

Journées Incertitudes et Simulation 2007

CEA DIF

3 et 4 octobre 2007

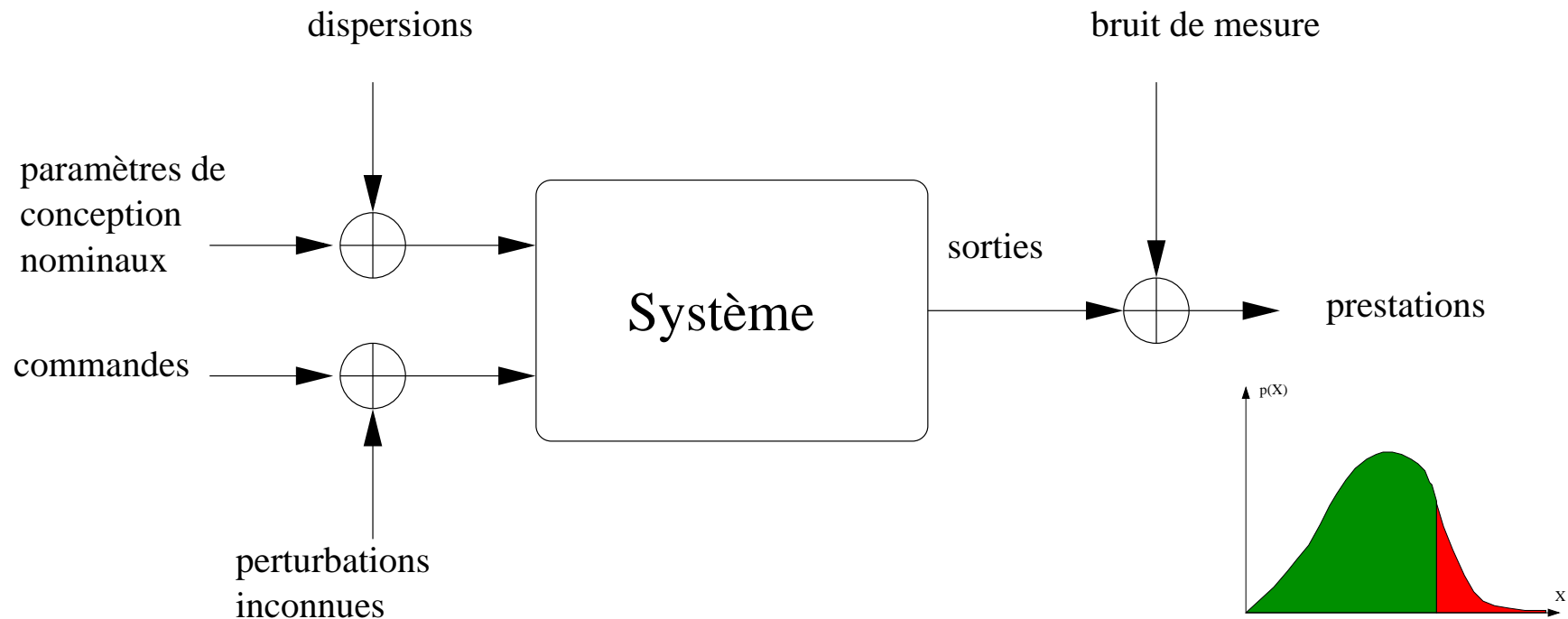
**Construction de modèles de code numérique par
méthodes à noyaux**

Emmanuel Vazquez

SUPELEC

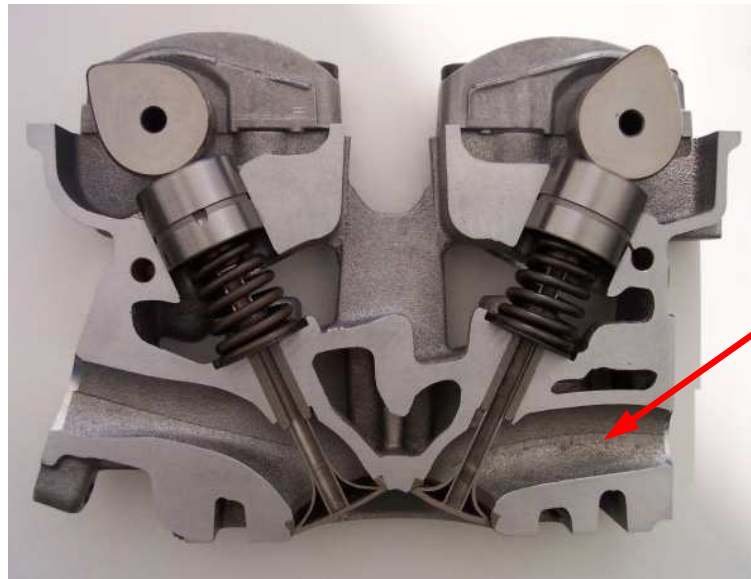
Gif-sur-Yvette

Contexte et objectifs



Modèles de systèmes sous forme de **programmes informatiques complexes** et **coûteux en ressources** (modèles physiques couplés, éléments finis...)

Exemple : simulation d'un conduit d'admission d'un moteur à explosion
(Thèse CIFRE RENAULT-SUPELEC de Julien Villemonteix)



conduit d'admission

- Forme du conduit paramétrée (5 à 20 facteurs)
- Importance de la forme du conduit :
 - impact sur la performance moteur (débit d'air)
 - impact sur les émissions polluantes (turbulences)
- 1 simulation = 5 à 24 heures de calcul

- **Objectifs**

- $\mathbb{X} \subseteq \mathbb{R}^d$: domaine des facteurs (entrées) du système
- $f : \mathbb{X} \rightarrow \mathbb{R}$: fonction de performance (une fonction des sorties)
- **Problème inverse** : recherche des facteurs qui réalisent une valeur donnée de la sortie : x^* t.q. $f(x^*) = u_0$
- **Optimisation** des paramètres de conception : maximiser des performances, minimiser des violations de cahier des charges...

$$x^* = \arg \max_{x \in \mathbb{X}} f(x)$$

- **Estimation de probabilité de défaillance**, c-à-d

$$\mathcal{P}_u = \mathbb{P}\{f(X) > u\}$$

avec $X \in \mathbb{X}$ de loi μ : facteurs aléatoires

Utilisation de modèles (ou méta-modèles)

- Construction d'une approximation f_n (peu coûteuse) de f à partir d'évaluations

$$f_1^{\text{obs}} = f(x_1), \dots, f_n^{\text{obs}} = f(x_n)$$

- Remplacer f par f_n pour optimiser, estimer la probabilité de défaillance, etc.
- Si f **suffisamment régulière**, convergence (attendue) rapide de f_n vers f , et par suite, des estimateurs

Comment construire f_n ?

⇒ Méthodes paramétriques ou non-paramétriques

⇒ Deux points de vue **équivalents** :

□ **Régression à noyau reproduisant**

- 1960 : splines, (**Schoenberg 1964, Duchon 1976–1979**)
- 1980 : RBF, (**Micchelli 1986, Powel 1987**)
- 1995 : SVM, (**Vapnik 1995**)
- 1997 : SVR, (**Smola 1997**)
- 1999 : SVR semi-paramétrique (**Smola 1999**)

□ **prédiction linéaire – Krigeage**

- 1950 : prédiction pour la recherche minière (**Krige 1951**)
- 1960 : krigeage, géostatistique (**Matheron 1963**) – École des Mines
- 1970 : krigeage intrinsèque (**Matheron 1971**)
- 1997 : prédiction par « processus gaussiens », (**Williams 1997, Neal 1997**)

Éléments sur la régression à noyau et le krigéage

Régression à noyau

- Espace de fonctions hilbertien à noyau reproduisant (RKHS)
 - Soit $k : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}$ une fonction symétrique définie positive (s.d.p.).
 - Considérons

$$\tilde{\mathcal{F}} = \text{vect} \left\{ f = \sum_{i=1}^n \lambda_i k(x_i, \cdot) \right\}.$$

- Munissons $\tilde{\mathcal{F}}$ du produit scalaire défini par

$$(k(x, \cdot), k(y, \cdot))_{\tilde{\mathcal{F}}} = k(x, y),$$

- alors le complété \mathcal{F} de $\tilde{\mathcal{F}}$ est un RKHS, puisque $\forall f \in \mathcal{F}$

$$(f, k(x, \cdot))_{\mathcal{F}} = f(x).$$

-
- Régression régularisée dans un RKHS (deux versions)
 - approximation

$$\begin{array}{l} \text{minimiser} \\ \underbrace{\|f_n\|_{\mathcal{F}}^2}_{\text{régularité}} + C \underbrace{\sum_{i=1}^n l(f_n(x_i) - f_i^{\text{obs}})}_{\text{adéquation aux données}} \end{array}$$

