ÉCOLE DE PHYSIQUE des HOUCHES

MaiMoSiNE

Groupement de Recherche MASCOT-NUM
Méthodes d'Analyse Stochastique pour les COdes et Traitements NUMériques

MEXICO

Optimisation par exploration de code numérique

Victor Picheny

INRA - Mathématiques et Informatique Appliquées - Toulouse

Les Houches, le 12 avril 2013

Optimisation par exploration de code numérique

Victor Picheny

INRA - Mathématiques et Informatique Appliquées - Toulouse

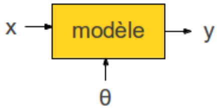
Les Houches, le 12 avril 2013

Plan

Qu'est-ce que l'optimisation de modèles (1/2) ?

Notations

- Facteurs x
- Paramètres θ
- Réponse(s) y



Objectif : trouver le jeu de facteurs pour lequel la réponse du système est la meilleure

- $x^* = \arg\min_{x \in D} y(x)$ ou $x^* = \arg\max_{x \in D} y(x)$
- paramètres θ fixes !

Qu'est-ce que l'optimisation de modèles (2/2) ?

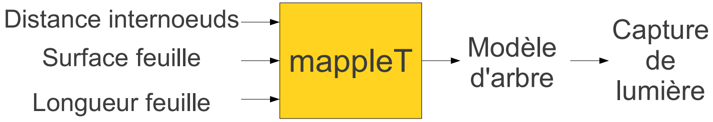
Optimisation = exploitation du modèle par **échantillonnage adaptatif**

- Trouver les décisions (facteurs) qui maximisent la performance (réponse)
- Plan d'expériences... focalisé sur la valeur de la réponse
- Nécessairement adaptatif (séquentiel)

Difficultés inhérentes

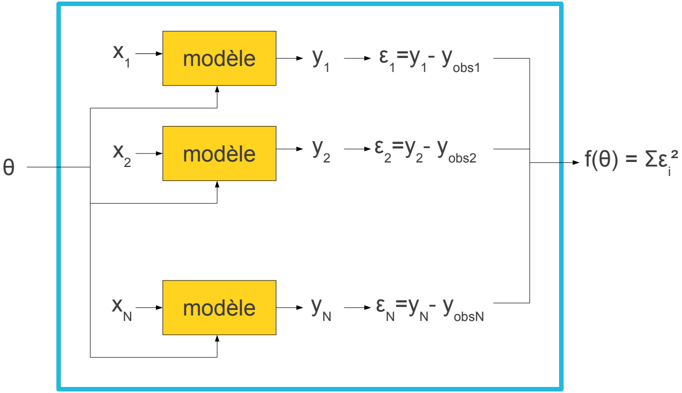
- Modèle = boîte noire : pas de dérivées ; au mieux informations de type régularité, etc.
- Temps de calcul élevé $\rightarrow$ parcimonie nécessaire

Un exemple



On peut agir sur certains traits géométriques pour obtenir des arbres avec une architecture optimale.

Un autre problème d'optimisation : le calage de paramètres



Objectif : $\theta^* = \arg\min f(\theta)$

Type de problème abordé dans cette présentation

- Toutes les variables sont des variables de décisions
- Variables continues, bornées
- Petite dimension (≤ 20)
- Réponse fortement non-linéaire

Plan

Optimisation locale vs. globale

Optimisation locale

- Départ : point initial x
- Direction de descente (plus forte pente : gradient)
- On améliore toujours la situation initiale, mais aucune garantie de trouver le meilleur point !

