Optimisation à l'aide de métamodèles

Victor Picheny

INRA - Mathématiques et Informatique Appliquées - Toulouse

La Rochelle, 5 novembre 2014

Plan de l'exposé

1. Introduction
   - Introduction à l'optimisation globale : l'algorithme DIRECT
   - Optimisation globale et métamodèles
2. Optimisation basée sur les modèles polynomiaux
3. Optimisation sur base de krigeage : l'algorithme EGO
4. Extensions du principe d'EGO à différents contextes
5. Conclusion

Plan de l'exposé

1. Introduction
   - Introduction à l'optimisation globale : l'algorithme DIRECT
   - Optimisation globale et métamodèles
2. Optimisation basée sur les modèles polynomiaux
3. Optimisation sur base de krigeage : l'algorithme EGO
4. Extensions du principe d'EGO à différents contextes
5. Conclusion

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

L'algorithme DIRECT

DIRECT : DIviding RECTangles

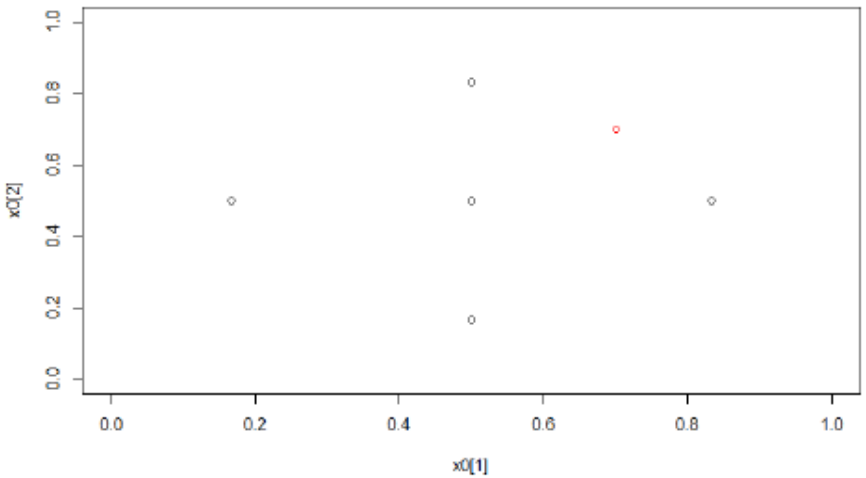
D. Jones, C. Perttunen, B. Stuckman (1993)
Lipschitzian optimization without the Lipschitz constant
Journal of Optimization Theory and Applications 79(1), 157-181

- 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

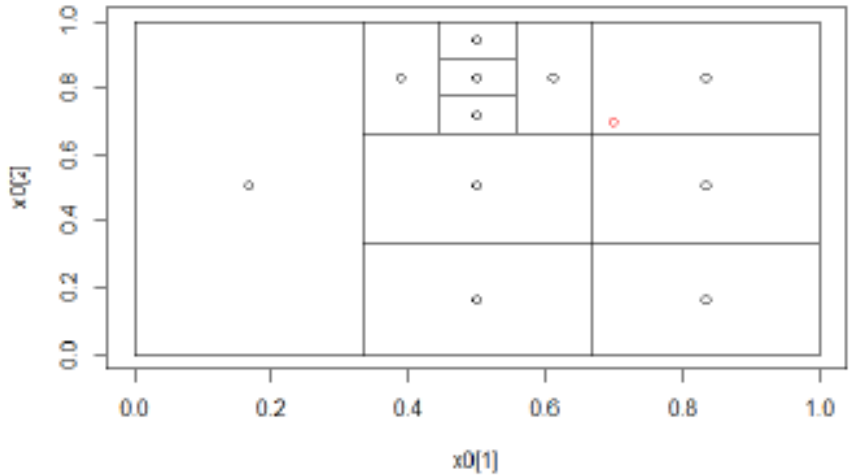
Quel compromis exploration / intensification ?

Pour (presque) **chaque** taille de rectangle, on divise celui qui a la meilleure valeur observée

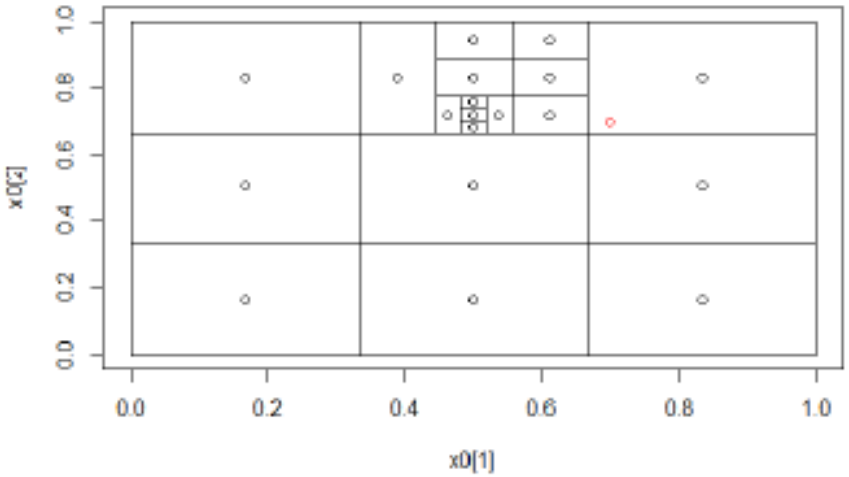
Exemple en dim 2



Après 2 itérations

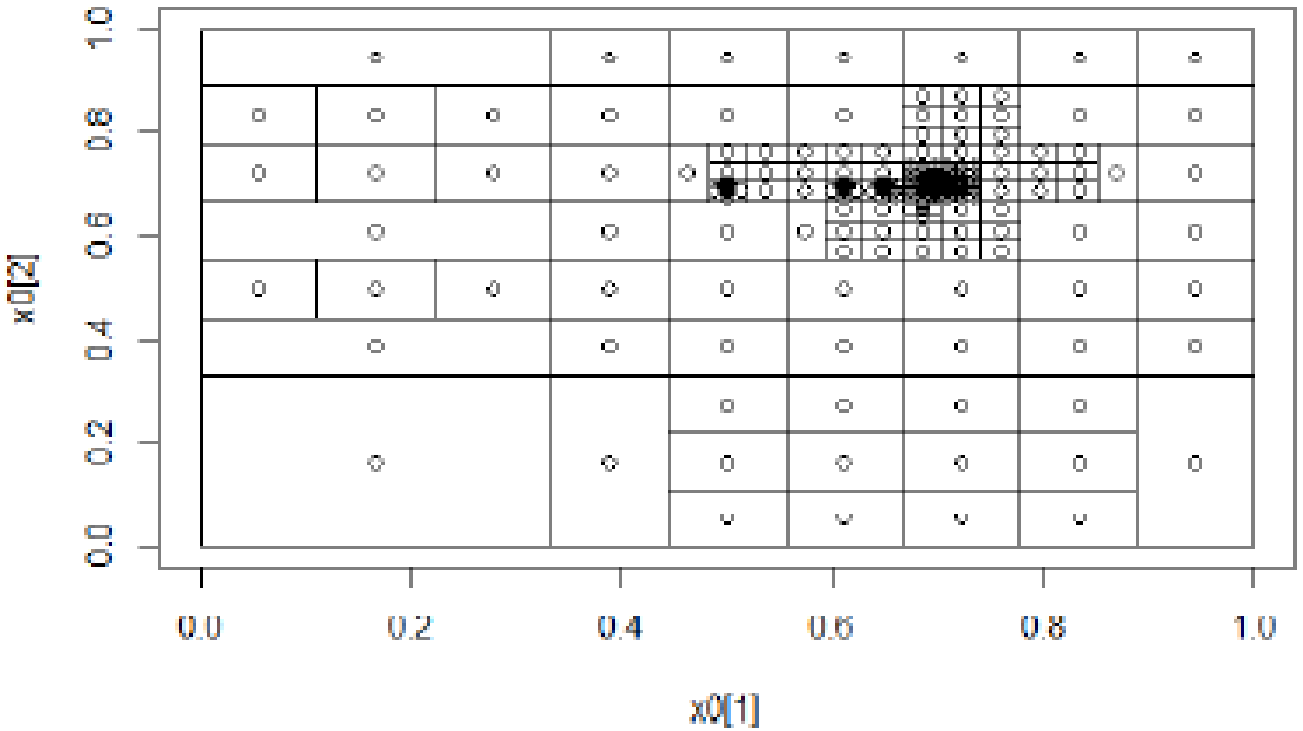


Après 3 itérations



Après 20 itérations

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

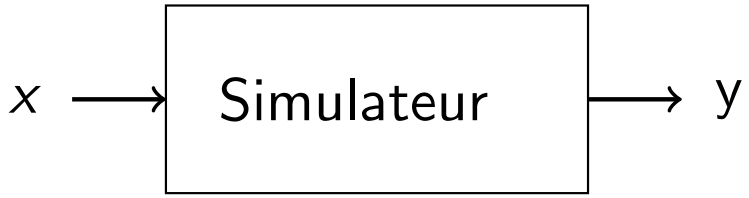


Optimisation sur base de métamodèles

Notations

Objectif : trouver

$$x^* = \arg\min_{x \in D} y(x)$$



Pourquoi faire ?

Simulation coûteuse $\rightarrow$ exploitation maximale de l'information $\rightarrow$ métamodèle

Principe général

- Echantillon initial (plan d'expériences)
- Construction d'un ou plusieurs métamodèles
- Validation $\Rightarrow$ capacité à prédire
- Et après ?

Ce qu'on est tenté de faire...

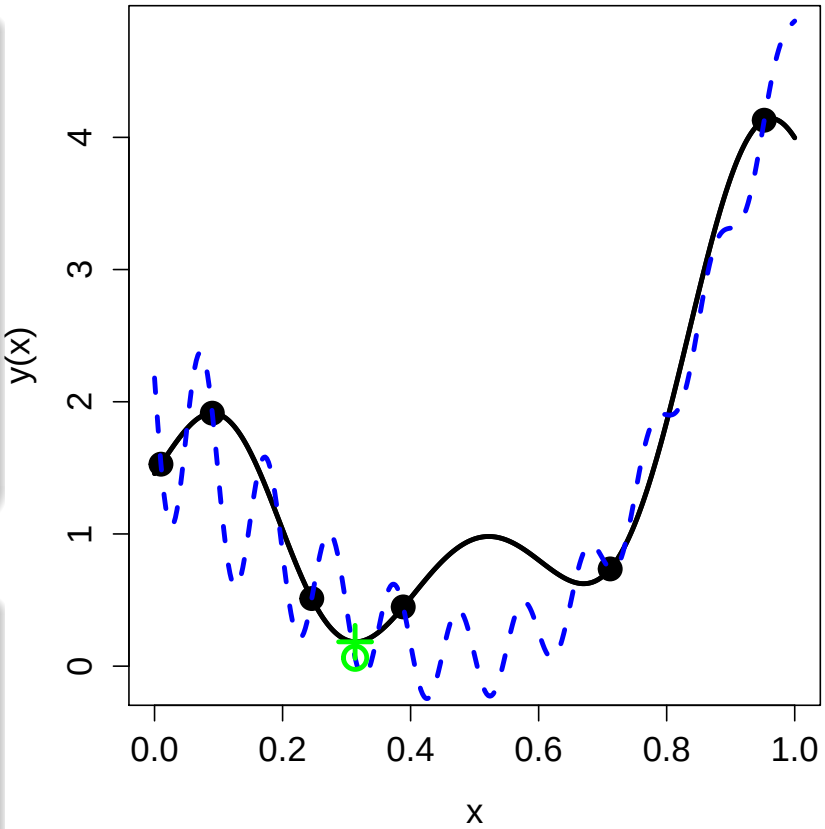
"Le métamodèle donne l'optimum"

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

$\Rightarrow$ C'est fini !

Compromis exploration / intensification ?

