Correction du critère EI pour la prise en compte de l'incertitude sur les paramètres d'une covariance

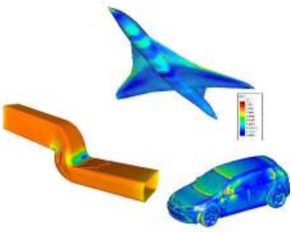
Romain Benassi Julien Bect Emmanuel Vazquez

SUPELEC

4 mai 2010
Atelier optimisation GDR MASCOT-NUM

Contexte

- Conception de systèmes à l'aide de simulations numériques
- Ici, problème de conception = problème d'optimisation mono-objectif
- Si les simulations prennent beaucoup de temps, le nombre de simulations qu'il est possible d'effectuer est limité (p.ex. < 100)
- Exemple : optimisation de formes



- Comment optimiser une fonction dont l'évaluation est coûteuse ?

1. Optimisation bayésienne par EI
2. Prise en compte de l'incertitude sur les paramètres
3. Exemple
4. Conclusions

1. Optimisation bayésienne par EI

2. Prise en compte de l'incertitude sur les paramètres

3. Exemple

4. Conclusions

Optimisation bayésienne

- $\mathbb{X} \subset \mathbb{R}^d$ compact : domaine de recherche
- $f : \mathbb{X} \to \mathbb{R}$ coûteuse à évaluer
- On cherche à maximiser f, c.-à-d. construire une suite de points d'évaluations (x_n) de façon à maximiser $M_N = f(x_1) \vee \cdots \vee f(x_N)$, avec N le budget d'évaluations de f

- Approche bayésienne : prise en compte d'informations a priori sur f
  $\Rightarrow$ modélisation de f par un processus aléatoire ξ

- Planification séquentielle des points d'évaluations (x_n)
  - à chaque nouvelle évaluation, utilisation de la loi *a posteriori* de ξ
  - construction d'un critère d'échantillonnage pour choisir les nouveaux points

Processus gaussien et EI [Mockus et al. 1978] 1/3

- Le plus souvent, ξ est pris gaussien
  - de moyenne nulle (ici, pour simplifier)
  - de fonction de covariance $k(x, y; \sigma, \underline{\theta}) = \sigma^2 r(x - y; \underline{\theta})$, avec $x, y \in \mathbb{X}$, $\sigma > 0$ et $\underline{\theta} \in \Theta$ vecteur de paramètres inconnus

- Principe d'optimisation à un pas si σ et $\underline{\theta}$ sont connus : choisir

$$x_{n+1} = \text{argmax}_{x \in \mathbb{X}} \mathbb{E}[M_n \vee \xi(x) \mid \underline{\xi}_n, \sigma, \underline{\theta}]$$

avec $M_n = \xi(x_1) \vee \cdots \vee \xi(x_n)$ et $\underline{\xi}_n = (\xi(x_1), \ldots, \xi(x_n))^\mathsf{T}$

Processus gaussien et EI [Mockus et al. 1978] 2/3

- On montre facilement que

$$x_{n+1} = \text{argmax}_{x \in \mathbb{X}}\, \rho_n(x; \sigma, \underline{\theta}) = \text{argmax}_{x \in \mathbb{X}}\, \mathbb{E}[(\xi(x) - M_n)_+ \mid \underline{\xi}_n, \sigma, \underline{\theta}]$$

et que l'amélioration moyenne (expected improvement)

$$\rho_n\left(x; \sigma, \underline{\theta}\right) = s_n(x)[u\Phi(u) + \Phi'(u)]$$

avec
- $u = \left(\hat{\xi}_n(x) - M_n\right)/s_n(x)$
- $\hat{\xi}_n(x) = \mathbb{E}[\xi(x) \mid \xi_n(x), \sigma, \underline{\theta}]$ (prédicteur par krigeage de $\xi(x)$)
- $s_n(x)$ l'écart-type *a posteriori* de l'erreur $\hat{\xi}_n(x) - \xi(x)$
- Φ fonction de répartition de $\mathcal{N}(0, 1)$

ρ_n se calcule très facilement !

