèles

Il n'y a pas *un* modèle meilleur que les autres dans toutes les situations.

Modèle aggrégé :

$$\hat{f}(x) = \sum_{k=1}^{s} w^k \tilde{f}^k(x)$$

- $\{\tilde{f}^1, \tilde{f}^2, \ldots, \tilde{f}^s\}$ est un ensemble de modèles de l'objectif f
- $\{w^1, w^2, \ldots, w^s\}$ est une collection de poids non négatifs dont la somme vaut 1

Idem pour c_j, $j \in \{1, 2, \ldots, m\}$:

$$\hat{c}_j(x) = \sum_{k=1}^{s} w_j^k \tilde{\sigma}_j^k(x)$$

Ensemble de modèles

Calcul des poids w^k :

Ensemble de modèles

Calcul des poids w^k :

$$w^k = g(\mathcal{E}^k)$$

Ensemble de modèles

Calcul des poids w^k :

$$w^k = g(\mathcal{E}^k)$$

- $\mathcal{E}^k$ est une mesure d'erreur du modèle $\tilde{f}^k$ calculée grâce à l'ensemble d'échantillonnage

Ensemble de modèles

Calcul des poids w^k :

$$w^k = g(\mathcal{E}^k)$$

- $\mathcal{E}^k$ est une mesure d'erreur du modèle $\tilde{f}^k$ calculée grâce à l'ensemble d'échantillonnage

- g est une fonction qui respecte $w^k \geq 0,\ \forall k \in \{1, \ldots, s\}$ et $\sum_{k=1}^{s} w^k = 1$, par exemple :

$$w^k \propto \mathcal{E}^{\text{tot}} - \mathcal{E}^k$$
$$w^k \propto \mathbb{1}_{\mathcal{E}^k = \mathcal{E}^{\min}}$$
$$w^k \propto (\mathcal{E}^k + \alpha \mathcal{E}^{\text{moy}})^{\beta}$$

Table des matières

Modèles stochastiques

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

Modèles stochastiques

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

Modèles stochastiques

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

Modèles stochastiques

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

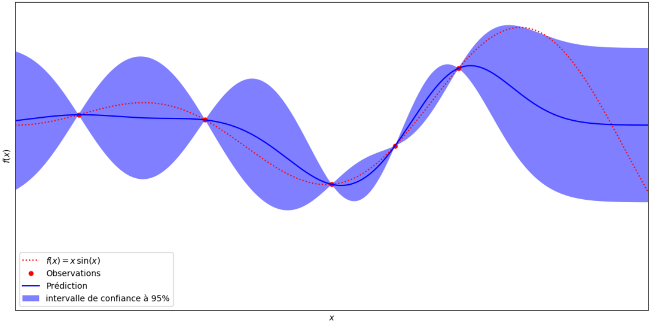


Figure: Processus gaussien sur la fonction $f : x \mapsto x \sin x$

Modèles stochastiques

Fonction d'acquisition : $EI(x) = \mathbb{E}[\max\{fmin - f(x), 0\}]$

Modèles stochastiques

Fonction d'acquisition : $EI(x) = \mathbb{E}[\max\{fmin - f(x), 0\}]$

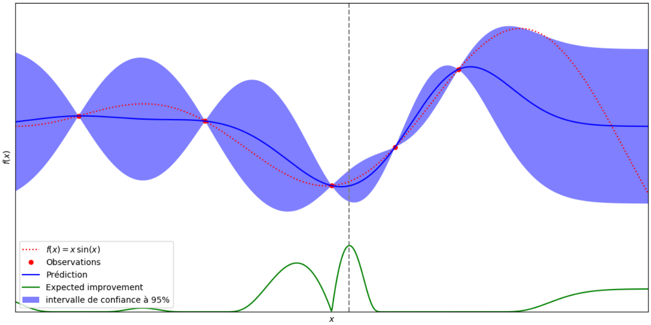


Figure: Processus gaussien et EI sur la fonction $f : x \mapsto x \sin x$

Modèles stochastiques

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

Modèles stochastiques

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

Modèles stochastiques

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

$$
\begin{array}{ll}
\min\limits_{x \in \mathbb{R}^n} & \tilde{f}(x) - \lambda \tilde{\sigma}(x) \\
s.c. & \tilde{c}_j(x) - \lambda \tilde{\sigma}_j(x) \leq 0, \quad \forall j \in \{1, 2, \ldots, m\}
\end{array}
\qquad (\tilde{P})
$$

avec $\lambda \in [0, 1]$

Modèles stochastiques

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

$$\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & \tilde{f}(x) - \lambda \tilde{\sigma}(x) \\ s.c. & \tilde{c}_j(x) - \lambda \tilde{\sigma}_j(x) \leq 0, \quad \forall j \in \{1, 2, \ldots, m\} \end{array} \qquad (\tilde{P})$$

avec $\lambda \in [0, 1]$

Voir [11] pour des exemples de telles formulations.