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



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

Ce qu'il faut faire

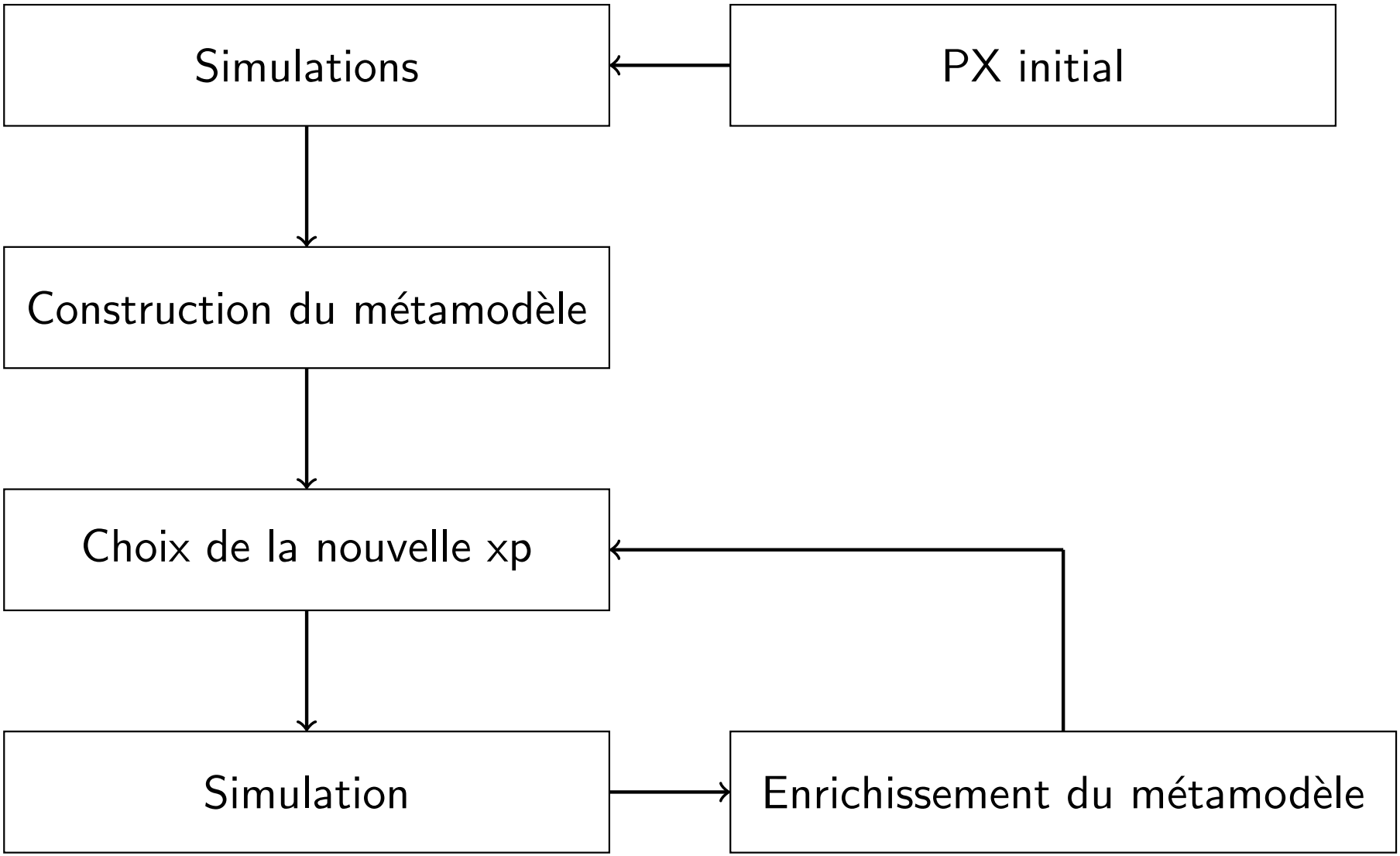
Si le budget est fixe

- 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

Schéma général : métamodèle = guide



Plan de l'exposé

1 Introduction
- Introduction à l'optimisation globale : l'algorithme DIRECT
- Optimisation globale et métamodèles

2 Optimisation basée sur les modèles polynomiaux

3 Optimisation sur base de krigeage : l'algorithme EGO

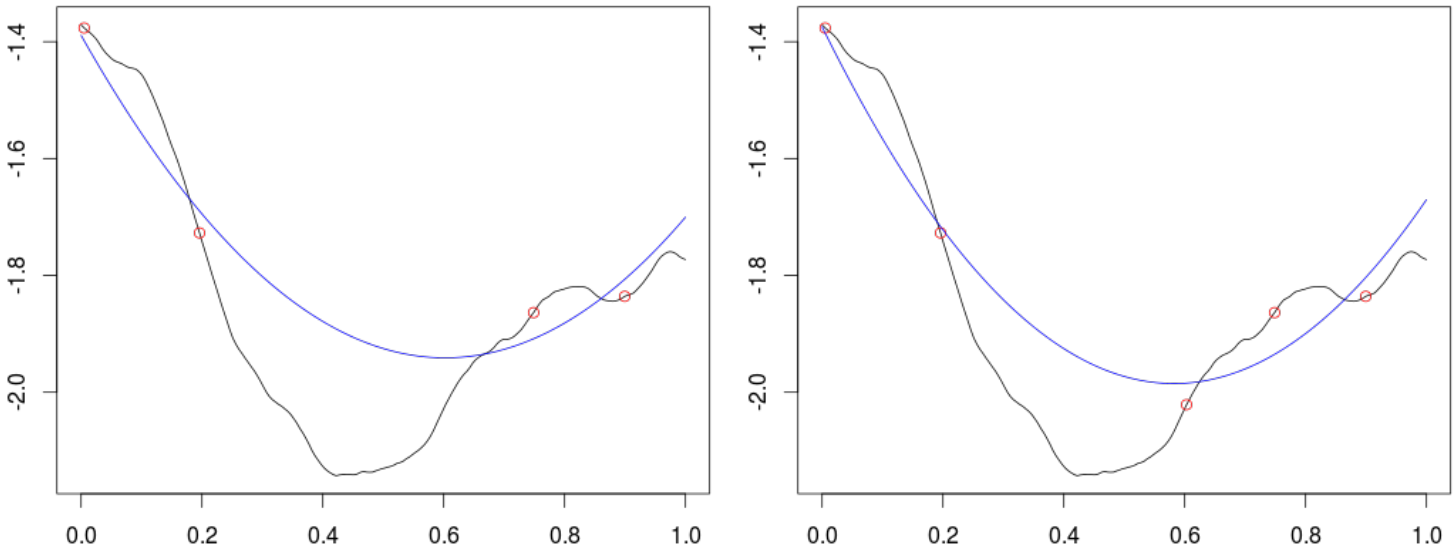
4 Extensions du principe d'EGO à différents contextes

5 Conclusion

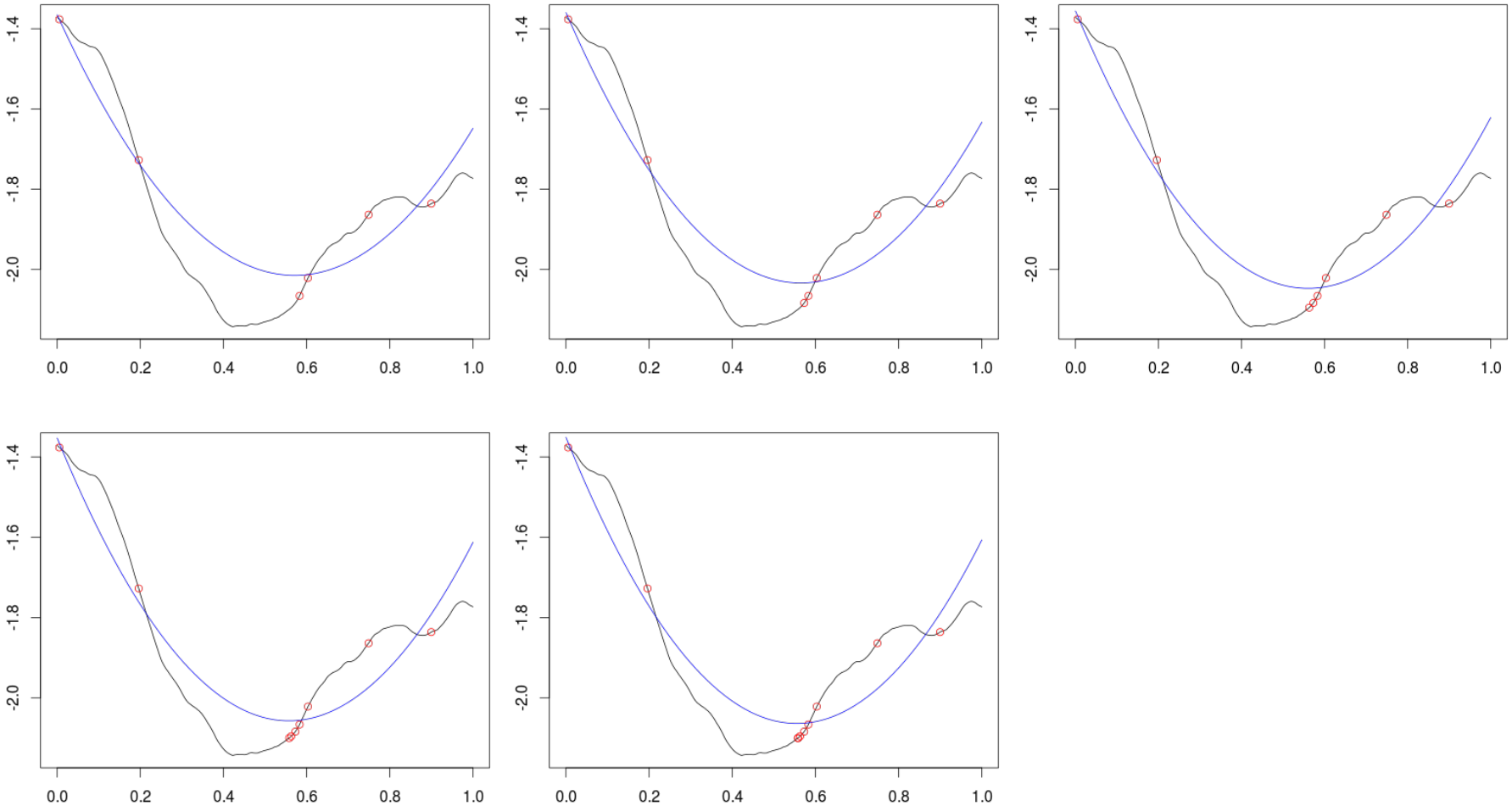
Optimisation basée sur les modèles polynomiaux

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**

cf. Présentation de B. Marteau, rencontres Mexico 2014

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)

