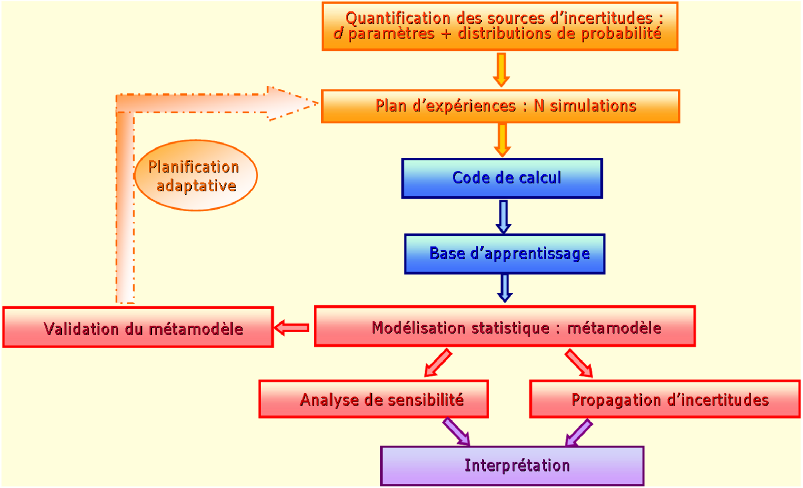
Stratégies séquentielles

Victor Picheny

Pec'num 2015

École Centrale de Lyon, 22 mai 2015

Cadre de l'exposé



Pourquoi une stratégie séquentielle ?

“Budget” ($= N$) nécessaire difficile à estimer

- Premier plan raisonnable
- Enrichissement jusqu'à satisfaction

Suite d'une première étude

- Premier plan $\Rightarrow$ métamodèle grossier $\Rightarrow$ analyse de sensibilité $\Rightarrow$ réduction de dimension
- Ajout d'expériences pour obtenir un métamodèle final précis

Planification ciblée $\Leftrightarrow$ utilisation du métamodèle

- Propagation d'incertitudes
- Analyse de sensibilité
- Optimisation / calibration

Approches géométriques vs. métamodèles

Comment ajouter des expériences à un plan existant ?

Approches “purement” géométriques

- Plan factoriels fractionnaires $\Rightarrow$ ajout d'une fraction
- Suite à faible discrépance $\Rightarrow$ éléments suivants
- Remplissage d'espace (critères *maximin*, *minimax*)

$\Rightarrow$ c.f. cours de W. Tinsson, H. Monod et L. Pronzato

Ici : plans séquentiels *orientés modèle*

Le métamodèle sert de guide pour choisir les nouvelles observations.

Objectifs

Cas 1 : métamodèle globalement précis

Utilisation générique $\Rightarrow$ le métamodèle doit remplacer fidèlement le modèle coûteux

Cas 2 : métamodèle = outil d'extraction d'une information

Intérêt guidé par la valeur des observations

- optimisation : recherche de minimum / maximum
- analyse de risque : dépassement de seuil

Une bonne précision partout n'est pas nécessaire !

Plan de l'exposé

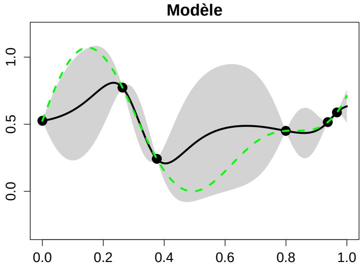
Point de départ de l'étude

On dispose de :

- Un plan initial (remplissage d'espace)
- Un premier métamodèle (krigeage)

Objectif

Ajouter des expériences pour augmenter la qualité du métamodèle



Peut-on s'aider du métamodèle ?

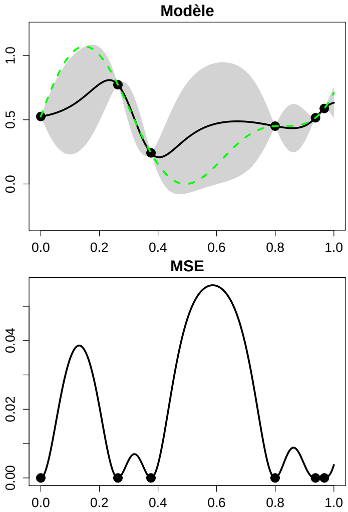
Hypothèse principale

Les paramètres de covariance sont connus avec précision.

Mesure de la qualité du modèle

Variance de prédiction (MSE)

Le métamodèle indique où l'erreur de prédiction est potentiellement la plus grande !



Enrichissement séquentiel guidé par la variance du modèle

On ajoute des observations une par une en suivant le schéma :

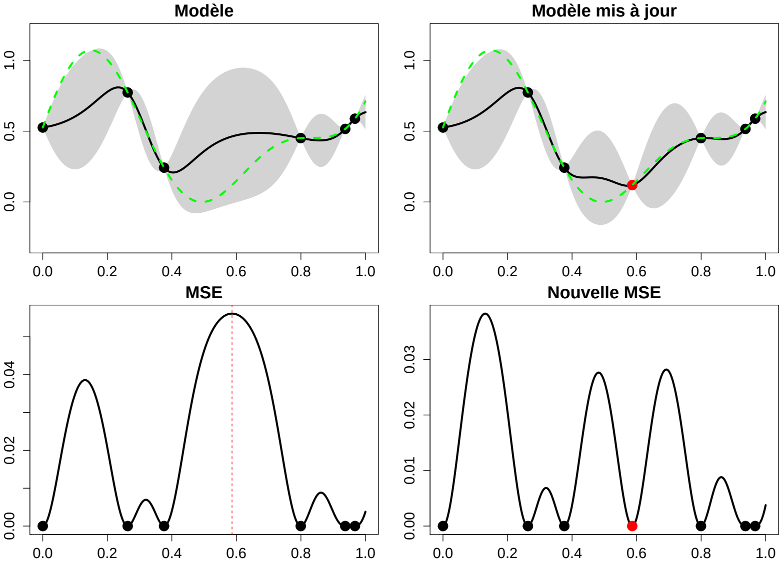
Algorithme (boucle)

1. On cherche le point où la variance est maximum :
   $\mathbf{x}* = \arg\max_D MSE(\mathbf{x})$
2. On ajoute une nouvelle observation en ce point : $\mathbf{x}_i = \mathbf{x}*$
3. On met à jour le métamodèle (plan d'expériences et observation, + éventuellement paramètres de covariance)

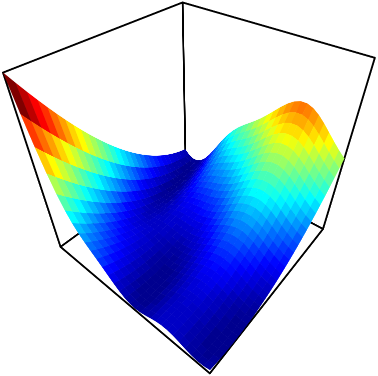
Attention !