Optimisation globale : le **compromis exploration / intensification**

- Exploration : recherche partout dans l'espace pour ne pas rater la zone optimale
- Intensification : une fois une zone identifiée : on recherche le minimum local

La plus forte pente : une stratégie discutable

L'algorithme DIRECT

DIRECT : DIviding RECTangles

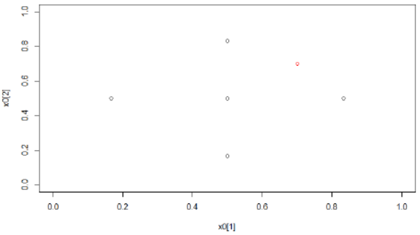
D. Jones, C. Perttunen, B. Stuckman (1993)
Lipschitzian optimization without the Lipschitz constant
Journal of Optimization Theory and Applications 79(1), 157-181

- Découpage de l'espace en (hyper)rectangles
- Un échantillon au centre de chaque rectangle
- On divise les rectangles les plus "intéressants" :
  - soit les plus grands (exploration)
  - soit ceux qui ont une valeur au centre basse (intensification)
- Pour diviser : ajout de 2 points, division en 3

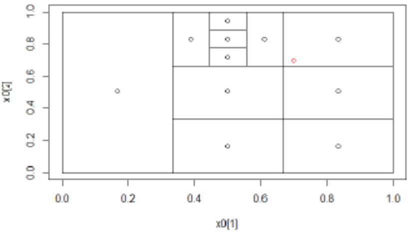
Quel compromis exploration / intensification ?

- Pour **chaque** taille de rectangle, on divise celui qui a la meilleure valeur de y
- Valeurs optimales au sens de Pareto !

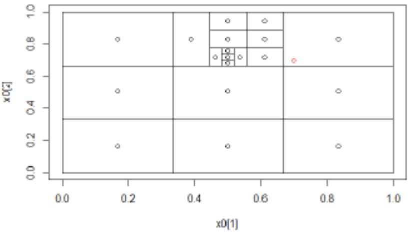
Exemple en dim 2



Après 2 itérations

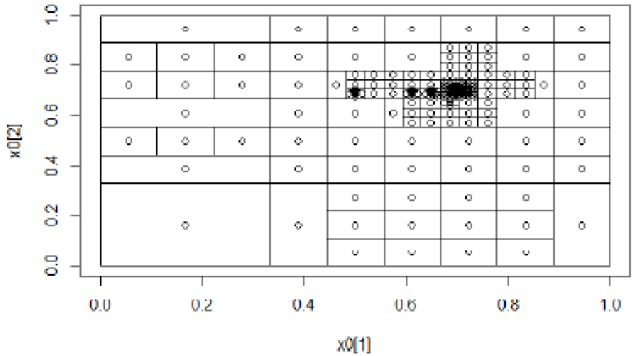


Après 3 itérations

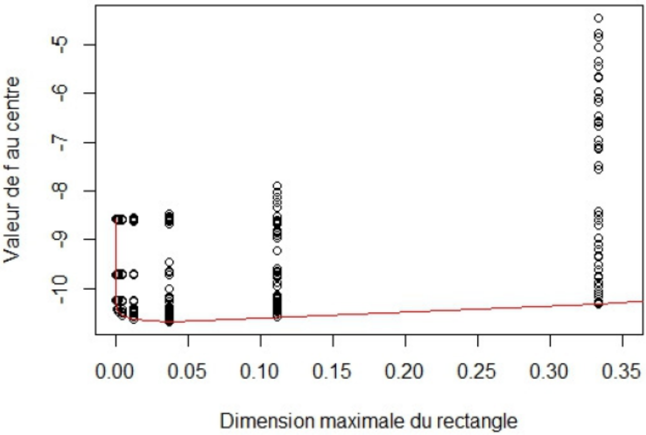


Après 20 itérations

- Echantillonnage intense dans la zone de l'optimum
- Bonne exploration



Front de Pareto



Plan

Optimisation sur base de métamodèles

Pourquoi faire ?

