

# Planification séquentielle pour l'estimation de probabilités de défaillance

Julien Bect

En collaboration avec Ling Li et Emmanuel Vazquez

SUPELEC

4 mai 2010

Atelier **événements rares**

GdR MASCOT-NUM

# Le problème à résoudre...

Données du problème

- un domaine  $\mathbb{X} \subset \mathbb{R}^d$  compact (espace des facteurs)
- une fonction  $f : \mathbb{X} \rightarrow \mathbb{R}$  coûteuse à évaluer
- une loi  $P_{\mathbb{X}}$  sur l'espace des facteurs
- un seuil  $T \in \mathbb{R}$

On veut calculer

$$\alpha(f) = P_{\mathbb{X}}\{x \in \mathbb{X} : f(x) > T\} = \int_{\mathbb{X}} \mathbb{1}_{f>T} dP_{\mathbb{X}}$$

# Le problème à résoudre...

Données du problème

- un domaine  $\mathbb{X} \subset \mathbb{R}^d$  compact (espace des facteurs)
- une fonction  $f : \mathbb{X} \rightarrow \mathbb{R}$  coûteuse à évaluer
- une loi  $P_{\mathbb{X}}$  sur l'espace des facteurs
- un seuil  $T \in \mathbb{R}$

On veut calculer

$$\alpha(f) = P_{\mathbb{X}}\{x \in \mathbb{X} : f(x) > T\} = \int_{\mathbb{X}} \mathbb{1}_{f>T} dP_{\mathbb{X}}$$

**Important**

Il ne s'agit pas (encore) d'un problème d'inférence statistique !

# Événement rare ?

## Fonction **coûteuse à évaluer**

- par ex., le calcul de  $f(x)$  fait appel à un code aux éléments finis
- **budget limité** d'évaluations
- « rare »  $\nrightarrow \alpha(f)$  *extrêmement* petit

# Événement rare ?

Fonction coûteuse à évaluer

- par ex., le calcul de  $f(x)$  fait appel à un code aux éléments finis
- budget limité d'évaluations
- « rare »  $\nrightarrow \alpha(f)$  *extrêmement* petit

État de l'art pré-2004 (thèses : Lagnoux Renaudie, 2006 ; Piera-Martinez, 2008)

- méthodes fondées sur des **approximations paramétriques**
  - ex : FORM/SORM, surface de réponse polynomiale
  - **économique** mais souvent **peu précis**
- méthodes utilisant l'**échantillonnage aléatoire**
  - Monte-Carlo, échantillonnage préférentiel, stratifié, etc.
  - **convergentes** mais **gourmand en évaluations**

# Événement rare ?

Fonction coûteuse à évaluer

- par ex., le calcul de  $f(x)$  fait appel à un code aux éléments finis
- budget limité d'évaluations
- « rare »  $\nrightarrow \alpha(f)$  *extrêmement* petit

État de l'art pré-2004 (thèses : Lagnoux Renaudie, 2006 ; Piera-Martinez, 2008)

- méthodes fondées sur des approximations paramétriques
  - ex : FORM/SORM, surface de réponse polynomiale
  - économique mais souvent peu précis
- méthodes utilisant l'échantillonnage aléatoire
  - Monte-Carlo, échantillonnage préférentiel, stratifié, etc.
  - convergentes mais gourmand en évaluations

Idée : approche bayésienne non-paramétrique

info. a priori sur  $f$   $\rightarrow$  **méthodes convergentes et économiques**

Apparition du **krigeage** en fiabilité (surface de réponse non-paramétrique)

- Romero, Swiler et Giunta ([Struct. Safety, 2004](#))
- Kaymaz ([Struct. Safety, 2005](#))
- Shan et Wang ([J. Mech. Design, 2006](#))

Apparition du krigage en fiabilité (surface de réponse non-paramétrique)

- Romero, Swiler et Giunta ([Struct. Safety, 2004](#))
- Kaymaz ([Struct. Safety, 2005](#))
- Shan et Wang ([J. Mech. Design, 2006](#))

Premiers travaux en **planification séquentielle**, point de vue bayésien

- Vazquez et Piera-Martinez ([JdS 2007](#)), Vazquez et Bect ([SYSID'09](#))
- Picheny, Ginsbourger, Roustant, Haftka ([ENBIS'08](#))  
Picheny, Ginsbourger, Roustant, Haftka et Kim ([J. Mech. Design \(tbp\)](#))
- Ranjan, Bingham et Michailidis ([Technometrics, 2008](#))
- Bichon, Eldred, Swiler, Mahadevan et McFarland ([J. AIAA 2008](#))



Apparition du krigage en fiabilité (surface de réponse non-paramétrique)

- Romero, Swiler et Giunta ([Struct. Safety, 2004](#))
- Kaymaz ([Struct. Safety, 2005](#))
- Shan et Wang ([J. Mech. Design, 2006](#))

Premiers travaux en planification séquentielle, point de vue bayésien

- Vazquez et Piera-Martinez ([JdS 2007](#)), Vazquez et Bect ([SYSID'09](#))
- Picheny, Ginsbourger, Roustant, Haftka ([ENBIS'08](#))  
Picheny, Ginsbourger, Roustant, Haftka et Kim ([J. Mech. Design \(tbp\)](#))
- Ranjan, Bingham et Michailidis ([Technometrics, 2008](#))
- Bichon, Eldred, Swiler, Mahadevan et McFarland ([J. AIAA 2008](#))

Dans cet exposé. . .

- une **introduction générale au sujet** (biaisée)
- les **premiers résultats obtenus par Ling Li** dans le cadre de sa thèse

- 1 Formalisme : théorie bayésienne de la décision
- 2 Processus gaussiens et modèles dérivés
- 3 Tour d'horizon des stratégies séquentielles existantes
- 4 Conclusion et perspectives

- 1 Formalisme : théorie bayésienne de la décision
- 2 Processus gaussiens et modèles dérivés
- 3 Tour d'horizon des stratégies séquentielles existantes
- 4 Conclusion et perspectives

# Cadre décisionnel (1/2)

Contrainte : on dispose d'un **budget limité de  $N$  évaluations**.

choisir  $x \in \mathbb{X}$   $\rightarrow$  obtenir  $f(x) \in \mathbb{R}$

## Questions

- 1 comment **choisir séquentiellement** les entrées  $x_1, \dots, x_N$  ?
- 2 comment **estimer  $\alpha(f)$**  à partir de  $f(x_1), \dots, f(x_N)$  ?

# Cadre décisionnel (1/2)

Contrainte : on dispose d'un budget limité de  $N$  évaluations.

choisir  $x \in \mathbb{X}$   $\rightarrow$  obtenir  $f(x) \in \mathbb{R}$

## Questions

- 1 comment choisir séquentiellement les entrées  $x_1, \dots, x_N$  ?
- 2 comment estimer  $\alpha(f)$  à partir de  $f(x_1), \dots, f(x_N)$  ?

## Réponse dans un cadre bayésien (pré-requis)

- choix d'une **fonction de coût**  $C(\alpha, \hat{\alpha}) = (\alpha - \hat{\alpha})^2$
- choix d'une **loi a priori** pour la fonction  $f$  inconnue  
 $\Rightarrow f$  est maintenant vue comme un processus aléatoire  $\xi$  !

## Cadre décisionnel (2/2)

Notations :  $X_1, \dots, X_N$  points d'évaluations

- $I_n = (X_1, \xi(X_1), \dots, X_n, \xi(X_n))$  info. disponible au temps  $n$
- $\mathcal{F}_n = \sigma(I_n)$ ,  $\alpha = \alpha(\xi)$ ,  $\hat{\alpha}_n = \hat{\alpha}_n(I_n)$

### Réponse dans un cadre bayésien (suite)

On cherche  $X_1, \dots, X_N$  et  $\hat{\alpha}_N$

- qui **minimisent**  $\mathbf{E} \left\{ (\alpha - \hat{\alpha}_N)^2 \right\}$
- avec  $X_{n+1}$   $\mathcal{F}_n$ -mesurable,  $\forall n \geq 1$

## Cadre décisionnel (2/2)

Notations :  $X_1, \dots, X_N$  points d'évaluations

- $I_n = (X_1, \xi(X_1), \dots, X_n, \xi(X_n))$  info. disponible au temps  $n$
- $\mathcal{F}_n = \sigma(I_n)$ ,  $\alpha = \alpha(\xi)$ ,  $\hat{\alpha}_n = \hat{\alpha}_n(I_n)$

### Réponse dans un cadre bayésien (suite)

On cherche  $X_1, \dots, X_N$  et  $\hat{\alpha}_N$

- qui minimisent  $\mathbf{E} \left\{ (\alpha - \hat{\alpha}_N)^2 \right\}$
- avec  $X_{n+1}$   $\mathcal{F}_n$ -mesurable,  $\forall n \geq 1$

Généralisations possibles :

- **évaluations en batches** (utile pour paralléliser)
- **évaluations bruités** (ex : code de calcul à base de MC)
- **décisions randomisées**, autres critères d'arrêt, etc.