La recherche de $\mathbf{x}*$ nécessite un algorithme d'optimisation.

Illustration



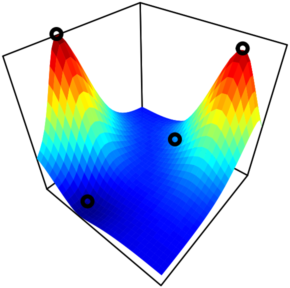
Exemple 2D : “vraie” fonction



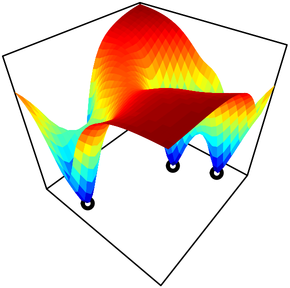
Exemple 2D

Plan de départ : 4 points

Moyenne

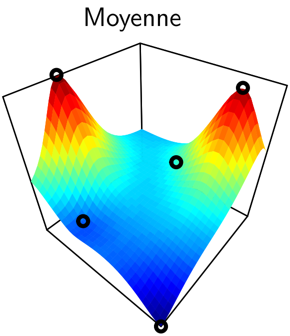


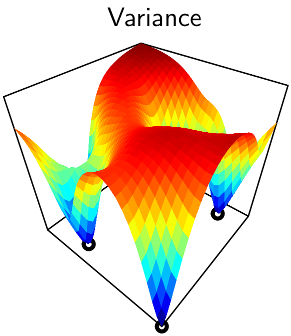
Variance



Exemple 2D

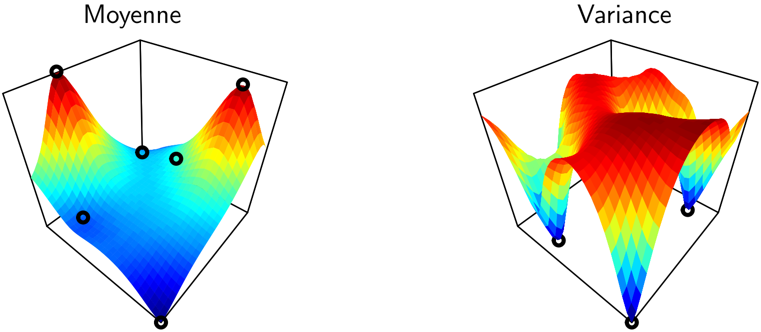
5 points





Exemple 2D

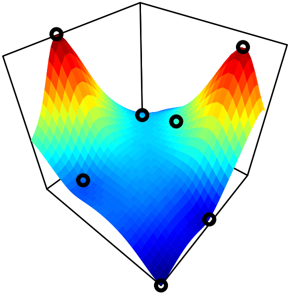
6 points



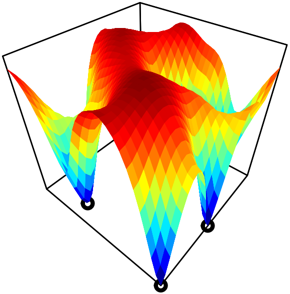
Exemple 2D

7 points

Moyenne



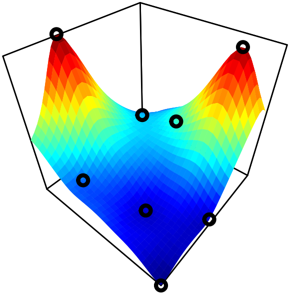
Variance



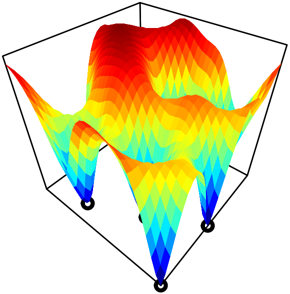
Exemple 2D

8 points

Moyenne



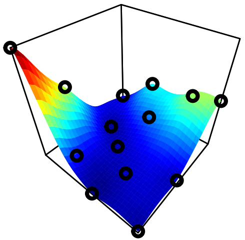
Variance



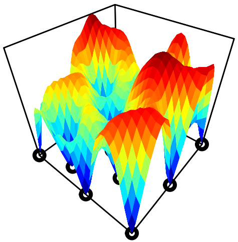
Exemple 2D

14 points

Moyenne

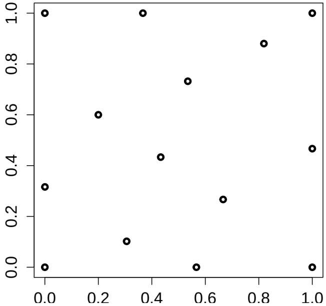


Variance



Exemple 2D : plan final à 14 points

- Bon remplissage d'espace
- Tendance à échantillonner sur les bords



Pour / contre

\+ Facile à mettre en place

– Vraiment efficace ?

Maximum vs. moyenne de la MSE

Critère IMSE *cf. cours de L. Pronzato*

$$IMSE = \int_D MSE(x)dx$$

Erreur **moyenne** du modèle

Sacks, Welch, Mitchell & Wynn
Design and analysis of computer experiments
Statistical science, 409-423 (1989)

Utilisation dans une démarche séquentielle

- MSE : une valeur par point $\rightarrow$ recherche du maximum
- IMSE : valeur unique

On cherche le point qui diminue le plus le critère **si on l'ajoute au métamodèle**

Comment évaluer le “futur” critère sans faire l’expérience ?

- Retour sur la formule : $MSE(x) = \sigma^2 - k(x)^T \Sigma^{-1} k(x)$
- Ne dépend pas des valeurs observées Z_S
- On peut ajouter x_{new} et choisir n’importe quelle valeur pour $Z(x_{new})$, l’IMSE reste la même.

Principe

Pour un point candidat x_{new}:

1. On ajoute x_{new} au plan d’expériences et une valeur quelconque aux observations
2. On met à jour le métamodèle sans changer la covariance*
3. On évalue le critère sur le nouveau modèle : $IMSE(x_{new})$

On cherche $\mathbf{x}* = \arg\min_D IMSE(\mathbf{x})$.

