

Krigeage et Algorithmes de la famille EGO: autour du critère d'EI multipoints.

David Ginsbourger

Stochastic Characterization Hydrogeology Group (équipe Renard)
[Université de Neuchâtel](#)

Séminaire IMPEC (06/10/2009)
CEA Saclay

Cadre général et objectifs

Problématiques industrielles

- Conception optimale: automobile, aéronautique, etc.
- Identification de paramètres: géosciences, imagerie

Cadre général et objectifs

Problématiques industrielles

- Conception optimale: automobile, aéronautique, etc.
- Identification de paramètres: géosciences, imagerie

Framework

Le simulateur numérique est vu comme une fonction numérique multivariable

$$y : \mathbf{x} \in D \subset \mathbb{R}^d \rightarrow \mathbb{R}$$

connue en quelques points de D :

$$\mathbf{X} = \{\mathbf{x}^1, \dots, \mathbf{x}^n\}$$

(Plan d'Expériences)

$$\mathbf{Y} = \{y(\mathbf{x}^1), \dots, y(\mathbf{x}^n)\}.$$

(Observations)

Cadre général et objectifs

Problématiques industrielles

- Conception optimale: automobile, aéronautique, etc.
- Identification de paramètres: géosciences, imagerie

Framework

Le simulateur numérique est vu comme une fonction numérique multivariable

$$y : \mathbf{x} \in D \subset \mathbb{R}^d \rightarrow \mathbb{R}$$

connue en quelques points de D :

$$\mathbf{X} = \{\mathbf{x}^1, \dots, \mathbf{x}^n\}$$

(Plan d'Expériences)

$$\mathbf{Y} = \{y(\mathbf{x}^1), \dots, y(\mathbf{x}^n)\}.$$

(Observations)

Apprentissage et optimisation de y

- Évaluer y de manière aussi parcimonieuse que possible
- Moyen: utiliser des représentations simplifiées de y (métamodèles)
- **But ultime**: sélectionner \mathbf{X} puis $\{\mathbf{x}^{n+1}, \dots, \mathbf{x}^{n+r}\}$ de manière à bien approximer et (presque) minimiser y .

Sommaire

- 1 Métamodèles de Krigeage pour l'approximation et l'optimisation
 - Bases du Krigeage et package DiceKriging
 - Expected Improvement et optimisation sur base de Krigeage
- 2 Parallélisation(s) de l'optimisation sur base de Krigeage
 - *EI* multipoints et heuristiques parallèles
 - DiceOptim. Application en dimension 6.
- 3 Derniers développements en optimisation et approximation
 - Optimisation séquentielle en temps fini
 - Quelques extensions du modèle de Krigeage

Métamodèles: *making new from old*

Métamodèles déterministes: approximation classique

- Chercher la meilleure approximation dans un "petit" ev:

$$\min_{f \in \mathcal{F}} \|f(\mathbf{X}) - y(\mathbf{X})\|_{\mathbb{R}^n}^2, \text{ où } \mathcal{F} = \text{vect}\{f_i, i \in I\}$$

Exemples: poly., Fourier, NN, Ondelettes, splines, etc.

- Régularisation de Tikhonov: $\min. \|f(\mathbf{X}) - y(\mathbf{X})\|^2 + \lambda P(f)$.

Métamodèles: *making new from old*

Métamodèles déterministes: approximation classique

- Chercher la meilleure approximation dans un "petit" ev:

$$\min_{f \in \mathcal{F}} \|f(\mathbf{X}) - y(\mathbf{X})\|_{\mathbb{R}^n}^2, \text{ où } \mathcal{F} = \text{vect}\{f_i, i \in I\}$$

Exemples: poly., Fourier, NN, Ondelettes, splines, etc.

- Régularisation de Tikhonov: $\min. \|f(\mathbf{X}) - y(\mathbf{X})\|^2 + \lambda P(f)$.

Métamodèles probabilistes: conditionnement de processus

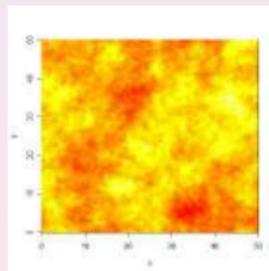
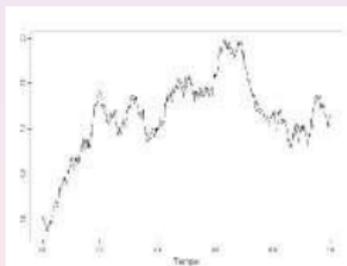
- Idée fondamentale: y est supposée être *une réalisation* d'un processus aléatoire $Y(x; \omega)$.
- On s'intéresse en chaque x à la loi $[Y(x) | Y(\mathbf{X}) = \mathbf{Y}]$.