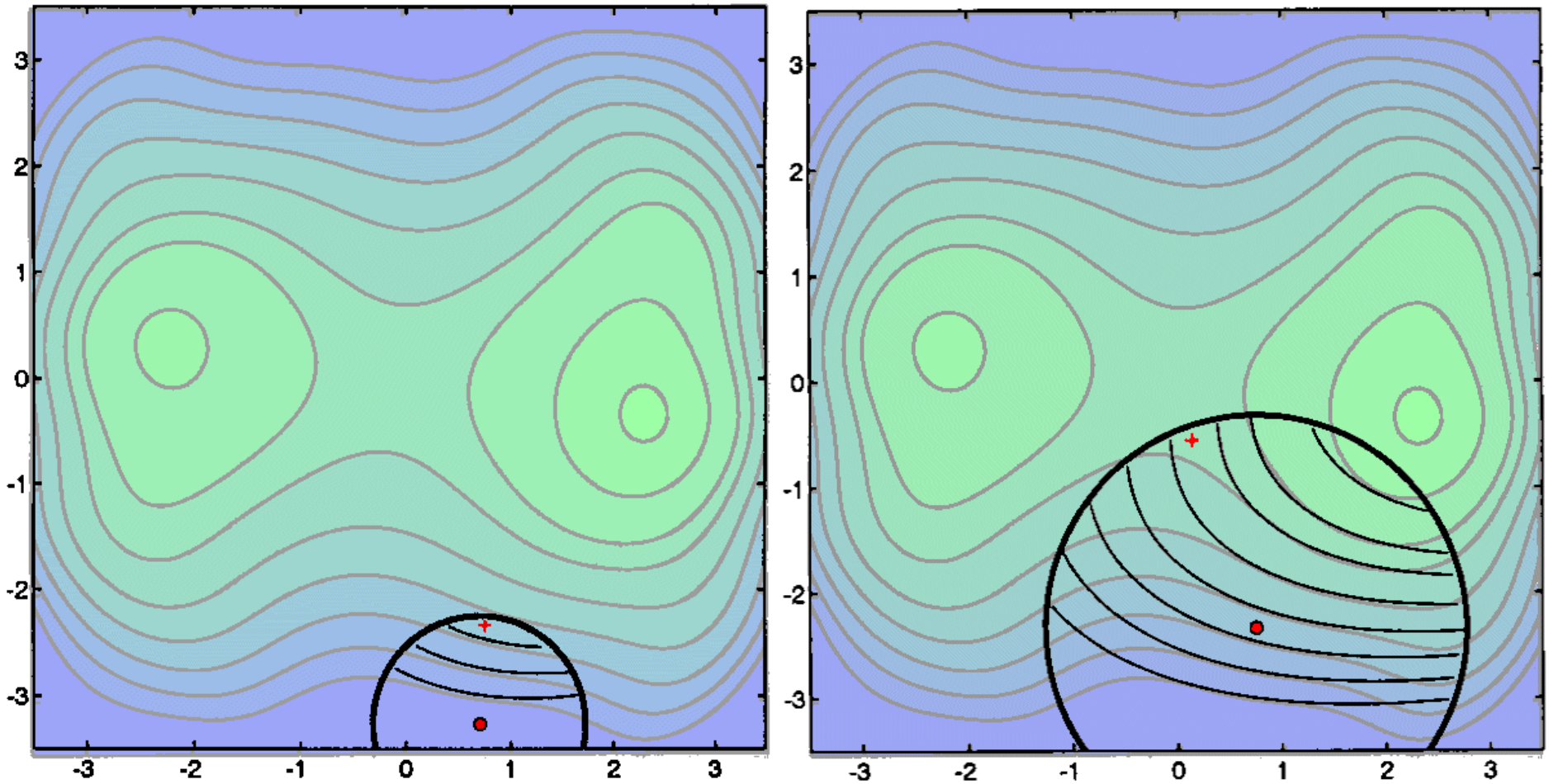
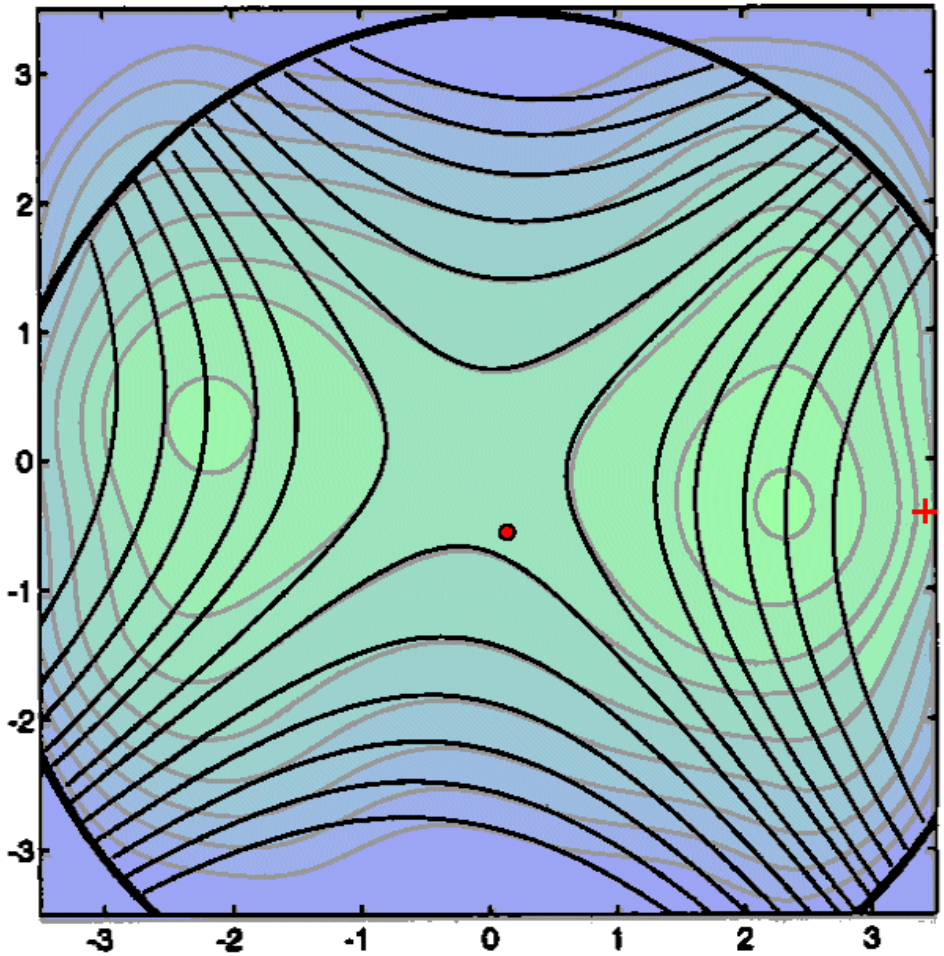


Illustration (source : F. Vanden Berghen)



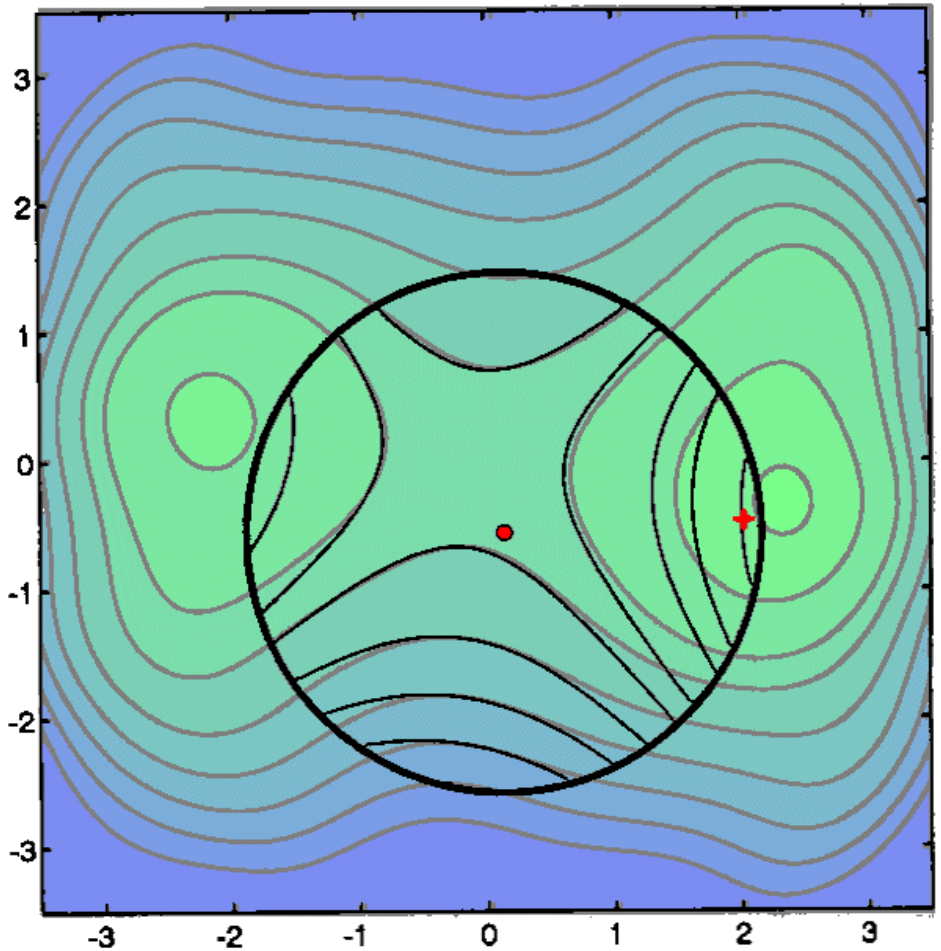
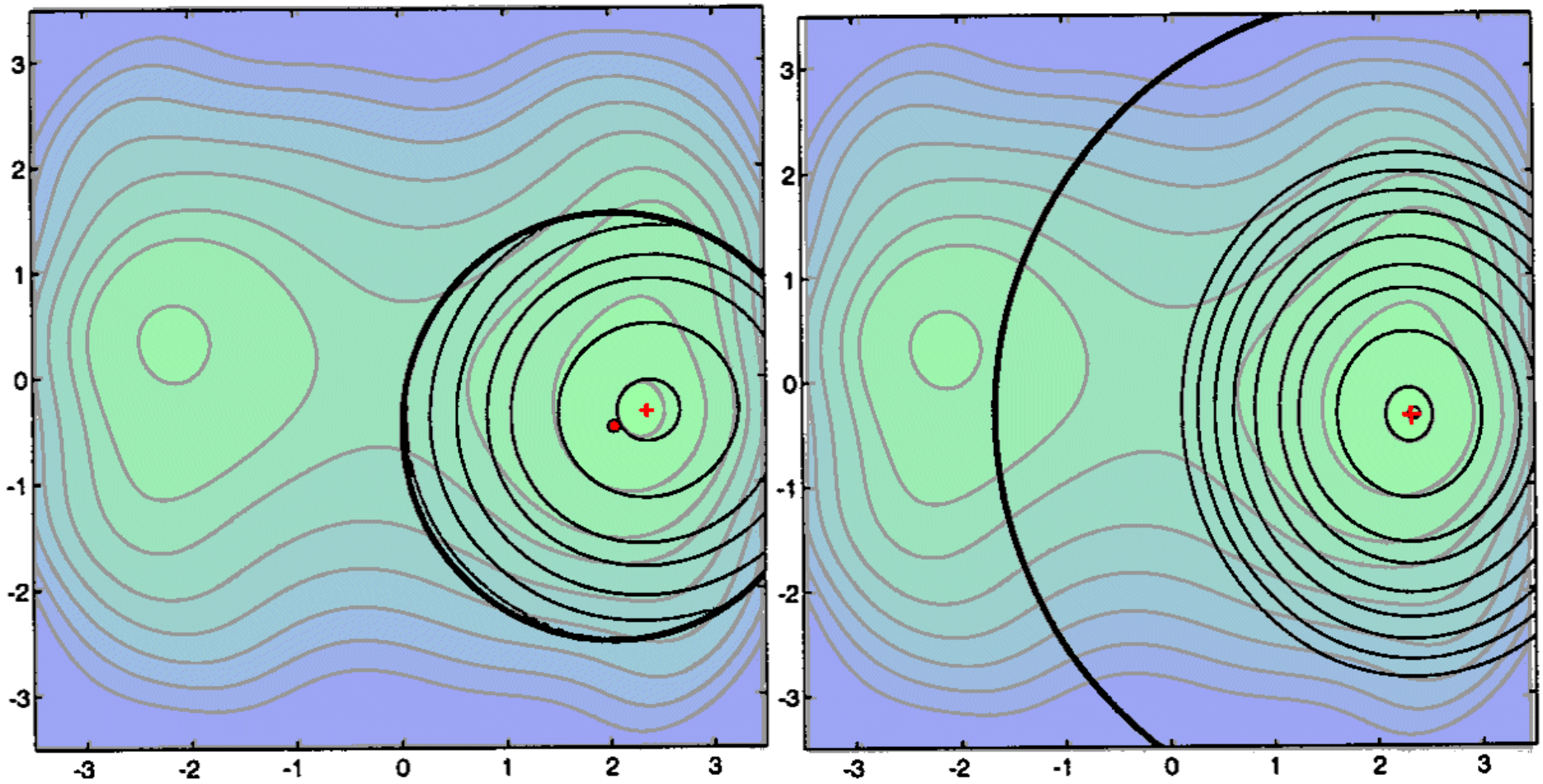


Illustration (source : F. Vanden Berghen)



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)

Où est passé le compromis exploration / intensification ?