Table des matières

Interprétation statistique d'un ensemble de modèles

Ensemble de modèles :

$$\hat{f}(x) = \sum_{k=1}^{s} w^k \tilde{f}^k(x)$$

$$\hat{c}_j(x) = \sum_{k=1}^{s} w_j^k \tilde{c}_j^k(x), \; j \in \{1, 2, \ldots, m\}$$

Interprétation statistique d'un ensemble de modèles

Ensemble de modèles :

$$\hat{f}(x) = \sum_{k=1}^{s} w^k \tilde{f}^k(x)$$

$$\hat{c}_j(x) = \sum_{k=1}^{s} w_j^k \tilde{c}_j^k(x), \; j \in \{1, 2, \ldots, m\}$$

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

Interprétation statistique d'un ensemble de modèles

Ensemble de modèles :

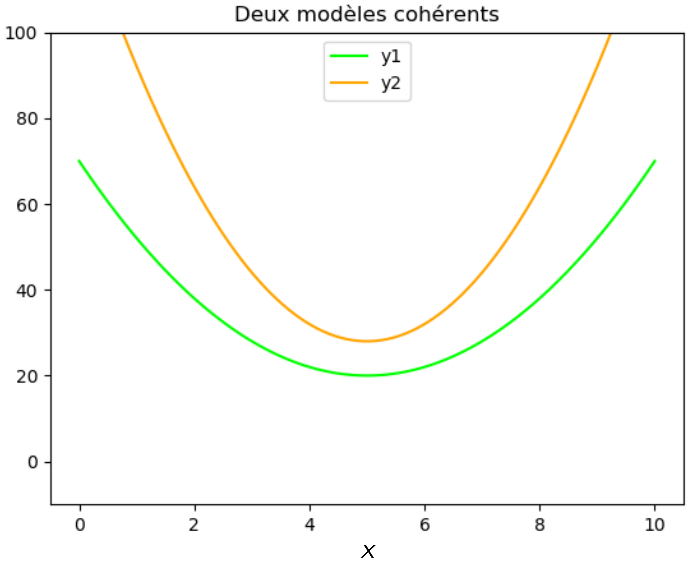
$$\hat{f}(x) = \sum_{k=1}^{s} w^k \tilde{f}^k(x)$$

$$\hat{c}_j(x) = \sum_{k=1}^{s} w_j^k \tilde{c}_j^k(x), \; j \in \{1, 2, \ldots, m\}$$

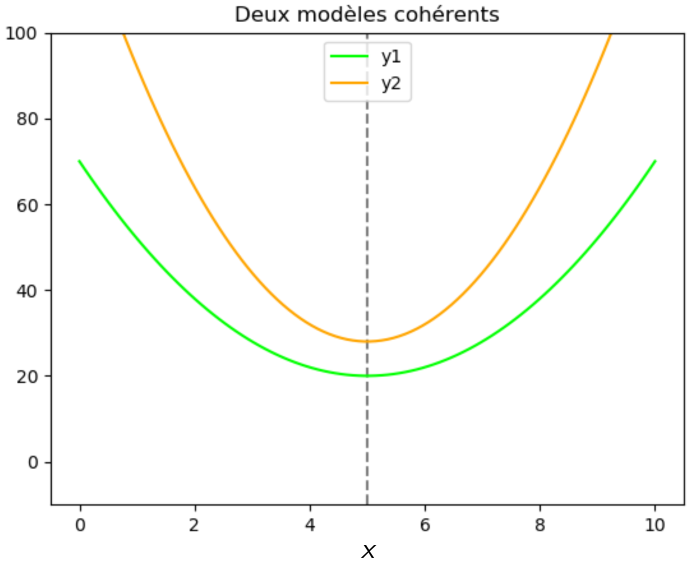
$\tilde{\sigma}(x)$? $\tilde{\sigma}_j(x)$?

Comment imiter un modèle *stochastique* avec plusieurs modèles déterministes ?

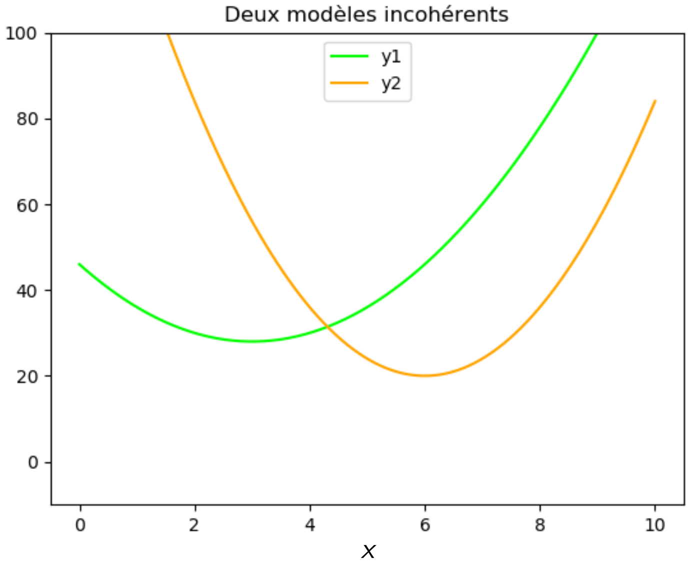
Interprétation statistique d'un ensemble de modèles



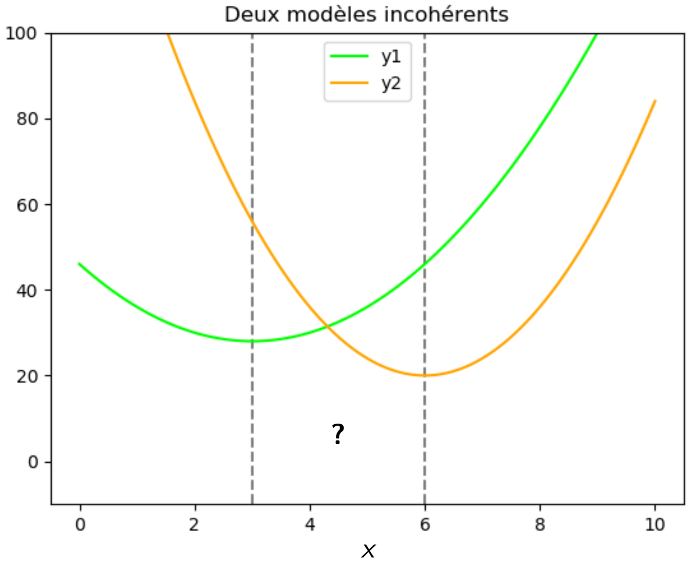
Interprétation statistique d'un ensemble de modèles