*ou, beaucoup plus efficace, on utilise des formules de mise à jour

Illustration : maxMSE vs. minIMSE

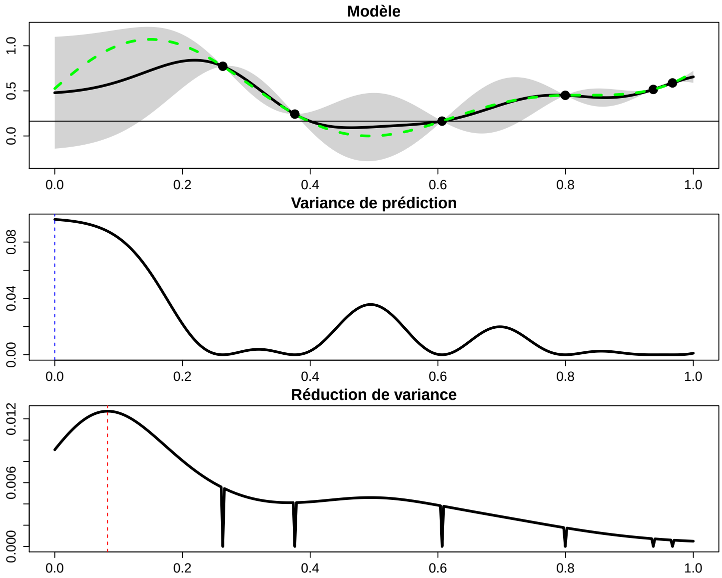
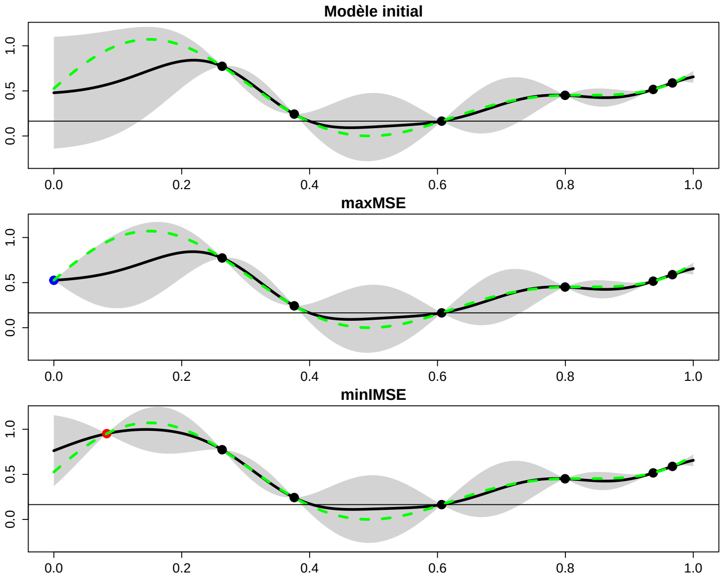


Illustration : modèles mis à jour



Enrichissement séquentiel IMSE-optimal

Algorithme

- Même boucle que pour maxMSE
- Recherche du meilleur point $\Rightarrow$ boucle d'optimisation emboîtée

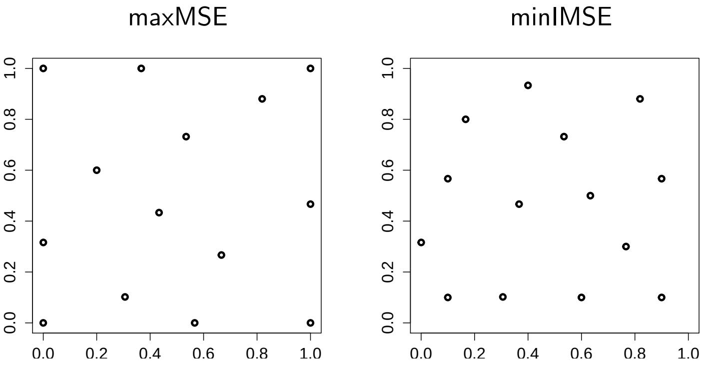
Critère sous forme intégrale

Pas de formule analytique $\rightarrow$ intégration numérique

Parallélisation

On peut aussi chercher un ensemble de points IMSE-optimaux.

Retour sur l'exemple 2D



Quelle stratégie choisir ?

minIMSE

+ Correspond au but poursuivi
+ Pas d'effet de bord
– Complexe : intégration numérique + mise à jour du modèle
– Coûteux à évaluer

maxMSE

Exactement l'inverse !

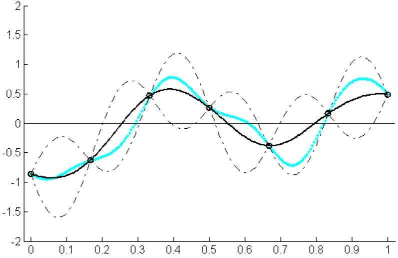
$\Rightarrow$ Le choix dépend essentiellement du budget de calcul et de la dimension.

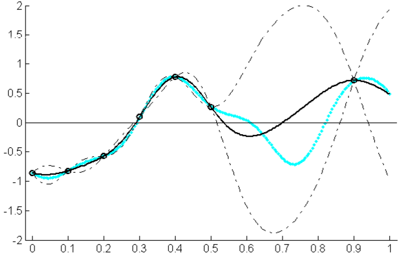
Plan de l'exposé

Motivation

Apprentissage global ou "ciblé"

- Le plan d'expérience a une influence très forte sur la précision locale
- Une bonne précision n'est pas nécessaire partout !





Dans cette section

Un *exemple* de plan séquentiel adapté à l'objectif

Propagation d'incertitudes et analyse de risque (1/2)

Un problème classique : dépassement de seuil

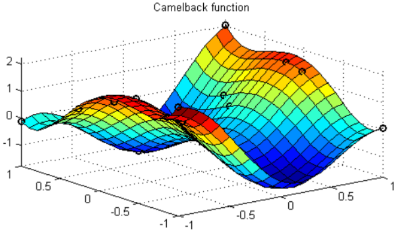
- La sortie du simulateur doit rester sous une valeur critique (contraintes mécaniques, température, etc.)
- On connaît la loi des paramètres d'entrée
- On veut estimer : $\mathbb{P}\left[y(X)\right] \geq \textit{seuil}$

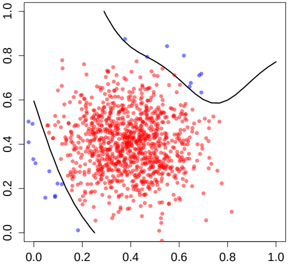
$$X \longrightarrow \boxed{\text{Simulateur}} \longrightarrow \text{y}$$

Propagation d'incertitudes et analyse de risque (2/2)

- Approche par Monte-Carlo : on génère $X_1, \ldots, X_N \Rightarrow$ $y(X_1), \ldots, y(X_N)$ et on compte

$$\frac{1}{N} \sum_{i=1}^{N} \mathbb{1}(y(X_i)) > T$$





- y coûteux $\Rightarrow$ métamodèle

Problème considéré et métamodèle

Recherche de lignes de niveau

- On veut savoir si $y(\mathbf{x}) \geq T$
- $y(\mathbf{x}) \ll T$ or $y(\mathbf{x}) \gg T$: un métamodèle imprécis suffit
- Région critique pour l'apprentissage : $X_T = \{\mathbf{x} / y(\mathbf{x}) \approx T\}$

On va enrichir le modèle pour qu'il soit précis dans la région critique.