# Estimateur optimal

Notations :  $\mathbf{P}_n \triangleq \mathbf{P}\{\cdot \mid \mathcal{F}_n\}$  et  $\mathbf{E}_n \triangleq \mathbf{E}\{\cdot \mid \mathcal{F}_n\}$

Meilleur estimateur  $\mathcal{F}_N$ -mesurable (risque quadratique) :

$$\hat{\alpha}_N = \mathbf{E}_N(\alpha)$$



# Estimateur optimal

Notations :  $\mathbf{P}_n \triangleq \mathbf{P} \{ \cdot \mid \mathcal{F}_n \}$  et  $\mathbf{E}_n \triangleq \mathbf{E} \{ \cdot \mid \mathcal{F}_n \}$

Meilleur estimateur  $\mathcal{F}_N$ -mesurable (risque quadratique) :

$$\hat{\alpha}_N = \mathbf{E}_N(\alpha)$$

En posant  $p_n(x) = \mathbf{P}_n \{ \xi(x) > T \}$ , on a

$$\hat{\alpha}_N = \mathbf{E}_N \left( \int_{\mathbb{X}} \mathbf{1}_{f > T} d\mathbf{P}_{\mathbb{X}} \right) = \int_{\mathbb{X}} p_N d\mathbf{P}_{\mathbb{X}},$$

## Important

On a transformé un problème de calcul d'intégrale en un autre problème de calcul d'intégrale  $\Rightarrow$  intéressant si la seconde est plus facile à calculer !

On peut aussi chercher un **estimateur « plugin »** de la forme

$$\hat{\alpha}_N = \int_{\mathbb{X}} g_N dP_{\mathbb{X}}, \quad \text{avec } g_N : \mathbb{X} \rightarrow \{0, 1\} \quad (1)$$

- $g_N$  est une **fonction de classification dure**
- ex : SVM dans la méthode SMART (thèse de F. Deheeger (LaMI))

On peut aussi chercher un estimateur « plugin » de la forme

$$\hat{\alpha}_N = \int_{\mathbb{X}} g_N dP_{\mathbb{X}}, \quad \text{avec } g_N : \mathbb{X} \rightarrow \{0, 1\} \quad (1)$$

- $g_N$  est une fonction de classification dure
- ex : SVM dans la méthode SMART (thèse de F. Deheeger (LaMI))

Meilleure fonction de classification à injecter dans (1) :

$$g_N = \mathbb{1}_{P_N > 1/2} = \mathbb{1}_{\hat{\xi}_N > T},$$

avec  $\hat{\xi}_N(x)$  la médiane a posteriori de  $\xi(x)$  (au temps  $N$ ).

# Stratégie bayésienne optimale (1/2)

On fixe un estimateur  $\hat{\alpha}_N$  (par ex :  $\hat{\alpha}_N = \mathbf{E}_N(\alpha)$ ).

Stratégie bayésienne optimale ?

- **stratégie** : suite  $X_1, \dots, X_N$  avec  $X_{n+1}$   $\mathcal{F}_n$ -mesurable
- **bayésienne optimale** : minimisant  $\mathbf{E} \{(\alpha - \hat{\alpha}_N)^2\}$

# Stratégie bayésienne optimale (1/2)

On fixe un estimateur  $\hat{\alpha}_N$  (par ex :  $\hat{\alpha}_N = \mathbf{E}_N(\alpha)$ ).

Stratégie bayésienne optimale ?

- stratégie : suite  $X_1, \dots, X_N$  avec  $X_{n+1}$   $\mathcal{F}_n$ -mesurable
- bayésienne optimale : minimisant  $\mathbf{E} \{(\alpha - \hat{\alpha}_N)^2\}$

Formellement, la solution s'obtient par **programmation dynamique**

- calcul par récurrence (rétrograde) du **risque bayésien**

$$R_N(I_N) = \mathbf{E}_N \left\{ (\alpha - \hat{\alpha}_N)^2 \right\}$$

$$R_n(I_n) = \min_{X_{n+1}} \mathbf{E}_n (R_{n+1}(I_{n+1})), \quad n = N - 1, \dots, 1$$

- stratégie optimale :  $X_{n+1} = \operatorname{argmin} \mathbf{E}_n (R_{n+1}(I_{n+1}))$

## Stratégie bayésienne optimale (2/2)

La stratégie optimale fait apparaître un **critère d'échantillonnage optimal** :

$$J_n(I_n, X_{n+1}) \triangleq \mathbf{E}_n(R_{n+1}(I_{n+1}))$$

## Stratégie bayésienne optimale (2/2)

La stratégie optimale fait apparaître un critère d'échantillonnage optimal :

$$J_n(I_n, X_{n+1}) \triangleq \mathbf{E}_n(R_{n+1}(I_{n+1}))$$

Malheureusement...

espace d'état **continu**, **dim.  $n$**   
espace d'actions **continu** également }  $\Rightarrow$  **calcul exact impossible!**

### Problème

Comment construire de bonnes solutions sous-optimales ?

- 1 Formalisme : théorie bayésienne de la décision
- 2 Processus gaussiens et modèles dérivés
- 3 Tour d'horizon des stratégies séquentielles existantes
- 4 Conclusion et perspectives



# Pourquoi des processus gaussiens ?

verrou

besoin de savoir **calculer** (ou **simuler**) efficacement la loi a posteriori  $\mathbf{P}^{\xi|\mathcal{F}_n}$

clé

modèles (a priori) construit à partir de **processus gaussiens** (GP)

# Pourquoi des processus gaussiens ?

## verrou

besoin de savoir calculer (ou simuler) efficacement la loi a posteriori  $\mathbf{P}^{\xi|\mathcal{F}_n}$

## clé

modèles (a priori) construit à partir de processus gaussiens (GP)

Le GP comme a priori sur une fonction inconnue : un peu d'histoire ...

- les **origines**
  - prédiction linéaire BLUP (Mathéron, Parzen ; années 60)
  - processus de Wiener en optimisation (Kushner, 1964)
  - splines & interprétation bayésienne (Kimeldorf & Wahba ; 70/71)

# Pourquoi des processus gaussiens ?

## verrou

besoin de savoir calculer (ou simuler) efficacement la loi a posteriori  $\mathbf{P}^{\xi|\mathcal{F}_n}$

## clé

modèles (a priori) construit à partir de processus gaussiens (GP)

Le GP comme a priori sur une fonction inconnue : un peu d'histoire ...

- les origines
  - prédiction linéaire BLUP (Mathéron, Parzen ; années 60)
  - processus de Wiener en optimisation (Kushner, 1964)
  - splines & interprétation bayésienne (Kimeldorf & Wahba ; 70/71)
- 1978, année bayésienne
  - en stats, trois articles dans JRSS (Wahba, O'Hagan, Leonard)
  - en optimisation : critère EI (Mockus, Tiesis, Zilinskas)

# Pourquoi des processus gaussiens ?

## verrou

besoin de savoir calculer (ou simuler) efficacement la loi a posteriori  $\mathbf{P}^{\xi|\mathcal{F}_n}$

## clé

modèles (a priori) construit à partir de processus gaussiens (GP)

Le GP comme a priori sur une fonction inconnue : un peu d'histoire ...

- les origines
  - prédiction linéaire BLUP (Mathéron, Parzen ; années 60)
  - processus de Wiener en optimisation (Kushner, 1964)
  - splines & interprétation bayésienne (Kimeldorf & Wahba ; 70/71)
- 1978, année bayésienne
  - en stats, trois articles dans JRSS (Wahba, O'Hagan, Leonard)
  - en optimisation : critère EI (Mockus, Tiesis, Zilinskas)
- **DACE** (Design & Analysis of Computer Experiments)
  - Sacks et al. (1989), Currin et al. (1991)

## Propriété

Si  $\xi \sim \text{GP}(m(x), k(x, x'))$ , alors  $\xi | \mathcal{F}_n \sim \text{GP}(\hat{\xi}_n(x), k_n(x, x'))$ ,  
avec  $\hat{\xi}_n$  et  $k_n$  donnés par les équations du krigeage (simple).

## Propriété

Si  $\xi \sim \text{GP}(m(x), k(x, x'))$ , alors  $\xi | \mathcal{F}_n \sim \text{GP}(\hat{\xi}_n(x), k_n(x, x'))$ ,  
avec  $\hat{\xi}_n$  et  $k_n$  donnés par les équations du krigeage (simple).