Processus Gaussiens et apprentissage fonctionnel

Approximation de fonctions déterministes par des PG (Krigeage)

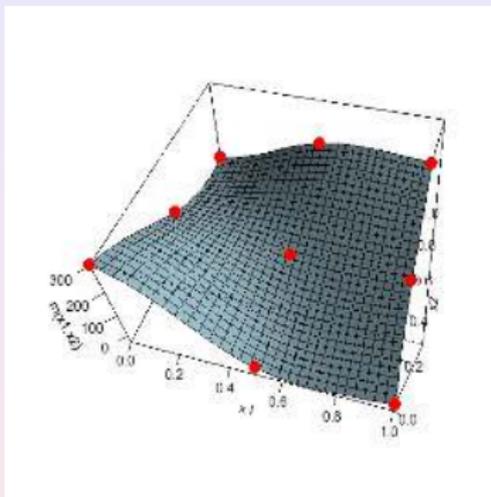
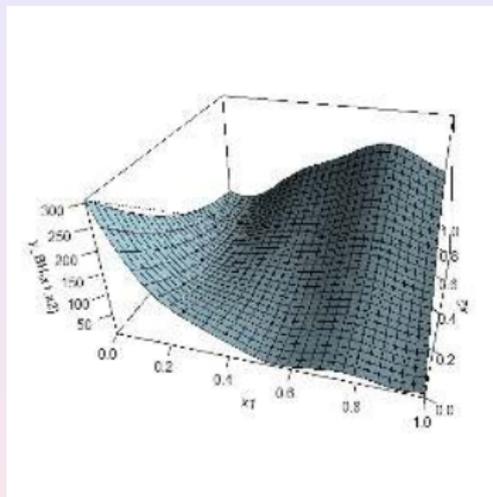
$$y : \mathbf{x} \in D \subset \mathbb{R}^d \rightarrow y(\mathbf{x}) \in \mathbb{R}$$

y est vu comme une réalisation d'un PG $Y_{\mathbf{x}}$ de moyenne $\mu(\mathbf{x})$ et de **noyau de covariance** $k \in (\mathbf{x}, \mathbf{x}') \in D \times D \rightarrow k(\mathbf{x}, \mathbf{x}') \in \mathbb{R}$.



- 1 Krigeage \approx approximer y en conditionnant $Y_{\mathbf{x}}$ aux observations en un plan d'expériences, notés $\{\mathbf{X}, \mathbf{Y}\} = \{(\mathbf{x}^1, \dots, \mathbf{x}^n), (y(\mathbf{x}^1), \dots, y(\mathbf{x}^n))\}$

Exemple d'interpolation par Krigeage Ordinaire

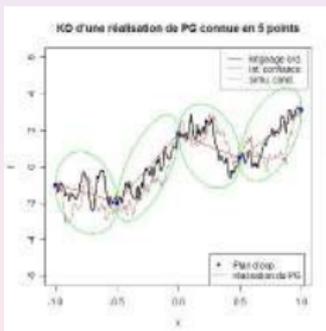
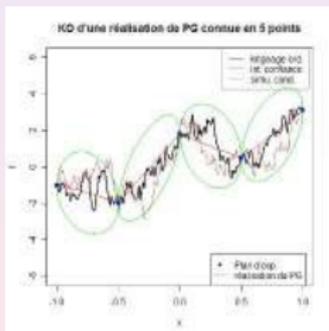


Interpolation (à droite) de la fonction de Branin function (à gauche), connue en 9 points (en rouge). La moyenne est ici supposée constante ($= \mu \in \mathbb{R}$).

Le noyau est $k(\mathbf{x}, \mathbf{x}') = \sigma^2 e^{-\sum_{j=1}^2 \left(\frac{x_j - x'_j}{\psi_j} \right)^2}$ avec (σ^2, ψ) estimé par MV

Equations de Krigeage Ordinaire

$$\left\{ \begin{array}{l} m(\mathbf{x}) = \hat{\mu} + \mathbf{k}(\mathbf{x})^T \mathbf{K}^{-1} (\mathbf{Y} - \hat{\mu} \mathbf{1}_n) \\ s^2(\mathbf{x}) = \sigma^2 - \mathbf{k}(\mathbf{x})^T \mathbf{K}^{-1} \mathbf{k}(\mathbf{x}) + \frac{(1 - \mathbf{1}_n^T \mathbf{K}^{-1} \mathbf{k}(\mathbf{x}))^2}{(\mathbf{1}_n^T \mathbf{K}^{-1} \mathbf{1}_n)} \end{array} \right.$$



Propriété fondamentale du KO, lorsque k est connu (et μ a un prior $\mathcal{U}(\mathbb{R})$...)

$$[Y(\mathbf{x}) | Y(\mathbf{X}) = \mathbf{Y}] \sim \mathcal{N}(m(\mathbf{x}), s^2(\mathbf{x}))$$

Sommaire

- 1 Métamodèles de Krigeage pour l'approximation et l'optimisation
 - Bases du Krigeage et package DiceKriging
 - Expected Improvement et optimisation sur base de Krigeage
- 2 Parallélisation(s) de l'optimisation sur base de Krigeage
 - *EI* multipoints et heuristiques parallèles
 - DiceOptim. Application en dimension 6.
- 3 Derniers développements en optimisation et approximation
 - Optimisation séquentielle en temps fini
 - Quelques extensions du modèle de Krigeage

Optimisation avec métamodèles: une bonne idée?