Un exemple de critère pour l'apprentissage ciblé

Exploitation de l'information du modèle

- Région critique : $X_T = \{\mathbf{x} / \, |y(\mathbf{x}) - T| \leq \varepsilon\}$
- On peut calculer la probabilité $P\left(\mathbf{x} \in X_T\right)$:

$$P\left(\mathbf{x} \in X_T\right) = \Phi\left(\frac{T + \varepsilon - m(\mathbf{x})}{s(\mathbf{x})}\right) - \Phi\left(\frac{T - \varepsilon - m(\mathbf{x})}{s(\mathbf{x})}\right)$$

Critère IMSE "ciblé"

Variance de prédiction pondérée par la prédiction d'appartenir à la région cible :

$$IMSE_T = \int_D MSE(\mathbf{x}) P\left(\mathbf{x} \in X_T\right) d\mathbf{x}$$

Illustration: modification du critère IMSE

$$\mathbf{x}_{n+1} = \arg\min IMSE_T(\mathbf{x}_{new})$$

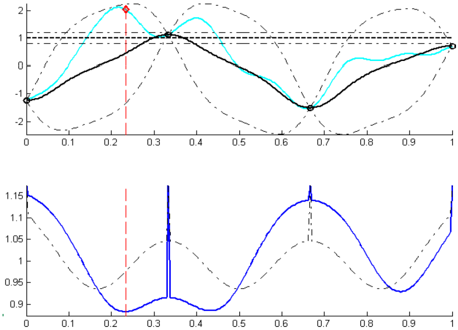
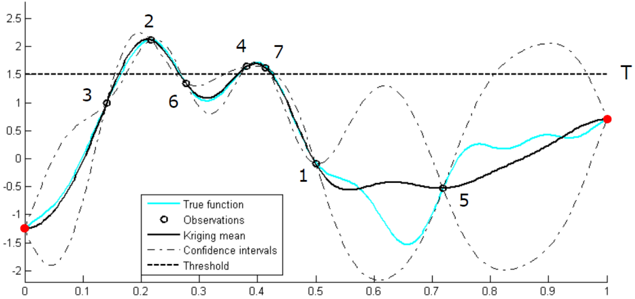


Illustration: 2 points initiaux + 7 itérations

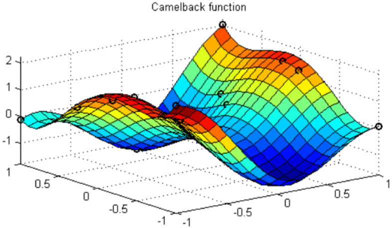


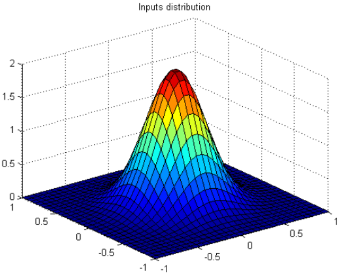
C. Chevalier, V. Picheny, D. Ginsbourger
KrigInv: An efficient and user-friendly implementation of batch-sequential inversion strategies based on kriging
Computational Statistics and Data Analysis, 71, 1021-1034 (2014)

Exploitation pour un calcul de probabilité de défaillance

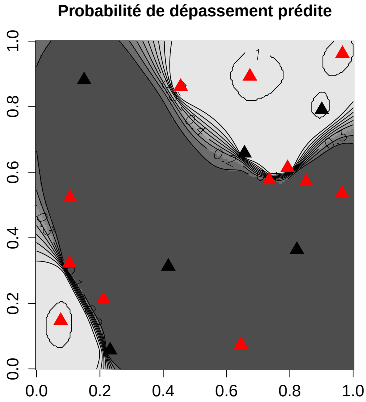
Problème jouet

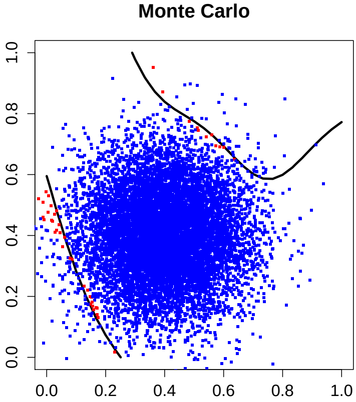
- Réponse dépendant de deux paramètres
- Entrées gaussiennes
- Seuil $= 0$





Exploitation pour un calcul de probabilité de défaillance





Probabilité exacte : 1,75%
Probabilité estimée : 1,69%

Plan de l'exposé

Contexte

Expériences numériques comme aide à la conception / décision

- Réponse du code de calcul = performance ou coût
- Recherche des paramètres optimaux :

$$x^* = \arg\min cout(x) \text{ ou } \arg\max perf(x)$$

- L'optimisation nécessite beaucoup d'appels au code
- Métamodèle : solution naturelle

Lien avec la problématique précédente

Le métamodèle doit être précis seulement dans les régions importantes (proche de l'extremum) $\Rightarrow$ répartition des expériences "ciblée"

Démarche nécessairement séquentielle

Il faut faire des expériences pour connaître les régions cibles !

Planification d'expériences et optimisation globale

Optimisation locale

Amélioration depuis un point initial

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

Dans un contexte de planification d'expériences

- Exploration : remplissage d'espace
- Intensification : "ciblage"

Introduction à l'optimisation globale : l'algorithme DIRECT

Garanti sans métamodèle !

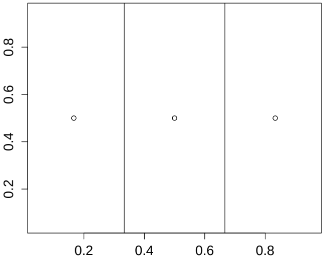
DIRECT : DIviding RECTangles

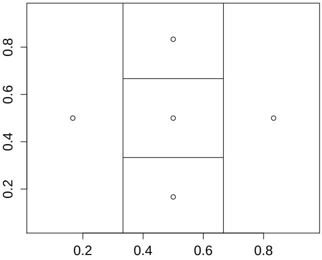
- Découpage de l'espace en (hyper)rectangles
- Un échantillon au centre de chaque rectangle
- On divise les rectangles les plus "intéressants" :
  - soit les plus grands (exploration)
  - soit ceux qui ont une valeur au centre basse (intensification)
- Pour diviser : ajout de 2 points, division en 3