Etat de l'art : **modèles hiérarchiques** construits autour du GP

- « DACE » : krigeage ordinaire + approche bayésienne empirique
- krigeage bayésien (ex : méthodologie BACCO par O'Hagan et collab.)
- *treed GP* (par Gramacy & Lee)

Note : krigeage bayésien, *treed GP*  $\Rightarrow$  **(RJ)MCMC**

# Modèle « DACE » (le plus répandu ?)

On considère :

- moyenne constante inconnue  $m(x) = m \in \mathbb{R}$
- covariance stationnaire de la forme  $k(x, x') = \sigma^2 r_\theta(x - x')$

propriété (en supposant  $\sigma^2$  et  $\theta$  connus)

$$\text{Si } \begin{cases} \xi \mid m \sim \text{GP}(m, k(x, x')) \\ m \sim \mathcal{U}_{\mathbb{R}} \end{cases} \text{ alors } \xi \mid \mathcal{F}_n \sim \text{GP}(\hat{\xi}_n(x), k_n(x, x'))$$

avec  $\hat{\xi}_n$  et  $k_n$  donnés par les équations du krigeage (ordinaire).

# Modèle « DACE » (le plus répandu ?)

On considère :

- moyenne constante inconnue  $m(x) = m \in \mathbb{R}$
- covariance stationnaire de la forme  $k(x, x') = \sigma^2 r_\theta(x - x')$

propriété (en supposant  $\sigma^2$  et  $\theta$  connus)

$$\text{Si } \begin{cases} \xi \mid m \sim \text{GP}(m, k(x, x')) \\ m \sim \mathcal{U}_{\mathbb{R}} \end{cases} \text{ alors } \xi \mid \mathcal{F}_n \sim \text{GP}(\hat{\xi}_n(x), k_n(x, x'))$$

avec  $\hat{\xi}_n$  et  $k_n$  donnés par les équations du krigeage (ordinaire).

**Approche bayésienne empirique** :  $\sigma^2$  et  $\theta$  sont estimés par MML  
(maximisation de la vraisemblance marginale)



# Modèle « DACE » (le plus répandu ?)

On considère :

- moyenne constante inconnue  $m(x) = m \in \mathbb{R}$
- covariance stationnaire de la forme  $k(x, x') = \sigma^2 r_\theta(x - x')$

propriété (en supposant  $\sigma^2$  et  $\theta$  connus)

$$\text{Si } \begin{cases} \xi \mid m \sim \text{GP}(m, k(x, x')) \\ m \sim \mathcal{U}_{\mathbb{R}} \end{cases} \text{ alors } \xi \mid \mathcal{F}_n \sim \text{GP}(\hat{\xi}_n(x), k_n(x, x'))$$

avec  $\hat{\xi}_n$  et  $k_n$  donnés par les équations du krigeage (ordinaire).

Approche bayésienne empirique :  $\sigma^2$  et  $\theta$  sont estimés par MML (maximisation de la vraisemblance marginale)

## Mise en garde

Les procédures séquentielles qui **ré-estiment** les paramètres à chaque étape par MML n'ont, à l'heure actuelle, aucun fondement théorique.

## Quelques quantités utiles pour la suite (et faciles à calculer)

Notation :  $\sigma_n^2(x) = k_n(x, x)$  ;  $\Phi$  la FR de la loi  $\mathcal{N}(0, 1)$ .

**Probabilité de dépassement** du seuil  $T$  au point  $x \in \mathbb{X}$  :

$$p_n(x) = \mathbf{P}_n \{ \xi(x) > T \} = 1 - \Phi \left( \frac{\hat{\xi}_n(x) - T}{\sigma_n(x)} \right)$$

# Quelques quantités utiles pour la suite (et faciles à calculer)

Notation :  $\sigma_n^2(x) = k_n(x, x)$  ;  $\Phi$  la FR de la loi  $\mathcal{N}(0, 1)$ .

Probabilité de dépassement du seuil  $T$  au point  $x \in \mathbb{X}$  :

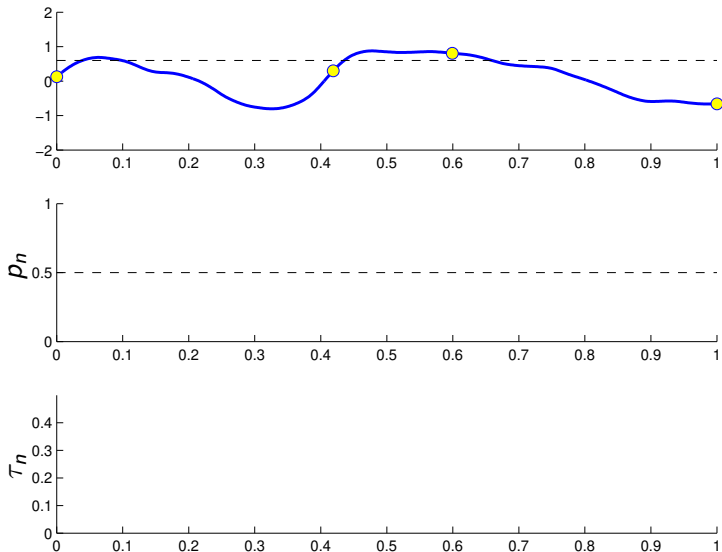
$$p_n(x) = \mathbf{P}_n \{ \xi(x) > T \} = 1 - \Phi \left( \frac{\hat{\xi}_n(x) - T}{\sigma_n(x)} \right)$$

Probabilité de mauvaise classification au point  $x \in \mathbb{X}$  :

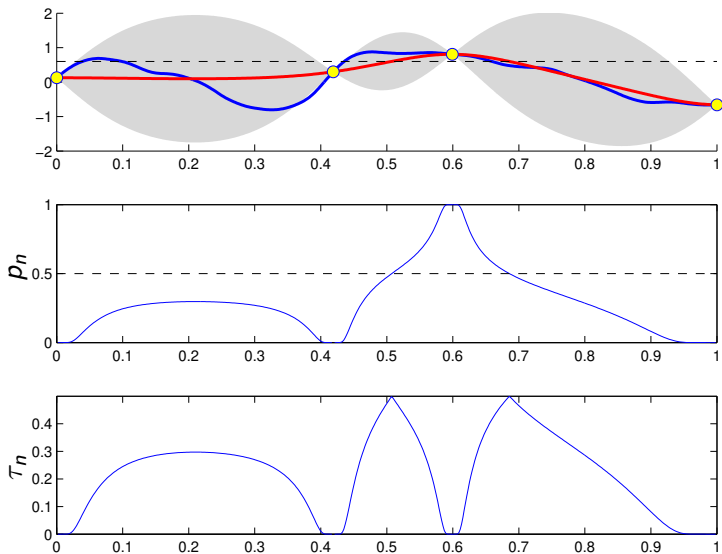
$$\tau_n(x) = \mathbf{P}_n \left\{ (\xi(x) - T)(\hat{\xi}_n(x) - T) < 0 \right\} = 1 - \Phi \left( \frac{|\hat{\xi}_n(x) - T|}{\sigma_n(x)} \right)$$

Remarque :  $\tau_n = \min(p_n, 1 - p_n)$ .

# Illustration



# Illustration



- 1 Formalisme : théorie bayésienne de la décision
- 2 Processus gaussiens et modèles dérivés
- 3 Tour d'horizon des stratégies séquentielles existantes**
- 4 Conclusion et perspectives

# Stratégies fondées sur une anticipation de $k$ évaluations

Principe général ( $k$ -step look-ahead) :

$$J_n(I_n, X_{n+1}) = \mathbf{E}_n \left( \min_{X_{n+2}} \mathbf{E}_{n+1} \left( \dots \min_{X_{n+k}} \mathbf{E}_{n+k-1} \left( \tilde{R}_{n+k}(I_{n+k}) \right) \right) \right)$$

avec  $\tilde{R}_{n+k}$  un substitut du risque bayésien  $R_{n+k}$ .

# Stratégies fondées sur une anticipation de $k$ évaluations

Principe général ( $k$ -step look-ahead) :

$$J_n(I_n, X_{n+1}) = \mathbf{E}_n \left( \min_{X_{n+2}} \mathbf{E}_{n+1} \left( \dots \min_{X_{n+k}} \mathbf{E}_{n+k-1} \left( \tilde{R}_{n+k}(I_{n+k}) \right) \right) \right)$$

avec  $\tilde{R}_{n+k}$  un substitut du risque bayésien  $R_{n+k}$ .

