

Prise en compte de la stochasticité dans les modèles :
optimisation robuste

Rodolphe Le Riche (CNRS & EMSE) & Victor Picheny (INRA)

La Rochelle, 5/11/2014

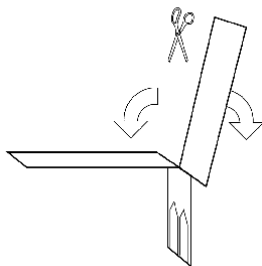
Plan

- 1 Introduction
- 2 Formulations pour l'optimisation sous incertitudes
- 3 Approches par métamodélisation
 - Le cas bruité (pas de contrôle de U)
 - Le cas où U est contrôlable
- 4 Algorithmes évolutionnaires

Un exemple introductif : l'hélicoptère en papier

Un problème d'optimisation simple...

- 4 paramètres (longueur, hauteur $\times 2$)
- Min et max pour chaque paramètre (doit tenir sur une feuille A4)
- Objectif : "maximiser le temps de chute"



Protocole experimental

Découpage, cage d'escalier + chronomètre
[fig2/helicovideo.mov](#)

Un exemple introductif (suite)

Un système hautement stochastique

- Erreur sur les paramètres (découpage)
- Erreur sur les mesures (chronométrage)
- Angle pales / corps aléatoire (pliage)
- Courant d'air...

⇒ Temps de chute toujours différent

Comment faire ?

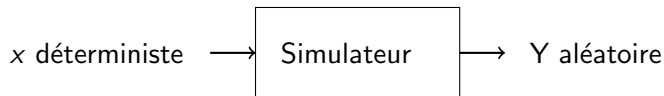
- Lancer plusieurs fois le même hélicoptère ?
- Découper plusieurs fois le même hélicoptère ?

Et surtout

Comment détermine-t-on le meilleur hélicoptère ?

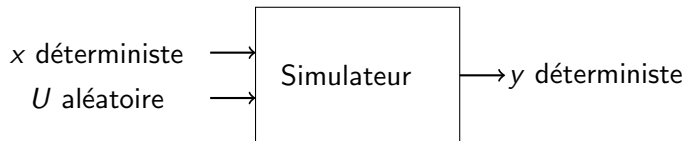
Stochasticité dans les modèles complexes

Modèle stochastique



Le simulateur contient des éléments stochastiques (tirages aléatoires non contrôlés, etc.)

Modèle déterministe, entrées stochastiques



U : environnement incertain (climat ?), paramètre méconnu, etc.

Plan

- 1 Introduction
- 2 Formulations pour l'optimisation sous incertitudes
- 3 Approches par métamodélisation
 - Le cas bruité (pas de contrôle de U)
 - Le cas où U est contrôlable
- 4 Algorithmes évolutionnaires

La paramétrisation double (x, U)

cf. Taguchi, 80's

- x : vecteur des variables d'optimisation déterministes
- U : vecteur de paramètres incertains

U sert à décrire :

- un bruit (de mesure)
- l'erreur de modèle (épistémique)
- une incertitude sur certains paramètres du simulateur

Une paramétrisation générale

3 cas (combinables) :

Bruit sur les variables d'optimisation

$f(x + U)$ (erreur de découpage)

Bruit exogène aux variables d'optimisation

$f(x, U)$ (courant d'air)

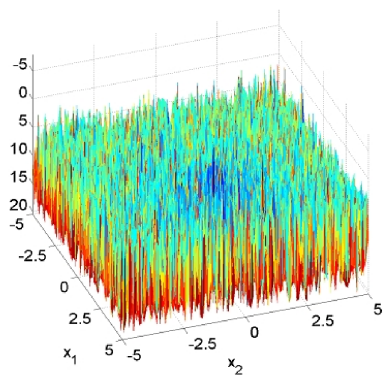
Bruit sur la sortie du modèle

$f(x) + U$ (chronométrage + courant d'air)

Le cas bruité

On ne prend pas en compte le bruit,
on essaie de résoudre :

$$\min_{x \in \mathbb{X}} f(x, U)$$



Pas de gradients, quels résultats pour l'optimisation ?