Métamodèles déterministes: approche limitée

- L'idée naïve de remplacer y par f et de minimiser f marche généralement très mal
- $\|f - y\|$ petit ne signifie *a priori* pas que $|\min(f) - \min(y)|$ petit ou $\|\operatorname{argmin}(f) - \operatorname{argmin}(y)\|$ petit.

Optimisation avec métamodèles: une bonne idée?

Métamodèles déterministes: approche limitée

- L'idée naïve de remplacer y par f et de minimiser f marche généralement très mal
- $\|f - y\|$ petit ne signifie *a priori* pas que $|\min(f) - \min(y)|$ petit ou $\|\operatorname{argmin}(f) - \operatorname{argmin}(y)\|$ petit.



D.R. Jones.

A taxonomy of global optimization methods based on response surfaces.

Journal of Global Optimization, 21, 345–383, 2001.

Optimisation avec métamodèles: une bonne idée?

Métamodèles déterministes: approche limitée

- L'idée naïve de remplacer y par f et de minimiser f marche généralement très mal
- $\|f - y\|$ petit ne signifie *a priori* pas que $|\min(f) - \min(y)|$ petit ou $\|\operatorname{argmin}(f) - \operatorname{argmin}(y)\|$ petit.



D.R. Jones.

A taxonomy of global optimization methods based on response surfaces.

Journal of Global Optimization, 21, 345–383, 2001.

Métamodèles probabilistes: un outil plus puissant

- $[Y(x)|Y(\mathbf{X}) = \mathbf{Y}]$ contient un approximateur, $f(x) = \mathbb{E}[Y(x)|Y(\mathbf{X}) = \mathbf{Y}]...$
- ...et beaucoup plus d'informations (ex: variance conditionnelle en chaque point!)
- Possibilité de construire des critères d'exploration 😊



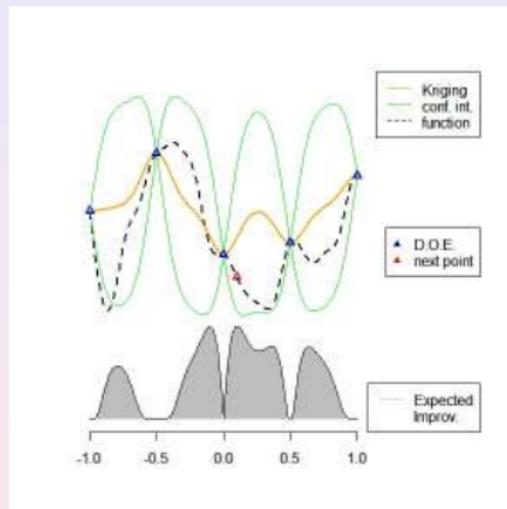
Le critère d'amélioration espérée (EI)

Expected Improvement

$$EI(\mathbf{x}) := \mathbb{E} [(\min(Y(\mathbf{X})) - Y(\mathbf{x}))^+ | Y(\mathbf{X}) = \mathbf{Y}]$$



M. Schonlau, W.J. Welch and D.R. Jones.
Efficient Global Optimization of Expensive
Black-box Functions
Journal of Global Optimization, 1998.



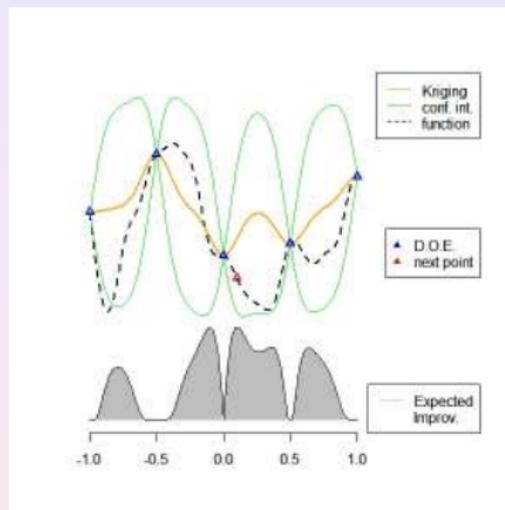
Le critère d'amélioration espérée (EI)

Expected Improvement

$$EI(\mathbf{x}) := \mathbb{E} [(\min(Y(\mathbf{X})) - Y(\mathbf{x}))^+ | Y(\mathbf{X}) = \mathbf{Y}]$$