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

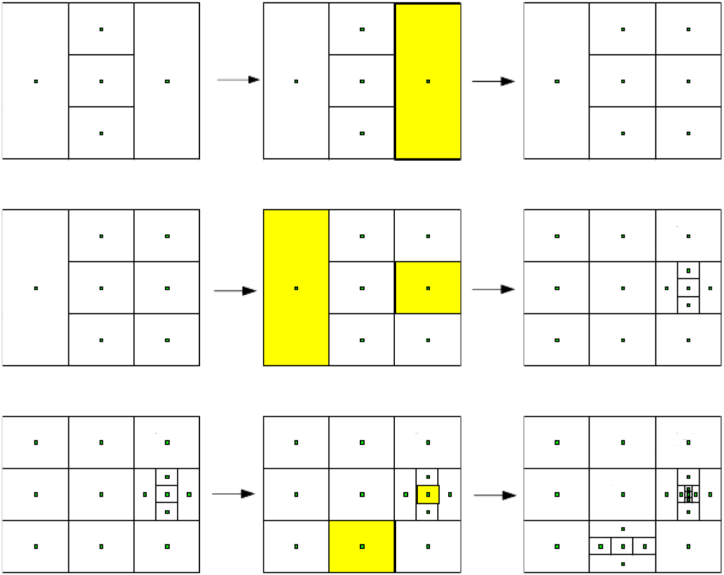
Exemple en dimension 2

- Départ : 3 points équirépartis dans une direction aléatoire
- On divise le rectangle ayant la meilleure observations



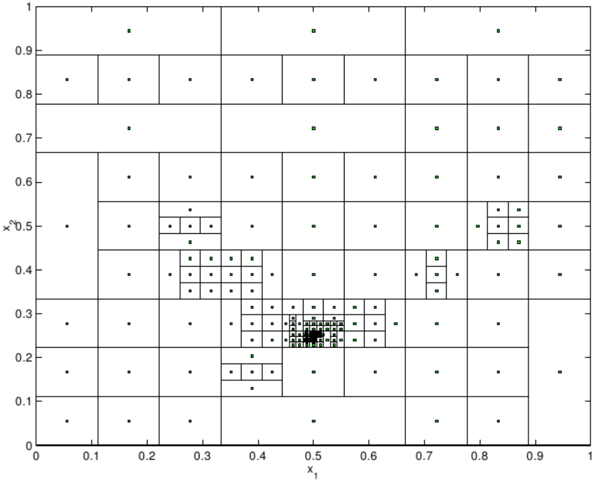


Exemple en dimension 2



Après 191 évaluations

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



Source figures :

D. E. Finkel
DIRECT Optimization
Algorithm User Guide (2003)

Intêret et limites

- $+$ Exploration de tout l'espace de recherche
- $+$ Stratégie robuste
- $-$ Limité aux petites dimensions
- $-$ Exploitation limitée de l'information

$\Rightarrow$ même principe général, avec un métamodèle ?

Optimisation et métamodèle : ce qu'on est tenté de faire...

"Le métamodèle donne l'optimum"

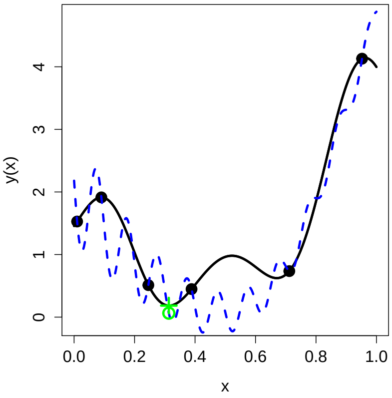
- On cherche le minimum x^* sur le métamodèle
- On évalue le vrai $y(x^*)$ sur le simulateur

$\Rightarrow$ C'est fini !

Répartition de l'effort

- Plan initial : 49 expériences
- 98% exploration, 2% exploitation

Que faire si x^* n'est pas bon ?



Optimisation et métamodèle : ce qu'il faut faire

Si le budget est fixe

- On divise le budget en 2
- Budget 1 : plan initial (LHS)
- Budget 2 : optimisation

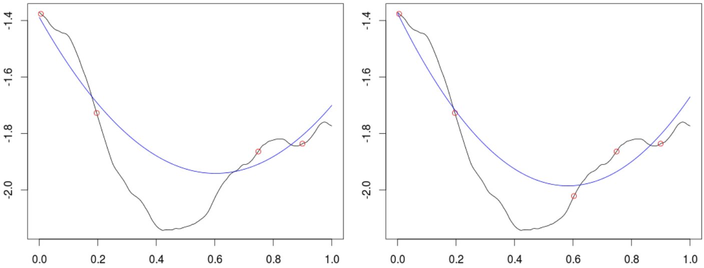
Utilisation **séquentielle** du métamodèle

- Métamodèle initial : *a priori* peu précis
- Le métamodèle sert à **choisir** pour les nouvelles observations
- A chaque nouvelle observation : amélioration du métamodèle

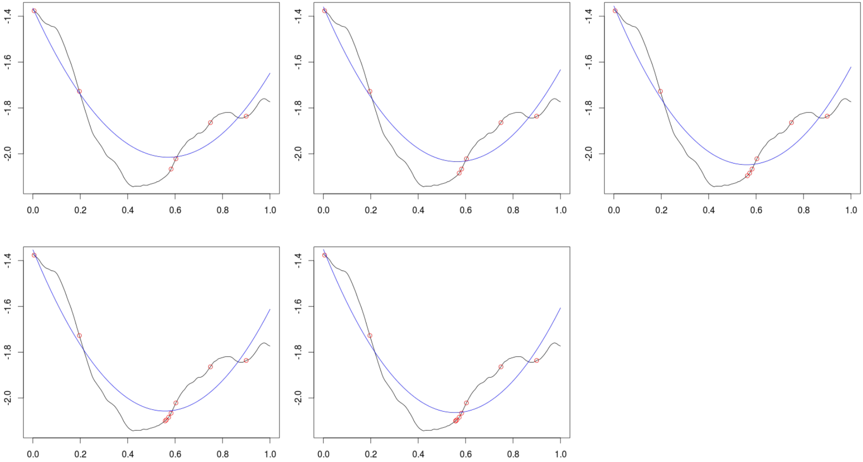
Optimisation basée sur les modèles polynomiaux

Principe

- On construit une surface de réponse $y = \beta_0 + \beta_1 x + \beta_2 x^2$
- 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 basée sur les modèles polynomiaux

Problème : modèle “rigide”

Le modèle ne s’ajuste pas aux données : $Y = \mathbf{X}\beta + \epsilon$
Pas de convergence vers un modèle précis, même localement

Solutions

1. Augmenter le dégré du polynôme
   $\Rightarrow$ risque de surapprentissage & d’instabilité !
2. Supprimer des points
   $\Rightarrow$ méthode **de région de confiance**

Régions de confiance : principe

Modèle quadratique "creux"

- Valide à l'intérieur d'une région de confiance (petite)
- Construit uniquement avec les points à l'intérieur de la région
- Selon les valeurs des simulations, on modifie la taille de la région

Gestion de la région de confiance

A chaque itération :

- $\hat{y}(x^*)$ bon $\Rightarrow$ confiance dans le modèle : on augmente la taille
- $\hat{y}(x^*)$ mauvais $\Rightarrow$ modèle peu fiable : on diminue la taille