Reformulation

On remplace le critère d'optimisation bruité par une mesure statistique :

$$\min_{x \in \mathbb{X}} y(x) = \mathbb{E} [f(x, U)]$$

Pour l'hélicoptère : on veut maximiser le temps de chute moyen.

Autres possibilités :

- quantile : $\min_{x \in \mathbb{X}} q_\alpha [f(x, U)]$ (*5e pire lancer / 100*)
- espérance + variance : $\min_{x \in \mathbb{X}} \mathbb{E} [f(x, U)] + \rho \sqrt{V [f(x, U)]}$

⇒ On retrouve un pb d'optimisation déterministe !

Approches directes

En pratique, les mesures statistiques sont estimées

Monte-Carlo simple :

$$\hat{y}(x) = \hat{\mathbb{E}}_U [f(x, U)] = \frac{1}{MC} \sum_{i=1}^{MC} f(x, u_i), \quad u_i \text{ i.i.d. } \sim \mathcal{U}$$

$$\hat{q}_\alpha(x) = [MC \times \alpha]\text{-ième plus petit parmi } f(x, u_1), \dots, f(x, u_{MC})$$

Deux problèmes majeurs :

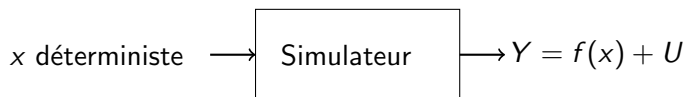
- **Double boucle** : propager les incertitudes sur U , optimiser sur $x \Rightarrow$ très coûteux !
- Objectifs estimés \Rightarrow **bruit**

Plan

- 1 Introduction
- 2 Formulations pour l'optimisation sous incertitudes
- 3 Approches par métamodélisation
 - Le cas bruité (pas de contrôle de U)
 - Le cas où U est contrôlable
- 4 Algorithmes évolutionnaires

Modèles polynomiaux

Modèle “bruité” : optimisation de la moyenne : $\min \mathbb{E}[Y] = f(x)$



Régression : prise en compte naturelle du bruit

$$Y = \mathbf{X}\beta + \epsilon$$

Le métamodèle ($\mathbf{X}\hat{\beta}$) donne une approximation **de la moyenne**

Optimisation : cadre favorable mais...

- Locale : plans d'expériences D-optimaux : algorithme de Kiefer \Rightarrow gourmand en pratique !
- Régions de confiance : peu de résultats théoriques de convergence

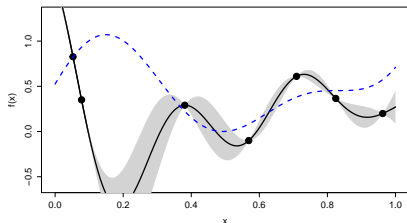
Krigeage et observations bruitées

On s'intéresse à : $f(x)$ (*temps de chute moyen*)

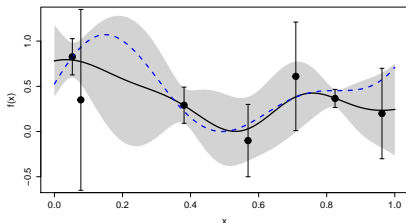
On a accès à : $Y = f(x) + U$ (*temps de chute mesuré une fois*)

Adaptation au bruit : ajout d'un terme dans la matrice de covariance

krigeage classique



krigeage avec "effet de pépite"



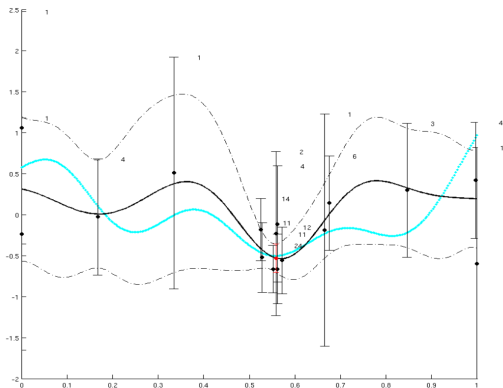
Modèle **non interpolant** : le krigeage prédit $f(x)$, pas $f(x) + U$!