- Algorithme DIRECT : (presque) sans hypothèse sur y
- Situation d'extrême parcimonie : exploitation maximale de l'information $\rightarrow$ métamodèle

Principe général

- Echantillon initial (plan d'expériences)
- Construction d'un métamodèle
- Utilisation du métamodèle pour choisir la prochaine observation

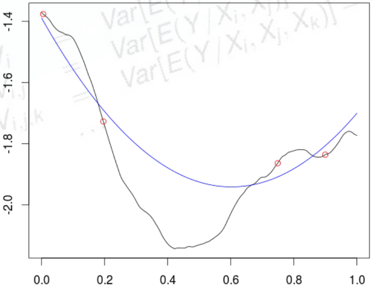
Approche séquentielle

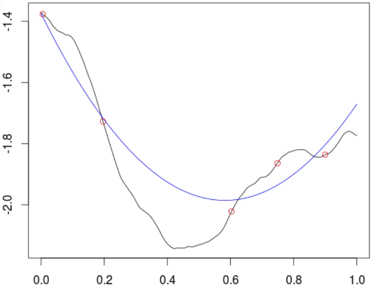
La nouvelle observation enrichit le plan d'expérience et le métamodèle

Optimisation sur base de modèle polynomial

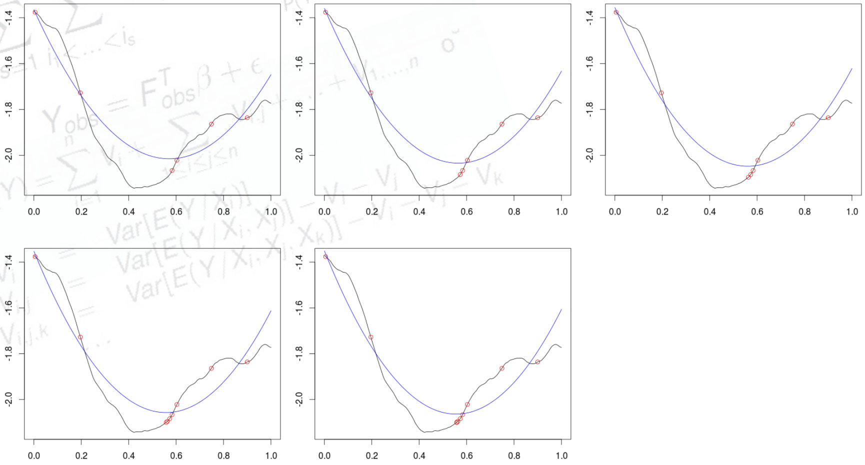
Principe

- On cherche le point qui minimise la surface de réponse
- On ajoute ce point
- On met à jour la surface de réponse
- On recommence...





Itérations 3 à 7



Optimisation sur base de modèle polynomial

Problème : modèle “rigide”

- Sélection des points et de la région de validité
  $\rightarrow$ régions de confiance
- Convergence locale : plans localement D-optimaux
  $\rightarrow$ algorithme de Kiefer

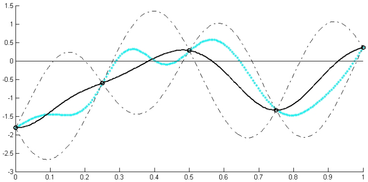
Où est passé le compromis exploration / intensification ?

Méthodes d’optimisation locale, proche des méthodes de gradient

Optimisation sur base de krigeage : l'algorithme EGO

Atouts du modèle (cf. cours de PG)

- Flexibilité
- Information riche :
  - Meilleur prédicteur $m(x)$
  - + Information ponctuelle d'erreur : variance $s^2(x)$



Compromis exploration / intensification

- Exploration : $s^2(x)$ grand ($\sim$ grands rectangles)
- Intensification : $m(x)$ petit ($\sim$ faibles valeurs)

Ou se placer sur le front de Pareto ?

Différents critères

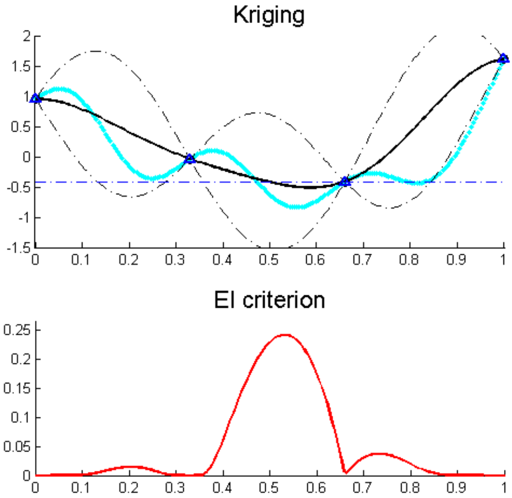
D. Jones (2001)
A taxonomy of global optimization methods based on response surfaces
Journal of global optimization 21(4), 345-383