$+$ beaucoup de règles pour sélectionner les points et enrichir le plan d'expériences

Illustration (source : F. Vanden Berghen)

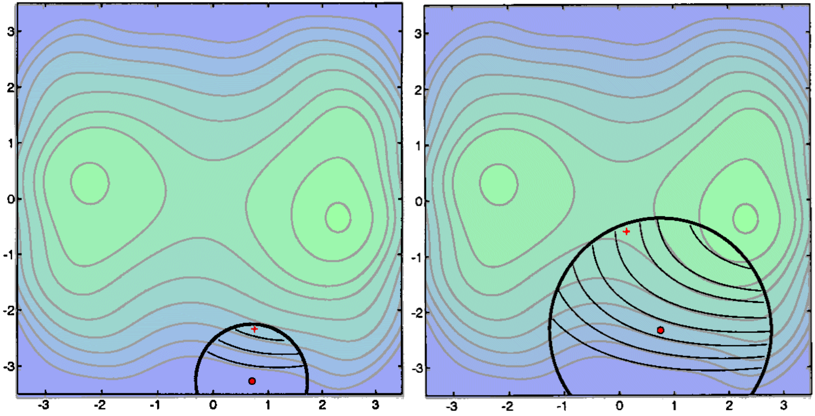


Illustration (source : F. Vanden Berghen)

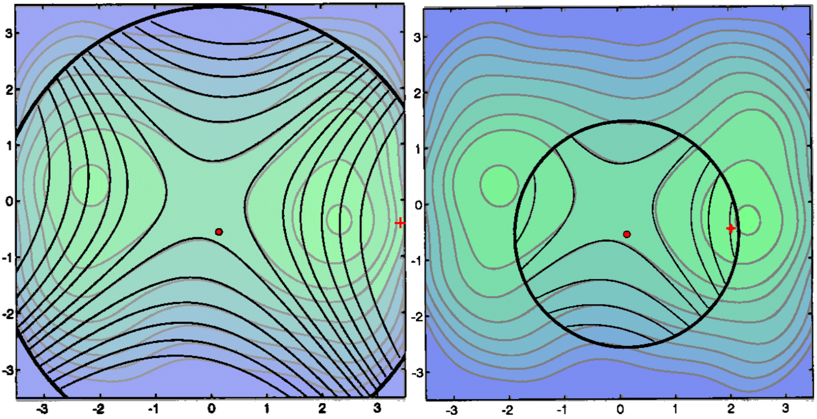
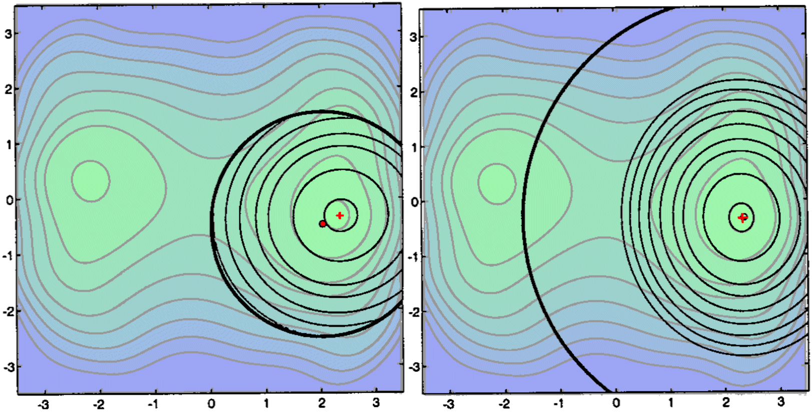


Illustration (source : F. Vanden Berghen)



Avantages et inconvénients

Avantages

- Garantie de convergence
- Méthodes assez parcimonieuses
- Robuste
- Accepte un assez grand nombre de variables

Conn, Scheinberg, and Vicente
Introduction to derivative-free optimization
MPS-SIAM Series on Optimization (2009)

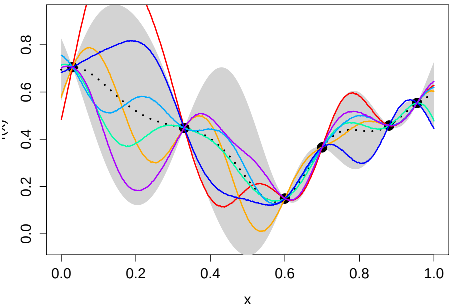
Méthode locale

- Pas de métamodèle final utilisable globalement
- Proche des méthodes de gradient

Optimisation à l'aide du krigeage

Atouts du modèle

- Global
- Flexible
- Information riche :
  - Réalisations conditionnelles
  - Moyenne : $m(x)$
  - Variance $s^2(x)$



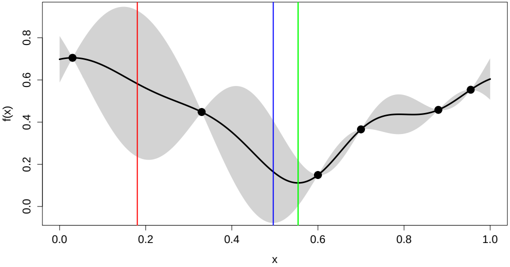
Critère d'échantillonnage

On utilise le métamodèle pour estimer "l'intérêt" d'une observation potentielle
$\Rightarrow$ On effectue la simulation la plus intéressante !

Différents critères

Cf. Jones (2001)

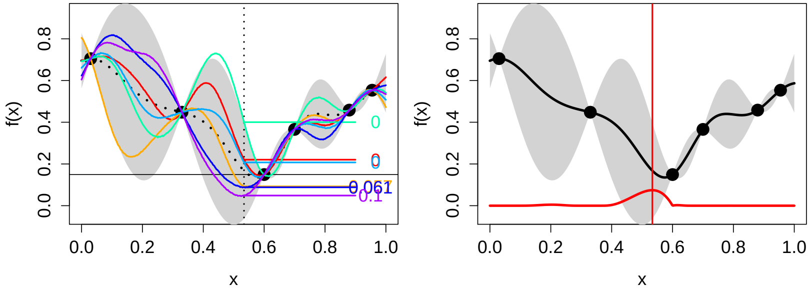
- Pure intensification : $\min m(x)$
- Pure exploration : $\max s^2(x)$
- Un compromis simple : $\min m(x) - \alpha s(x)$



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

(ou *EI: Expected Improvement*)

- n observations : meilleur choix $y_{\min} = \min(y_1, \ldots, y_n)$
- Amélioration à $n+1$ = gain sur la fonction coût : $(y_{\min} - Y(x))^+$
- Au temps n : l'amélioration est une v.a., *mais* on connaît sa loi
- **Espérance de l'amélioration :** $EI(x) = \mathbb{E}\left[(y_{\min} - Y(x))^+\right]$



L'amélioration espérée

Forme analytique simple

$$EI(x) = s(x)\left(m(x)\Phi(\xi(x)) + \phi(\xi(x))\right)$$

avec $\xi(x) = (y_{\min} - m(x))/s(x)$.