Méthodes d'optimisation locale, proche des méthodes de gradient

Plan de l'exposé

1 Introduction
- Introduction à l'optimisation globale : l'algorithme DIRECT
- Optimisation globale et métamodèles

2 Optimisation basée sur les modèles polynomiaux

3 Optimisation sur base de krigeage : l'algorithme EGO

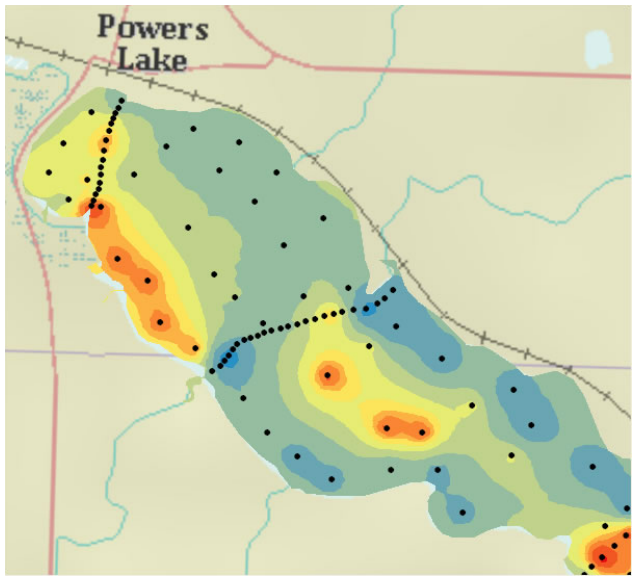
4 Extensions du principe d'EGO à différents contextes

5 Conclusion

Krigeage

Métamodèlisation par processus gaussiens

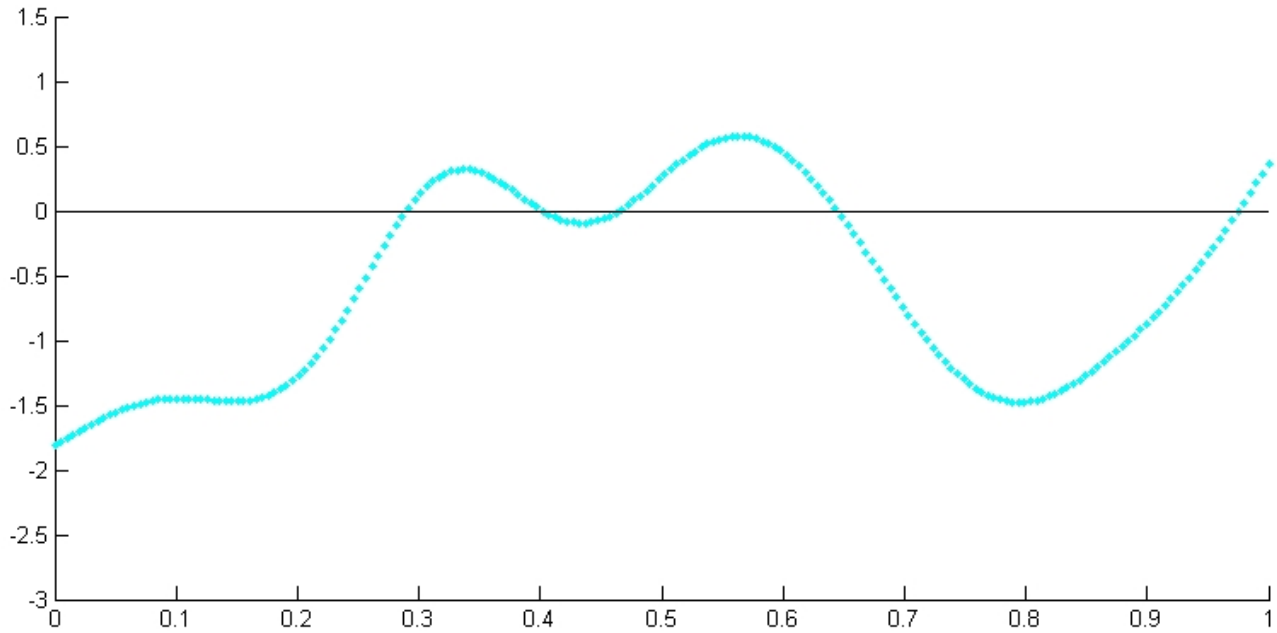
- Origine : modèles géostatistiques (reconstruction de sous-sols à partir de données de forages)
- Ajout d'information (fausse !) : forme simple pour la réponse



Ici : incertitude spatiale $\leftrightarrow$ absence d'expérience numérique

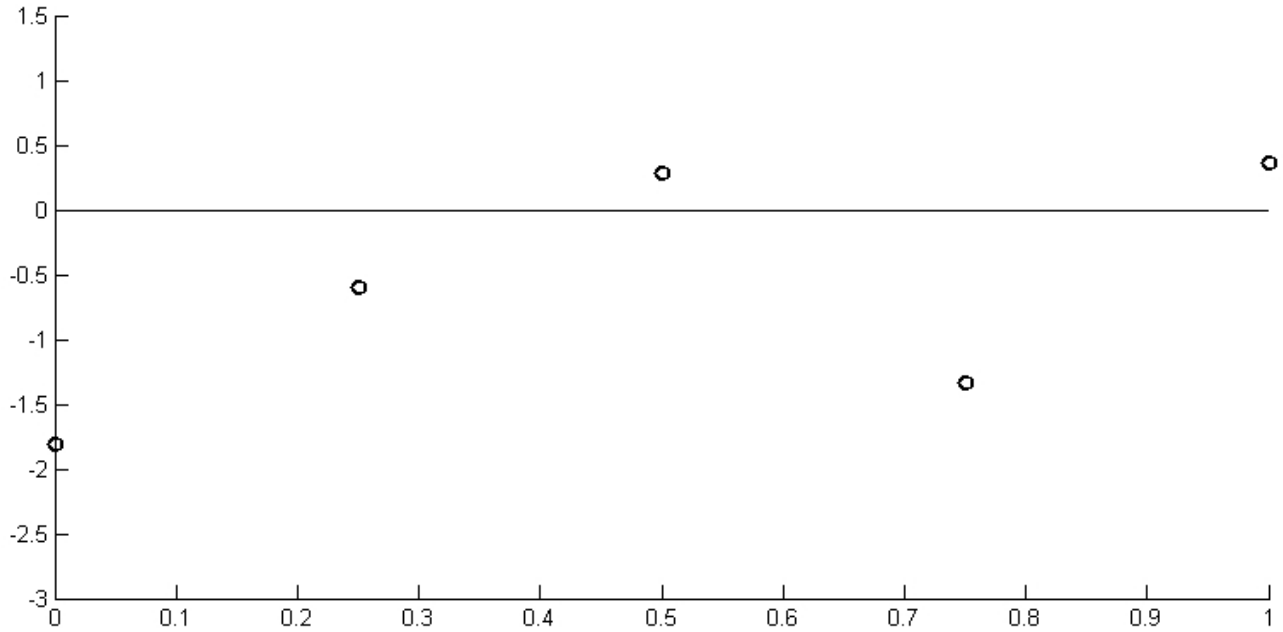
Krigeage 1/4

Vraie réponse (fonction d'un seul facteur)



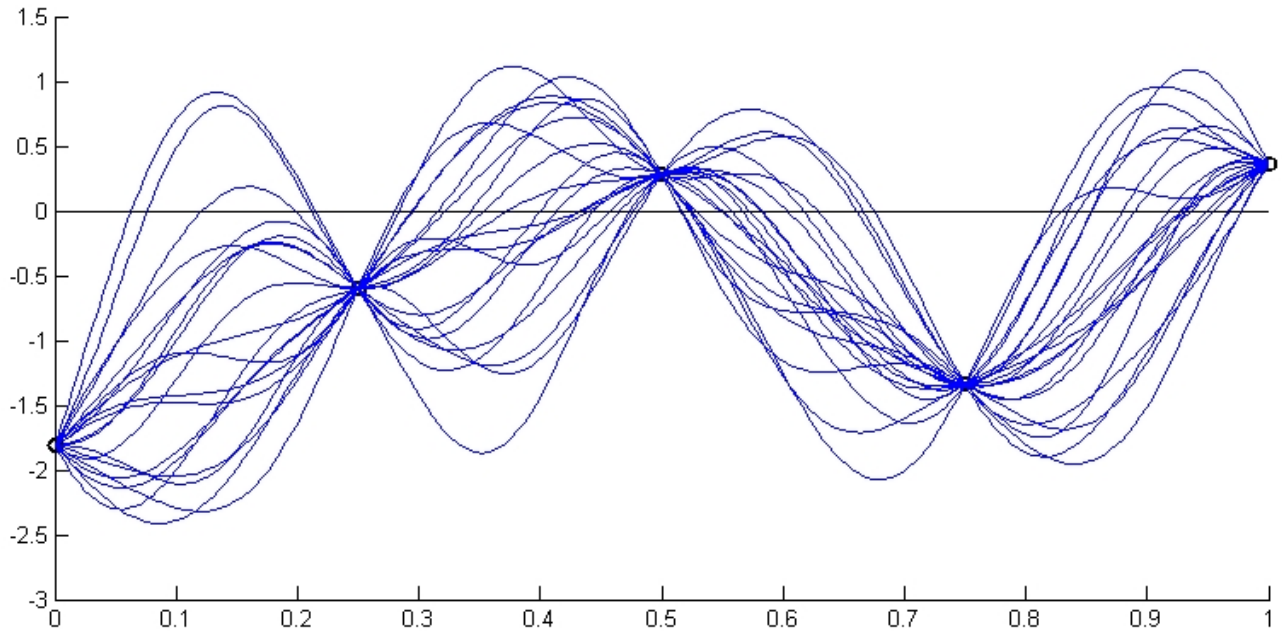
Krigeage 2/4

Ce qu'on a : un jeu de réponses



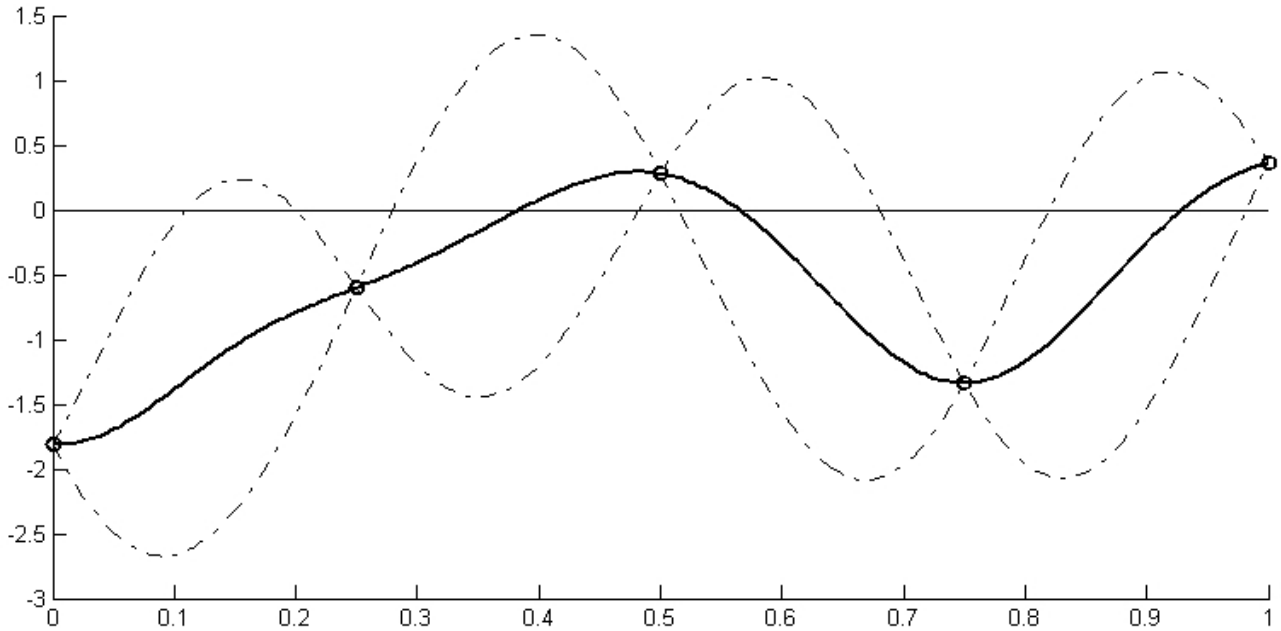
Krigeage 3/4

Chemins possibles (sous certaines hypothèses = fonction de covariance)



Krigeage 4/4

Modèle : prédicteur + incertitudes



Krigeage en pratique

Hypothèses

- Fonction de covariance paramétrique (θ portée, σ amplitude)
- "Tendance" ($X\beta$)

Apprentissage : maximum de vraisemblance

- Plan d'expérience initial
- Estimations de θ, σ, β (optimisation)