(Tikhonov et Arsenin 1977)

- interpolation

$$\begin{array}{l} \text{minimiser} \\ \text{sous contraintes} \end{array} \quad \begin{array}{l} \|f_n\|_{\mathcal{F}}^2, \quad f_n \in \mathcal{F} \\ f_n(x_i) = f_i^{\text{obs}}, \quad \forall i = 1, \dots, n. \end{array}$$

- En principe, dans le cadre de l'approximation de codes numériques, on préférera une interpolation pour construire f_n (bruit d'observation négligeable)

- **Théorème du représentant** (Kimeldorf, Wahba 1970)

Dans un RKHS \mathcal{F} , le programme

$$\begin{array}{ll} \text{minimiser} & \|f_n\|_{\mathcal{F}}^2, \quad f_n \in \mathcal{F} \\ \text{sous contraintes} & f_n(x_i) = f_i^{\text{obs}}, \quad \forall i = 1, \dots, n \end{array}$$

admet une solution **unique** qui s'écrit

$$f_n(x) = \sum_{i=1}^n a_i k(x, x_i)$$

où $\mathbf{a}_n = [a_1, \dots, a_n]^T$ est solution de

$$\mathbf{K}_n \mathbf{a}_n = \mathbf{f}_n^{\text{obs}}$$

avec \mathbf{K}_n matrice $n \times n$ d'éléments $k(x_i, x_j)$ et $\mathbf{f}_n^{\text{obs}}$ vecteur des évaluations $f_1^{\text{obs}}, \dots, f_n^{\text{obs}}$.

Prédiction linéaire – krigage

- $f(x)$ modélisé par un **processus aléatoire** $\xi(x)$
 - de *moyenne nulle* et de covariance $k(x, y)$ (s.d.p)
 - les $f(x_i)$ sont des **réalisations** des variables aléatoires $\xi(x_i)$
 - $\boldsymbol{\xi}_n = (\xi(x_1), \dots, \xi(x_n))^\top$
- on cherche le BLP : **prédicteur linéaire** de $\xi(x)$

$$\xi_n(x) = \boldsymbol{\lambda}_n(x)^\top \boldsymbol{\xi}_n$$

minimisant

$$\text{Var}[\xi(x) - \xi_n(x)] = \mathbb{E} [(\xi(x) - \xi_n(x))^2]$$

-
- $\boldsymbol{\lambda}_n(x)$ solution de

$$\mathbf{K}_n \boldsymbol{\lambda}_n(x) = \mathbf{k}_n(x)$$

avec \mathbf{K}_n matrice de covariance du vecteur aléatoire $\boldsymbol{\xi}_n$ et $\mathbf{k}_n(x)$ vecteur de covariances entre $\boldsymbol{\xi}_n$ et $\xi(x)$

- Variance d'erreur de prédiction :

$$\sigma_n^2(x) = \text{Var}[\xi(x) - \xi_n(x)] = \text{Var}[\xi(x)] - \boldsymbol{\lambda}_n(x)^\top \mathbf{k}_n(x)$$

→ construction d'**intervalles de confiance**

- On utilise ensuite le prédicteur pour construire l'approximation

$$f_n(x) = \boldsymbol{\lambda}_n(x)^\top \mathbf{f}_n^{\text{obs}}$$

→ **moyenne a posteriori** de $\xi(x)$

-
- **Équivalence interpolation à noyau – krigeage** : il suffit de remarquer que

$$\begin{aligned} f_n(x) &= \boldsymbol{\lambda}_n(x)^\top \mathbf{f}_n^{\text{obs}} \\ &= (\mathbf{K}_n^{-1} \mathbf{k}_n(x))^\top \mathbf{f}_n^{\text{obs}} = (\mathbf{K}_n^{-1} \mathbf{f}_n^{\text{obs}})^\top \mathbf{k}_n(x) \\ &= \mathbf{a}_n^\top \mathbf{k}_n(x) \end{aligned}$$

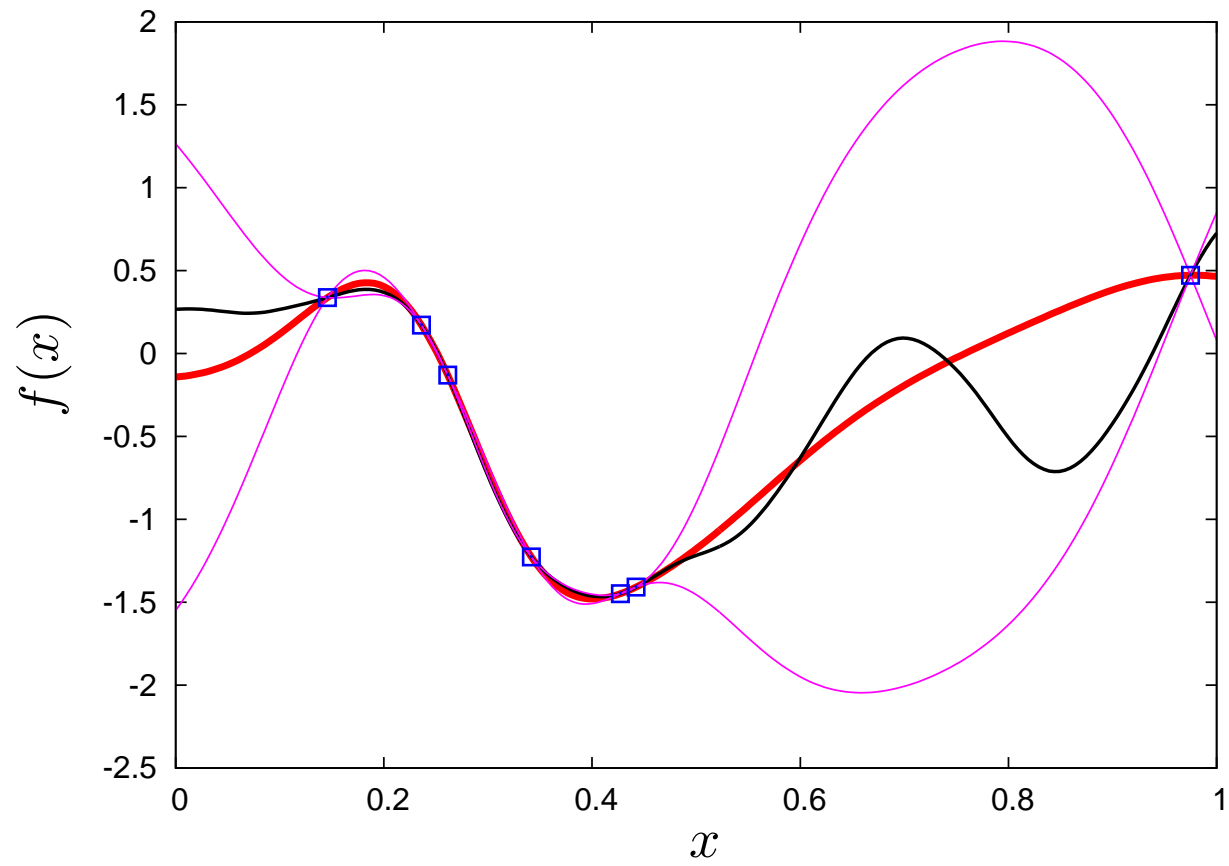
- en pratique, les noyaux choisis sont des fonctions s.d.p. **paramétrées** et **invariantes par translation**, c-à-d

$$k(x, y) = k_{\text{param}}(x - y)$$

avec, par exemple,

$$k_{\text{param}}(h) = \exp \left(- \left(\sum_{i=1}^d (h_i / \rho_i)^2 \right)^\alpha \right), \quad h = (h_1, \dots, h_d)^\top$$

- les paramètres des noyaux peuvent être estimés par **maximum de vraisemblance**, par **MAP**, par **validation croisée**...



Convergence du BLP

- L'estimateur est **consistant** pour (presque) tous les noyaux que l'on considère habituellement
- **Q** : **vitesse de décroissance de $\sigma_n^2(x)$** lorsque la densité des observations au voisinage de x augmente ?