Cas particulier :  $k = 1$  (**one-step look-ahead**)

$$J_n(I_n, X_{n+1}) = \mathbf{E}_n \left( \tilde{R}_{n+1}(I_{n+1}) \right)$$



# Stratégies fondées sur une anticipation de $k$ évaluations

Principe général ( $k$ -step look-ahead) :

$$J_n(I_n, X_{n+1}) = \mathbf{E}_n \left( \min_{X_{n+2}} \mathbf{E}_{n+1} \left( \dots \min_{X_{n+k}} \mathbf{E}_{n+k-1} \left( \tilde{R}_{n+k}(I_{n+k}) \right) \right) \right)$$

avec  $\tilde{R}_{n+k}$  un substitut du risque bayésien  $R_{n+k}$ .

Cas particulier :  $k = 1$  (one-step look-ahead)

$$J_n(I_n, X_{n+1}) = \mathbf{E}_n \left( \tilde{R}_{n+1}(I_{n+1}) \right)$$

Stratégies OSL proposées dans la littérature :

- **SUR** (Piera-Martinez & Vazquez, JdS'07 ; Vazquez & Bect, SYSID'09)
- **targeted IMSE** (Picheny et. al, ENBIS8 (2008), J. Mech. Design (tbp))

# SUR : Stepwise Uncertainty Reduction (1/2)

Principe (cf. algorithme IAGO, thèse Villemonteix, 2008)

Chercher à **réduire au plus l'incertitude résiduelle.**

# SUR : Stepwise Uncertainty Reduction (1/2)

Principe (cf. algorithme IAGO, thèse Villemonteix, 2008)

Chercher à réduire au plus l'incertitude résiduelle.

Mise en œuvre

- 1 **Approximation gloutonne** du risque :

$$\tilde{R}_{n+1} = \mathbf{E}_{n+1} \left( (\alpha - \hat{\alpha}_{n+1})^2 \right) \Rightarrow J_n = \mathbf{E}_n \left( (\alpha - \hat{\alpha}_{n+1})^2 \right)$$

# SUR : Stepwise Uncertainty Reduction (1/2)

Principe (cf. algorithme IAGO, thèse Villemonteix, 2008)

Chercher à réduire au plus l'incertitude résiduelle.

Mise en œuvre

- 1 Approximation gloutonne du risque :

$$\tilde{R}_{n+1} = \mathbf{E}_{n+1} \left( (\alpha - \hat{\alpha}_{n+1})^2 \right) \Rightarrow J_n = \mathbf{E}_n \left( (\alpha - \hat{\alpha}_{n+1})^2 \right)$$

- 2 Calcul d'une **borne supérieure**

$$J_n \leq \mathbf{E}_n \left( \left( \int_{\mathbf{X}} \gamma_n dP_{\mathbf{X}} \right)^2 \right) \quad (2)$$

- 3 Discrétisation et minimisation de la borne (2).

# SUR : Stepwise Uncertainty Reduction (2/2)

Expression(s) de la fonction  $\gamma_n$  :

$$\hat{\alpha}_{n+1} = \mathbf{E}_{n+1}(\alpha) \quad \rightsquigarrow \quad \gamma_n^2(x) = p_n(x) (1 - p_n(x))$$

variance de l'indicatrice de dépassement

$$\hat{\alpha}_{n+1} = \int_{\mathbb{X}} \mathbb{1}_{\hat{\xi}_n > T} dP_{\mathbb{X}} \quad \rightsquigarrow \quad \gamma_n^2(x) = \min(p_n(x), 1 - p_n(x)) = \tau_n(x)$$

probabilité de mauvais classement

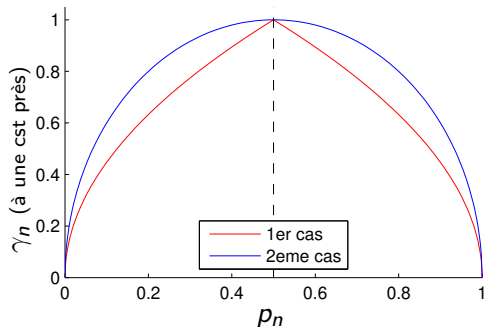


Illustration :  $\mathbb{X} = [0; 1]$ ,  $P_{\mathbb{X}} = \mathcal{U}_{[0;1]}$ ,  $\alpha(f) = 2\%$

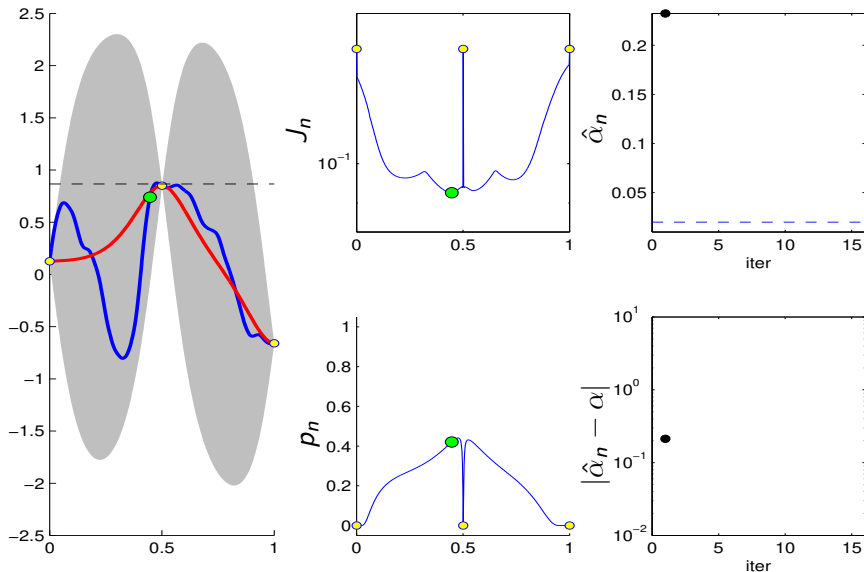


Illustration :  $\mathbb{X} = [0; 1]$ ,  $P_{\mathbb{X}} = \mathcal{U}_{[0;1]}$ ,  $\alpha(f) = 2\%$

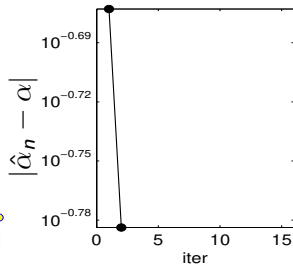
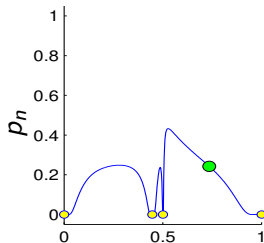
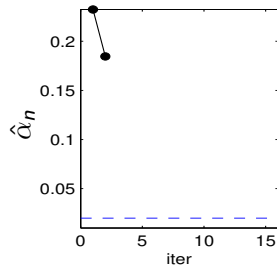
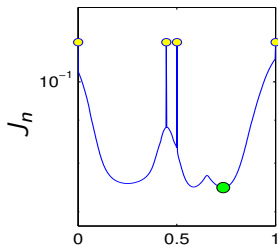
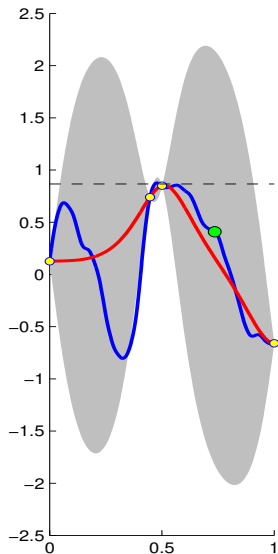


Illustration :  $\mathbb{X} = [0; 1]$ ,  $P_{\mathbb{X}} = \mathcal{U}_{[0;1]}$ ,  $\alpha(f) = 2\%$

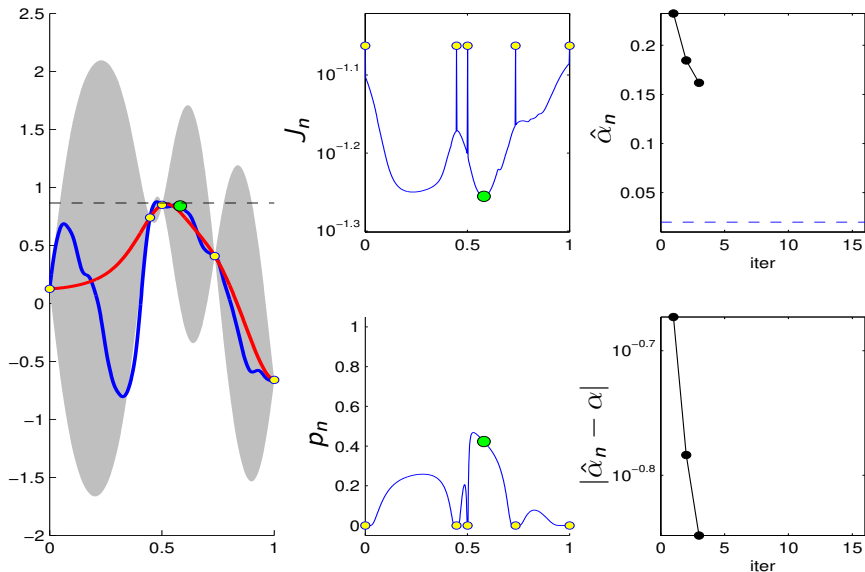




Illustration :  $\mathbb{X} = [0; 1]$ ,  $P_{\mathbb{X}} = \mathcal{U}_{[0;1]}$ ,  $\alpha(f) = 2\%$