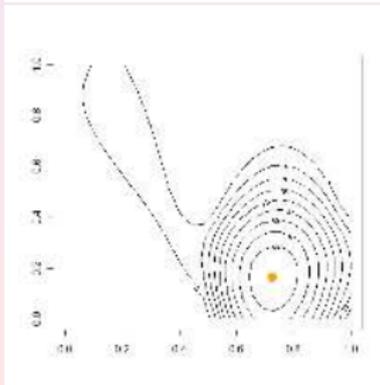
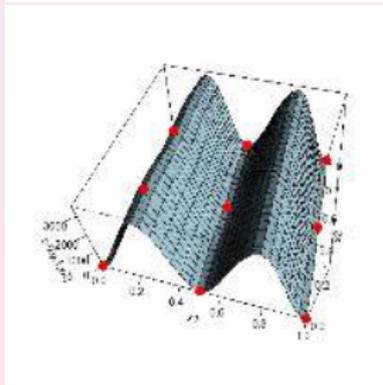
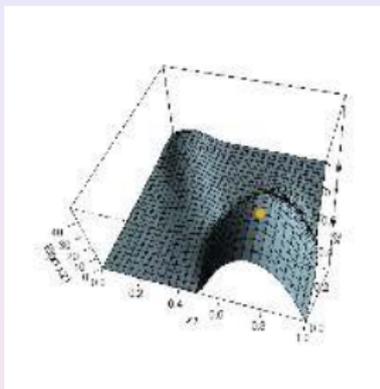
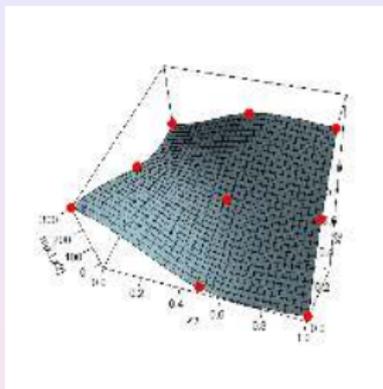
M. Schonlau, W.J. Welch and D.R. Jones.
Efficient Global Optimization of Expensive
Black-box Functions
Journal of Global Optimization, 1998.



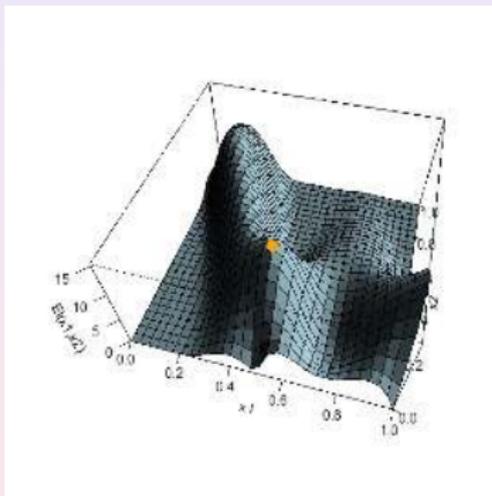
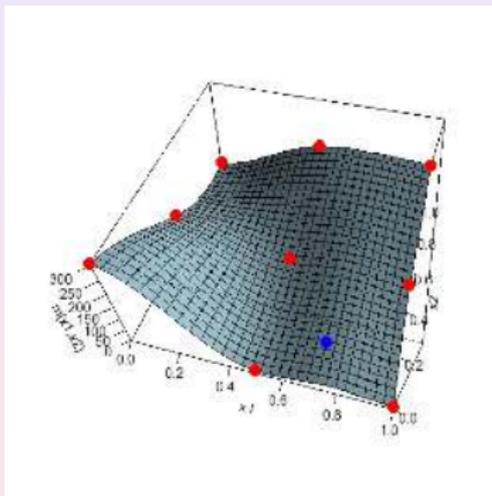
$$EI(\mathbf{x}) = (\min(y(\mathbf{X})) - m(\mathbf{x})) \Phi \left(\frac{\min(y(\mathbf{X})) - m(\mathbf{x})}{s(\mathbf{x})} \right) + s(\mathbf{x}) \phi \left(\frac{\min(y(\mathbf{X})) - m(\mathbf{x})}{s(\mathbf{x})} \right),$$

où Φ et ϕ sont respectivement la densité et la f.d.r. de la loi $\mathcal{N}(0, 1)$.

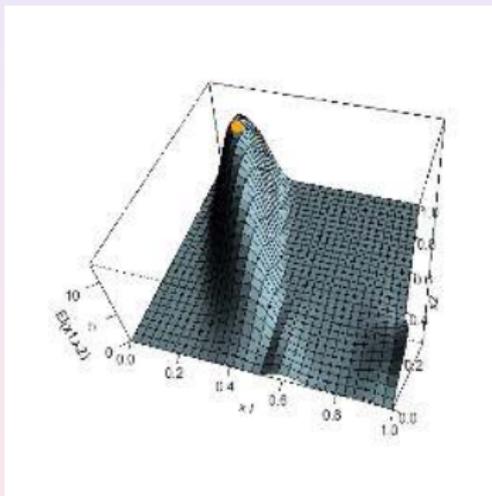
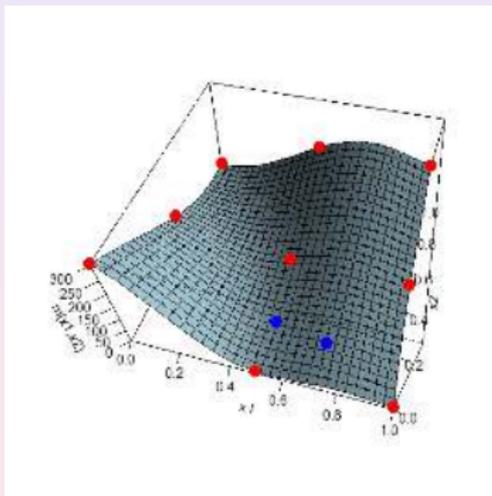
Optimisation sur base de Krigeage (Algorithme EGO)