Processus gaussien et EI [Mockus et al. 1978] 3/3

- EI : critère utilisé par l'algorithme EGO [Jones et al. 1998]
- Problème : σ et $\underline{\theta}$ ne sont pas connus !
- Solution possible : estimer σ et $\underline{\theta}$ à chaque nouvelle évaluation

1 Optimisation bayésienne par EI

2 Prise en compte de l'incertitude sur les paramètres

3 Exemple

4 Conclusions

Estimation avec MLE

- Les paramètres $(\sigma^2, \underline{\theta})$ peuvent être estimés par maximum de vraisemblance

$$\hat{\sigma}_n, \hat{\underline{\theta}}_n = \text{argmax} \frac{1}{(2\pi\sigma^2)^{n/2} |R(\underline{\theta})|^{1/2}} \exp\left[-\frac{\underline{\xi}_n^T R(\underline{\theta})^{-1} \underline{\xi}_n}{2\sigma^2} \right],$$

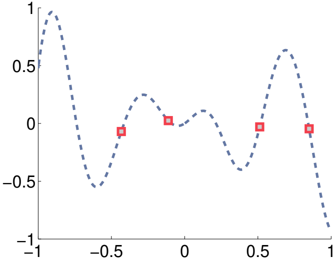
où $R(\underline{\theta})$ est la matrice de corrélation de $\underline{\xi}_n$

- L'algorithme EGO : $x_{n+1} = \text{argmax}_x\, \rho_n \left(x; \hat{\sigma}_n, \hat{\underline{\theta}}_n \right)$

- Facile à calculer mais ne fonctionne pas bien pour les *deceptive functions*

Deceptive functions 1/2

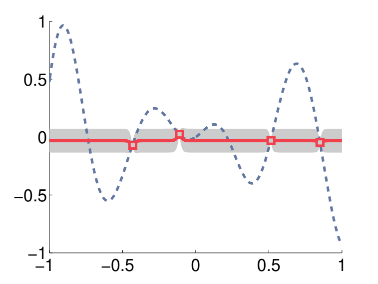
- Les points d'évaluation choisis peuvent ne pas être représentatifs des variations de la fonction objectif f



- Ces situations "trompeuses" entraînent une mauvaise estimation de l'erreur de prédiction
- L'estimation des paramètres par MLE y est souvent peu satisfaisante

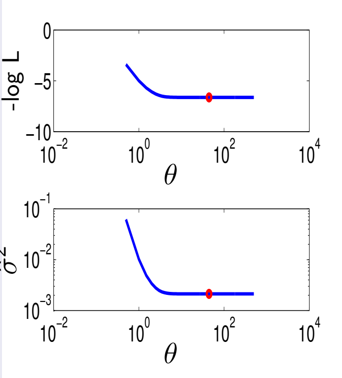
Deceptive functions 2/2

Prédiction par MLE



Estimation de $\hat{\sigma}^2$ par MLE

Valeur de $-\log L$ et de $\hat{\sigma}^2$



Estimation par approche bayésienne

- Proposition : approche bayésienne pour prendre en compte l'incertitude sur les paramètres de la covariance
- Choix de lois *a priori* sur les paramètres
- La nouvelle expression de l'EI est

$$\tilde{\rho}_n(x) = \mathbb{E}\left[\rho_n(x; \sigma^2, \underline{\theta}) \mid \underline{\xi}_n\right]$$

- Lors du choix d'un nouveau point d'évaluation, l'incertitude associée aux paramètres sera ainsi prise en compte

Choix des lois *a priori* pour les paramètres

- Pour σ^2 : loi inverse gamma
- Pour $\underline{\theta}$: on fait le choix d'une discrétisation

$\Rightarrow$ nombre fini de valeurs $\underline{\theta}_k$ avec des probabilités *a priori* $p(\underline{\theta}_k)$ choisies par l'utilisateur

Correction du critère EI

1. Calcul analytique de

$$\rho'_n(x; \underline{\theta}) = \mathbb{E}[(\xi(x) - M_n)_+ \mid \xi_n, \underline{\theta}]$$

(comme ξ est gaussien, $\xi(x) \mid \underline{\xi}_n, \underline{\theta}$ suit une loi de Student)