La moyenne de krigeage filtre le bruit

Optimisation bruitée avec krigage

Ne pas utiliser l'EI car f_{\min} est bruité (potentiellement trompeur)

Plusieurs stratégies adaptées en redéfinissant un EI : par exemple EQI (amélioration sur les *quantiles du krigage*)



V. Picheny, D. Ginsbourger (2013)

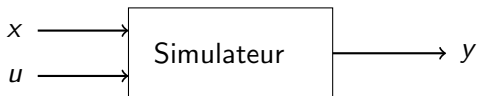
Noisy kriging-based optimization methods : A unified implementation within the DiceOptim package

Computational Statistics & Data Analysis

Le bruit doit être connu ou estimé

Le cas où U est contrôlable

C'est notamment le cas d'un simulateur numérique dont les entrées x et u sont choisies.



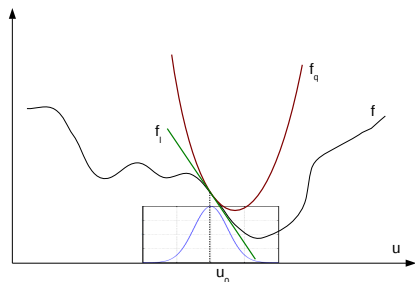
Hypothèse : U suit une distribution connue.

Il faut choisir pour chaque simulation un couple (x, u)

- MC possible (un x , N u)
- Méthode des moments
- Enorme gain à considérer l'espace joint $\mathbb{D} = \{X, U\}$

Méthode des moments

- Si les U peuvent être contrôlés, on peut faire une approximation à l'ordre 1 ou 2 de $f(x, u)$ en u par exemple par différences finies en u .
- Si $U \sim \mathcal{N}(u_0, \Gamma^2)$, alors les moments (moyenne, variance, ...) de f peuvent être approchés analytiquement.
- AN : cas assez général car de nombreuses distributions peuvent être ramenées aux distributions gaussiennes par transformations de Nataf ou Rosenblatt.



Moments d'une forme quadratique gaussienne :



[D. Slepian \(1958\)](#)

Fluctuations of Random Noise Power,
[Bell System Technical Journal](#)

Transformations de Rosenblatt et Nataf :



[R. E. Melchers \(1999\)](#)

Structural Reliability Analysis and
Prediction, [Wiley \(2nd Ed.\)](#)

Méthode des moments (approximation linéaire)

$$f(x, U) \approx f_L(x, U) \equiv f(x, u_0) + \nabla_U f(x, u_0)(U - u_0)$$

(simplification des notations puisque x fixe ici)

$$f_L(x, U) \Rightarrow f_L(U) \equiv f(u_0) + \nabla f(u_0)(U - u_0)$$

$f_L(U)$ est une transformation linéaire de U , on a directement

$$\begin{aligned} \mathbb{E}f_L(U) &= f(u_0) \\ \nabla f_L(U) &= \nabla f(u_0) \Gamma^2 \nabla f(u_0)^\top \end{aligned}$$

Méthode des moments (approximation quadratique)

$$f(U) \approx f_q(U) \equiv f(u_0) + \nabla f(u_0)(U - u_0) + \frac{1}{2}(U - u_0)^\top \nabla^2 f(u_0)(U - u_0)$$

mais aussi

$$f_q(U) = a + \frac{1}{2}(U - u^*)^\top \nabla^2 f(u_0)(U - u^*)$$

où $u^* = u_0 - (\nabla^2 f(u_0))^{-1} \nabla f(u_0)^\top$ et

$$a = f(u_0) - \nabla f(u_0)(\nabla^2 f(u_0))^{-1} \nabla f(u_0)^\top.$$

Suivent des changements de variables pour se ramener à des v.a. normales centrées réduites :

$$(U - u^*)^\top \nabla^2 f(u_0)(U - u^*) = V^\top \underbrace{\Gamma^\top \nabla^2 f(u_0) \Gamma}_B V$$

où $\Gamma V = U - u^*$, $V \sim \mathcal{N}(\Gamma^{-1}(u_0 - u^*), I_n)$.