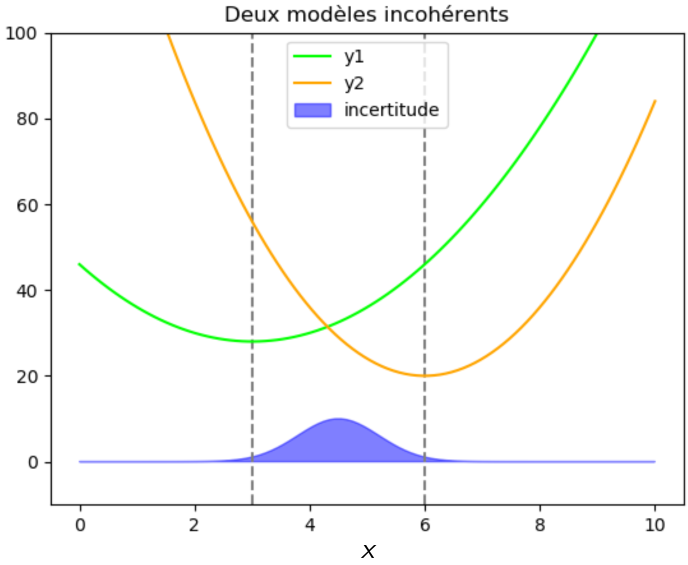
Interprétation statistique d'un ensemble de modèles



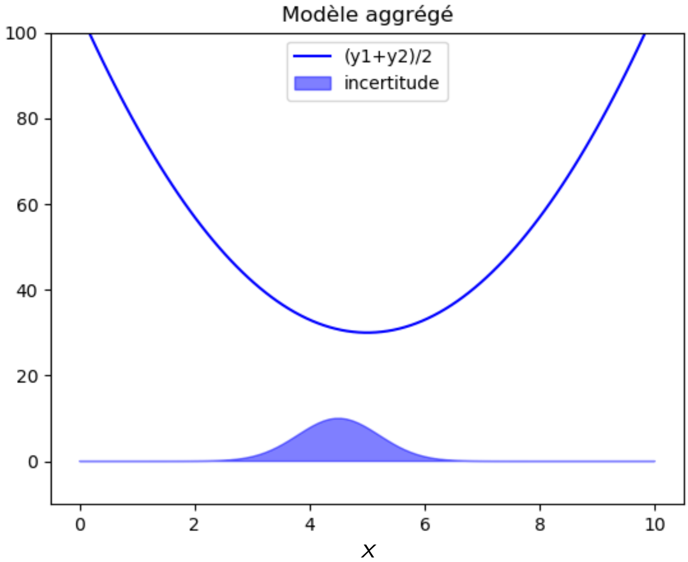
Interprétation statistique d'un ensemble de modèles



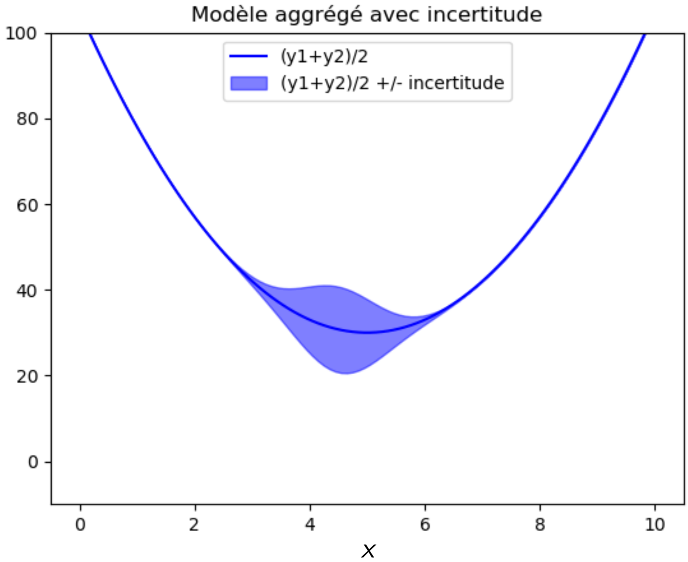
Interprétation statistique d'un ensemble de modèles



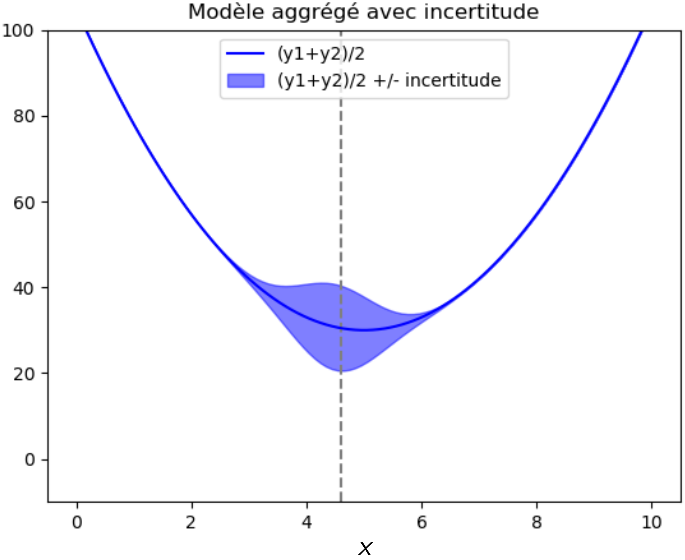
Interprétation statistique d'un ensemble de modèles