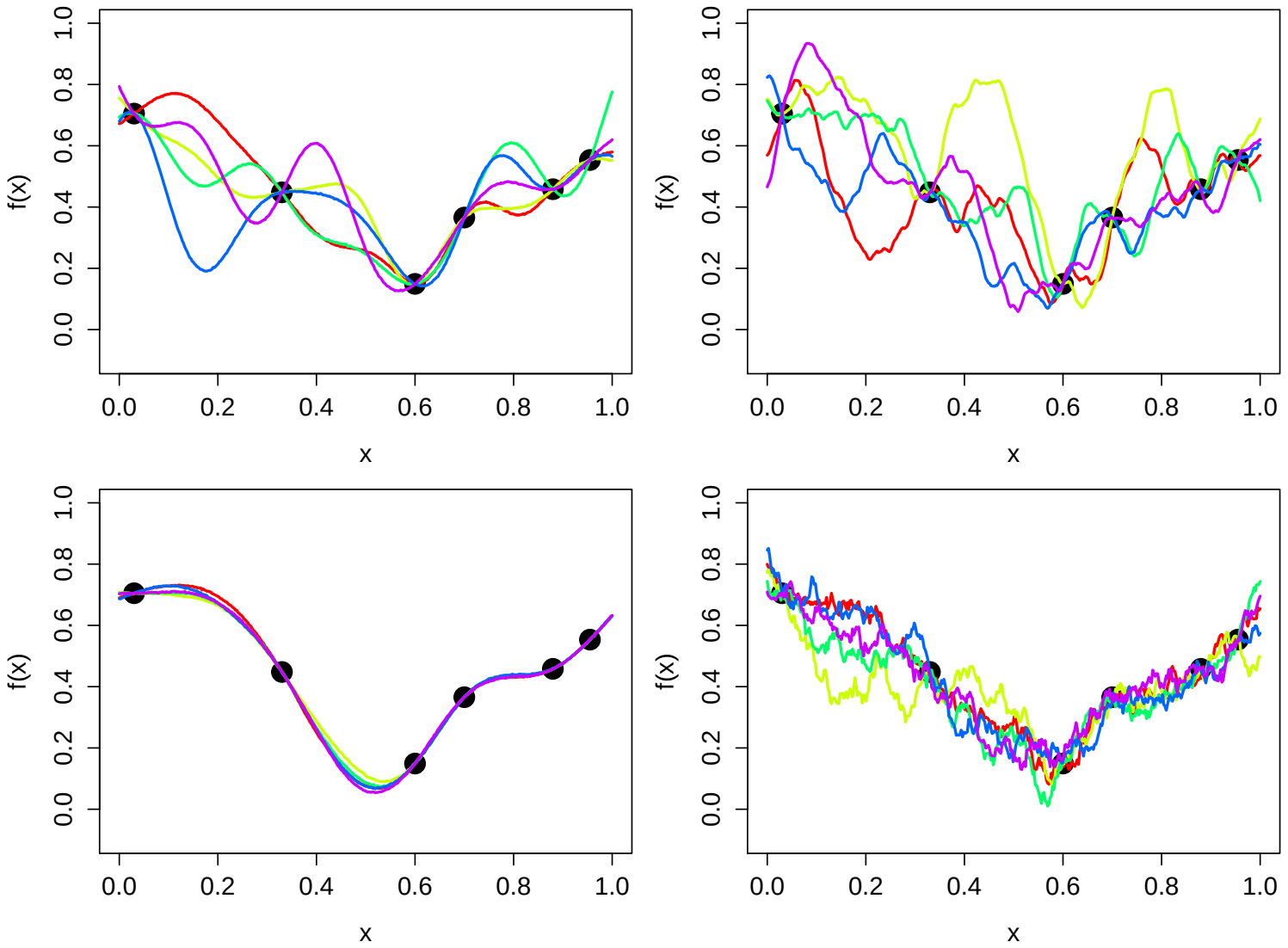
Validation

Si non concluant : changement de covariance, tendance, transformation des données

(c.f. présentation de R. Lardy)

Packages `R` : `DiceKriging`, `laGP`, `gpr`...

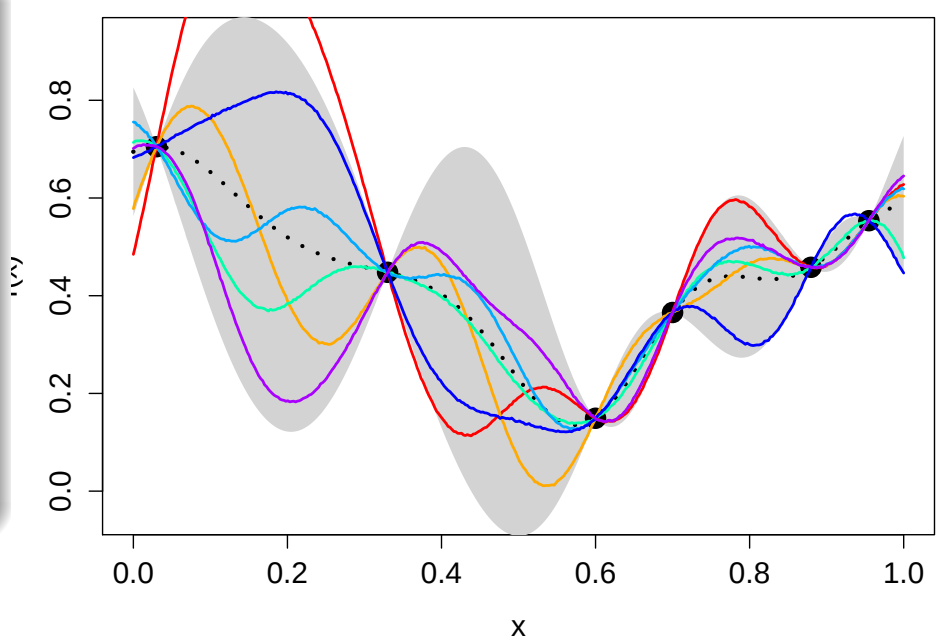
Même jeu d'observations, différentes covariances



Comment utiliser le krigeage pour optimiser ?

Atouts du modèle

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



Critère d'échantillonnage

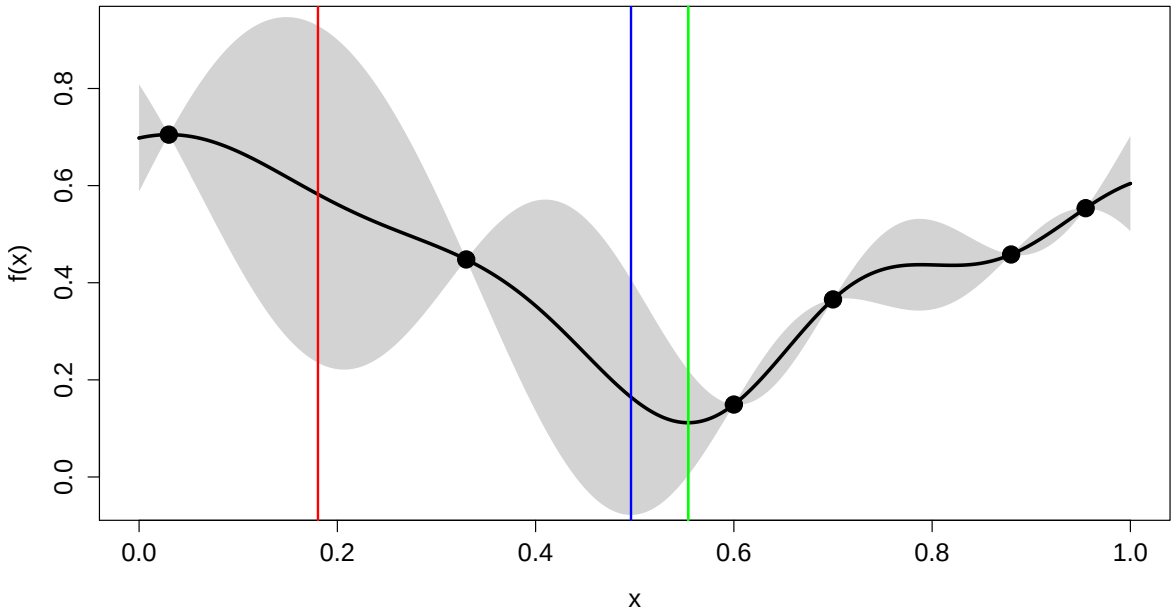
On utilise le métamodèle pour estimer "l'intérêt" d'une observation potentielle

$\Rightarrow$ On effectue la simulation la plus intéressante !

Différents critères

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

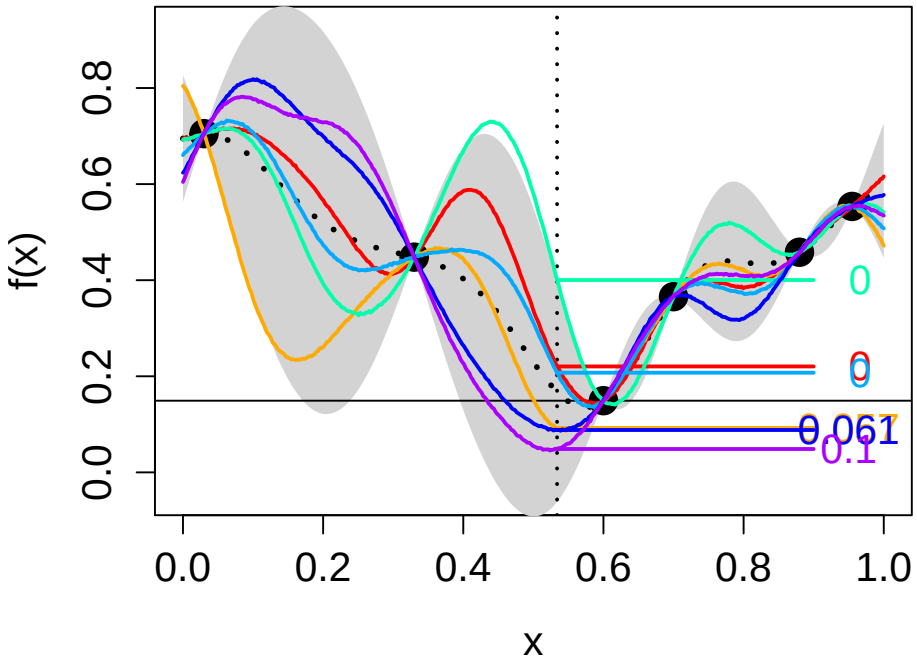
- Pure intensification : $\min m(x)$
- Pure exploration : $\max s^2(x)$
- Un compromis simple : $\min m(x) - \alpha s(x)$

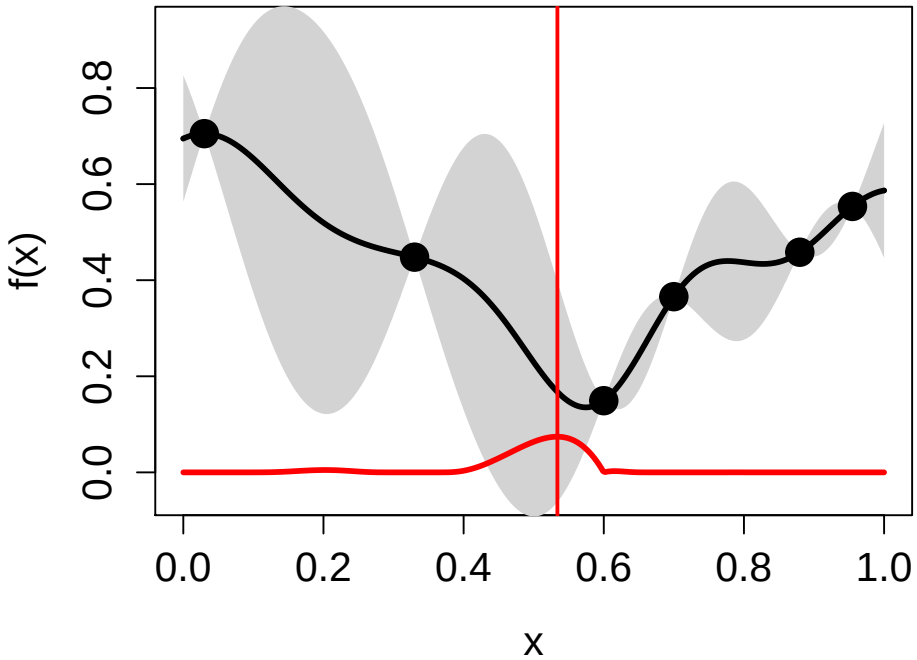


Le meilleur compromis : amélioration espérée

(ou *EI : Expected Improvement*)

- 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'algorithme *Efficient Global Optimization*