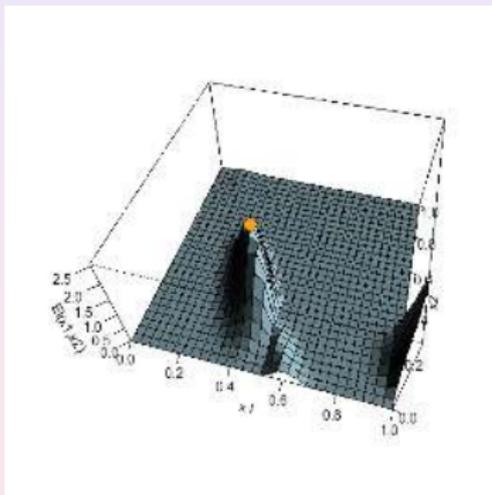
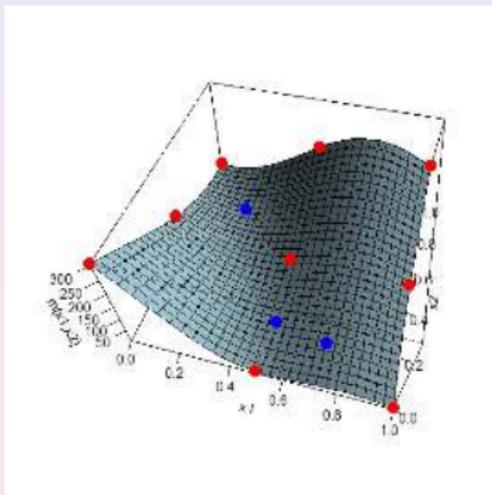
Optimisation sur base de Krigeage (Algorithme EGO)



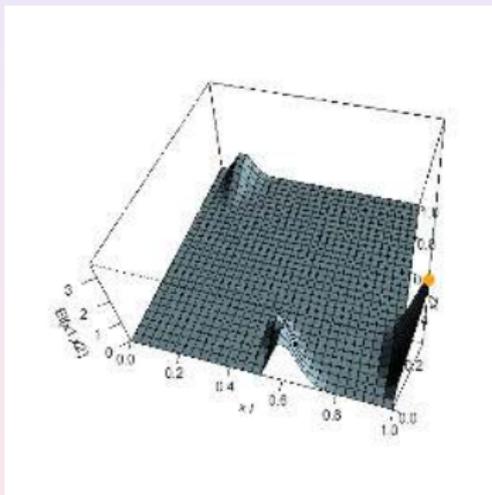
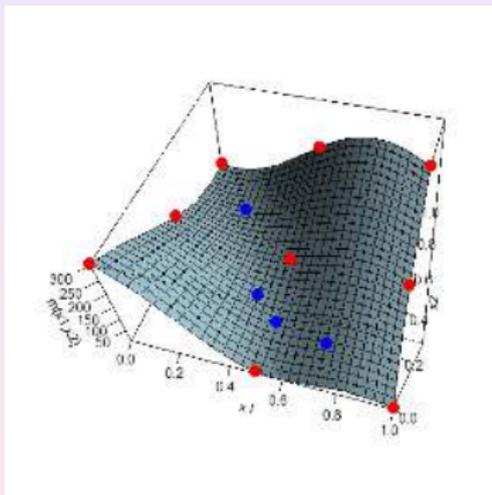
Optimisation sur base de Krigeage (Algorithme EGO)



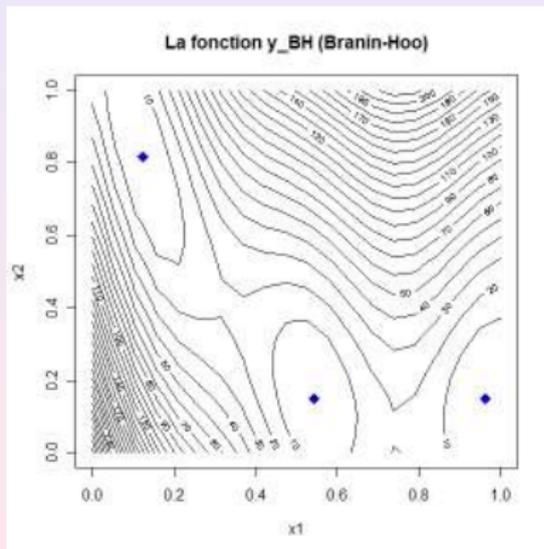
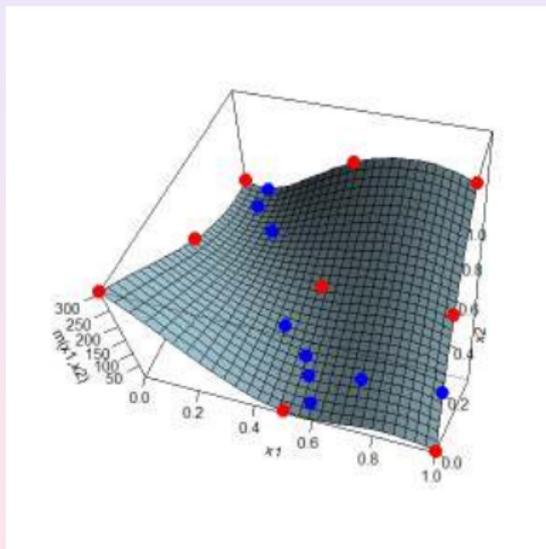
Optimisation sur base de Krigeage (Algorithme EGO)