Wu et Schaback 93 (et d'autres) :

- si ξ stationnaire, de covariance $k(h) \in L^2(\mathbb{R}^d)$, \mathbb{X} domaine borné de \mathbb{R}^d , et si la transformée de Fourier de $k(h)$ satisfait

$$c_1(1 + \|\omega\|_2^2)^{-\nu} \leq \tilde{k}(\omega) \leq c_2(1 + \|\omega\|_2^2)^{-\nu} .$$

avec $\nu > d/2$, alors

$$\sup_{\mathbb{X}} \sigma_n(x) \leq Ch_n^{\nu-d/2}$$

où $h_n = \sup_{y \in \mathbb{X}} \min_i \|y - x_i\|_2$ mesure la densité de (x_1, \dots, x_n) dans \mathbb{X} .

Autrement dit,

- f_n est une **interpolation** de f
- l'erreur d'approximation en dehors des points d'évaluation de la fonction **décroît polynomialement** avec le nombre de points
- plus la fonction à approximer est **régulière**, plus **la vitesse de convergence augmente**
- problème de la **dimension de l'espace des facteurs** : « la régularité de la fonction à approximer compense la dimension de l'espace des facteurs »

Éléments de krigeage intrinsèque (RBF – splines)

- **Origine** : Matheron 1973, *The intrinsic random functions, and their applications*

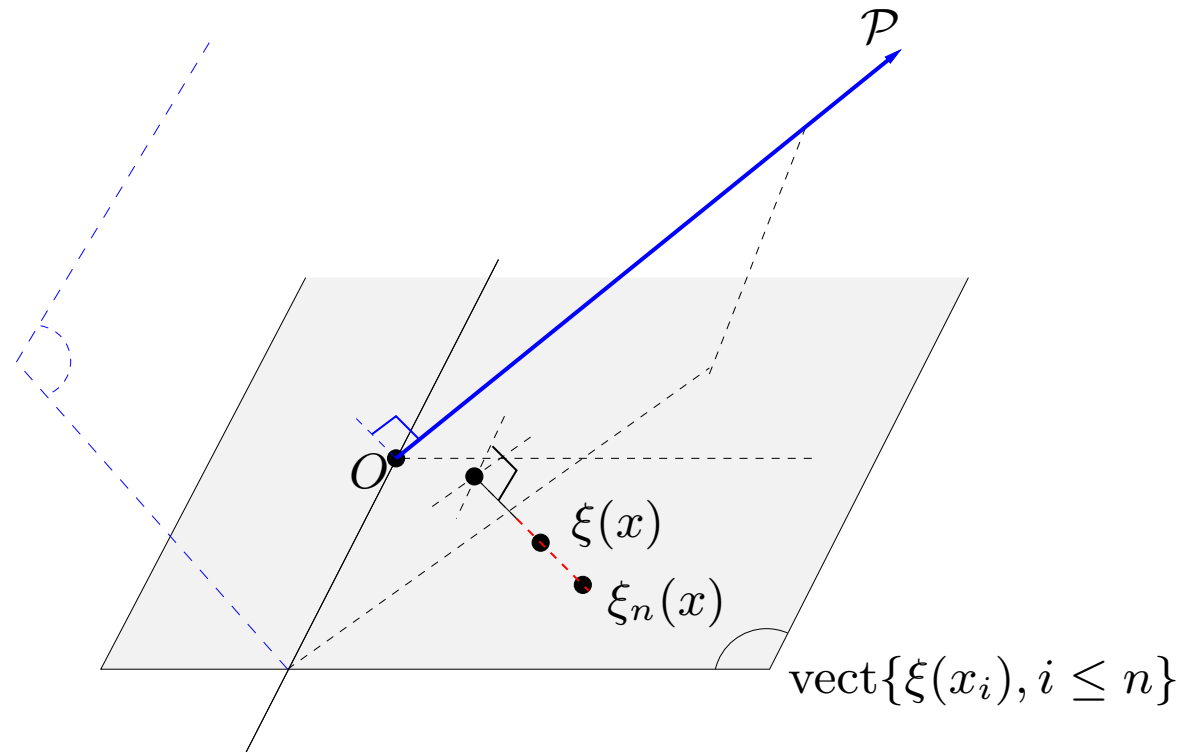
Krigeage intrinsèque \leftrightarrow BLUP d'un processus aléatoire de **moyenne inconnue** dans un espace de fonctions

$$\mathcal{P} = \{b^\top p(x), b \in \mathbb{R}^q\}$$

avec $p(x) : \mathbb{X} \rightarrow \mathbb{R}^q$ tel \mathcal{P} stable par translations (typiquement espace de polynômes).

- **IRF** : processus aléatoires indexés par des mesures à support fini orthogonales à \mathcal{P}
 \Rightarrow classes d'équivalence de processus à moyenne dans \mathcal{P} .

-
- $\forall x$, BLUP $\xi_n(x)$ de $\xi(x) =$ projection linéaire de $\xi(x)$ sur $\text{vect}\{\xi(x_1), \dots, \xi(x_n)\}$ orthogonalement à \mathcal{P}



- Linéarité $\implies \xi_n(x) = \lambda_n(x)\xi_n$

-
- $\lambda_n(x)$ solution de

$$\begin{pmatrix} \mathbf{K}_n & \mathbf{P}^\top \\ \mathbf{P} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \lambda_n(x) \\ \alpha_n(x) \end{pmatrix} = \begin{pmatrix} \mathbf{k}_n(x) \\ \mathbf{p}(x) \end{pmatrix}$$

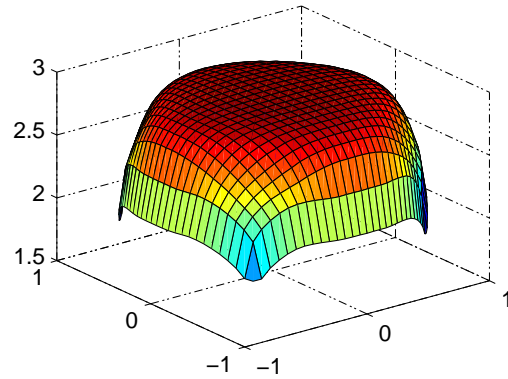
- \mathbf{K}_n matrice de covariance de ξ_n
 - $\mathbf{P} = (p(x_1), \dots, p(x_n))$ matrice $q \times n$
 - $\alpha_n(x)$ vecteur de coefficients de Lagrange
 - $\mathbf{k}_n(x)$ covariance de $\xi(x)$ et ξ_n
- **variance de prédiction :**
 $\sigma_n(x)^2 = \text{Var}[\xi(x) - \xi_n(x)] = k(x, x) - \lambda_n(x)^\top \mathbf{k}_n(x) - \mathbf{p}(x)^\top \alpha_n(x)$

- **remarques**

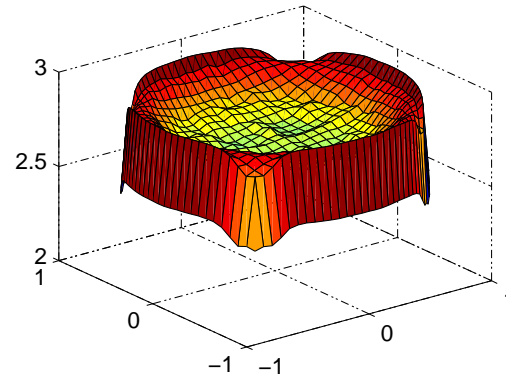
- (i) $\mathbb{E}[\xi(x) - \xi_n(x)]$ **inconnu** ($\rightarrow 0$ asymptotiquement)
- (ii) Il n'est pas nécessaire d'estimer la moyenne de ξ pour calculer ξ_n
- (iii) Possibilité d'utiliser des **covariances généralisées** (fonctions conditionnellement positives), p.ex. $k(x, y) = (-1)^{l+1} \|x - y\|^{2l+1}$ (type de splines plaque mince)

-
- Utilisation du krigeage intrinsèque pour la prise en compte de connaissances a priori
 - Exemple : estimation d'un profil de vitesse dans une conduite d'eau à partir de mesures ponctuelles de la vitesse du fluide

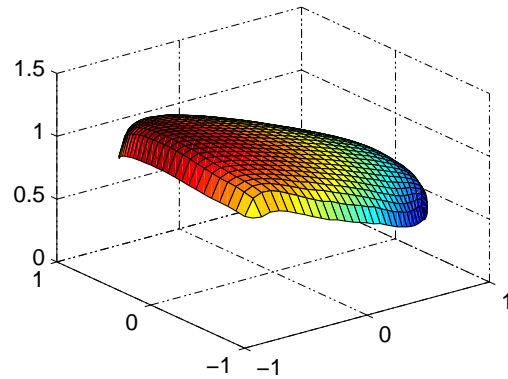
Simulation d'un certain nombre de profils, dans différentes configurations