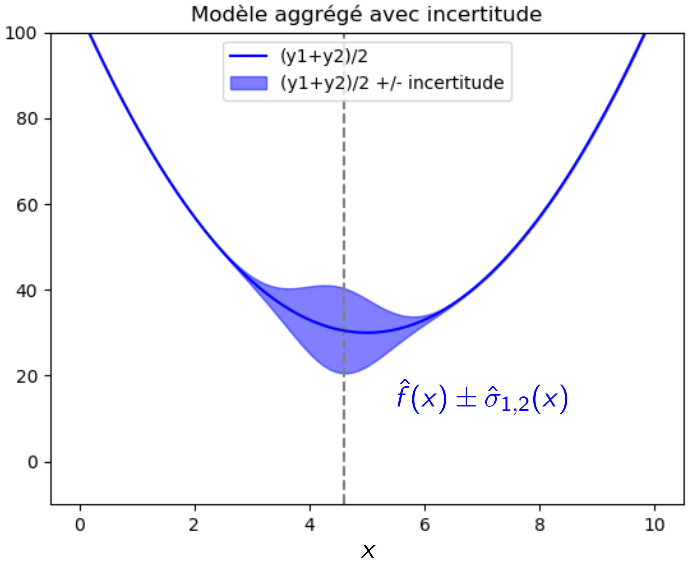
Interprétation statistique d'un ensemble de modèles



Interprétation statistique d'un ensemble de modèles



Interprétation statistique d'un ensemble de modèles



Interprétation statistique d'un ensemble de modèles

1) Quelle expression pour $\hat{\sigma}_{1,2}(x)$?

2) Comment généraliser à $s \geq 2$ modèles ?

Interprétation statistique d'un ensemble de modèles

1) Quelle expression pour $\hat{\sigma}_{1,2}(x)$?

Interprétation statistique d'un ensemble de modèles

1) Quelle expression pour $\hat{\sigma}_{1,2}(x)$?

- Version *lisse* :

$$\hat{\sigma}_{1,2}(x) = \frac{1}{2}\left(1 - \cos\left\langle \nabla_S \tilde{f}^1(x),\ \nabla_S \tilde{f}^2(x) \right\rangle\right)$$

où $\nabla_S \tilde{f}$ désigne le gradient simplexe de $\tilde{f}$

- Version *non-lisse* :

$$\hat{\sigma}_{1,2}(x) = \frac{1}{|\mathcal{D}|} \sum_{d \in \mathcal{D}} \mathrm{xor}\left(\tilde{f}^1(x+td) < \tilde{f}^1(x)\ ,\ \tilde{f}^2(x+td) < \tilde{f}^2(x) \right)$$

où $\mathcal{D}$ est un ensemble générateur positif et t est un paramètre

Interprétation statistique d'un ensemble de modèles

2) Comment généraliser à $s \geq 2$ modèles ?

Interprétation statistique d'un ensemble de modèles

2) Comment généraliser à $s \geq 2$ modèles ?

$$\hat{\sigma}(x) = \alpha \, \frac{\displaystyle\sum_{k=1}^{s-1} \sum_{\ell=k+1}^{s} w^k w^\ell \times \hat{\sigma}_{k,\ell}(x)}{\displaystyle\sum_{k=1}^{s-1} \sum_{\ell=k+1}^{s} w^k w^\ell}$$

Interprétation statistique d'un ensemble de modèles

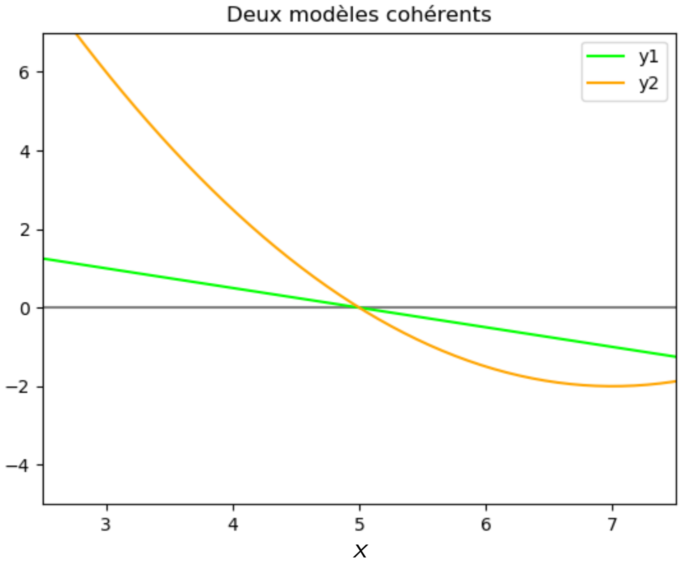
2) Comment généraliser à $s \geq 2$ modèles ?

$$\hat{\sigma}(x) = \alpha \frac{\sum_{k=1}^{s-1} \sum_{\ell=k+1}^{s} w^k w^\ell \times \hat{\sigma}_{k,\ell}(x)}{\sum_{k=1}^{s-1} \sum_{\ell=k+1}^{s} w^k w^\ell}$$