Méthode des moments (approximation quadratique)

B est symétrique et réelle, elle peut être diagonalisée, $B = PDP^T$,

$$V^T B V = W^T D W$$

où $W = P^T V$, $W \sim \mathcal{N}(\underbrace{P^T \Gamma^{-1}(u_0 - u^*)}_{\equiv m}, I_n)$ car $P^T P = I_n$. Enfin,

$$W^T D W = (Z + m)^T D (Z + m) = m^T D m + 2m^T D Z + Z^T D Z$$

où $Z = W - m$, $Z \sim \mathcal{N}(0, I_n)$. Après les changements de variables, l'approximation quadratique devient,

$$\begin{aligned} f_q(Z) &= a + \frac{1}{2} m^T D m + m^T D Z + \frac{1}{2} Z^T D Z \\ &= a + \frac{1}{2} m^T D m + \sum_{i=1}^n D_i m_i Z_i + \frac{1}{2} \sum_{i=1}^n D_i Z_i^2 \end{aligned}$$

Méthode des moments (approximation quadratique)

$$Z_i^2 \sim \chi_1^2 \Rightarrow \mathbb{E}Z_i^2 = 1 \text{ et } \mathbb{V}Z_i^2 = 2$$

$$\mathbb{E}f_q = a + m^\top Dm + \frac{1}{2}\text{Tr}(D)$$

$$\begin{aligned} \mathbb{V}f_q &= \sum_{i=1}^n D_i^2 m_i^2 \mathbb{V}(Z_i) + \frac{1}{4} \sum_{i=1}^n D_i^2 \mathbb{V}(Z_i^2) \\ &\quad + \sum_{i \neq j} D_i D_j m_i m_j \text{Cov}(Z_i, Z_j) + \frac{1}{4} \sum_{i \neq j} D_i D_j m_i \text{Cov}(Z_i, Z_j^2) \end{aligned}$$

$$\text{Cov}(Z_i, Z_j) = \text{Cov}(Z_i, Z_j^2) = 0 \text{ car } Z_i \perp\!\!\!\perp Z_j \text{ donc}$$

$$\mathbb{V}f_q = m^\top D^2 m + \frac{1}{2}\text{Tr}(D^2)$$

Approche par krigeage



Janusevskis & Le Riche (2012)

Simultaneous kriging-based estimation and optimization of mean response

Journal of Global Optimization

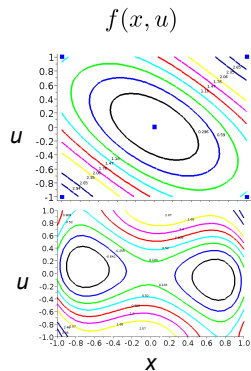
Algorithme en deux temps

- 1 Trouver x^* qui donne la meilleure amélioration pour $\mathbb{E}_U f(x, U)$
- 2 Trouver (x, u) qui donne le plus d'information sur $\mathbb{E}_U f(x^*, U)$

Outils

- “Processus moyen” $Z(x) = \mathbb{E}_U[Y(x, U)] + \text{amélioration espérée}$
- Réduction d'incertitude : $\min_{x,u} \text{VAR}(Z_{n+1}(x^*))$

Illustration (1/3)



integration

$\mathbb{E}_U[f(x, U)]$

Krigeage

f

Exact

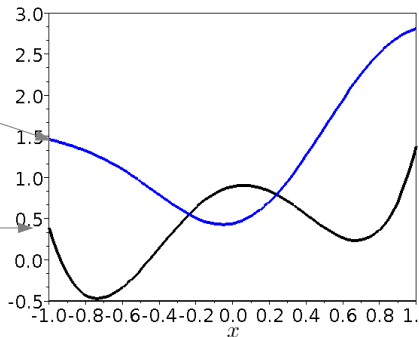


Illustration (2/3)