- Pure intensification : $\min m(x)$
- Pure exploration : $\max s^2(x)$
- Un compromis simple : $\min m(x) - \alpha s(x)$

Le meilleur compromis : amélioration espérée (ou *EI: Expected Improvement*)

- Amélioration = gain sur la fonction coût : $(\min(y_1, \ldots, y_n) - Y(x))^+$
- Au temps n : l'amélioration est une v.a., *mais* on connaît sa loi
- Critère 1 : probabilité d'amélioration
- **Critère 2 : espérance de l'amélioration :**
  $EI(x) = \mathbb{E}\left[(\min(y_1, \ldots, y_n) - Y(x))^+\right]$
- Ecriture analytique en fonction de $m(x)$ et $s^2(x)$!

Illustration



L'algorithme *Efficient Global Optimization*

Initialisation

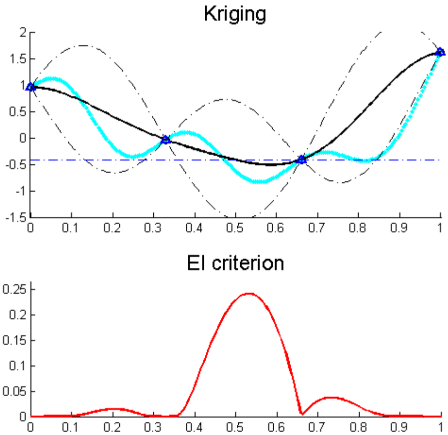
- Réalisation d'un plan initial
- Construction d'un krigeage

Boucle d'optimisation

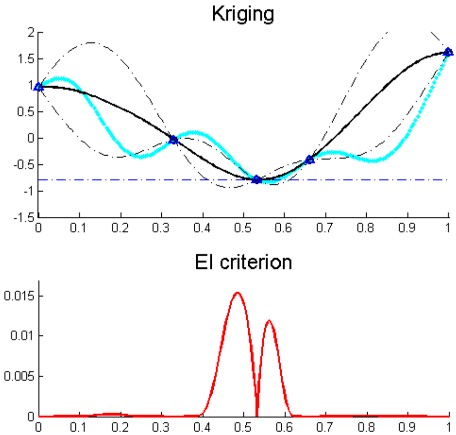
- Recherche du point qui maximise l'amélioration espérée
- Calcul du vrai modèle en ce point
- Ajout du point au plan et mise à jour du modèle

D. Jones, M. Schonlau, W. Welch
Efficient global optimization of expensive black-box functions
Journal of Global optimization 13 (4), 455-492 (1998)