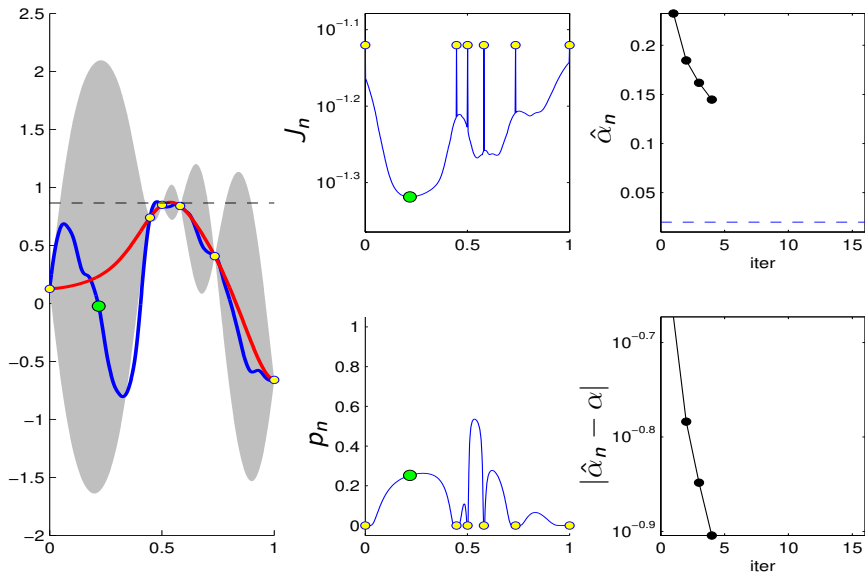


Illustration :  $\mathbb{X} = [0; 1]$ ,  $P_{\mathbb{X}} = \mathcal{U}_{[0;1]}$ ,  $\alpha(f) = 2\%$

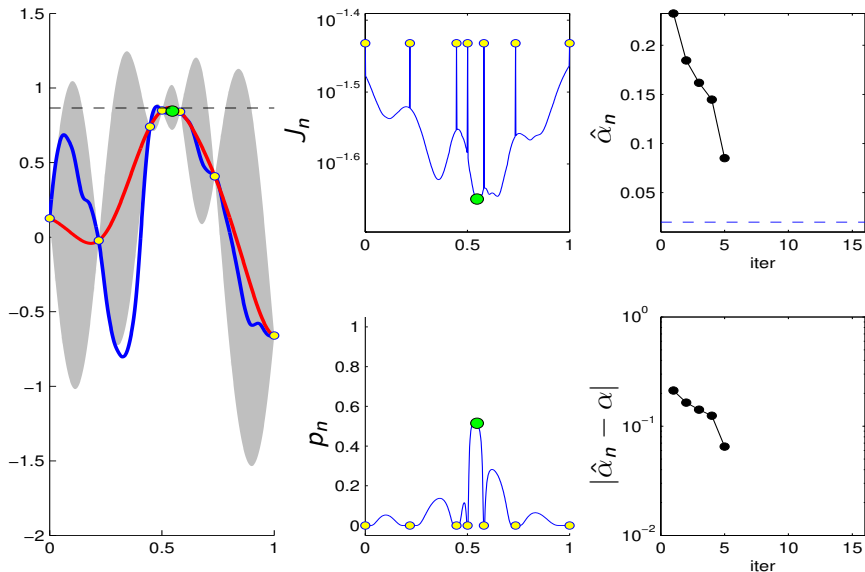


Illustration :  $\mathbb{X} = [0; 1]$ ,  $P_{\mathbb{X}} = \mathcal{U}_{[0;1]}$ ,  $\alpha(f) = 2\%$

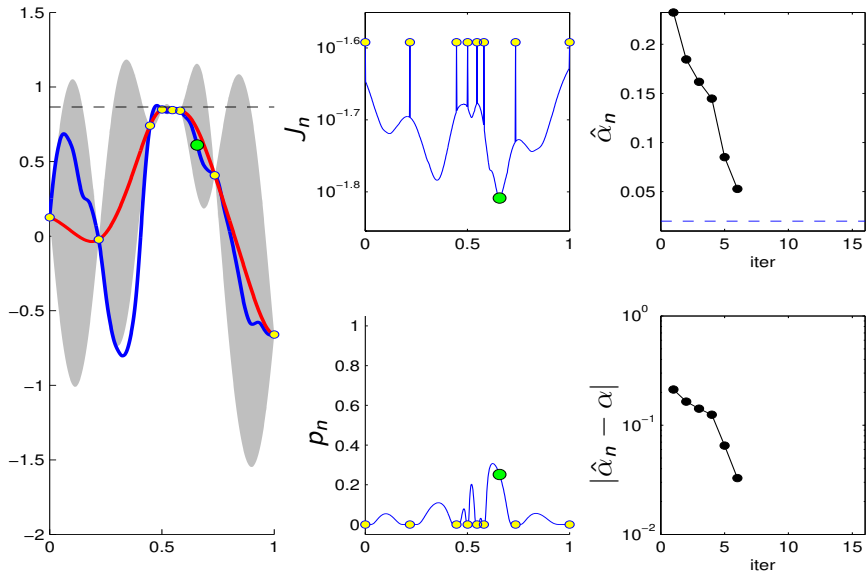


Illustration :  $\mathbb{X} = [0; 1]$ ,  $P_{\mathbb{X}} = \mathcal{U}_{[0;1]}$ ,  $\alpha(f) = 2\%$

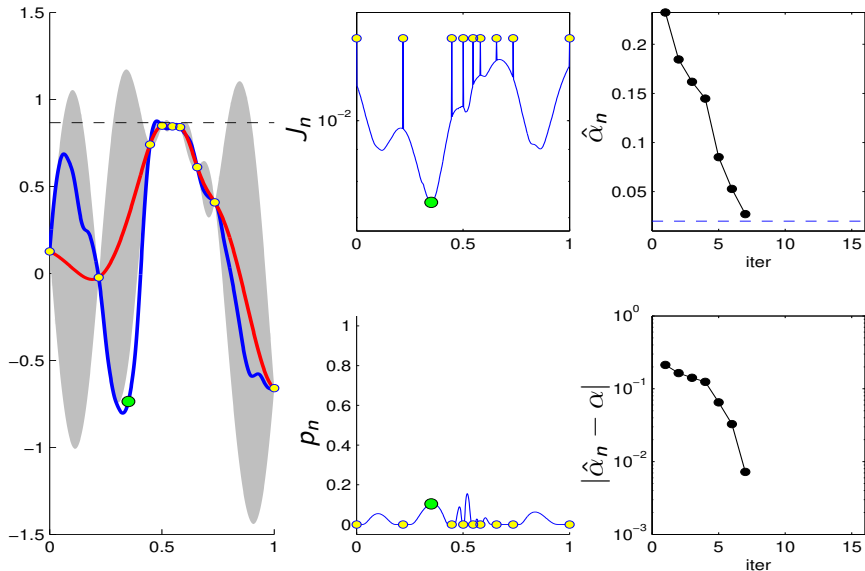


Illustration :  $\mathbb{X} = [0; 1]$ ,  $P_{\mathbb{X}} = \mathcal{U}_{[0;1]}$ ,  $\alpha(f) = 2\%$

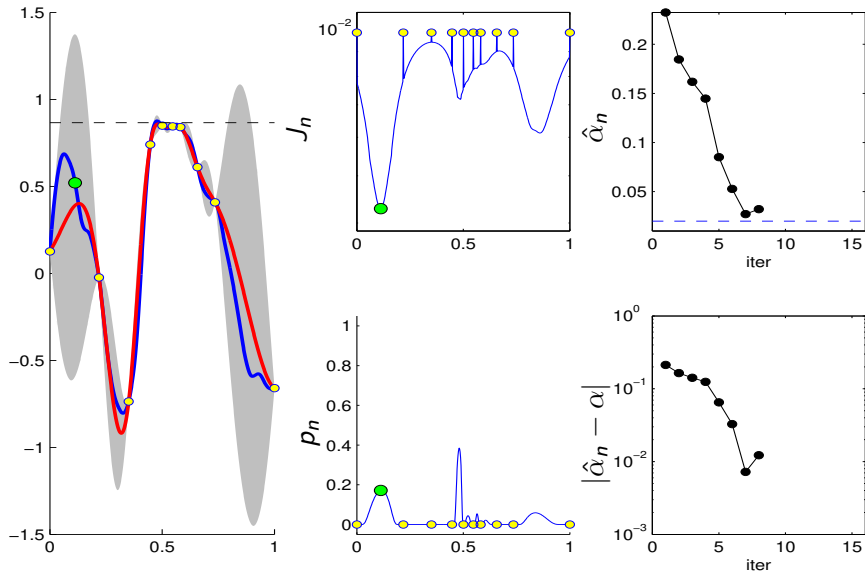


Illustration :  $\mathbb{X} = [0; 1]$ ,  $P_{\mathbb{X}} = \mathcal{U}_{[0;1]}$ ,  $\alpha(f) = 2\%$