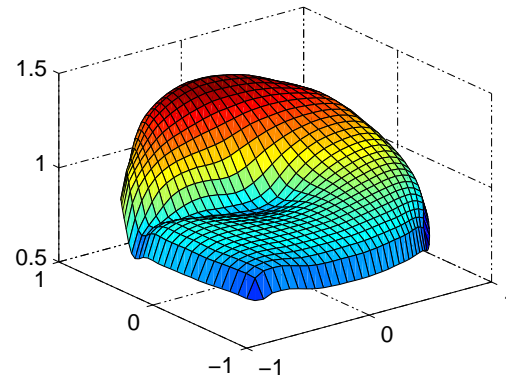
(A)



(B)

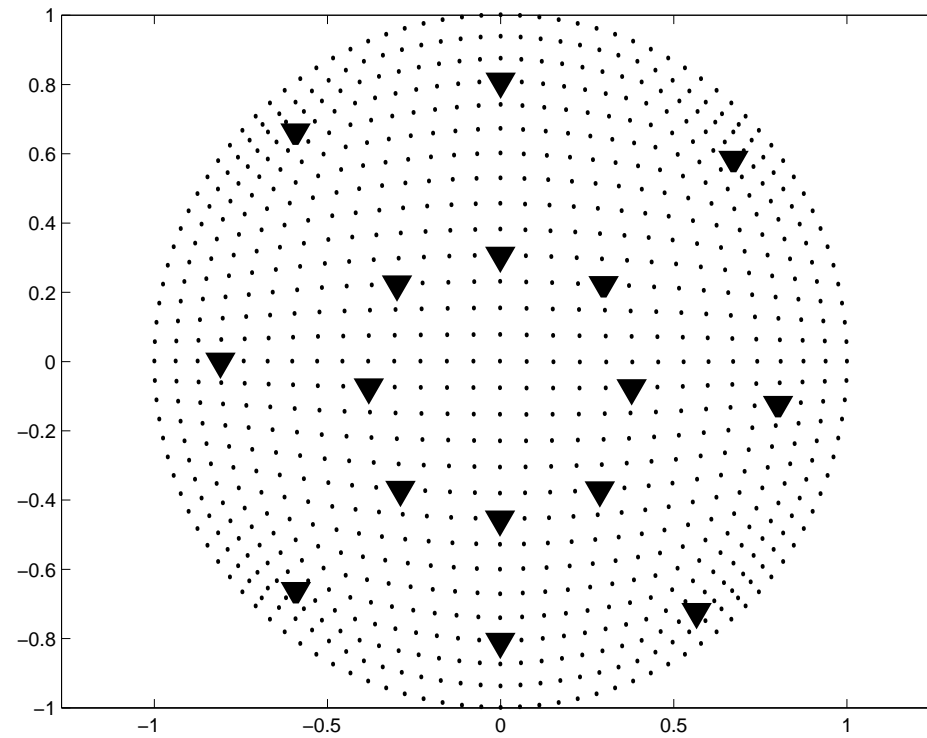


(C)

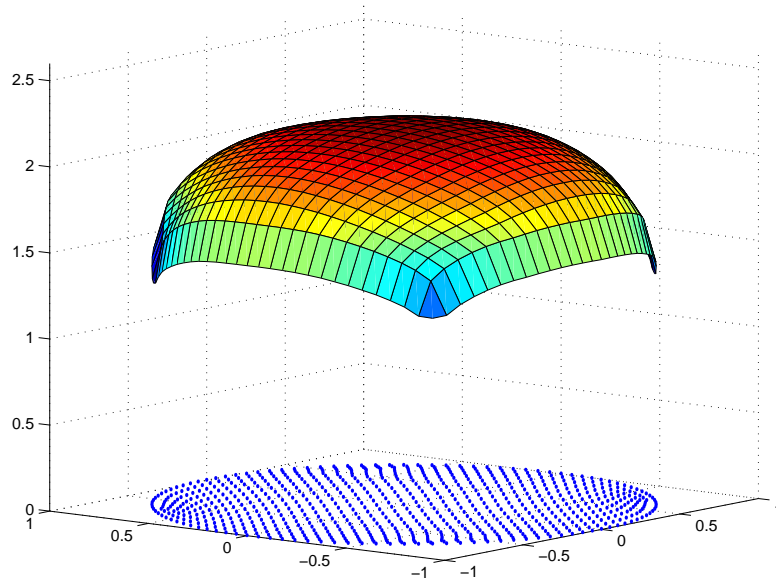


(D)

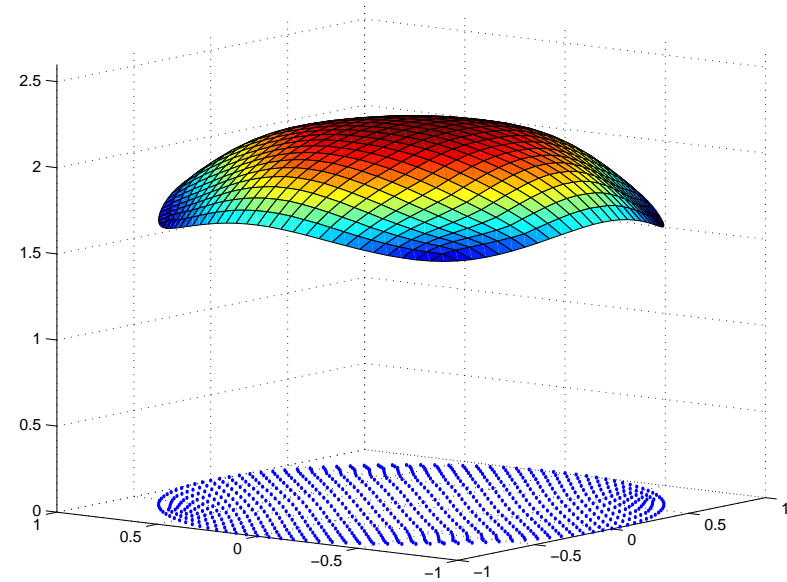
Mesures ponctuelles de la vitesse sur une section de conduite



L'interpolation directe des mesures n'est pas suffisante



profil simulé

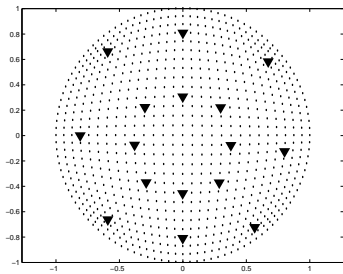


profil interpolé

Prise en compte d'a priori :

- Ajouter de **nouveaux facteurs** x^+ correspondant à de l'information a priori.
- Par exemple, à toute coordonnée spatiale $x \in \mathbb{R}^2$, associer un scalaire x^+ donnant une **vitesse du fluide au premier ordre** en x
- On choisit une moyenne (inconnue) du processus sous la forme

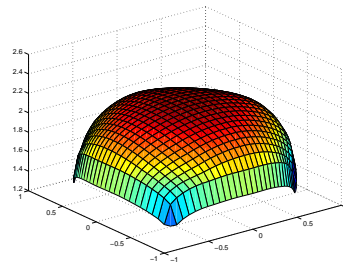
$$m(x, x^+) = \mathbb{E}[\xi(x, x^+)] = b_0 + b_1 x^+$$



observations

$$f(x_i)$$

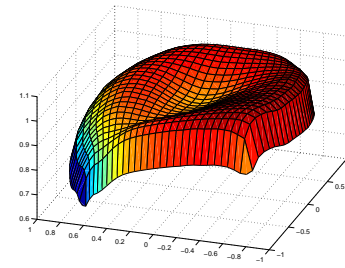
+



a priori

$$x^+$$

→



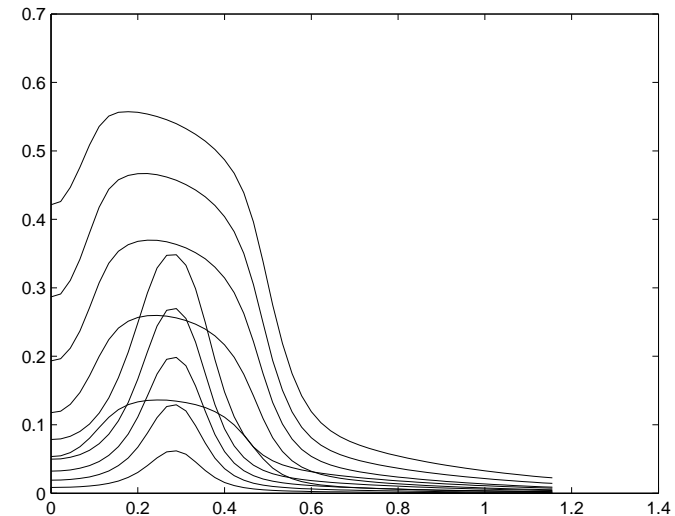
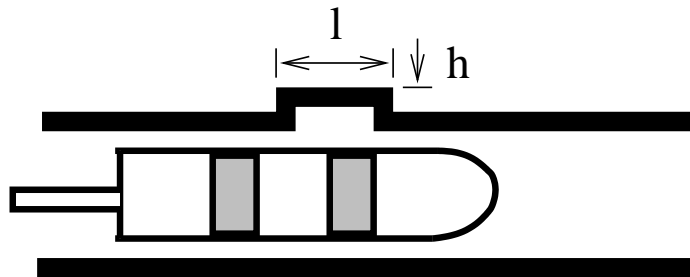
estimation

$$f_n(x, x^+)$$

Utilisation des approximations

Modélisation boîte noire

- Construction d'un modèle peu coûteux d'un code numérique
- Application extrêmement classique pour les **problèmes inverses**
- Exemple en CND