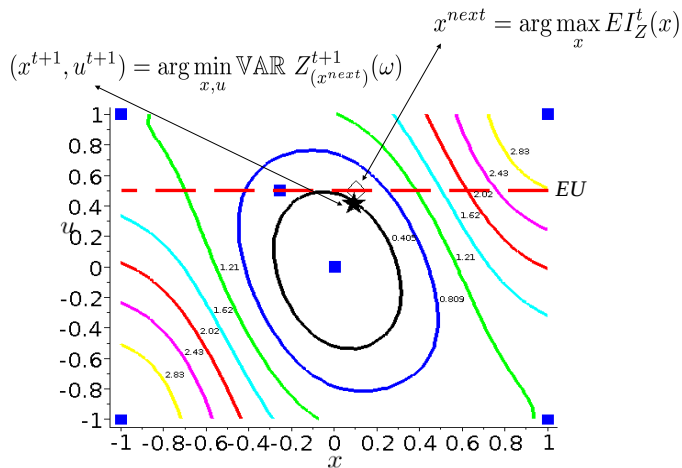
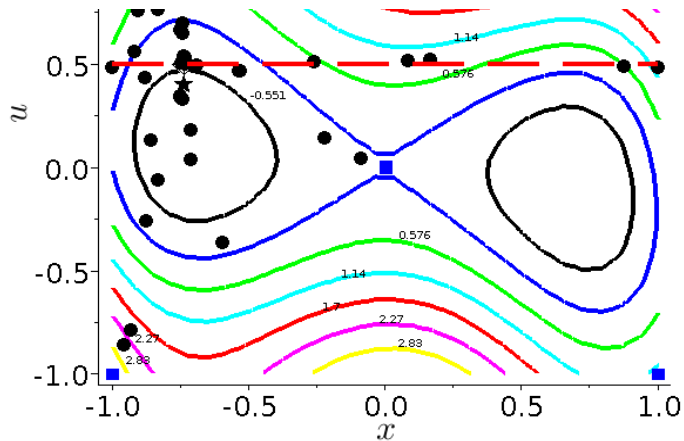


Illustration (3/3)



Conclusion partielle

Optimisation robuste et métamodèles

- Minimisation d'espérance, deux cas standards :

$$\min_{x \in \mathbb{X}} \mathbb{E}[f(x) + U] \quad \text{et} \quad \min_{x \in \mathbb{X}} \mathbb{E}_U[f(x, U)]$$

- $U = \text{bruit}$: le métamodèle, bâti avec hyp. de bruit additif, agit comme un filtre
- U contrôlable :
 - le métamodèle lie les deux espaces
 - Calcul analytique des moments

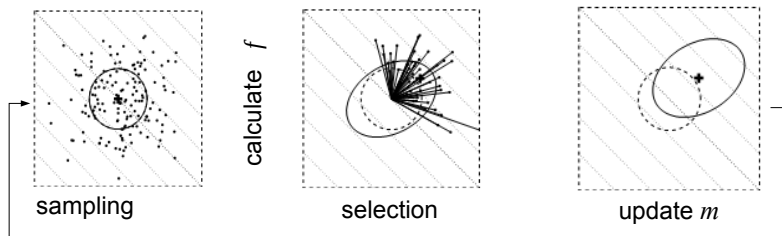
Limitations

- Dimension $(x+u) \leq 20$
- Complexité !

Plan

- 1 Introduction
- 2 Formulations pour l'optimisation sous incertitudes
- 3 Approches par métamodélisation
 - Le cas bruité (pas de contrôle de U)
 - Le cas où U est contrôlable
- 4 Algorithmes évolutionnaires

CMA-ES



$CMA - ES(\mu, \lambda)$ peut optimiser de nombreuses fonctions bruitées car :

- non élitiste
- choix de la moyenne de la prochaine itération filtre le bruit (échantillonnage spatial comme proxy pour l'échantillonnage de U)

$$m^{t+1} = m^t + \frac{1}{\mu} \sum_{i=1}^{\mu} y^{i:\lambda}$$

CMA-ES et optimisation bruitée

Ré-échantillonner ou augmenter la taille de la population ?