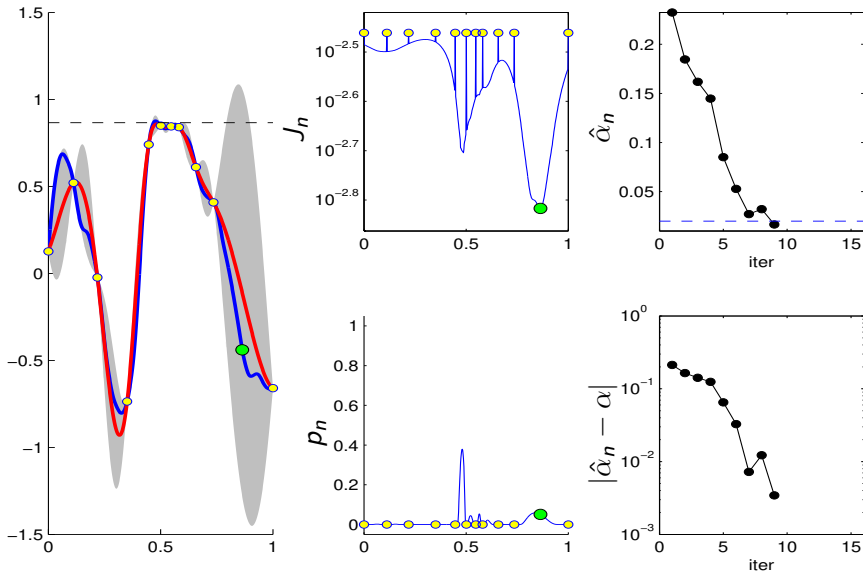


Illustration :  $\mathbb{X} = [0; 1]$ ,  $P_{\mathbb{X}} = \mathcal{U}_{[0;1]}$ ,  $\alpha(f) = 2\%$

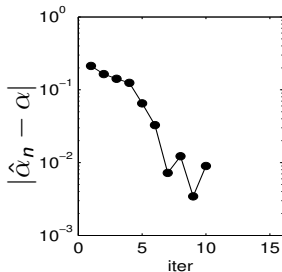
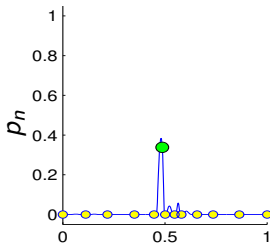
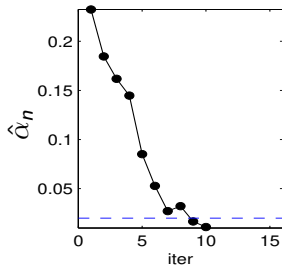
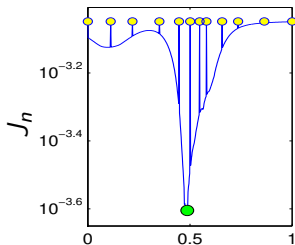
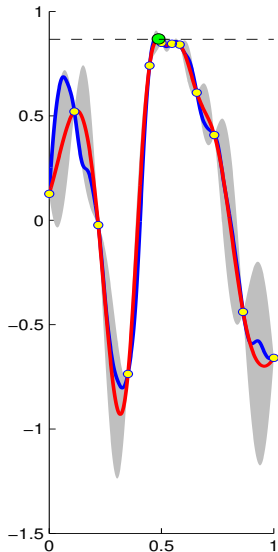


Illustration :  $\mathbb{X} = [0; 1]$ ,  $P_{\mathbb{X}} = \mathcal{U}_{[0;1]}$ ,  $\alpha(f) = 2\%$

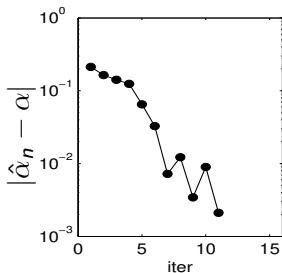
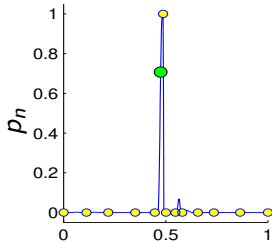
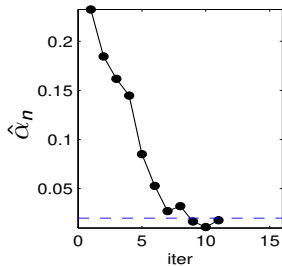
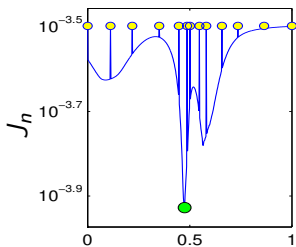
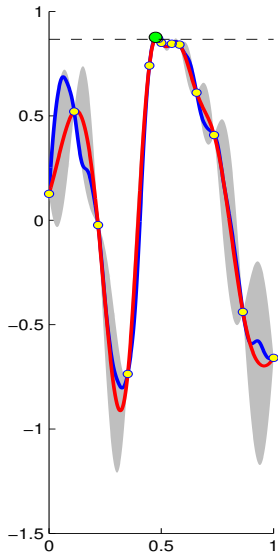




Illustration :  $\mathbb{X} = [0; 1]$ ,  $P_{\mathbb{X}} = \mathcal{U}_{[0;1]}$ ,  $\alpha(f) = 2\%$

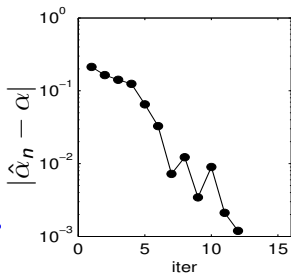
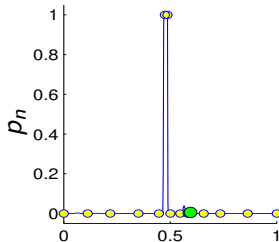
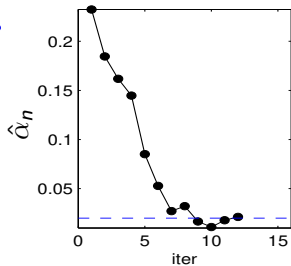
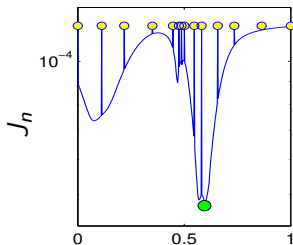
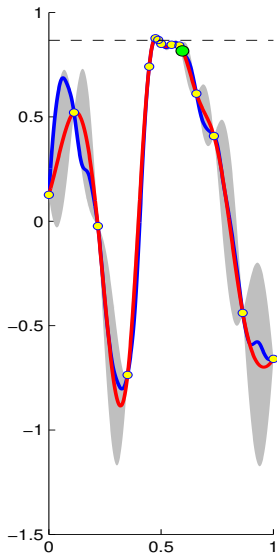


Illustration :  $\mathbb{X} = [0; 1]$ ,  $P_{\mathbb{X}} = \mathcal{U}_{[0;1]}$ ,  $\alpha(f) = 2\%$

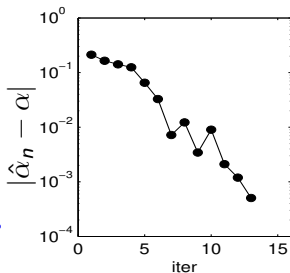
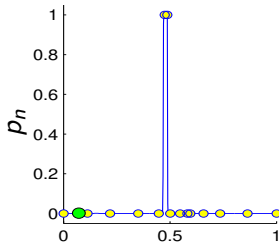
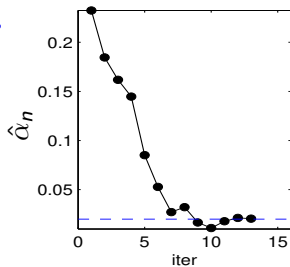
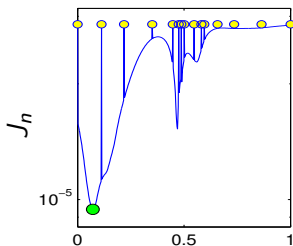
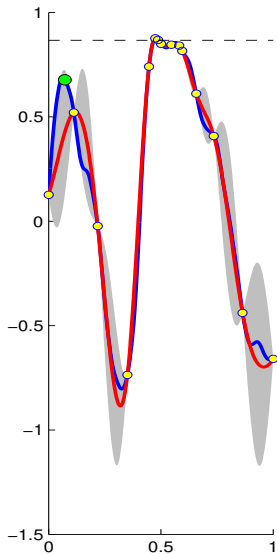