- défaut $x \in \mathbb{R}^2$
- réponse du capteur $y \in \mathbb{R}^d$, $d > 100$
- données simulées $(x_i, y_i)_{i=1, \dots, n}$

-
- **Objectif** : étant donné y , estimer x
 - Deux solutions
 1. construire le **modèle direct** $y = f_n(x)$ puis résoudre le problème
$$x^* \text{ t.q. } f_n(x^*) = y$$
 2. construire le **modèle inverse** $x = f_n^-(y)$ et obtenir directement une estimation de x
 - Problème de **planification d'expériences** pour construire le modèle (direct ou inverse)
 - Limites : problème de la **dimension** de y , sensibilité des solutions
 - À explorer : modélisation **multi-niveau**
 - utilisation conjointe d'une simulation coûteuse et précise avec une simulation approximative peu coûteuse

Application à l'optimisation

- Utilisation d'une approximation dans les problèmes d'optimisation
- Exemple : **Expected Improvement** (Mockus 1978, Schonlau 96)
 - Algorithme de recherche séquentielle du maximum
 - À chaque étape, recherche du point qui maximise la moyenne des dépassements de maximum courant, c-à-d

$$x_{n+1} = \arg \max_{x \in \mathbb{X}} \rho_n(x)$$

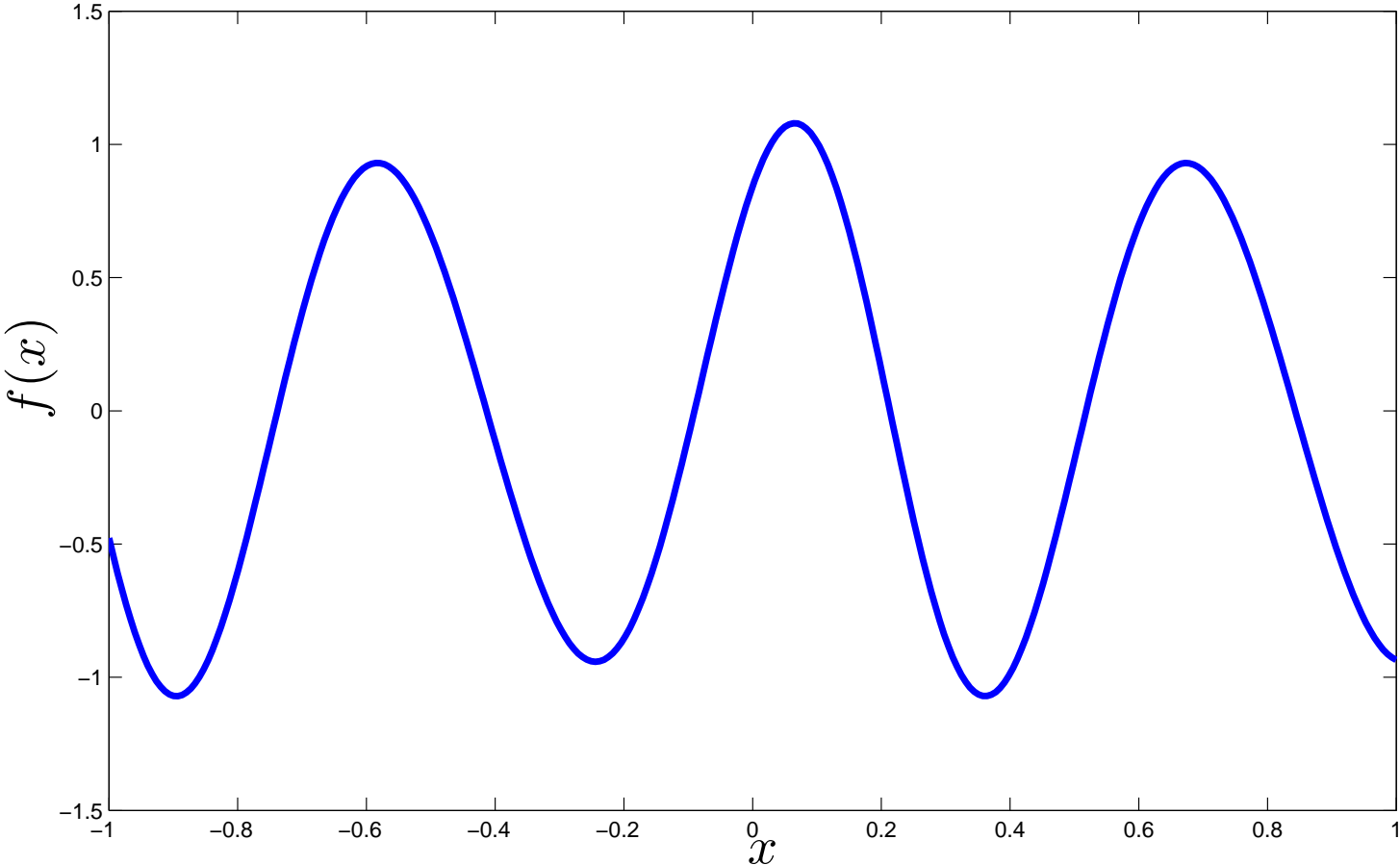
avec

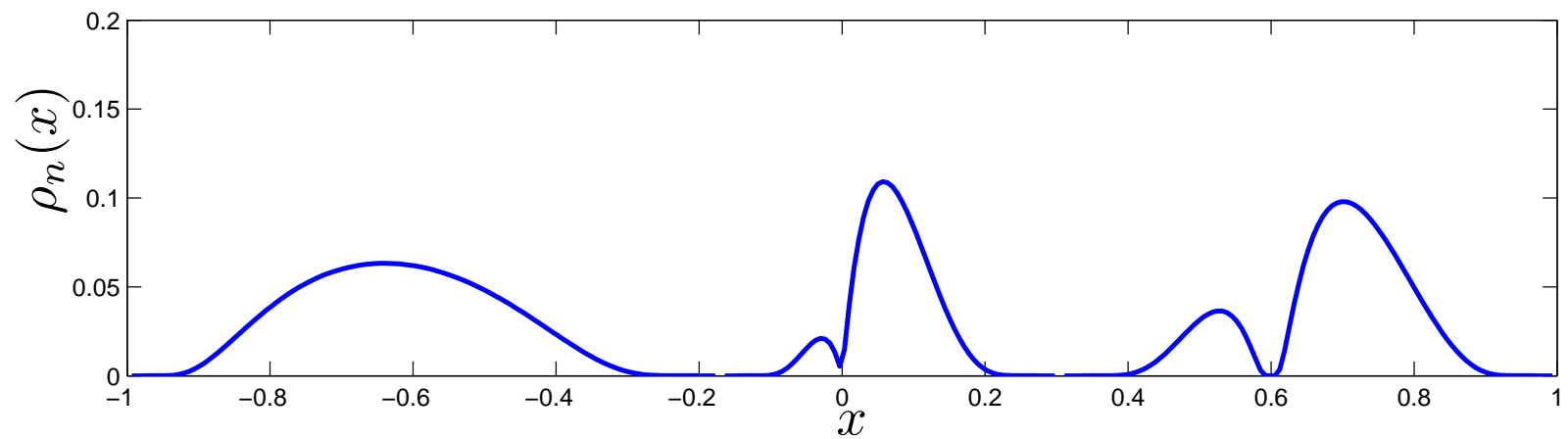
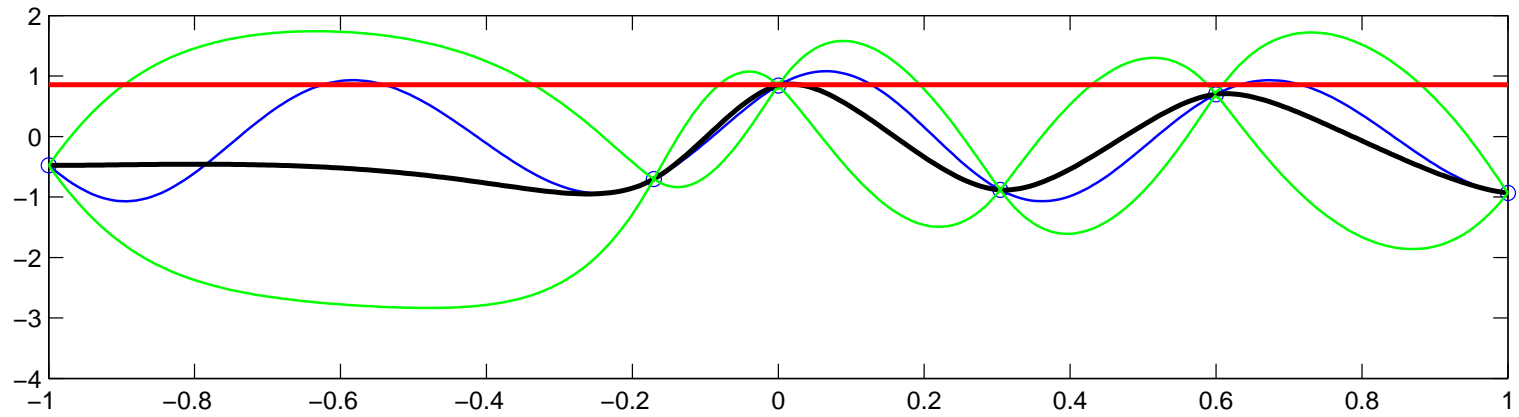
$$\rho_n(x) = \mathbb{E} [(\xi(x) - M_n)_+ \mid \boldsymbol{\xi}_n = \mathbf{f}_n^{\text{obs}}]$$