Optimisation sur base de Krigeage (Algorithme EGO)



Optimisation sur base de Krigeage: résultats



Sommaire

- 1 Métamodèles de Krigeage pour l'approximation et l'optimisation
 - Bases du Krigeage et package DiceKriging
 - Expected Improvement et optimisation sur base de Krigeage
- 2 **Parallélisation(s) de l'optimisation sur base de Krigeage**
 - ***EI* multipoints et heuristiques parallèles**
 - DiceOptim. Application en dimension 6.
- 3 Métamodèles de Krigeage pour l'approximation et l'optimisation
 - Optimisation séquentielle en temps fini
 - Quelques extensions du modèle de Krigeage

Introduction au critère d'El *multipoints*

Point de départ: $\mathbf{x}^{n+1}, \dots, \mathbf{x}^{n+q}$ doivent être choisis de manière à ce que l'amélioration obtenue après évaluation de y aux q points soit maximale.

Par définition, on vise ainsi à ce qu'ils maximisent (*a posteriori*):

$$i(\mathbf{x}^{n+1}, \dots, \mathbf{x}^{n+q}) := \max\{[\min(y(\mathbf{X})) - y(\mathbf{x}^{n+1})]^+, \dots, [\min(y(\mathbf{X})) - y(\mathbf{x}^{n+q})]^+\} \\ = \left[\min(y(\mathbf{X})) - \min(y(\mathbf{x}^{n+1}), \dots, y(\mathbf{x}^{n+q})) \right]^+$$

Introduction au critère d'El *multipoints*

Point de départ: $\mathbf{x}^{n+1}, \dots, \mathbf{x}^{n+q}$ doivent être choisis de manière à ce que l'amélioration obtenue après évaluation de y aux q points soit maximale.

Par définition, on vise ainsi à ce qu'ils maximisent (*a posteriori*):

$$i(\mathbf{x}^{n+1}, \dots, \mathbf{x}^{n+q}) := \max\{[\min(y(\mathbf{X})) - y(\mathbf{x}^{n+1})]^+, \dots, [\min(y(\mathbf{X})) - y(\mathbf{x}^{n+q})]^+\} \\ = \left[\min(y(\mathbf{X})) - \min(y(\mathbf{x}^{n+1}), \dots, y(\mathbf{x}^{n+q})) \right]^+$$

Cela mène directement à l'El à q points:

Le critère q -El

$$\mathbb{E}I(\mathbf{x}^{n+1}, \dots, \mathbf{x}^{n+q}) := \\ \mathbb{E} \left[\left(\min(y(\mathbf{X})) - \min(Y(\mathbf{x}^{n+1}), \dots, Y(\mathbf{x}^{n+q})) \right)^+ \mid Y(\mathbf{X}) = \mathbf{Y} \right]$$

Remarque: Les $Y(\mathbf{x}^{n+i})$'s sont des v.a. **dependantes** (aussi $\mid Y(\mathbf{X}) = \mathbf{Y}$).