Illustration :  $\mathbb{X} = [0; 1]$ ,  $P_{\mathbb{X}} = \mathcal{U}_{[0;1]}$ ,  $\alpha(f) = 2\%$

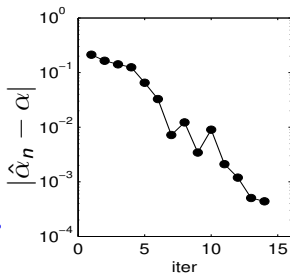
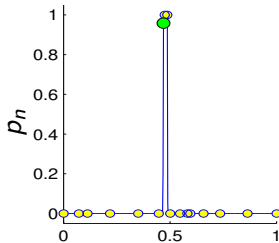
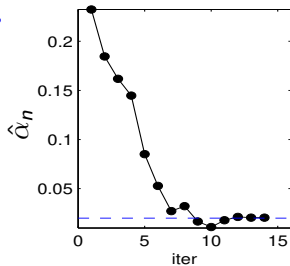
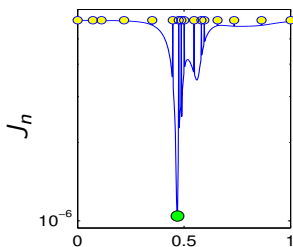
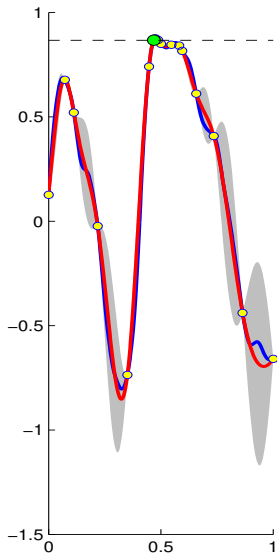
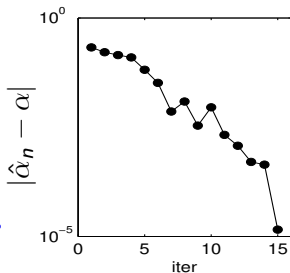
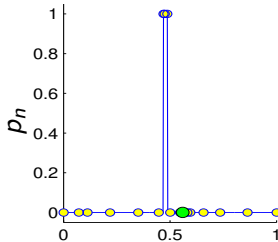
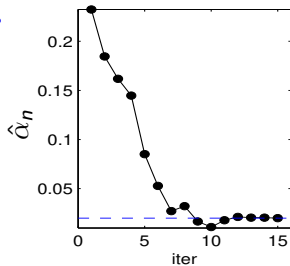
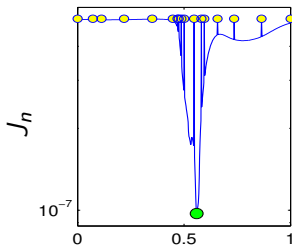
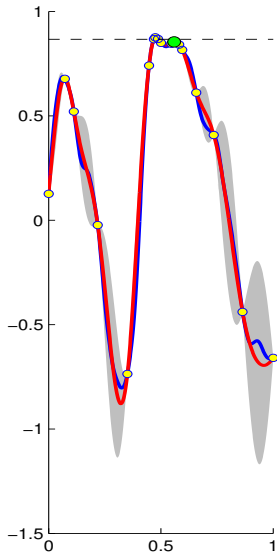


Illustration :  $\mathbb{X} = [0; 1]$ ,  $P_{\mathbb{X}} = \mathcal{U}_{[0;1]}$ ,  $\alpha(f) = 2\%$



Méthode fondée sur l'**IMSE pondéré** :

$$\tilde{R}_{n+1}(I_{n+1}) = \int_{\mathbb{X}} \sigma_{n+1}^2(x) W_{n+1}(x) dP_{\mathbb{X}}$$

où  $g$  est une fenêtre  $\geq 0$  et  $W_{n+1}(x) = \mathbf{E}_{n+1}(g(\xi(x) - T))$

Méthode fondée sur l'IMSE pondéré :

$$\tilde{R}_{n+1}(I_{n+1}) = \int_{\mathbb{X}} \sigma_{n+1}^2(x) W_{n+1}(x) dP_{\mathbb{X}}$$

où  $g$  est une fenêtre  $\geq 0$  et  $W_{n+1}(x) = \mathbf{E}_{n+1}(g(\xi(x) - T))$

Le **critère d'échantillonnage** correspondant s'écrit :

$$J_n(I_n, X_{n+1}) = \int_{\mathbb{X}} \sigma_{n+1}^2(x) W_n(x) dP_{\mathbb{X}}$$

Pondération recommandée par les auteurs :

$$g(z) = e^{-\frac{z^2}{2\sigma_\varepsilon^2}} \Rightarrow W_n(x) \propto \frac{1}{\sqrt{\sigma_\varepsilon^2 + \sigma_n^2(x)}} e^{-\frac{1}{2} \frac{(\xi_n(x) - T)^2}{\sigma_\varepsilon^2 + \sigma_n^2(x)}}$$

# Une comparaison des critères existants : protocole

- $\mathbb{X} = [0; 1]$  avec  $P_{\mathbb{X}} = \mathcal{U}_{[0;1]}$
- comparaison **en moyenne sur 4000 trajectoires d'un GP**
  - GP centré, covariance de Matérn ( $\sigma^2 = 1, \nu = 2, \rho = 0.1$ )
  - les algos utilisent la vraie fonction de covariance
  - $T$  de façon à ce que  $\alpha = 2\%$ .

# Une comparaison des critères existants : protocole

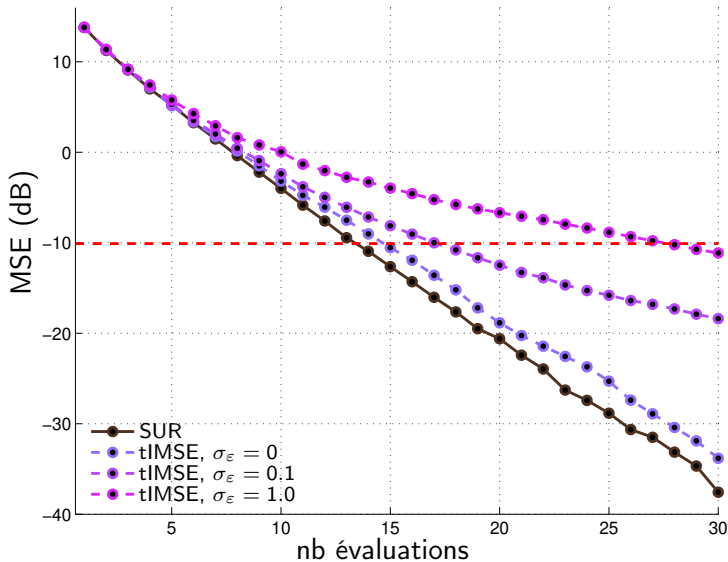
- $\mathbb{X} = [0; 1]$  avec  $P_{\mathbb{X}} = \mathcal{U}_{[0;1]}$
- comparaison en moyenne sur 4000 trajectoires d'un GP
  - GP centré, covariance de Matérn ( $\sigma^2 = 1, \nu = 2, \rho = 0.1$ )
  - les algos utilisent la vraie fonction de covariance
  - $T$  de façon à ce que  $\alpha = 2\%$ .
- **version discrétisée du problème** (« méta-estimation »)
  - $Y_1, \dots, Y_m \stackrel{\text{iid}}{\sim} P_{\mathbb{X}}$  ( $m = 500$ )
  - on veut **estimer l'estimateur**  $\alpha^m = \frac{1}{m} \sum_j \mathbb{1}_{\xi(Y_j) > T}$
  - on travaille sur  $\tilde{\mathbb{X}}^m = \{Y_1, \dots, Y_m\}$  avec  $\tilde{P}_{\mathbb{X}}^m = \frac{1}{m} \sum_j \delta_{Y_j}$