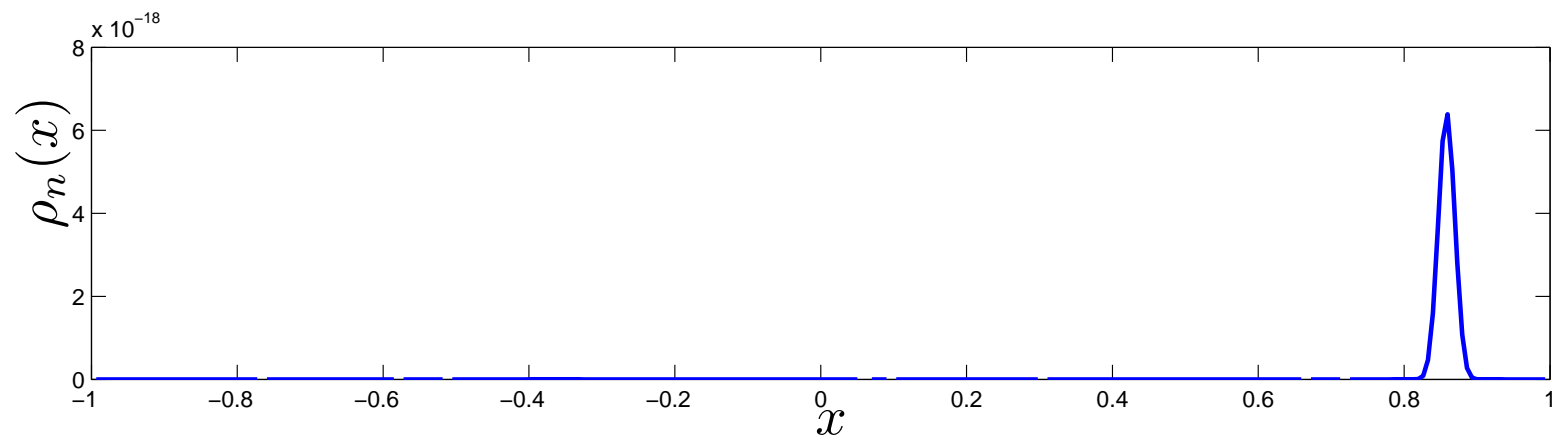
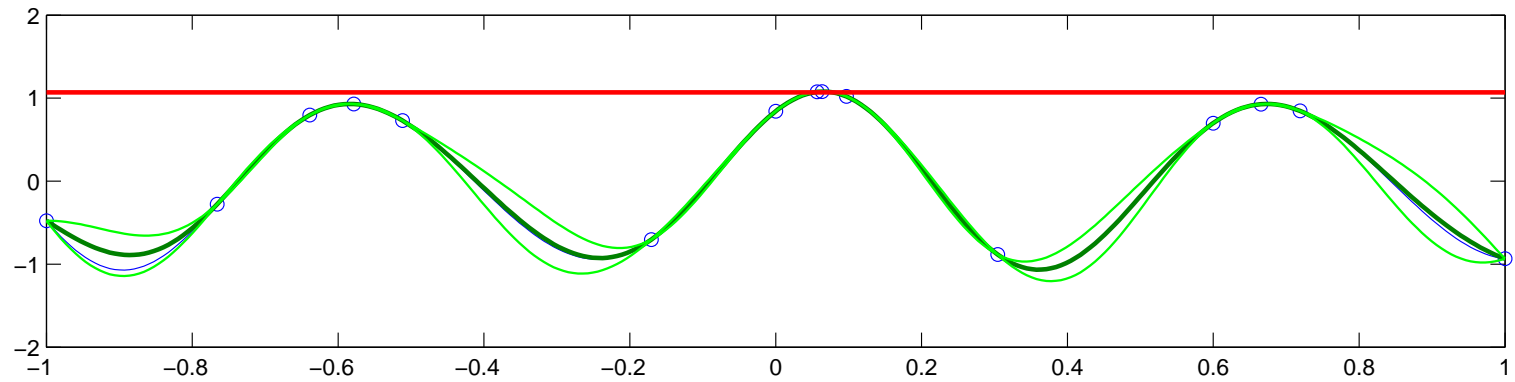
$$M_n = \max(\xi(x_1), \dots, \xi(x_n))$$

- $\rho_n(x)$ se calcule analytiquement en fonction de $f_n(x)$ et $\sigma_n^2(x)$

Exemple







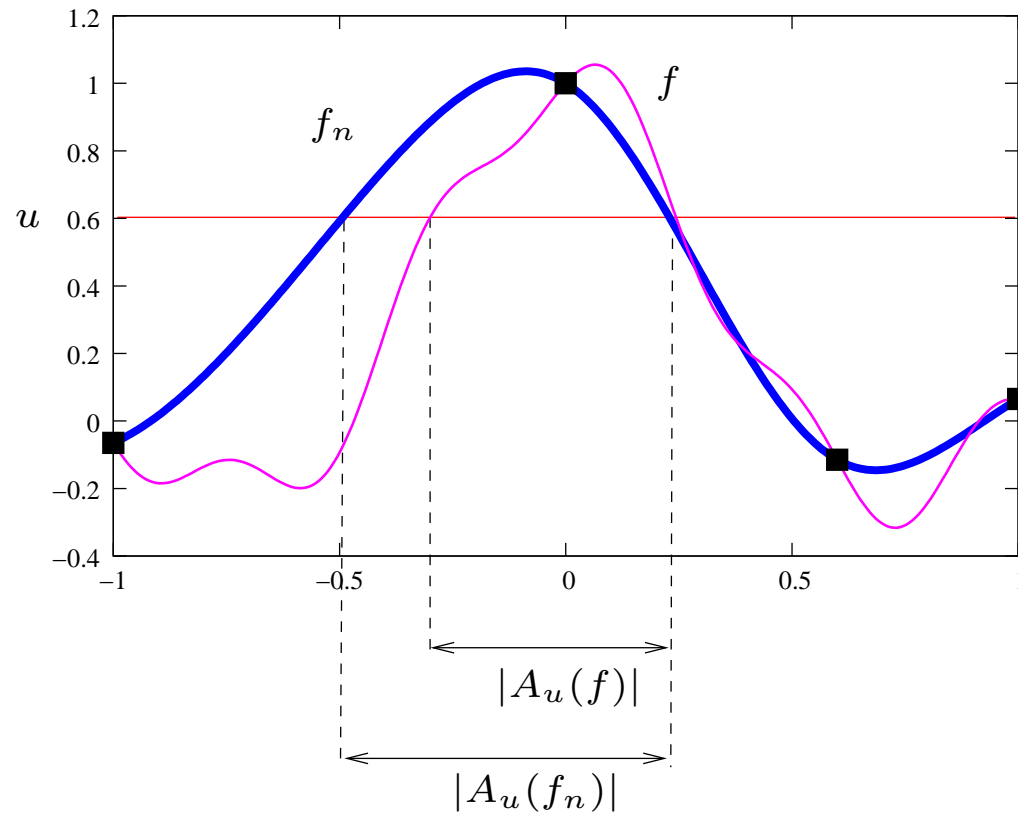
-
- Avantages :
 - optimisation globale
 - quantification de l'incertitude restante sur la position du maximum
 - Algorithmes similaires :
 - P-Algorithm (Zilinskas 1992)
 - IAGO : Informational Approach for Global Optimization (Villemonaix, Vazquez, Walter 2007)