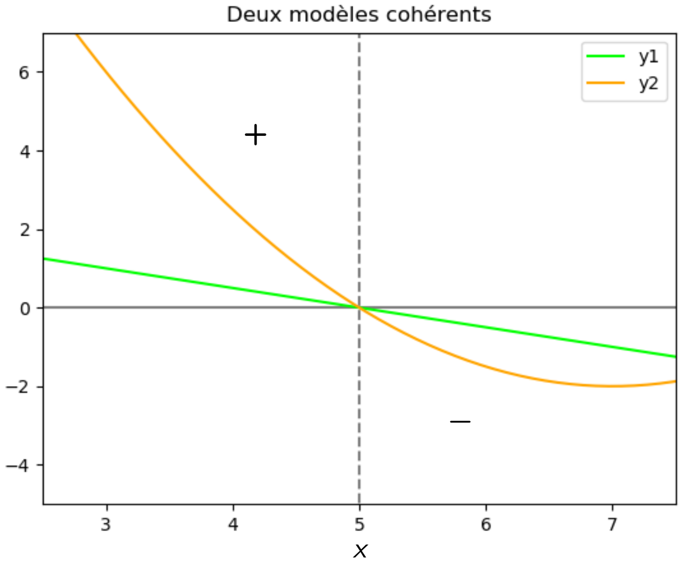
$$= \alpha \frac{w^\top \Sigma(x) w}{w^\top T w}$$

- $\Sigma(x) \in \mathbb{R}^{s \times s}$ est la matrice telle que $[\Sigma(x)]_{k,\ell} = \hat{\sigma}_{k,\ell}(x)$ si $k < \ell$ et 0 sinon
- $T \in \mathbb{R}^{s \times s}$ est la matrice telle que $[T]_{k\ell} = 1$ si $k < \ell$ et 0 sinon
- $w = [w^1, \ldots, w^s]^\top$
- $\alpha := \mathrm{Var}(f(\mathbb{X}))$ où $\mathbb{X}$ est la cache

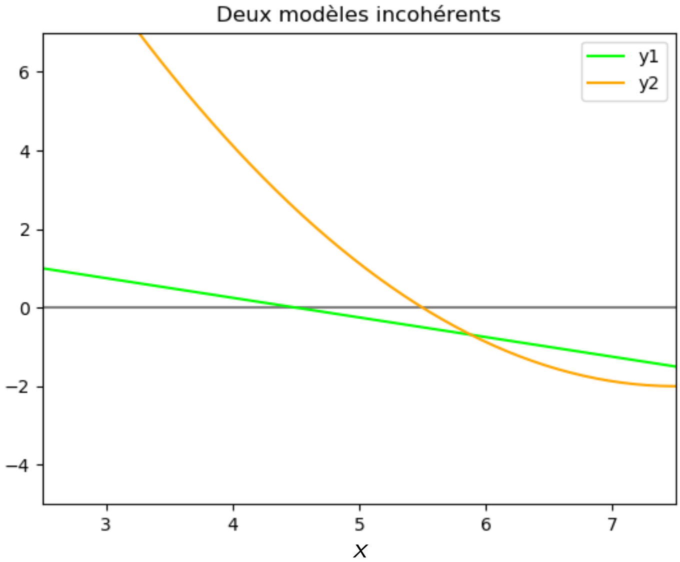
Interprétation statistique d'un ensemble de modèles



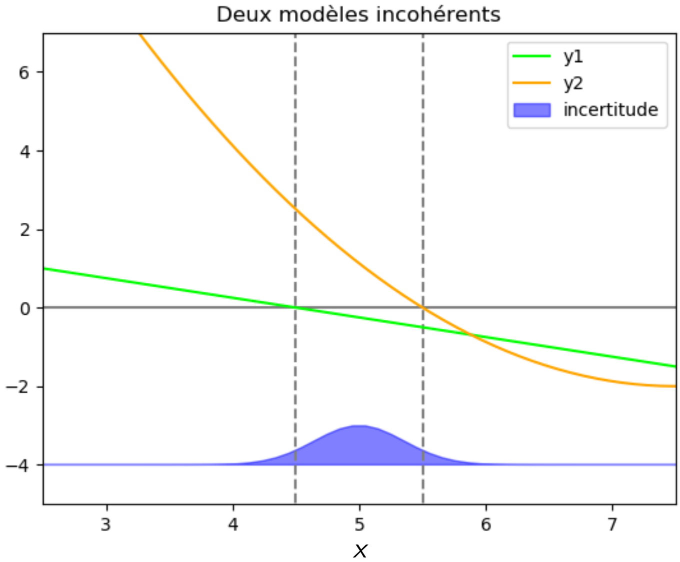
Interprétation statistique d'un ensemble de modèles



Interprétation statistique d'un ensemble de modèles



Interprétation statistique d'un ensemble de modèles



Interprétation statistique d'un ensemble de modèles

Pour la contrainte c_j, $j \in \{1, 2, \ldots, m\}$:

- Version *lisse* :

$$\hat{\sigma}_{k,\ell}(x) = \mathrm{sigm}\left(-\tilde{c}_j^k(x) \times \tilde{c}_j^\ell(x)\right)$$

  où $\mathrm{sigm}(\cdot)$ est la fonction sigmoïde

- Version *non lisse* :

$$\hat{\sigma}_{k,\ell}(x) = \mathrm{xor}\left(\tilde{c}_j^k(x) \leq 0\ ,\ \tilde{c}_j^\ell(x) \leq 0\right)$$

Interprétation statistique d'un ensemble de modèles

Pour la contrainte c_j, $j \in \{1, 2, \ldots, m\}$:

- Version *lisse* :

$$\hat{\sigma}_{k,\ell}(x) = \mathrm{sigm}\left(-\tilde{c}_j^k(x) \times \tilde{c}_j^\ell(x)\right)$$