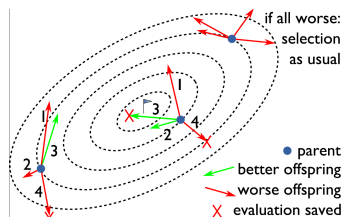
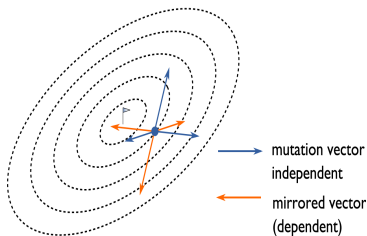
Il est possible de faire : $\hat{f}(x) = \frac{1}{\kappa} \sum_{i=1}^{\kappa} Y^i(x)$

⇒ Augmenter la population est plus efficace [Beyer and Sendhoff 2007, Arnold and Beyer 2006]

Améliorations spécifiques au bruit

- Observation : $(1 + 1) - \text{CMA} - \text{ES}$ avec redémarrage très efficace ← petites populations avantageuses ?
- “Elitisme” pénalisant pour les fonctions bruitées : un échantillon chanceux va attirer l’optimiseur dans une mauvaise région
- Comment définir un algorithme rapide, local et non élitiste ?

Echantillonnage en miroir et sélection séquentielle



Dérandomisation et échantillonnage en miroir

- un aléa \rightarrow deux enfants
- souvent bons / mauvais opposés

Sélection séquentielle

Arrêt des évaluations dès qu'une solution meilleure qu'un parent est trouvée

En combinant

Quand un enfant est meilleur que son parent, son symétrique est moins bon et inversement \rightarrow on évalue dans l'ordre $m + y_1, m - y_1, m + y_2, m - y_2, \dots$

Résultats

(1, 4) – *ES* avec miroir et sélection séquentielle

Testé dans BBOB'2010 (Black Box Optimization Benchmarking)

<http://coco.gforge.inria.fr/bbob2010-downloads>

Meilleure performance sur tous les algorithmes testés à ce jour sur certaines fonctions bruitées



[Brockhoff, Auger, Hansen, Arnold, Hohm \(2010\)](#)

Mirrored Sampling and Sequential Selection for Evolution Strategies

[PPSN XI](#)

Conclusion (1/2)

Pour récapituler

- Optimisation → difficile en présence de bruit → formulation adaptée
→ fonctions bruitées
- ⇒ Métamodèles + statistiques
- ⇒ Optimiseurs stochastiques

Domaines d'application

- CMA-ES : robuste (cf. cours du 4/11) mais inapplicables aux simulateurs coûteux
- Krigeage : petite dimension, simulateur \gg calculs de recherche de point

Conclusion (2/2)

Beaucoup d'approches non traitées

- Approches fiabilistes (FORM/SORM, contraintes probabilistes)
- “Anti-optimisation” (pire cas pour U)
- Chaos polynomial, ...

Beaucoup de choses à développer

- Métamodèles en grande dimension
- Approches Bayésiennes ou apparentées lorsque l'on ne connaît pas la loi de U
- Incertitudes en cascade dans les couplages de codes
- ...

Livres, articles de synthèse



R. E. Melchers (1999)
Structural Reliability Analysis and Prediction
Wiley (2nd Ed.)



A. Ben-Tal, L. El Ghaoui, A. Nemirovski (2009)
Robust Optimization
Princeton Univ. Press



M. Lemaire, A. Chateauneuf, J.-C. Mitteau (2009)
Structural Reliability
Wiley



P. Breitkopf and R. Filomeno Coehlo Eds. (2010)
Multidisciplinary Design Optimization in Computational Mechanics
Wiley/ISTE



H.G. Beyer, B. Sendhoff (2007)
Robust Optimization – A comprehensive survey
Comput. Methods Appl. Mech. Engrg, 196, pp. 3190-3218