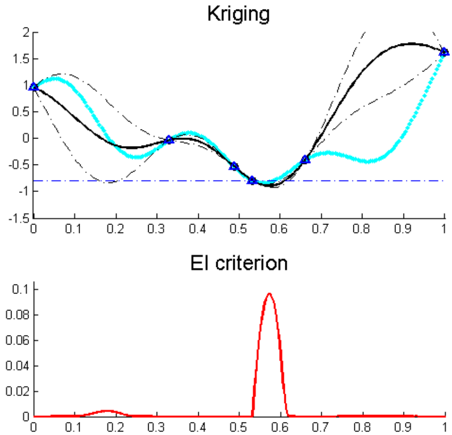
EGO: illustration 1/6



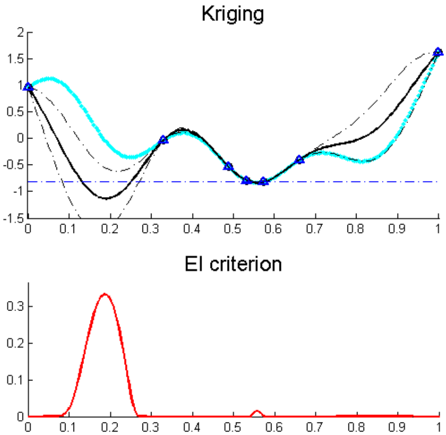
EGO: illustration 2/6



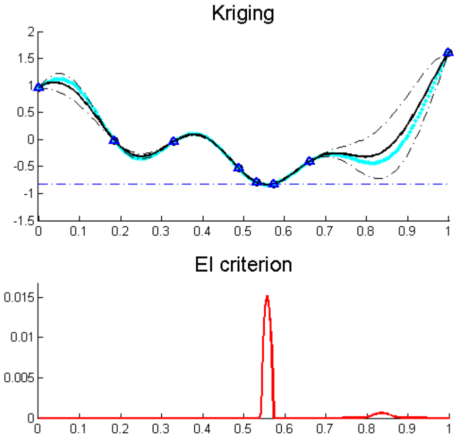
EGO: illustration 3/6



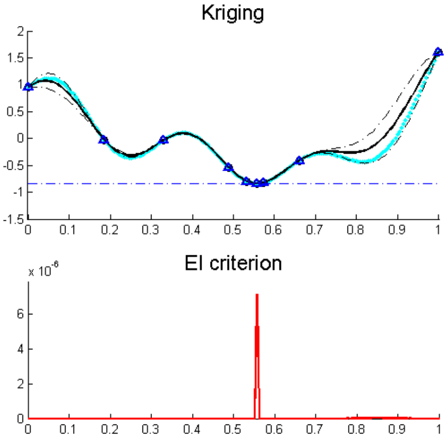
EGO: illustration 4/6



EGO: illustration 5/6



EGO: illustration 6/6



EGO en pratique

Maximisation de l'amélioration espérée

- Sous-problème d'optimisation (globale) !
- *EI* "gratuit" $\rightarrow$ méthodes intensives
- De plus : gradients et hessiens analytiques

cf. R package `DiceOptim`

O. Roustant D. Ginsbourger, Y. Deville (2010)
DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization

V. Picheny, D. Ginsbourger (2013)
Noisy kriging-based optimization methods: A unified implementation within the DiceOptim package
Computational Statistics & Data Analysis

Limites

- Dimension
- Capacité du métamodèle à représenter le système

Deux propriétés d'EGO

Maximiseur de l'amélioration espérée : stratégie optimale

Donne le meilleur gain (en espérance) sur y

EGO : stratégie myope

Optimale... pour une itération

J. Mockus (1989)
Bayesian approach to global optimization: theory and applications

Plan

Extensions du principe d'EGO à différents contextes

Principe général

- Critère traduisant notre objectif
- Recherche du maximisateur du critère
- Ajout séquentiel de points

Un domaine foisonnant

- Parallélisation
- Modèles stochastiques, optimisation avec paramètres incertains
- Optimisation sous contraintes, multi-objectif, multi-fidélité
- Alternatives à l'EI
- Autres modèles : par ex. Bayesian additive regression trees

Chipman et al. (2009)
Sequential Design for Computer Experiments with a Flexible Bayesian Additive Model

Parallélisation

Ajout de plusieurs points simultanément au lieu d'un seul

L'amélioration est apportée par le meilleur des points :

$$EI(x_1, \dots, x_p) = \mathbb{E}\left[(y_{min} - \min(Y(x_1), \dots, Y(x_p))^+ \right]$$

Difficultés

Pas de forme analytique !