Illustrations de l'EI à 2 points

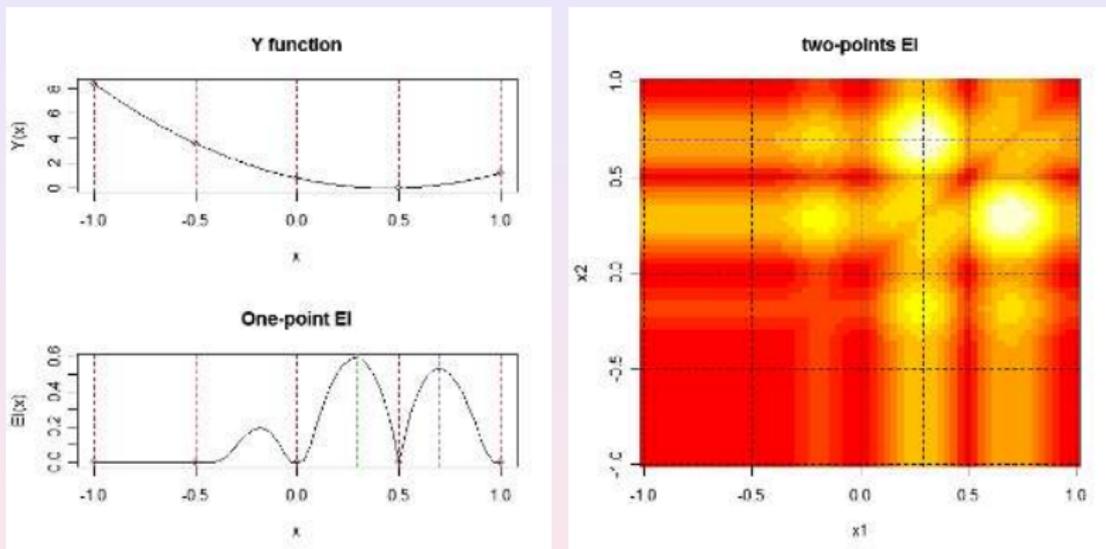


Figure: 1 et 2-EI associés à un polynôme de degré 2 en dim. 1

Le couple 2-EI-optimal est ici formé de 2 maxima locaux de l'EI à un point

Illustrations de l'EI à 2 points

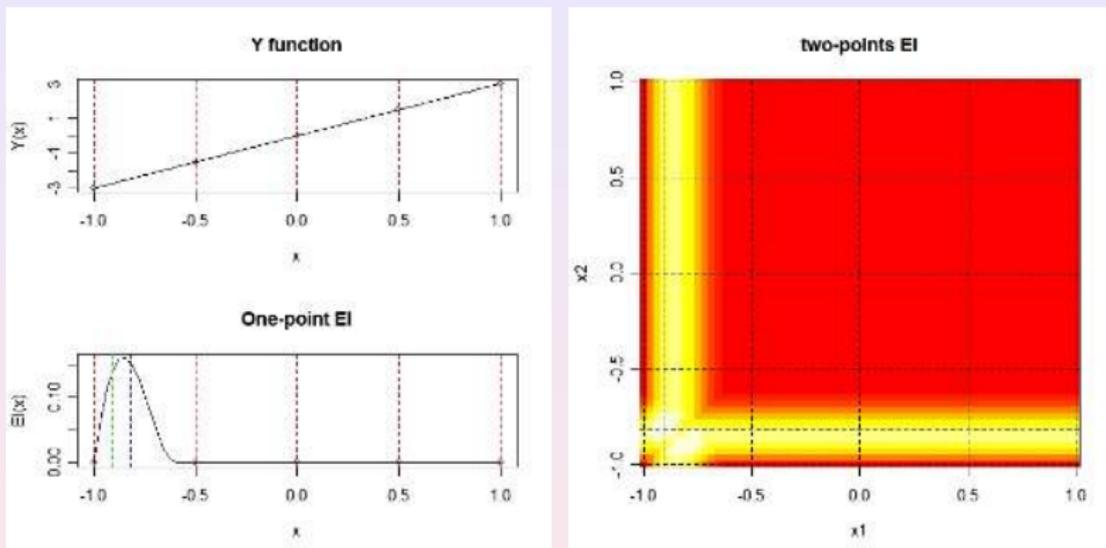


Figure: 1 et 2-EI associés à une application affine en dim. 1

Le couple 2-EI-optimal est maintenant formé de 2 points **autour** du maximiseur de l'EI.

Calcul du q-El par simulation stochastique

```
1: function Q-El(X, Y, Xnew)
2:    $L = \text{chol}(\text{Var}[Y(\mathbf{X}^{\text{new}}) | Y(\mathbf{X}) = \mathbf{Y}])$ 
3:   for  $i \leftarrow 1, n_{\text{sim}}$  do
4:      $N \sim \mathcal{N}(0, I_q)$ 
5:      $M_i = m_K(\mathbf{X}^{\text{new}}) + LN$ 
6:      $ql_{\text{sim}}(i) = [\min(\mathbf{Y}) - \min(M_i)]^+$ 
7:   end for
8:    $qEl_{\text{sim}} = \frac{1}{n_{\text{sim}}} \sum_{i=1}^{n_{\text{sim}}} ql_{\text{sim}}(i)$ 
9: end function
```

- ▷ Cholesky decomposition of $K_{\text{cond}}^{\text{new}}$
- ▷ Drawing a vector N at random
- ▷ Simulating Y at \mathbf{X}^{new}
- ▷ Simulating the improvement at \mathbf{X}^{new}
- ▷ Empirical Expected Improvement

Calcul du *q-El* par simulation stochastique

```

1: function q-El(X, Y, Xnew)
2:   L = chol(Var[Y(Xnew)|Y(X) = Y])
3:   for i ← 1, nsim do
4:     N ~  $\mathcal{N}(0, I_q)$ 
5:     Mi = mK(Xnew) + LN
6:     qIsim(i) = [min(Y) - min(Mi)]+
7:   end for
8:   qElsim =  $\frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} qI_{sim}(i)$ 
9: end function
    
```

- ▷ Cholesky decomposition of K_{cond}^{new}
- ▷ Drawing a vector N at random
- ▷ Simulating Y at \mathbf{X}^{new}
- ▷ Simulating the improvement at \mathbf{X}^{new}
- ▷ Empirical Expected Improvement

Résultats asymptotiques (Monte-Carlo)

$$qEl_{sim} = \sum_{i=1}^{n_{sim}} \frac{[\min(\mathbf{Y}) - \min(M_i)]^+}{n_{sim}} \xrightarrow[n_{sim} \rightarrow +\infty]{} El(\mathbf{x}^{n+1}, \dots, \mathbf{x}^{n+q}) \text{ a.s.}$$

$$\sqrt{n_{sim}} \left(\frac{qEl_{sim} - El(\mathbf{x}^{n+1}, \dots, \mathbf{x}^{n+q})}{\sqrt{\text{Var}[I(\mathbf{x}^{n+1}, \dots, \mathbf{x}^{n+q})]}} \right) \xrightarrow[n_{sim} \rightarrow +\infty]{} \mathcal{N}(0, 1) \text{ in law}$$