Estimation de probabilités de défaillance

- Probabilité qu'une fonction de performance du système dépasse un seuil donné
 - $f : \mathbb{X} \rightarrow \mathbb{R}$: fonction de performance (une fonction des sorties)
 - $X \in \mathbb{X} \subseteq \mathbb{R}^d$ de loi μ : facteurs aléatoires

$$\mathcal{P}_u = \mathbb{P}\{f(X) > u\}?$$

- **Nombreuses méthodes classiques**
 - Monte Carlo sans ou avec échantillonnage d'importance, p.ex. cross-entropy (Rubinstein 99), méthode des sous-ensembles (Au et Beck 01)...
 - théorie des valeurs extrêmes
 - approximations FORM/SORM/FOSPA



- $A_u(f) := \{x \in \mathbb{X} : f(x) \geq u\}$, ensemble d'excursion de f au dessus d'un niveau u
- $|A_u(f)| := \mu(A_u(f))$, volume d'excursion
- Estimation de $|A_u(f)|$ par $|A_u(f_n)|$.

Questions

- Vitesse de convergence de l'approximation du volume d'excursion, c-à-d de $|A_u(f_n)|$ vers $|A_u(f)|$?

Sous certaines hypothèses, avec un tirage uniforme des points x_i , $\|\sigma_n\|_\infty \rightarrow 0$ et

$$\mathbb{E} \left[(|A_u(\xi)| - |A_u(\xi_n)|)^2 \right] = O\left(\|\sigma_n\|_\infty (\log 1/\|\sigma_n\|_\infty)^{1/2}\right)$$

- Comment augmenter cette vitesse de convergence en choisissant les points x_i en fonction des évaluations précédentes (planification d'expériences séquentielle) ?

Contrôle de la convergence

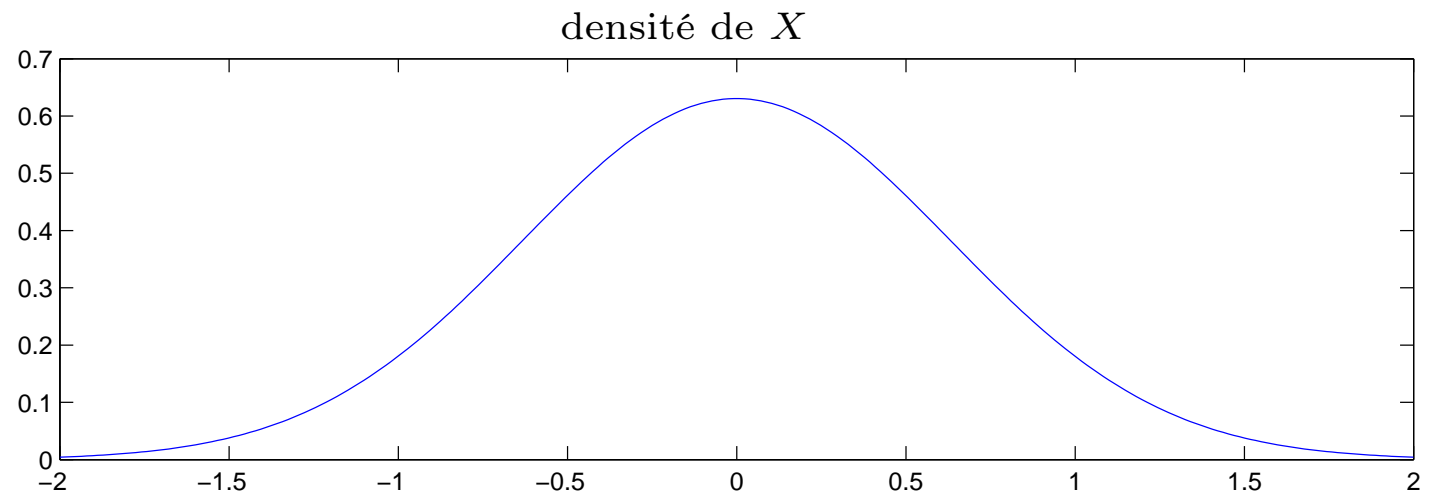
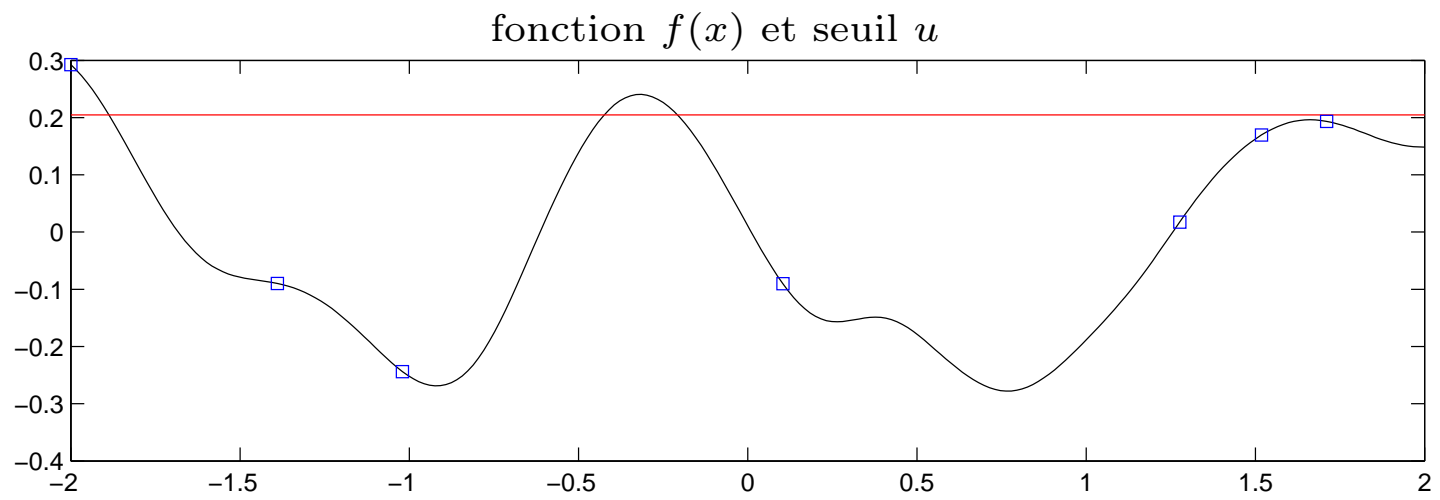
Choix du point qui **réduit le plus possible l'erreur sur le volume d'excursion, à un pas** \Leftrightarrow

$$x_n = \operatorname{argmin}_{x_n \in \mathbb{X}} \Upsilon_n(x_n) := \mathbb{E} [(|A_u(\xi)| - |A_u(\xi_n)|)^2 \mid \xi_{n-1}], \quad (1)$$