2. Calcul de

$$\begin{aligned} \tilde{\rho}_n(x) &= \mathbb{E}(\rho'_n(x, \underline{\theta}) \mid \underline{\xi}_n) \\ &= \int \rho'_n(x, \underline{\theta}) P(\,\mathrm{d}\underline{\theta} \mid \underline{\xi}_n) \\ &= \sum_k \rho'_n(x, \underline{\theta}_k) p(\underline{\theta}_k \mid \underline{\xi}_n) \end{aligned}$$

avec

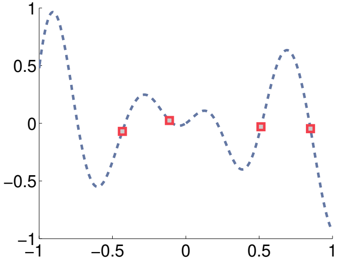
$$p(\underline{\theta}_k \mid \underline{\xi}_n) \propto p(\underline{\xi}_n \mid \underline{\theta}_k) p(\underline{\theta}_k)$$

et $p(\underline{\xi}_n \mid \underline{\theta}_k)$ qui suit une loi de Student multivariée

1. Optimisation bayésienne par EI
2. Prise en compte de l'incertitude sur les paramètres
3. Exemple
4. Conclusions

Fonction objectif

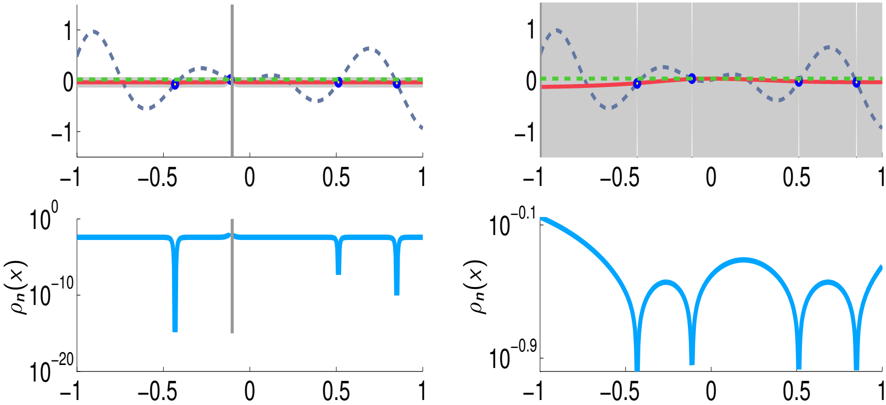
La fonction objectif et le design initial choisis (deceptive function)



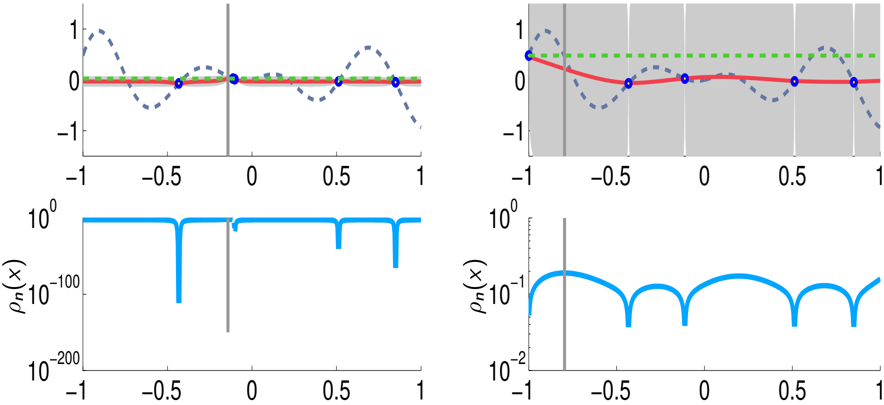
Conditions de test

- Pour l'approche bayésienne, on suppose que σ^2 suit une loi *a priori* inverse gamma $IG(0.2, 12)$
- On utilise des covariances de Matèrn de paramètre $[\sigma^2;\ \nu;\ 1/\alpha]$
- La régularité ν est prise égale à 2
- On a $\underline{\theta} = [\alpha]$ (l'inverse de la portée)
- On considère $m = 100$ valeurs possibles pour $\underline{\theta}$ réparties exponentiellement
- Les lois *a priori* sont choisies uniformes