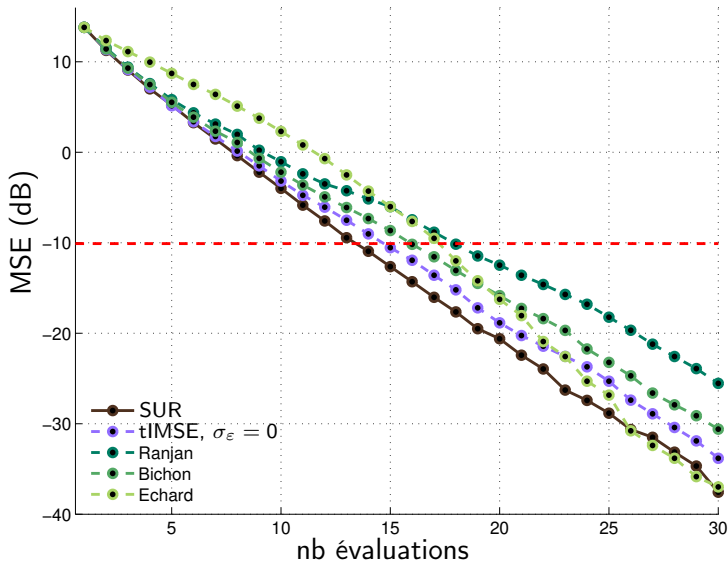
# Une comparaison des critères existants : protocole

- $\mathbb{X} = [0; 1]$  avec  $P_{\mathbb{X}} = \mathcal{U}_{[0;1]}$
- comparaison en moyenne sur 4000 trajectoires d'un GP
  - GP centré, covariance de Matérn ( $\sigma^2 = 1$ ,  $\nu = 2$ ,  $\rho = 0.1$ )
  - les algos utilisent la vraie fonction de covariance
  - $T$  de façon à ce que  $\alpha = 2\%$ .
- version discrétisée du problème (« méta-estimation »)
  - $Y_1, \dots, Y_m \stackrel{\text{iid}}{\sim} P_{\mathbb{X}}$  ( $m = 500$ )
  - on veut estimer l'estimateur  $\alpha^m = \frac{1}{m} \sum_j \mathbb{1}_{\xi(Y_j) > T}$
  - on travaille sur  $\tilde{\mathbb{X}}^m = \{Y_1, \dots, Y_m\}$  avec  $\tilde{P}_{\mathbb{X}}^m = \frac{1}{m} \sum_j \delta_{Y_j}$
- **autres critères** comparés (en plus de SUR et targeted IMSE)
  - Ranjan, Bingham et Michailidis ([Technometrics 2008](#))
  - Bichon, Eldred, Swiler, Mahadevan et McFarland ([J. AIAA 2008](#))
  - Echard, Gayton et Lemaire ([Journées Nat. de Fiabilité, 2010](#))

# Une comparaison des critères existants : résultats (1/2)



# Une comparaison des critères existants : résultats (2/2)



A retenir :

- Prise en compte d'**information a priori** sur  $f$ , dans un cadre **bayésien non-paramétrique** → méthodes qui **convergent rapidement**
- L'approche déplace le problème d'intégration mais ne le supprime pas → complément aux méthodes existantes (par ex. Monte Carlo)
- Les premières études empiriques montrent le potentiel des stratégies *one-step look-ahead* (SUR, targeted IMSE)

A retenir :

- Prise en compte d'information a priori sur  $f$ , dans un cadre bayésien non-paramétrique → méthodes qui convergent rapidement
- L'approche **déplace le problème d'intégration mais ne le supprime pas** → complément aux méthodes existantes (par ex. Monte Carlo)
- Les premières études empiriques montrent le potentiel des stratégies *one-step look-ahead* (SUR, targeted IMSE)

A retenir :

- Prise en compte d'information a priori sur  $f$ , dans un cadre bayésien non-paramétrique → méthodes qui convergent rapidement
- L'approche déplace le problème d'intégration mais ne le supprime pas → complément aux méthodes existantes (par ex. Monte Carlo)
- Les premières études empiriques montrent le potentiel des **stratégies *one-step look-ahead*** (SUR, targeted IMSE)

- Mieux exploiter la complémentarité bayésienne/fréquentiste
  - **méta-estimation** → **intervalles de confiance**
    - papier en cours de réaction avec E. Vazquez (bientôt sur arXiv...)
  - **combinaison GP / techniques d'échantillonnage aléatoire**
    - cf. exposé de Pierre Barbillon à MASCOT-NUM 2010

- Mieux exploiter la complémentarité bayésienne/fréquentiste
  - méta-estimation → intervalles de confiance
    - papier en cours de réaction avec E. Vazquez (bientôt sur arXiv...)
  - combinaison GP / techniques d'échantillonnage aléatoire
    - cf. exposé de Pierre Barbillon à MASCOT-NUM 2010
- Poursuivre l'analyse des stratégies de planification
  - **en simulation & sur cas tests industriels**
    - collaboration SUPELEC / B. Iooss (code Cathare)
    - D. Ginsbourger et collab. : projet de toolbox libre (R) ?
  - **sur le plan théorique**
    - convergence, vitesse de convergence



- Mieux exploiter la complémentarité bayésienne/fréquentiste
  - méta-estimation → intervalles de confiance
    - papier en cours de réaction avec E. Vazquez (bientôt sur arXiv...)
  - combinaison GP / techniques d'échantillonnage aléatoire
    - cf. exposé de Pierre Barbillon à MASCOT-NUM 2010
- Poursuivre l'analyse des stratégies de planification
  - en simulation & sur cas tests industriels
    - collaboration SUPELEC / B. Iooss (code Cathare)
    - D. Ginsbourger et collab. : projet de toolbox libre (R) ?
  - sur le plan théorique
    - convergence, vitesse de convergence
- En cours : thèse de Li Ling (financement projet CSDL)

## Références traitant de la stratégie SUR :

- E. Vazquez et M. Piera-Martinez, [Estimation du volume des ensembles d'excursion d'un processus gaussien par krigeage intrinsèque](#), 39èmes Journées de Statistiques (JdS 2007), CD-ROM proceedings, Angers (France), 11–15 juin 2007 [\[clickme\]](#)
- M. Piera-Martinez, [Modélisation des comportements extrêmes en ingénierie](#), Thèse de l'Université Paris-Sud 11 / Supélec, 29 septembre 2008 [\[clickme\]](#)
- E. Vazquez et J. Bect, [A sequential Bayesian algorithm to estimate a probability of failure](#), 15<sup>th</sup> IFAC Symposium on System Identification (SYSID 2009), Saint-Malo, France. IFAC-PapersOnLine / Elsevier, 2009 [\[clickme\]](#)

## Références traitant de la stratégie *targeted IMSE* :

- V. Picheny, D. Ginsbourger, O. Roustant et R.T. Haftka, [Iterative Designs of experiments for constraint approximation](#), 8th European Network for Business and Industrial Statistics conference (ENBIS8), CD-ROM proceedings, Athènes (Grèce), 21–25 septembre 2008
- V. Picheny, D. Ginsbourger, O. Roustant, R.T. Haftka et N.-H. Kim, [Adaptive designs of experiments for accurate approximation of target regions](#), [Journal of Mechanical Design](#), to be published [\[clickme\]](#)
- V. Picheny, [Improving accuracy and compensating for uncertainty in surrogate modeling](#), Thèse de l'École Nationale Supérieure des Mines de Saint-Etienne, 15 décembre 2009 [\[clickme\]](#)

Autres références décrivant des stratégies séquentielles :

- P. Ranjan, D. Bingham et G. Michailidis, [Sequential experiment design for contour estimation from complex computer codes](#), *Technometrics*, 50(4), 527–541, 2008 [\[clickme\]](#)
- B. J. Bichon, M. S. Eldred, L. P. Swiler, S. Mahadevan et J. M. McFarland, [Efficient global reliability analysis for nonlinear implicit performance functions](#), *AIAA Journal*, 46(10), 2459–2468, 2008 [\[clickme\]](#)
- B. Echart, N. Gayton et M. Lemaire, [Kriging based Monte Carlo simulation to compute the probability of failure efficiently : the AK-MCS methods](#), *6èmes Journées Nationales de Fiabilité*, Toulouse, France, 2010

## Apparition du krigeage en fiabilité :

- V. J. Romero, L. P. Swiler et A. A. Giunta, [Construction of response surfaces based on progressive-lattice-sampling experimental designs with application to uncertainty propagation](#), [Structural Safety](#), 26(2), 201–219, 2004 [\[clickme\]](#)
- I. Kaymaz, [Application of kriging method to structural reliability problems](#), [Structural Safety](#), 27(2), 133–151, 2005 [\[clickme\]](#)
- S. Shan et G. G. Wang, [Failure surface frontier for reliability assessment on expensive performance function](#), [Journal of Mechanical Design](#), 128(6), 1227–1236, 2006 [\[clickme\]](#)