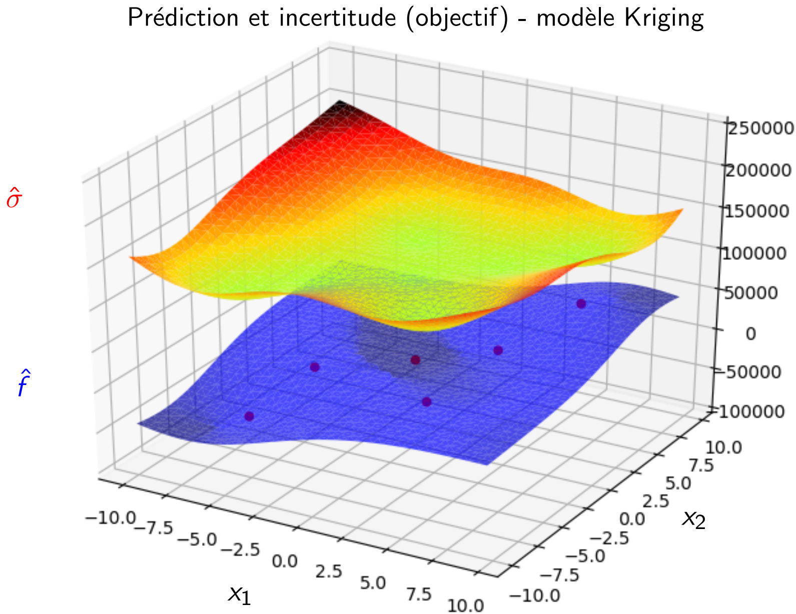
  où $\mathrm{sigm}(\cdot)$ est la fonction sigmoïde

- Version *non lisse* :

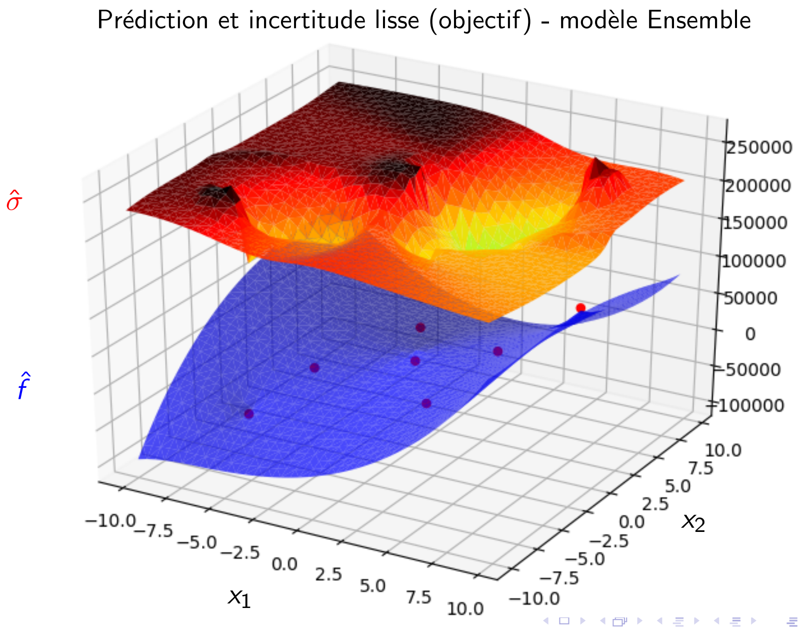
$$\hat{\sigma}_{k,\ell}(x) = \mathrm{xor}\left(\tilde{c}_j^k(x) \leq 0 \ , \ \tilde{c}_j^\ell(x) \leq 0\right)$$

$$\longrightarrow \qquad \hat{\sigma}(x) = \alpha \frac{w^\top \Sigma(x) w}{w^\top T w}$$

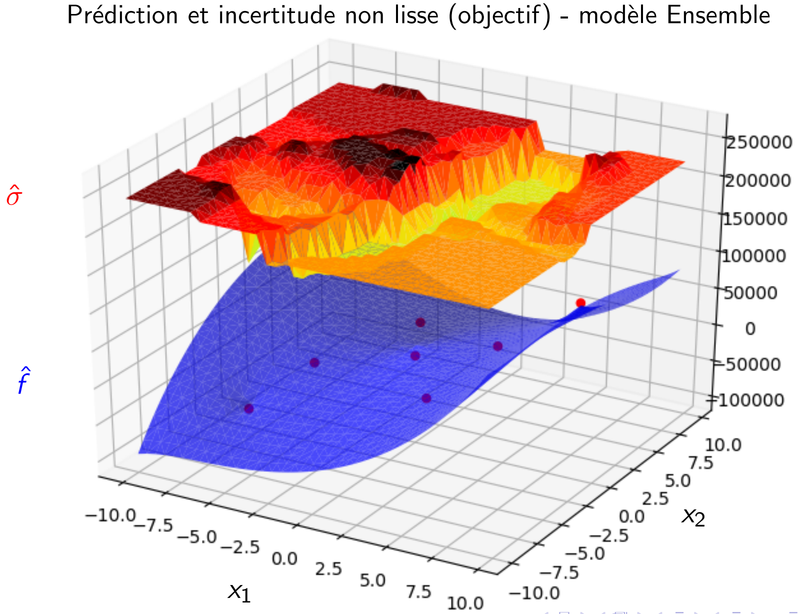
Interprétation statistique d'un ensemble de modèles