Initialisation

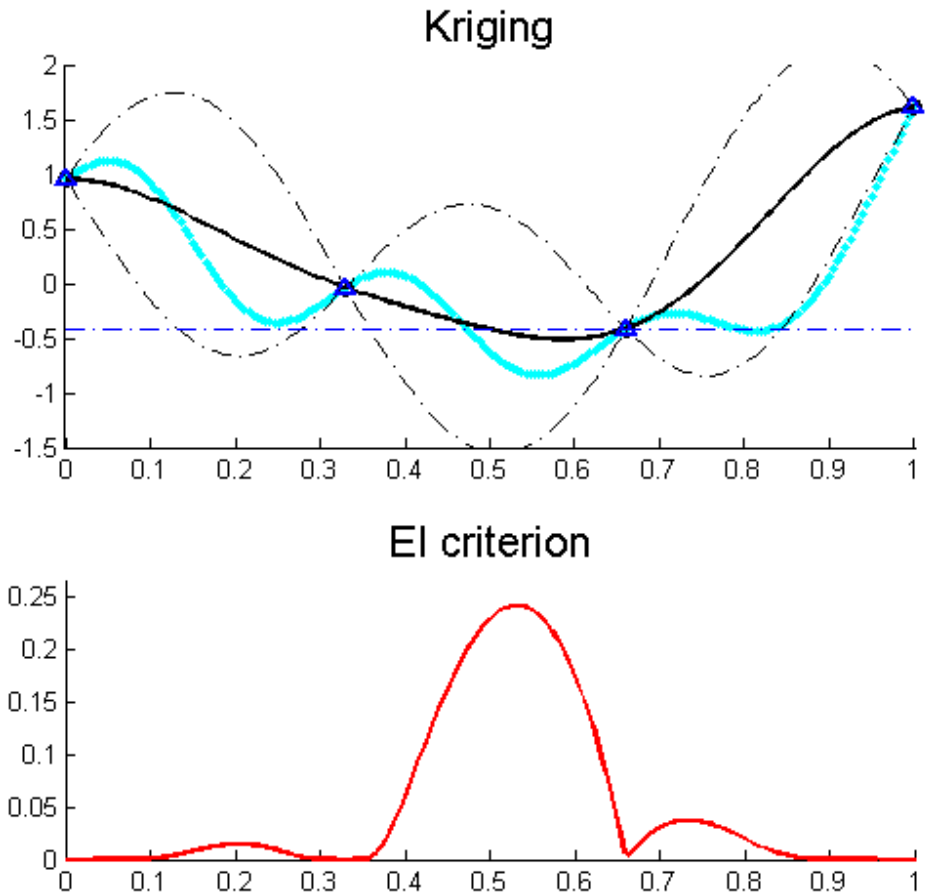
- Réalisation d'un plan initial
- Construction d'un krigeage

Boucle d'optimisation

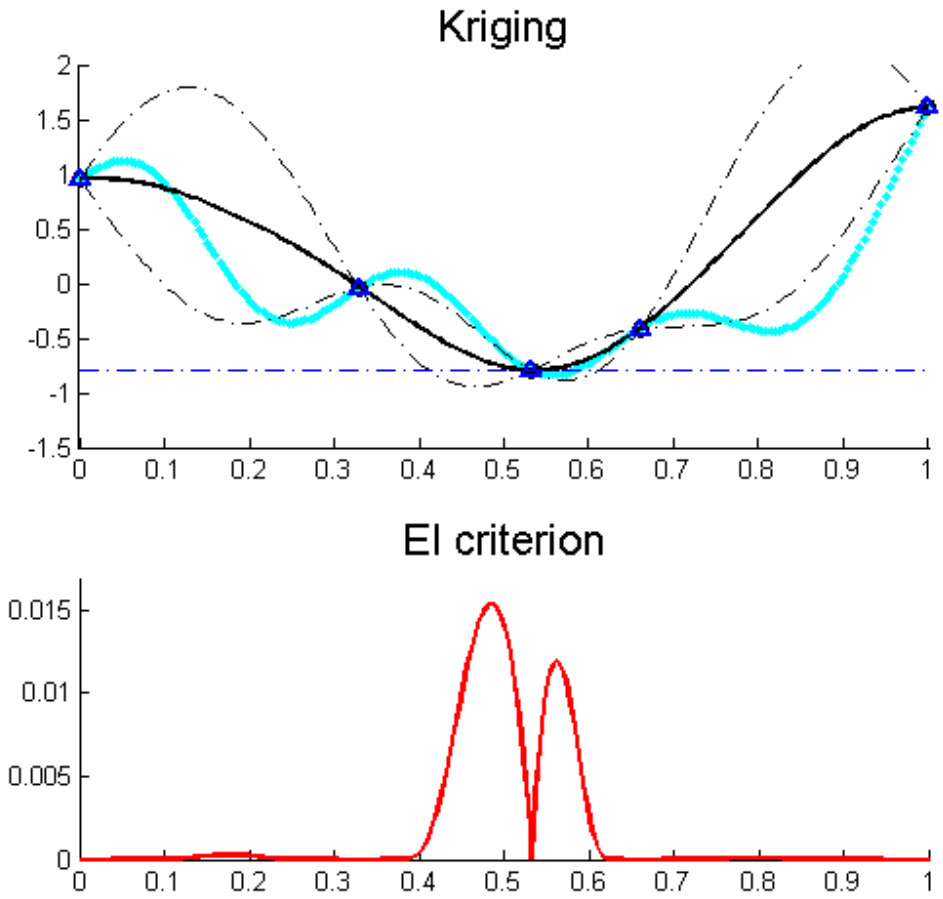
- Recherche du point qui maximise l'amélioration espérée
- Calcul du vrai modèle en ce point
- Ajout du point au plan et mise à jour du modèle