Quelques propriétés

- $EI(x) = 0$ aux points déjà visités, positif partout ailleurs
- Fonction multimodale
- Correspond à la minimisation du regret espéré
- Stratégie optimale... à un pas

L'algorithme *Efficient Global Optimization*

Initialisation

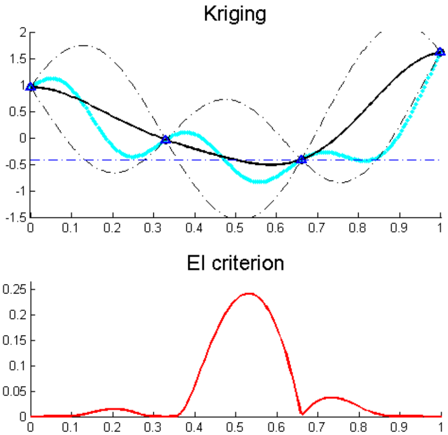
- Réalisation d'un plan initial
- Construction d'un krigeage

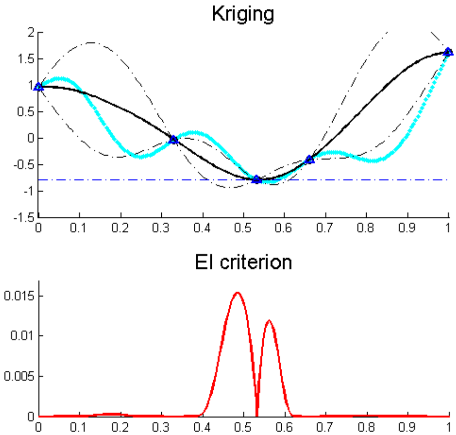
Boucle d'optimisation

- 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

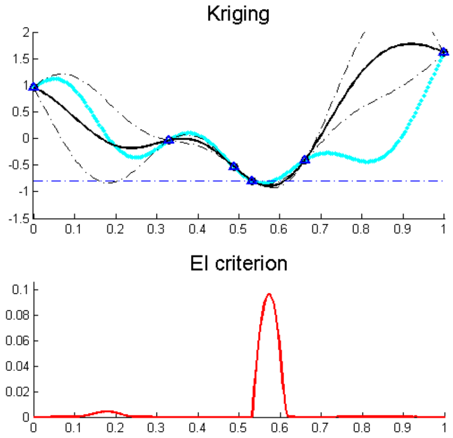
D. Jones, M. Schonlau, W. Welch
Efficient global optimization of expensive black-box functions
Journal of Global optimization 13 (4), 455-492 (1998)

EGO: illustration 1/6

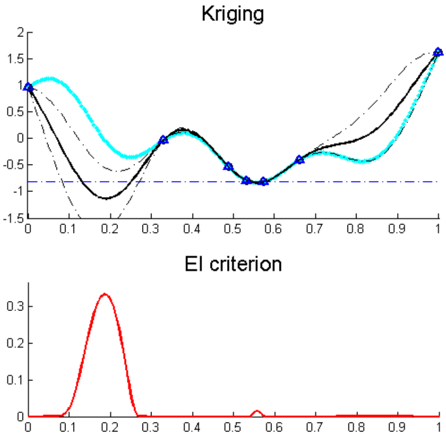



Kriging
2
1.5
1
0.5
0
-0.5
-1
-1.5
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
EI criterion
0.015
0.01
0.005
0
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

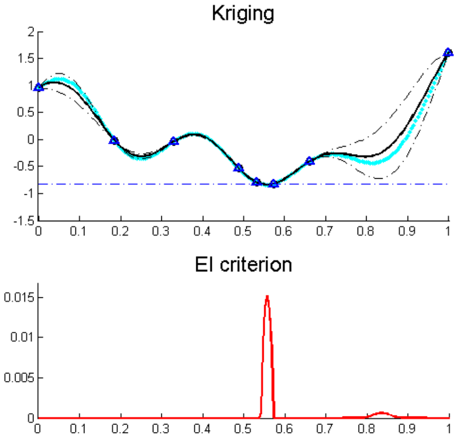
EGO: illustration 3/6



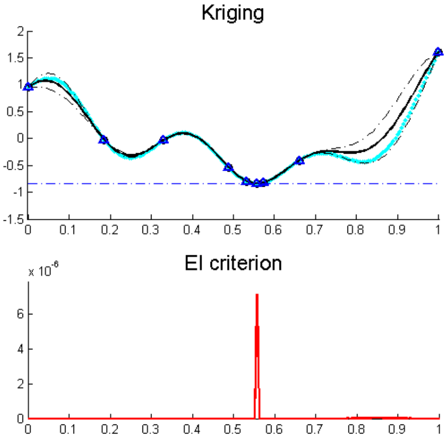
EGO: illustration 4/6



EGO: illustration 5/6

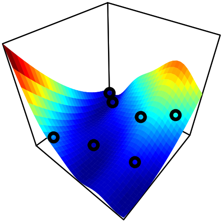


EGO: illustration 6/6

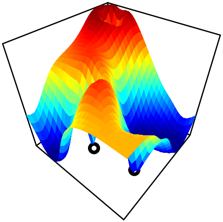


Retour sur l'exemple 2D

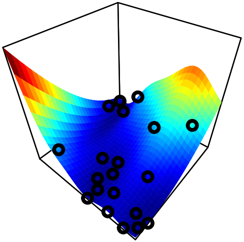
Vraie fonction + 7 expériences



EI



Après 12 itérations



Remarques

Compromis exploration / intensification

Géré automatiquement par le critère

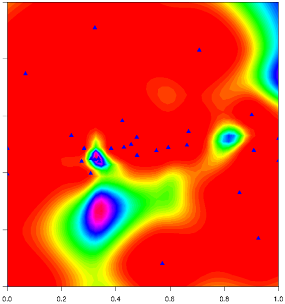
En pratique

- Très performant en dimension petite ($\leq$50)
- Stratégie coûteuse (boucle d'optimisation emboîtée) : réservée aux modèles coûteux

EGO en pratique : maximisation du critère

Problème critique, valable aussi pour les sections précédentes !