Quelques références

M. Taddy, H. Lee, G. Gray, J. Griffin (2009)
Bayesian guided pattern search for robust local optimization
Technometrics, 51(4), 389-401

D. Ginsbourger, R. Le Riche, L. Carraro (2010)
Kriging is well-suited to parallelize optimization
Computational intelligence in expensive optimization problems, 131-162

C. Chevalier, D. Ginsbourger (2013)
Fast computation of the multipoint Expected Improvement with applications in batch selection

Optimisation de modèles stochastiques

Définition

- On veut optimiser : $y(x)$
- On a accès seulement à : $y(x) + \varepsilon$

Difficultés avec EGO

- y_{min} est inconnu
- La future valeur $Y(x)$ ne sera pas observée
- Il faut définir une “amélioration” adaptée

Optimisation de modèles stochastiques

Définition

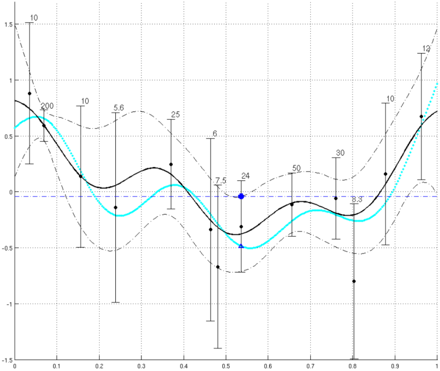
- On veut optimiser : $y(x)$
- On a accès seulement à : $y(x) + \varepsilon$

Difficultés avec EGO

- y_{min} est inconnu
- La future valeur $Y(x)$ ne sera pas observée
- Il faut définir une “amélioration” adaptée

Prise de décisions sur données bruitées

Relations d'ordre non garanties à cause du bruit → comment choisir le meilleur point ?

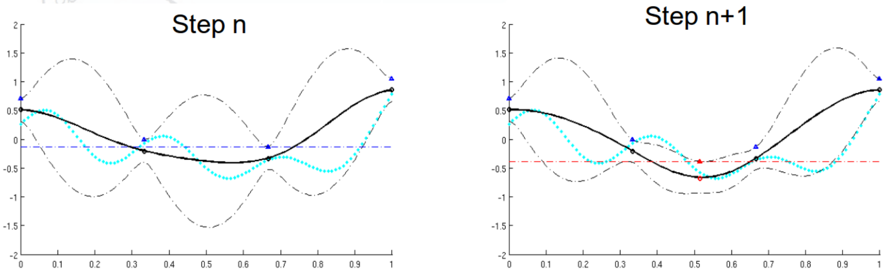


Proposition : utiliser le krigeage comme un filtre → : quantile de krigeage

Un critère séquentiel d'échantillonnage

D'une décision statique à un critère séquentiel

- Performance : plus petit quantile
- Amélioration : diminution du plus petit quantile
- $EQI = \mathbb{E}\left[q_{min} - Q(x)\right]$



V. Picheny, D. Ginsbourger, Y. Richet, G. Caplin (2013)
Quantile-based optimization of noisy computer experiments with tunable precision
Technometrics 55, 2-13

Alternatives à l'EI : critères spatialisés

Prise en compte de toute l'information fournie par une observation

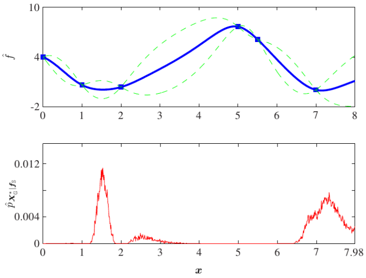
IAGO : Informal approach to Global Optimization

Réduction de l'entropie du minimum

IECI : Integrated Expected Conditional Improvement

Gain estimé sur tout l'espace

Distribution du minimum (IAGO) :



J. Villemonteix, E. Vazquez, E. Walter (2009)
An informational approach to the global optimization of expensive-to-evaluate functions
Journal of Global Optimization, 44(4), 509-534

R. Gramacy, H. Lee (2011)
Optimization under unknown constraints
Bayesian Statistics 9

Plan

Conclusion

Optimisation sur base de métamodèle

- Nécessité d'un compromis exploration / intensification
- Métamodèle : outil pertinent pour piloter les observations
- Une stratégie de référence : EGO

Applications à différents contextes

- Soit en adaptant le modèle (cf. présentation de D. Ginsbourger par ex.)
- Soit en adaptant le critère