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

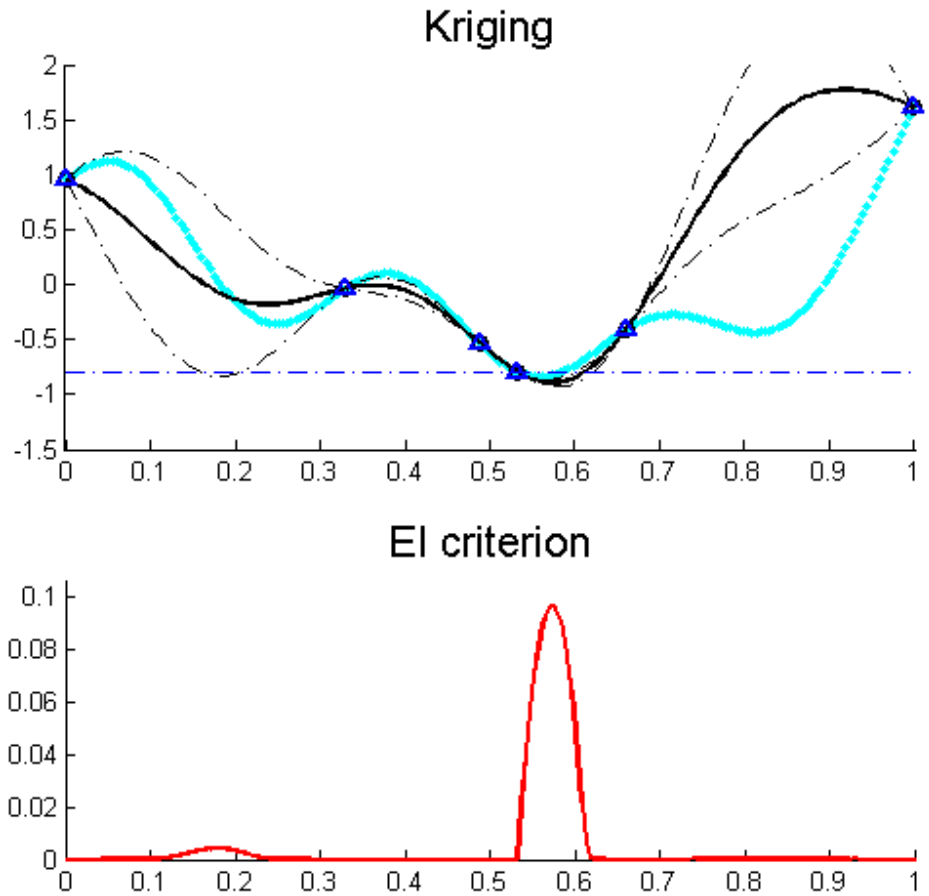
EGO : illustration 1/6



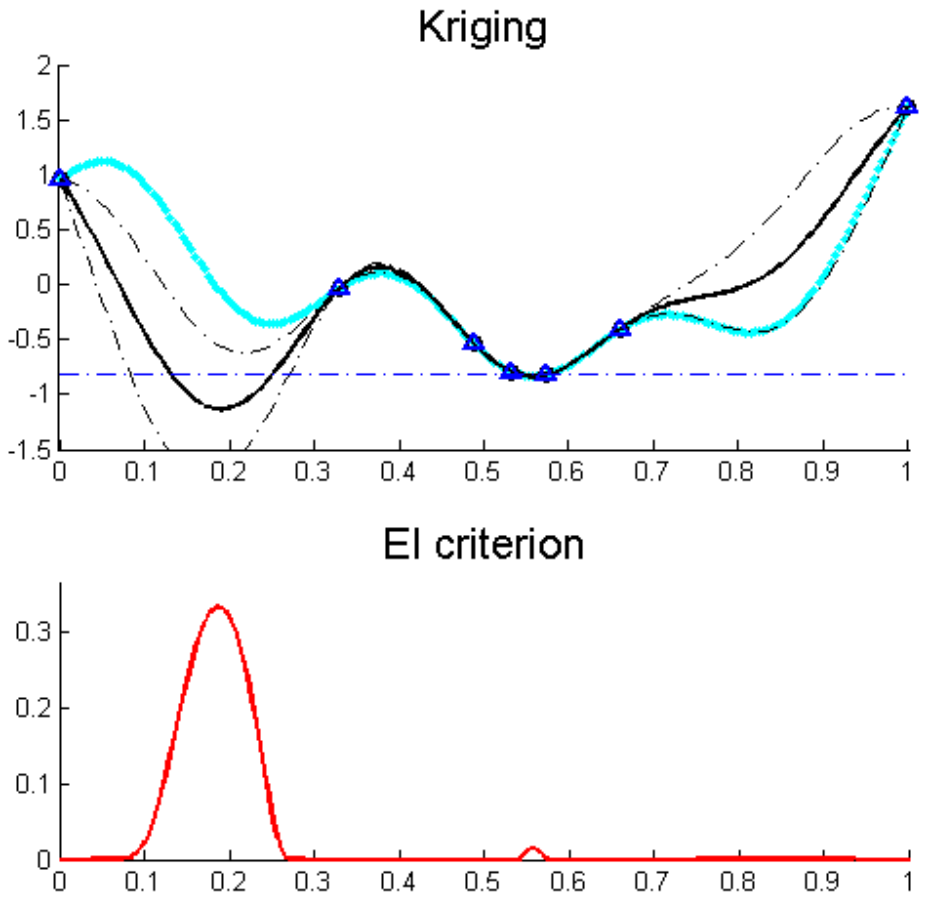
EGO : illustration 2/6



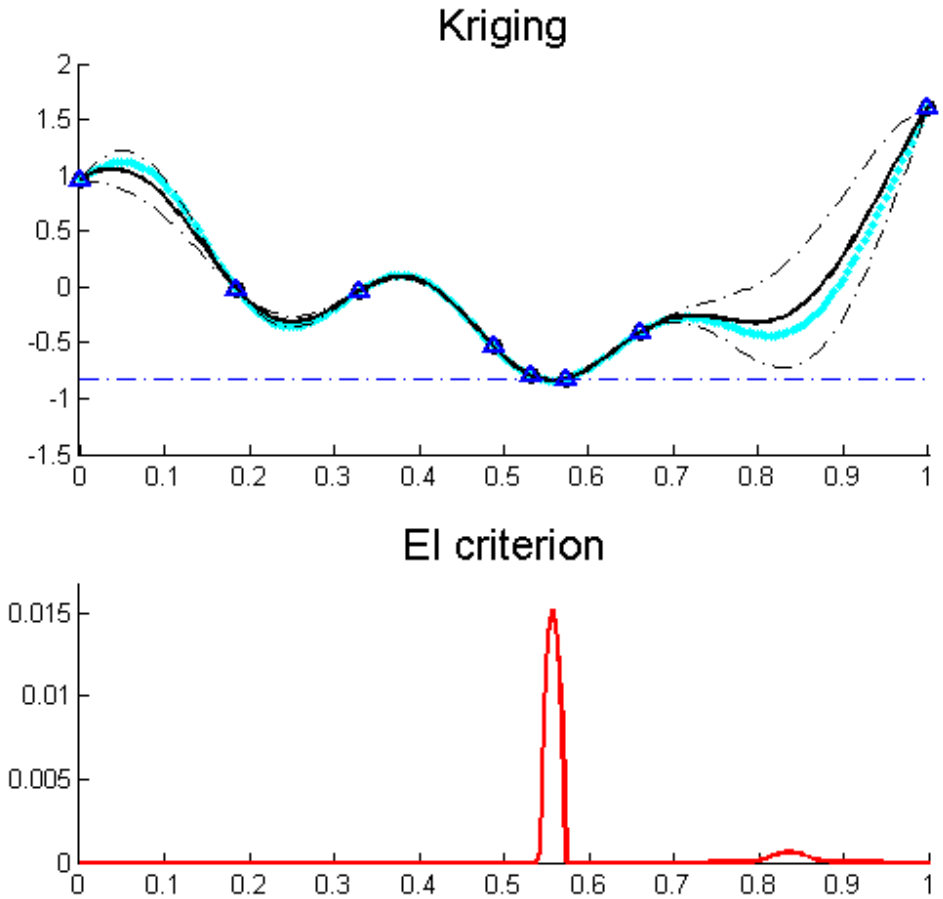
EGO : illustration 3/6



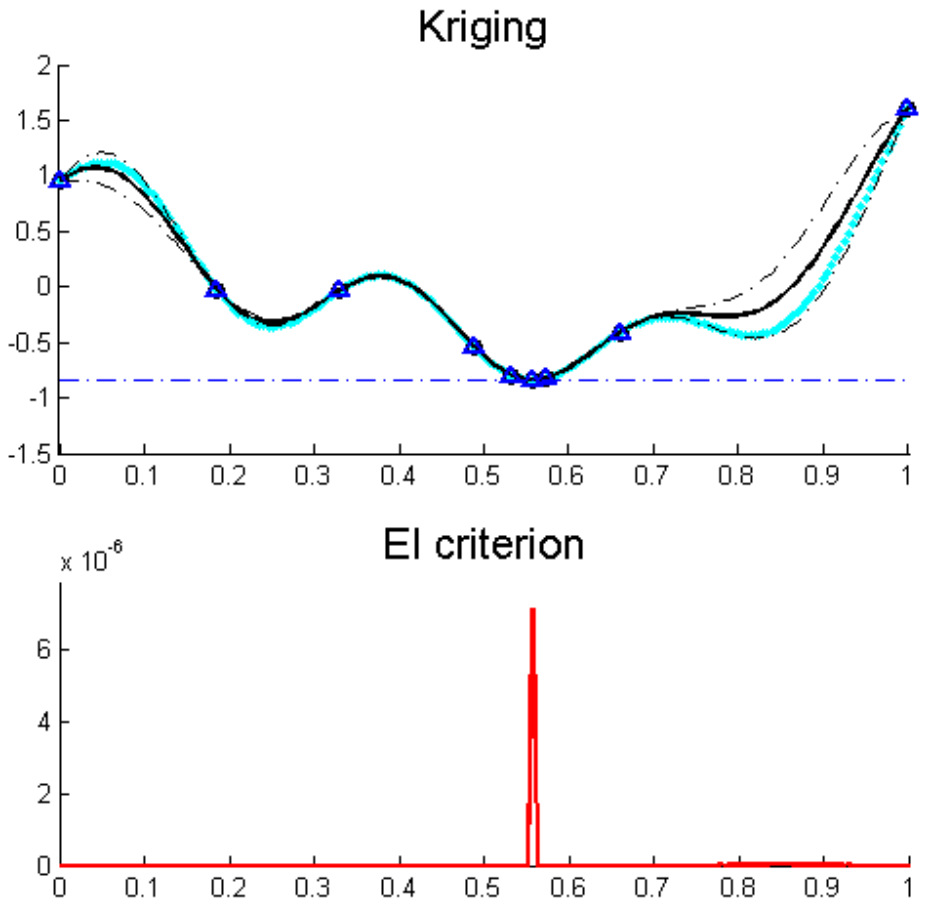
EGO : illustration 4/6



EGO : illustration 5/6



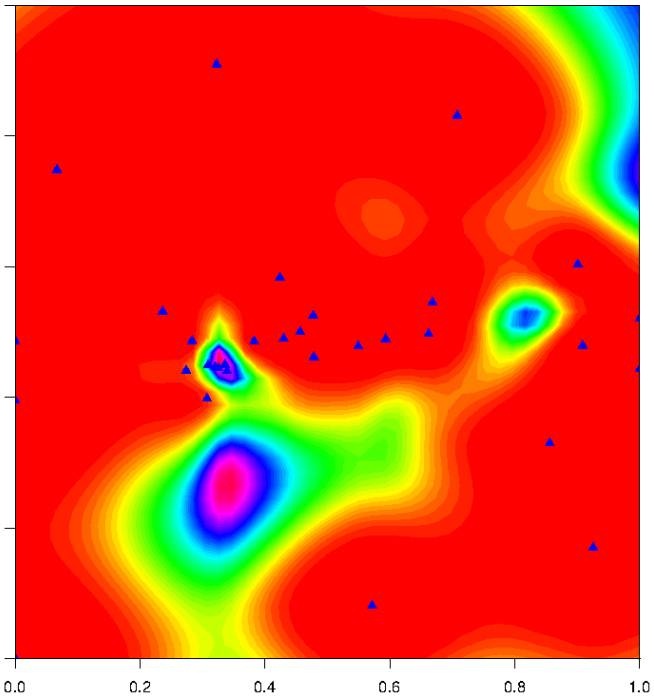
EGO : illustration 6/6



EGO en pratique : deux points critiques (1/2)

1- Maximisation de l'amélioration espérée

- Sous-problème d'optimisation (globale) difficile !
- *EI* "gratuit" ($\approx 1/100s$) $\rightarrow$ 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
Journal of Computational Software

EGO en pratique : deux points critiques (2/2)

2- Apprentissage du modèle de krigeage

- Quelle proportion plan initial / optimisation ?
- Choix a priori du modèle
- Paramètres de covariance : ré-estimation ?
  - coûteux → dépend du temps de calcul pour une simulation !
  - parfois instable

Quelques résultats... à prendre avec précaution

Picheny, Wagner, Ginsbourger (2013)
A benchmark of kriging-based infill criteria for noisy optimization
Structural and Multidisciplinary Optimization

- Budget du plan initial (20% - 50%) & choix du noyau : peu critiques
- Paramètres de covariance & capacité du modèle à représenter la fonction : influence majeure sur le comportement !

Plan de l'exposé

1. Introduction
   - Introduction à l'optimisation globale : l'algorithme DIRECT
   - Optimisation globale et métamodèles
2. Optimisation basée sur les modèles polynomiaux
3. Optimisation sur base de krigeage : l'algorithme EGO
4. Extensions du principe d'EGO à différents contextes
5. Conclusion

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 :

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

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

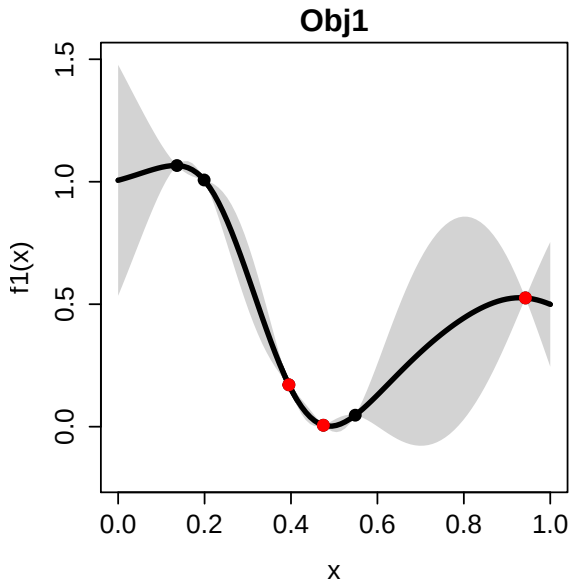
D. Ginsbourger, R. Le Riche, L. Carraro (2010)
Kriging is well-suited to parallelize optimization
Computational intelligence in expensive optimization problems, 131-162

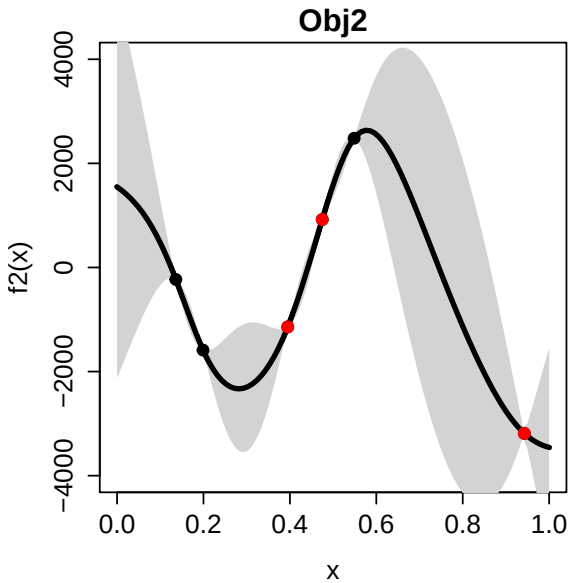
Multi-objectifs : un métamodèle par objectif

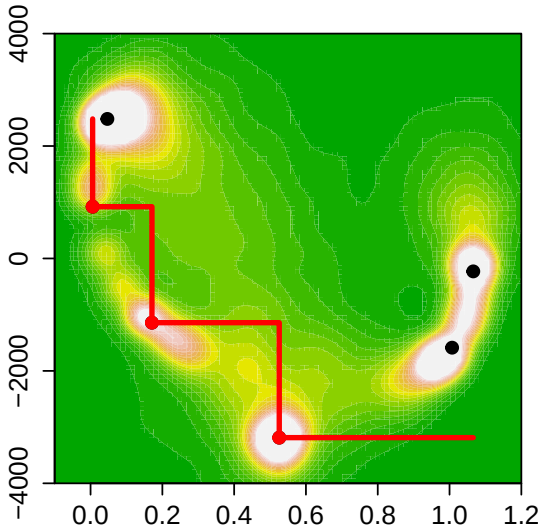
"Amélioration" sur le front de Pareto ?

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







Prochainement : package `R` `GPareto`

Optimisation avec contraintes : 3 cas (1/2)