Interprétation statistique d'un ensemble de modèles

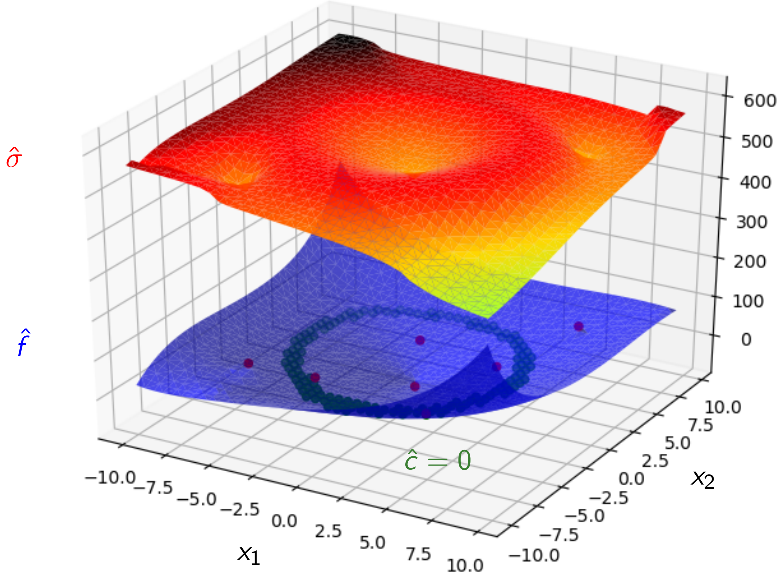


Interprétation statistique d'un ensemble de modèles



Interprétation statistique d'un ensemble de modèles

Prédiction et incertitude lisse (contrainte) - modèle Ensemble



Interprétation statistique d'un ensemble de modèles

Prédiction et incertitude non lisse (contrainte) - modèle Ensemble

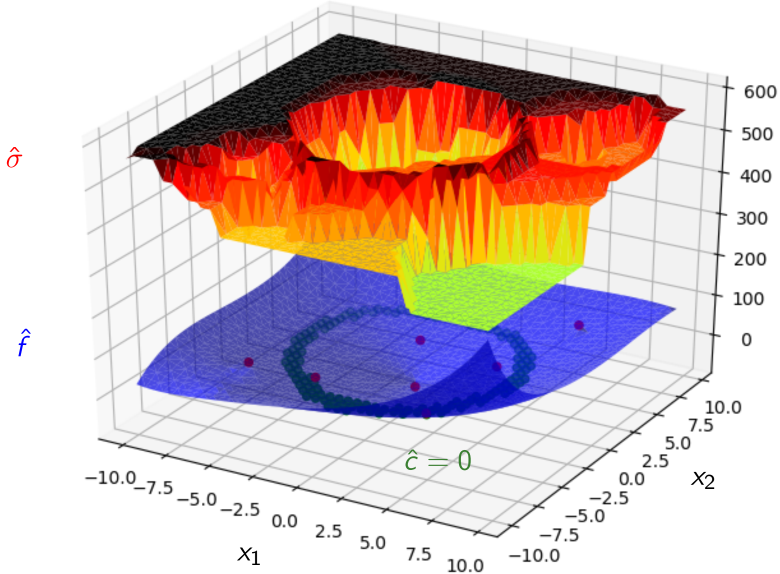


Table des matières

Résultats

- Problèmes analytiques

- Incertitude lisse *vs.* non lisse $\left(\hat{\sigma}(x) = \alpha \frac{w^\top \Sigma(x) w}{w^\top T w}\right)$ *vs.* basée sur la distance aux point d'échantillonnage :

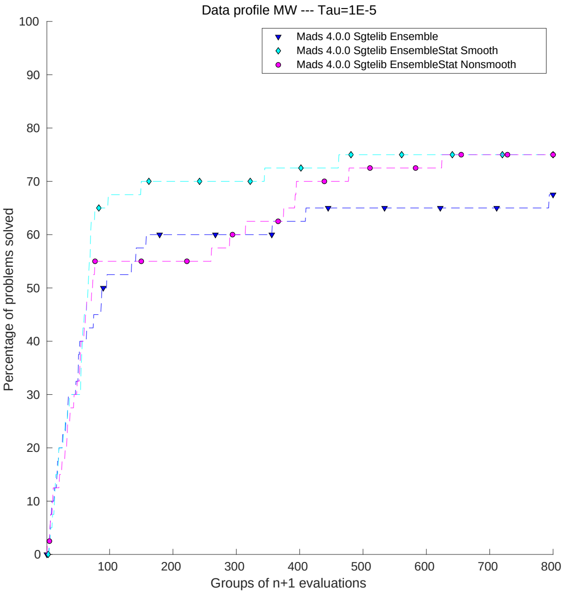
$$\hat{\sigma}(x) \propto \min_{y \in \mathbb{X}} ||x - y||_1$$

- Formulation du sous-problème $(\tilde{P})$ de la *search* :

$$\begin{aligned} \min_{x \in \mathcal{X}} \quad & \hat{f}(x) - 0.1\hat{\sigma}(x) \\ s.c. \quad & \tilde{c}_j(x) - 0.1\tilde{\sigma}_j(x) \leq 0, \quad \forall j \in \{1, 2, \ldots, m\} \end{aligned}$$