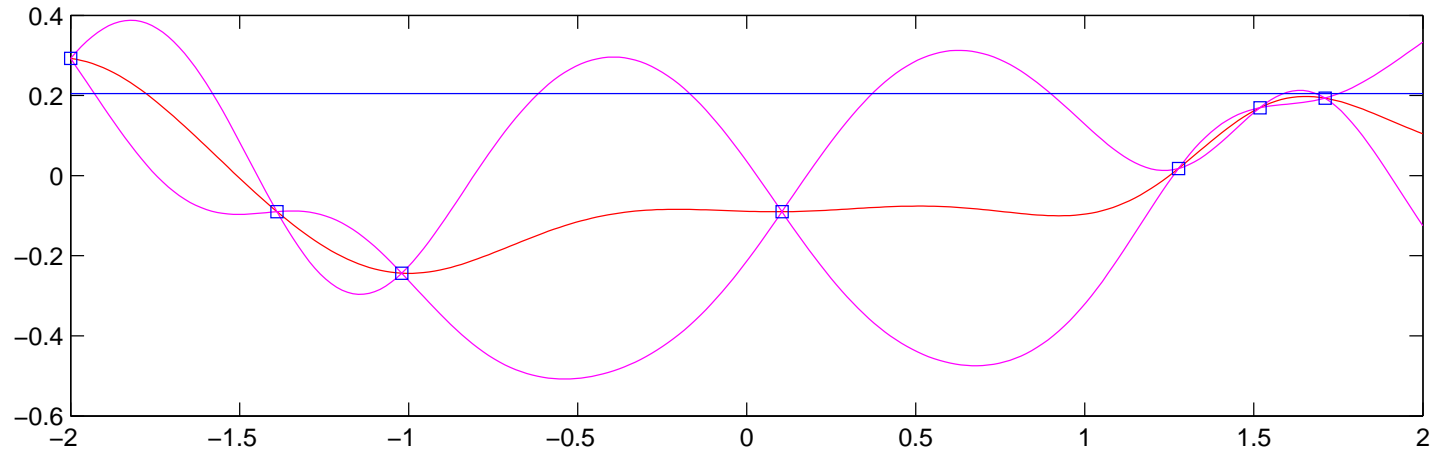
où $\xi_n = (\xi(x_1), \dots, \xi(x_n))$.

Il est possible de trouver une approximation analytique de (2).

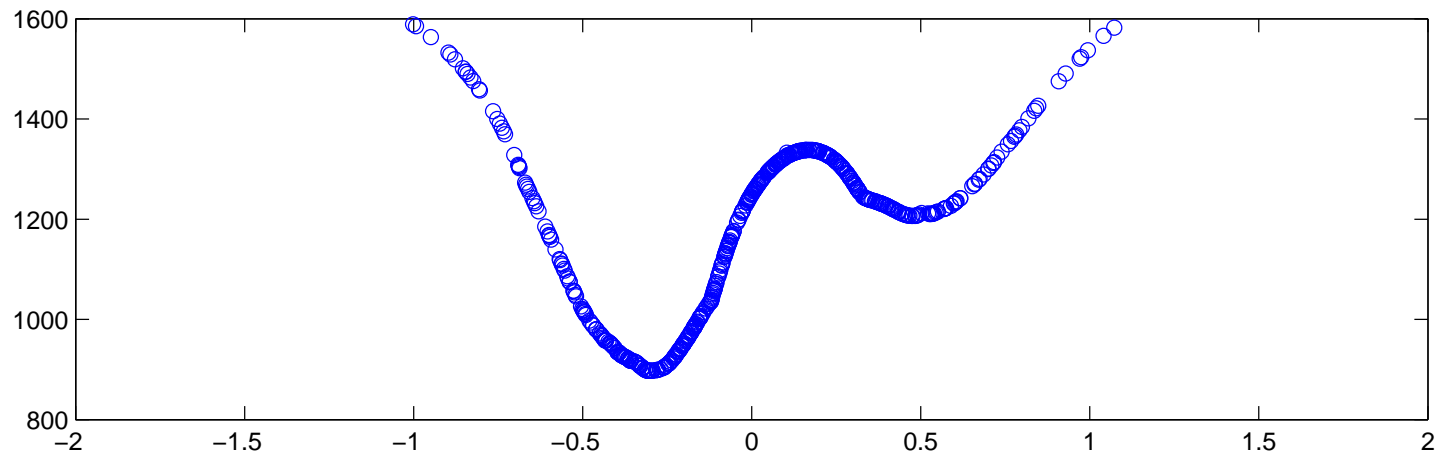
Exemple



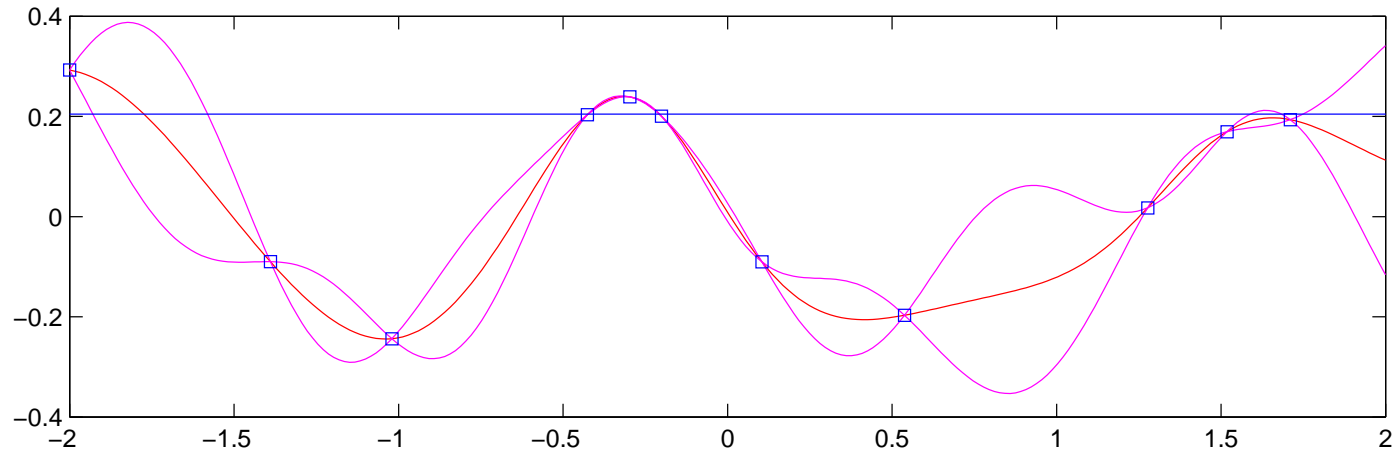
Prédiction et intervalles de confiance à 95%



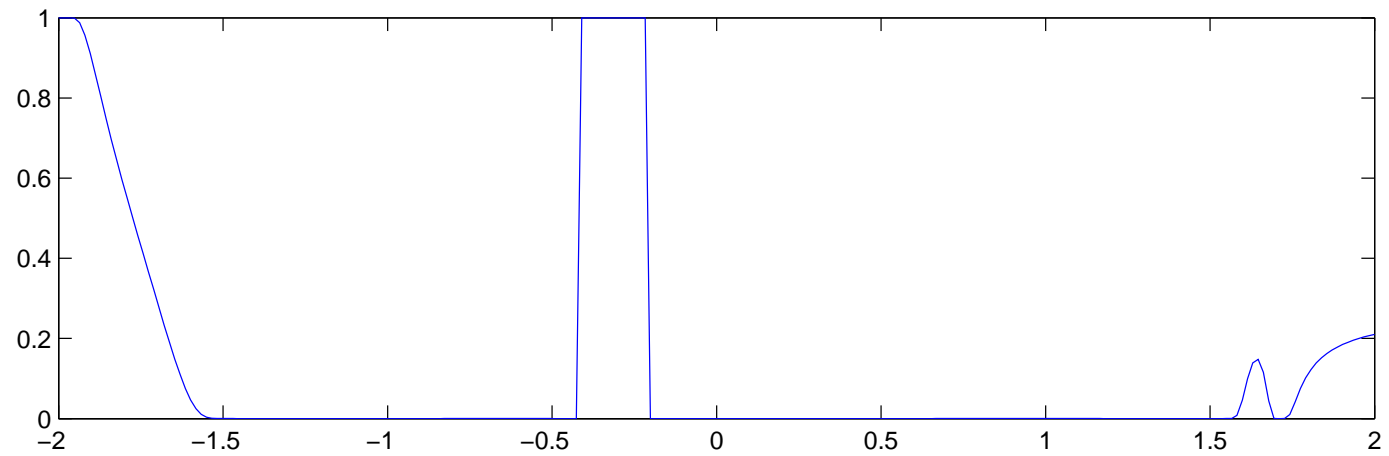
graphe de $\Upsilon_{n,m,Q}(x)$, $x \in S$



Prédiction et intervalles de confiance à 95%



Probabilité de dépassement de seuil



Conclusions

- Avantages
 - interpolation des simulations
 - vitesses de convergence fonction de la régularité
 - possibilité de prendre en compte de l'information a priori
- Limites
 - interpolation avec peu d'évaluations
 - interpoler en grande dimension ($d > 20 \sim 100$)
- Souvent, il n'est pas nécessaire de construire l'approximation sur tout le domaine des facteurs (planification d'expériences adaptée au résultat recherché)