Objectif coûteux, contrainte rapide

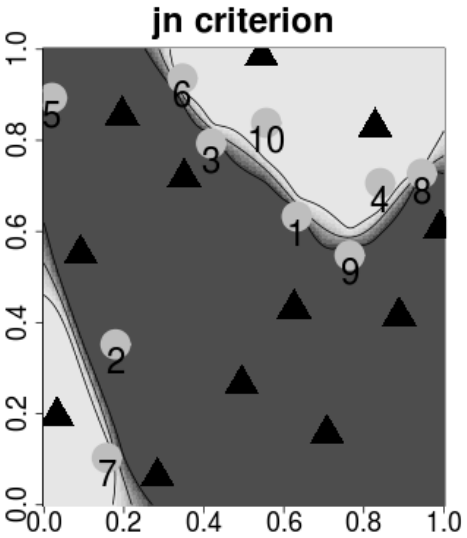
max Rendement
s.c. ITK réalisable

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

Objectif rapide, contrainte coûteuse

min Poids
s.c. Contrainte méca. $\leq$ Seuil

- Métamodèle pour la contrainte
- Apprentissage de la frontière

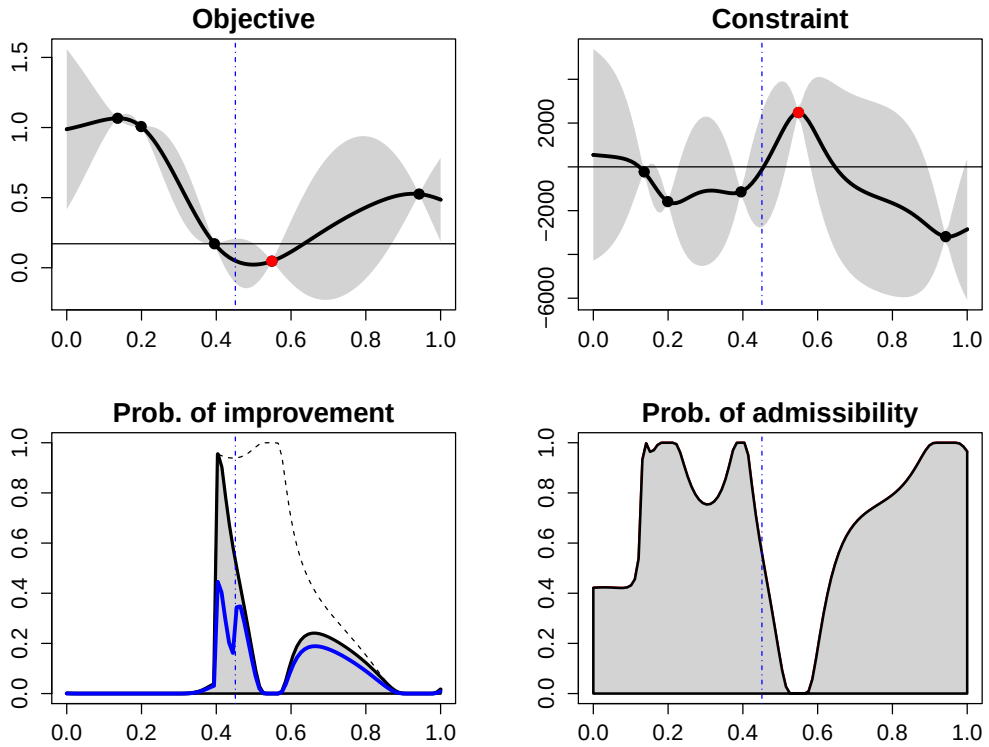


Chevalier, Picheny, Ginsbourger (2014)
KrigInv : An efficient and user-friendly implementation of batch-sequential inversion strategies based on kriging,
CSDA, vol.71, pp.1021-1034

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

IAGO : Informal Approach to Global Optimization

Réduction de l'entropie du minimum

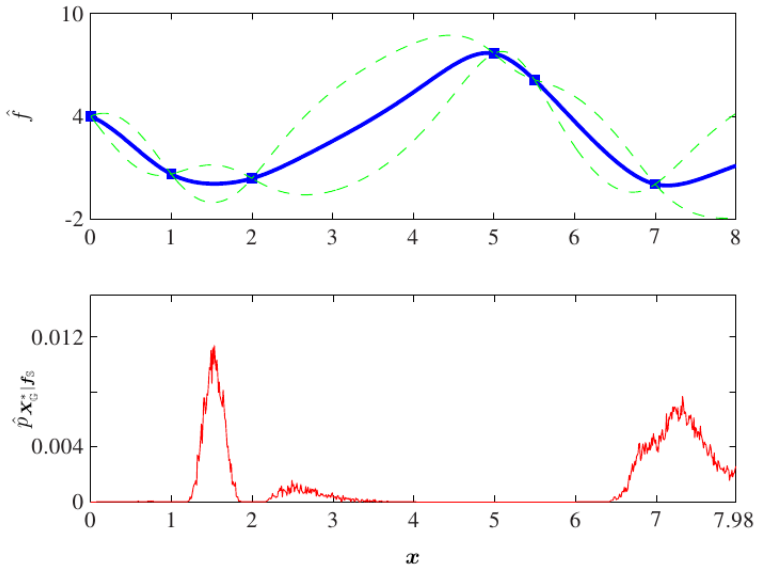
IECI : Integrated Expected Conditional Improvement

Gain estimé sur tout l'espace

EEV : Expected Volume Reduction

Réduction du volume sous le minimum

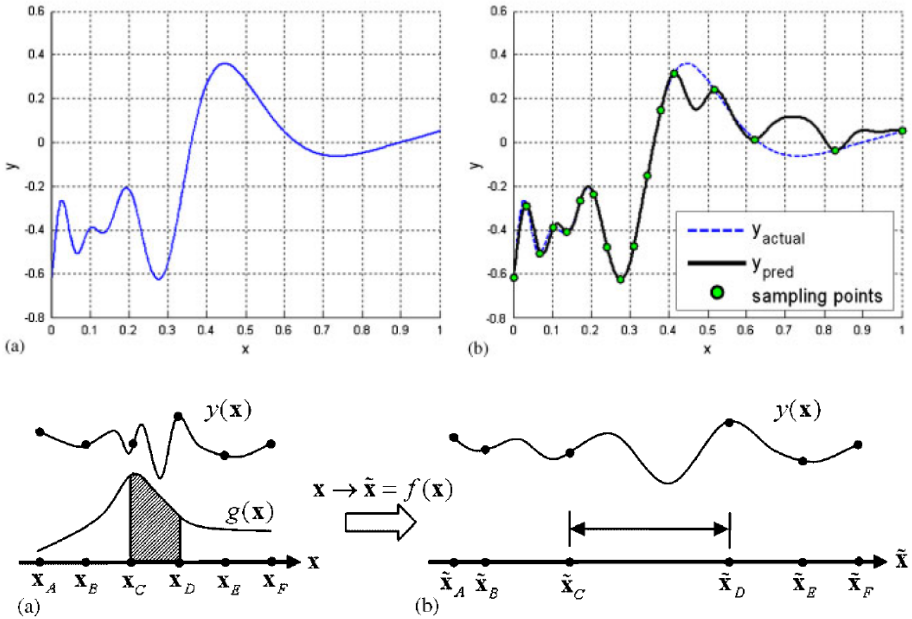
Distribution du minimum (IAGO) :



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

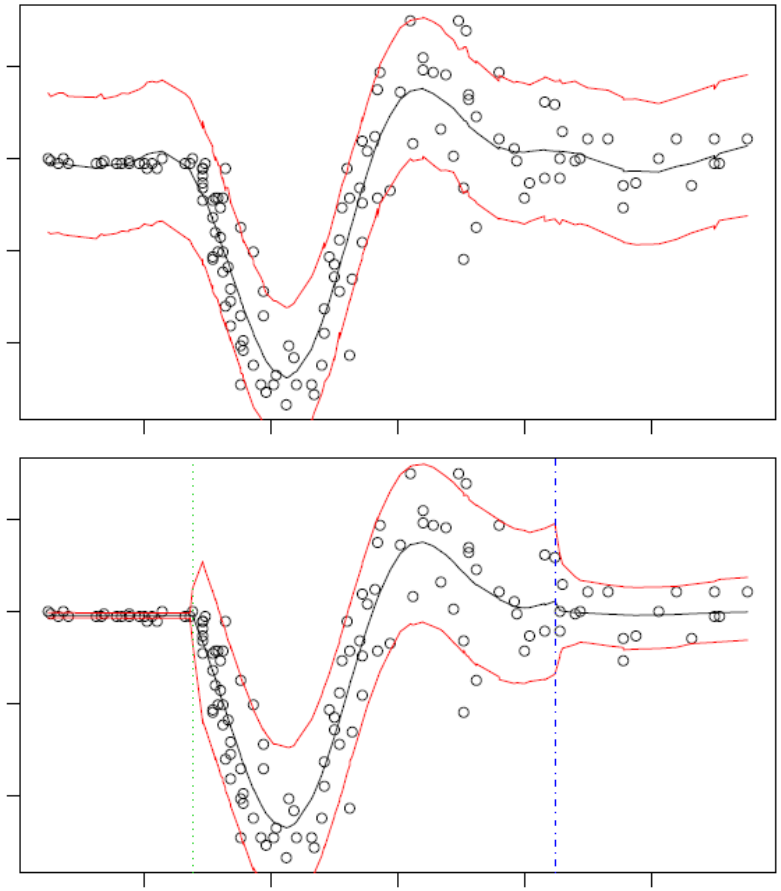
Modèles complexes (1/2) : krigeages non stationnaires

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.

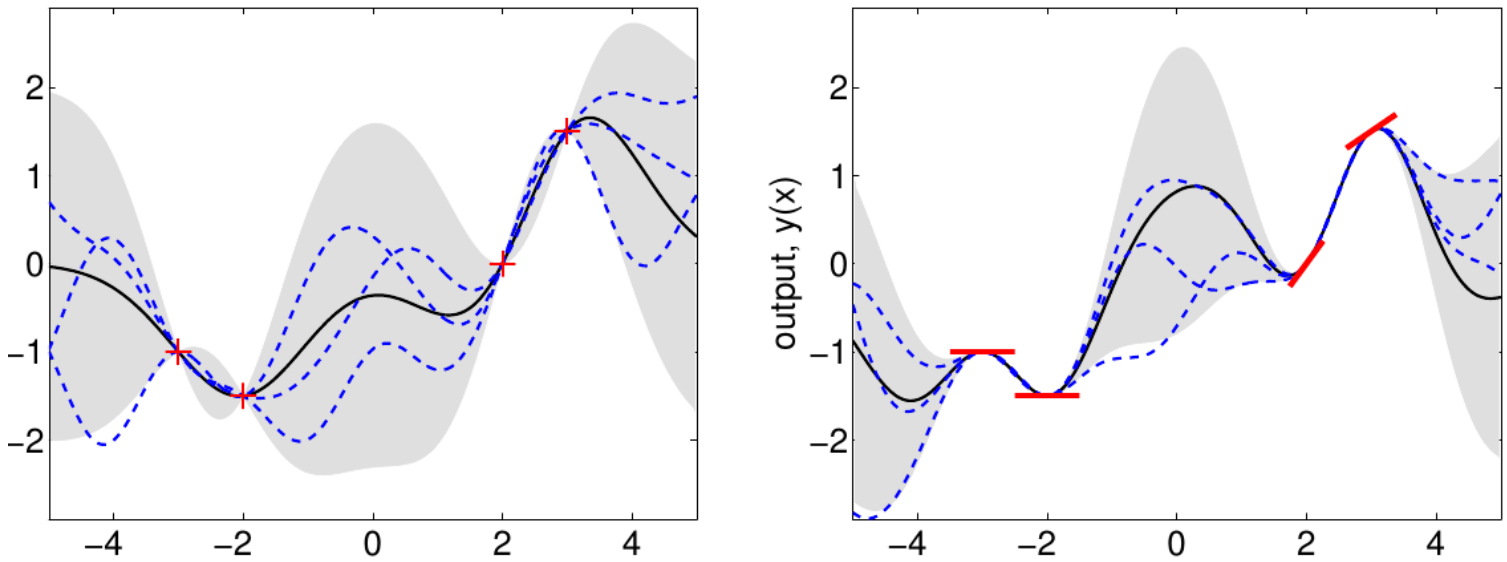
Découpage de l'espace : treed GP



⇒ 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

Plan de l'exposé

1. Introduction
   - Introduction à l'optimisation globale : l'algorithme DIRECT
   - Optimisation globale et métamodèles
2. Optimisation basée sur les modèles polynomiaux
3. Optimisation sur base de krigeage : l'algorithme EGO
4. Extensions du principe d'EGO à différents contextes
5. Conclusion

Conclusion (1/2)

Optimiser à l'aide de métamodèles

- Simulateurs très coûteux, dimension restreinte
- Métamodèle = exploitation maximale de l'information disponible

Approches non séquentielles (PX + 1 point)

\+ Simple

– Mauvais compromis exploration / intensification

– Pas de plan B !

Approches séquentielles

- Minimiser de manière répétée le métamodèle ne fonctionne pas !
- La bonne stratégie dépend du modèle

Conclusion (2/2)

Polynômes + régions de confiance

- \+ Convergence, robustesse
- \+ Grande dimension
- − Optimisation locale
- − Pas si parcimonieux

Domaine d'application

- Dimension ≤ 1000
- Nombre d'observations ≤ 10000
- Temps de simulation $\leq 10min$
- Objectif unimodal (ou bien : démarrages multiples)

Krigeage + EGO

- \+ Optimiseur global
- \+ Très parcimonieux
- \+ Concept exportable
- − Robustesse
- − Modèle pas toujours adapté !

Domaine d'application

- Dimension ≤ 20
- Nombre d'observations ≤ 1000
- Temps de simulation $\leq 1j$
- Fonctions multimodales

Pour finir : de nombreux sujets non abordés...

...par manque de temps !

- Régions de confiance dans différents contextes
- Optimisation robuste (à venir !)
- Multi-fidélité
- Budget fini
- Approches bayésiennes pour les modèles
- Grande dimension

Merci de votre attention !