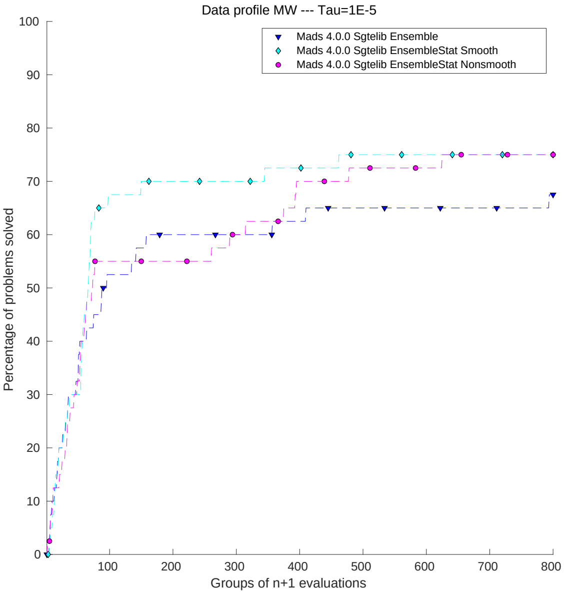
Résultats

20 problèmes non contraints



Résultats

20 problèmes non contraints



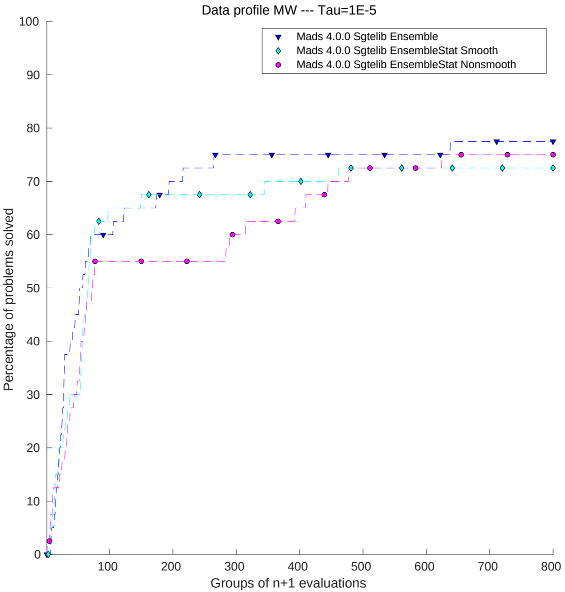
$w^k \propto \mathcal{E}^{\text{tot}} - \mathcal{E}^k$

$w^k \propto \mathcal{E}^{\text{tot}} - \mathcal{E}^k$

$w^k \propto \mathcal{E}^{\text{tot}} - \mathcal{E}^k$

Résultats

20 problèmes non contraints



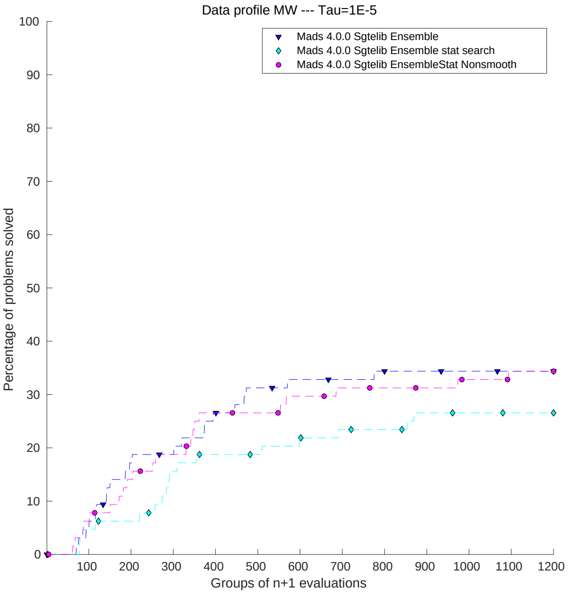
$$w^k \propto \mathcal{E}^{\text{tot}} - \mathcal{E}^k$$

$$w^k \propto \mathcal{E}^{\text{tot}} - \mathcal{E}^k$$

$$w^k = \mathbb{1}_{\mathcal{E}^k = \mathcal{E}^{\min}}$$

Résultats

16 problèmes contraints



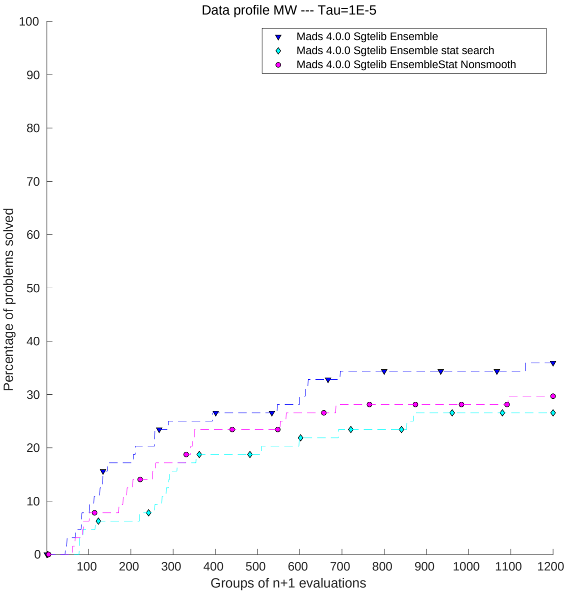
$w^k \propto \mathcal{E}^{\text{tot}} - \mathcal{E}^k$

$w^k \propto \mathcal{E}^{\text{tot}} - \mathcal{E}^k$

$w^k \propto \mathcal{E}^{\text{tot}} - \mathcal{E}^k$

Résultats

16 problèmes contraints



$$w^k \propto \mathcal{E}^{\mathrm{tot}} - \mathcal{E}^k$$

$$w^k \propto \mathcal{E}^{\mathrm{tot}} - \mathcal{E}^k$$

$$w^k = \mathbb{1}_{\mathcal{E}^k = \mathcal{E}^{\min}}$$

Résultats

Travail à venir :

- Tester plusieurs formulations de $\tilde{P}$
- Comparer à d'autres solveurs

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.