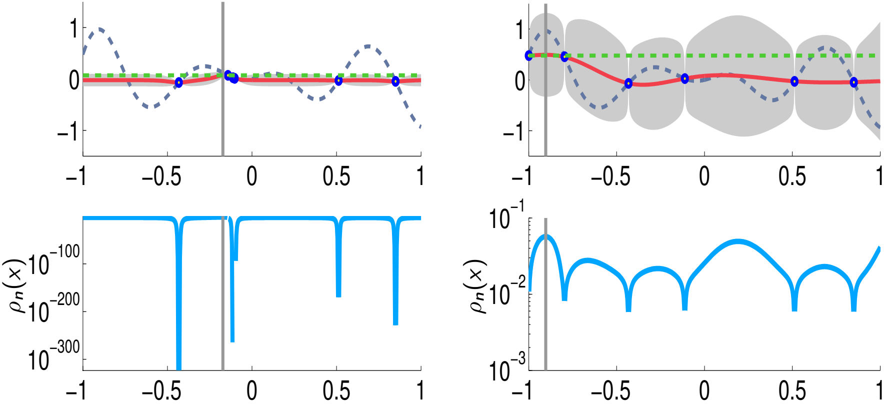
Résultats : itération 1



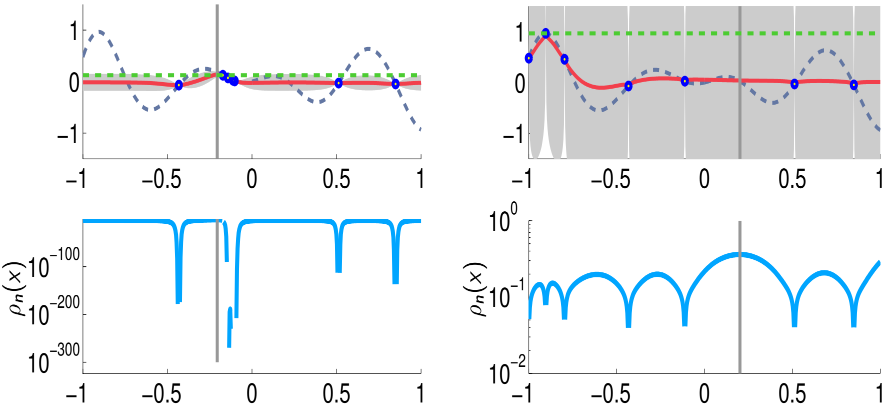
Résultats : itération 2



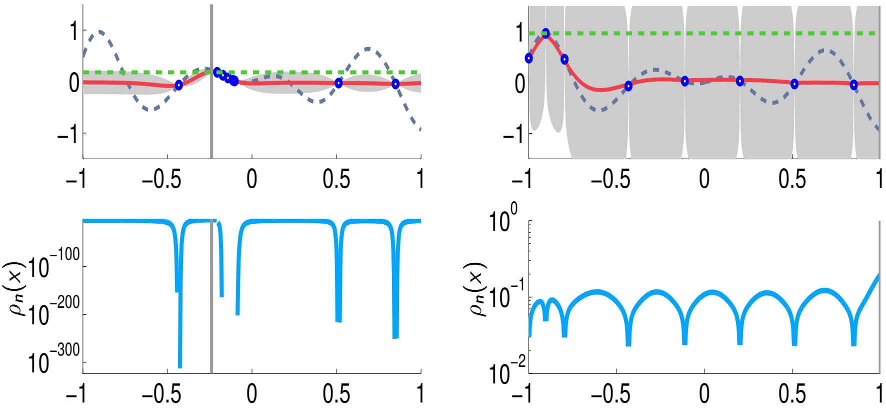
Résultats : itération 3



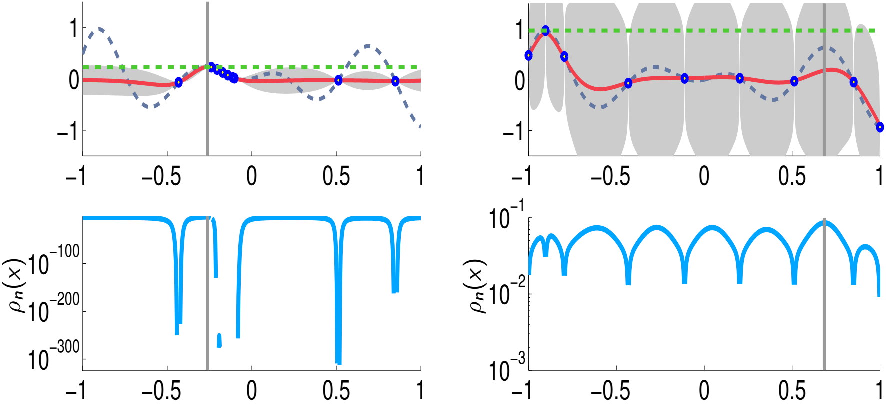
Résultats : itération 4



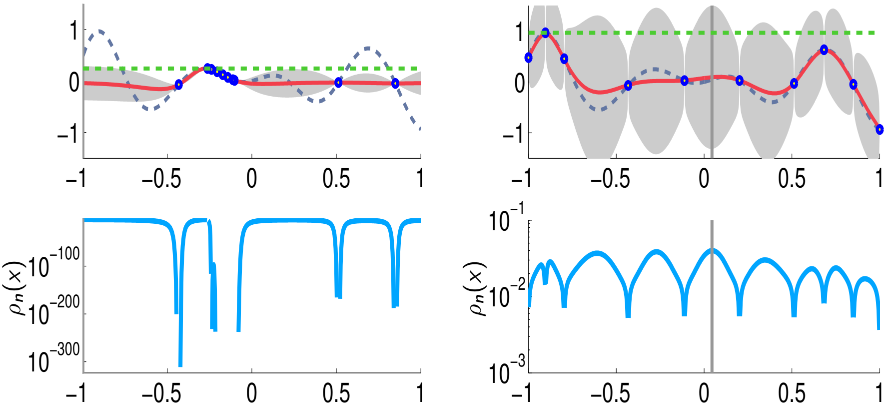
Résultats : itération 5



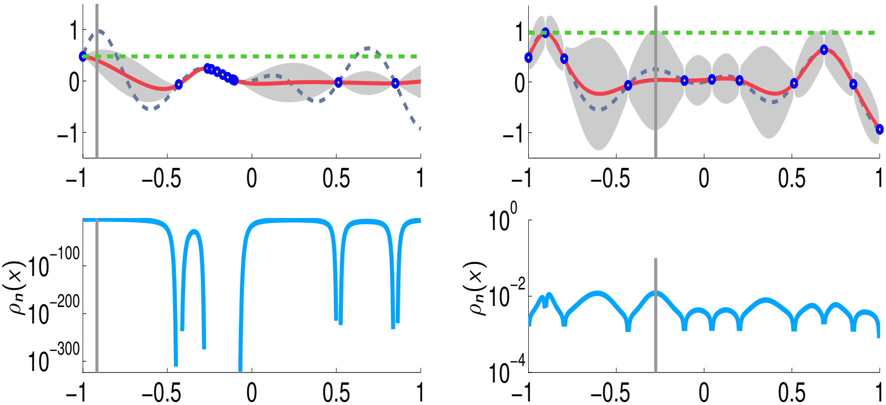
Résultats : itération 6



Résultats : itération 7



Résultats : itération 8



1 Optimisation bayésienne par EI

2 Prise en compte de l'incertitude sur les paramètres

3 Exemple

4 Conclusions

- Dans certaines situations (*deceptive functions* mais pas seulement...), estimer les paramètres de la covariance par MLE à chaque itération conduit à des résultats très décevants

- L'approche bayésienne prend en compte l'incertitude sur les paramètres et évite les phénomènes de clustering des points d'évaluations

- Les résultats obtenus montrent que cette nouvelle approche est robuste, même dans des situations très défavorables