- Sous-problème d'optimisation (globale) difficile !
- *EI* "gratuit" ($\approx 1/100s$) $\rightarrow$ méthodes intensives
- De plus : gradients et hessiens analytiques
- Algorithmes : stratégies évolutionnaires + Newton



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
Journal of Computational Software

Extensions du principe d'EGO à différents contextes

Principe général

- Critère *(statistique)* 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 :

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

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

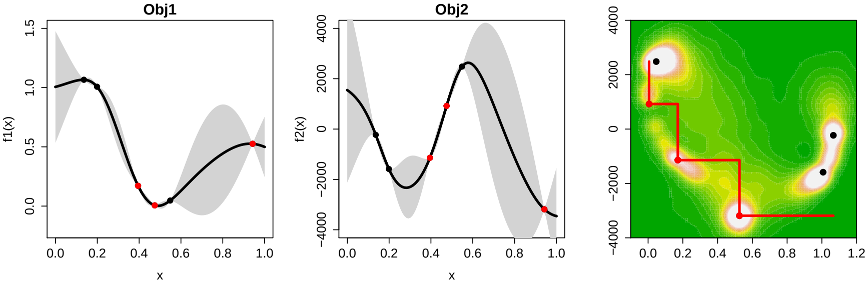
D. Ginsbourger, R. Le Riche, L. Carraro (2010)
Kriging is well-suited to parallelize optimization
Computational intelligence in expensive optimization problems, 131-162

Multi-objectifs : un métamodèle par objectif

"Amélioration" sur le front de Pareto ?

J. Svenson (2011)
Computer experiments: multiobjective optimization and sensitivity analysis, PhD thesis, Ohio State University

V. Picheny (2014)
Multiobjective optimization using Gaussian process emulators via stepwise uncertainty reduction, Statistics and Computing



R package GPareto

Optimisation avec contraintes : 3 cas (1/2)

Objectif coûteux, contrainte rapide

$$\begin{array}{ll} \max & \text{Rendement} \\ s.c. & \text{ITK réalisable} \end{array}$$

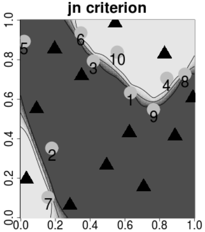
- Métamodèle pour l'objectif
- EGO classique avec contrainte sur la maximisation de l'EI

Chevalier, Picheny, Ginsbourger (2014)
KrigInv: An efficient and user-friendly implementation of batch-sequential inversion strategies based on kriging,
CSDA, vol.71, pp.1021-1034

Objectif rapide, contrainte coûteuse

$$\begin{array}{ll} \min & \text{Poids} \\ s.c. & \text{Contrainte méca.} \leq \text{Seuil} \end{array}$$

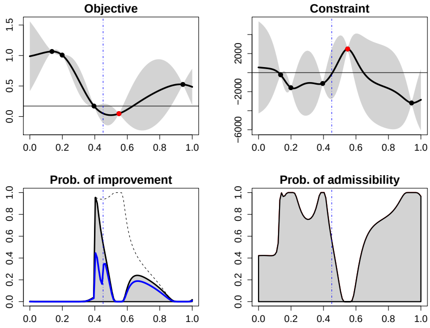
- Métamodèle pour la contrainte
- Apprentissage de la frontière



Optimisation avec contraintes : 3 cas (2/2)

Objectif & contraintes coûteux *(ou : contrainte de crash)*

- Un métamodèle pour chaque fonction
- "Amélioration faisable" ?

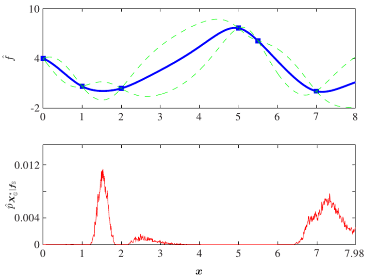


Alternatives à l'EI : critères spatialisés

Distribution du minimum (IAGO) :

IAGO : Informal Approach to Global Optimization

Réduction de l'entropie du minimum



IECI : Integrated Expected Conditional Improvement

Gain estimé sur tout l'espace

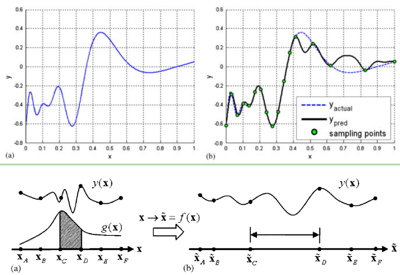
EEV : Expected Volume Reduction

Réduction du volume sous le minimum

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

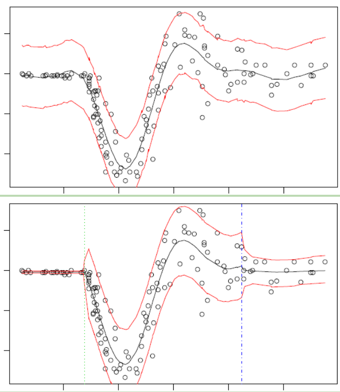
Modèles complexes (1/2) : krigeages non stationnaires

Déformation de l'espace : *scaling*



Xiong, Chen, Apley, Ding (2007)
A non-stationary covariance-based Kriging method for metamodelling in engineering design
Int. J. for Num. Meth. in Eng.

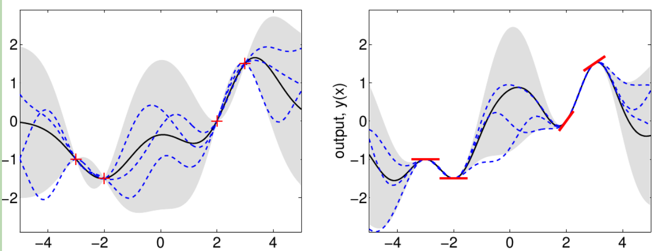
Découpage de l'espace : treed GP



⇒ EGO directement applicable

Modèles complexes (2/2) : ajout d'informations

Les processus gaussiens peuvent inclure beaucoup d'information !



source : Gaussian Process for Machine Learning (2006)

Mais également conditions au bord, symétries...

Presque tout est envisageable si on trouve la distance / covariance adaptée

Pour finir : de nombreux sujets non abordés...

...par manque de temps !

- Optimisation robuste (ou fiabiliste)
- Optimisation bruitée
- Multi-fidélité
- Grande dimension
- ...

Conclusion

Pourquoi utiliser une stratégie séquentielle ?

- Flexibilité / contraintes imposées par l'étude
- Ajout possible d'objectifs guidés par les observations effectuées

Stratégie séquentielle et métamodèle

- On exploite l'information donnée par le métamodèle pour choisir les observations
- L'information utile s'adapte à l'objectif poursuivi !

Augmentation de la complexité !

- Boucles d'optimisation imbriquées
- Mise à jour de modèles, etc.

Généralisation de l'approche séquentielle

Un schéma unique

- Définition d'un critère (MSE, IMSE, $IMSE_T$, EI)
- Recherche du point optimal au sens de ce critère
- Enrichissement du métamodèle

Stratégie séquentielle "sur mesure"

- Il suffit de définir un critère correspondant au besoin !
- Beaucoup de travaux existants...
- ... un plus grand nombre encore de besoins spécifiques