Plans d'expérience q -EI optimaux?

$$\mathbf{x}^{new*} = \operatorname{argmax}_{\mathbf{x}^{n+1}, \dots, \mathbf{x}^{n+q} \in D} \mathbb{E}I(\mathbf{x}^{n+1}, \dots, \mathbf{x}^{n+q})$$

L'optimisation est en dimension dq . Typiquement, $dq \geq 100$

Plans d'expérience q -EI optimaux?

$$\mathbf{x}^{new*} = \operatorname{argmax}_{\mathbf{x}^{n+1}, \dots, \mathbf{x}^{n+q} \in D} \mathbb{E}I(\mathbf{x}^{n+1}, \dots, \mathbf{x}^{n+q})$$

L'optimisation est en dimension dq . Typiquement, $dq \geq 100$

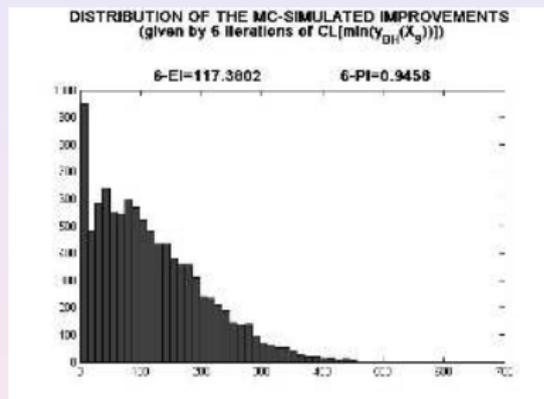
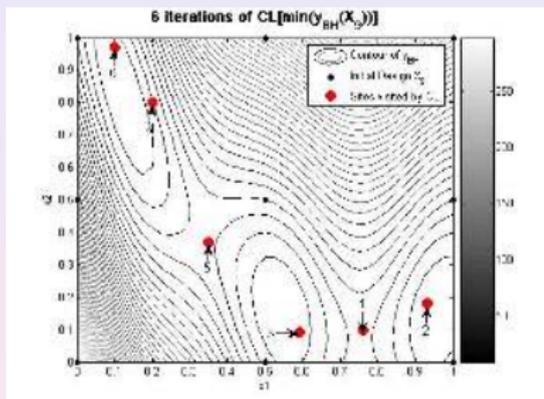
Constant Liar

Le Krigeage est renseigné avec un mensonge $L \in \mathbb{R}$ fixé par l'utilisateur:

```
1: function CL(X, Y,  $L$ ,  $q$ )
2:   for  $i \leftarrow 1, q$  do
3:      $\mathbf{x}^{n+i} = \operatorname{argmax}_{\mathbf{x} \in D} \mathbb{E}I(\mathbf{x})$  ▷ with updated X and Y
4:      $\mathbf{X} = \mathbf{X} \cup \{\mathbf{x}^{n+i}\}$ 
5:      $\mathbf{Y} = \mathbf{Y} \cup \{L\}$ 
6:   end for
7: end function
```

La constante L permet de contrôler la répulsion créée par les points virtuellement visités ($L = \max(\mathbf{Y})$ pour une répulsion forte, $L = \min(\mathbf{Y})$ pour une répulsion modérée)

Utilisation du q-El pour filtrer les heuristiques



Left: Fonction de Branin-Hoo avec plan \mathbf{X}_9 (petits points noirs) et les 6 premiers points donnés par la stratégie $CL[\min(f_{BH}(\mathbf{X}_9))]$ (cercle).

Right: Histogramme de 10^3 valeurs de l'amélioration associée aux 6-points donnée par la stratégie $CL[\min(f_{BH}(\mathbf{X}_9))]$, obtenus par simulation Monte Carlo. Les valeurs de PI et El correspondantes sont données au-dessus.

Sommaire

- 1 Métamodèles de Krigeage pour l'approximation et l'optimisation
 - Bases du Krigeage et package DiceKriging
 - Expected Improvement et optimisation sur base de Krigeage
- 2 **Parallélisation(s) de l'optimisation sur base de Krigeage**
 - *EI* multipoints et heuristiques parallèles
 - **DiceOptim. Application en dimension 6.**
- 3 Métamodèles de Krigeage pour l'approximation et l'optimisation
 - Optimisation séquentielle en temps fini
 - Quelques extensions du modèle de Krigeage

Exemple de fonction objectif en 6D: "Hartmann 6"

$$y(x) = - \sum_{j=1}^4 c_j \times \exp \left(- \sum_{i=1}^6 a_{i,j} \times (x_i - p_{i,j})^2 \right)$$

$$a = \begin{pmatrix} 10.00 & 0.05 & 3.00 & 17.00 \\ 3.00 & 10.00 & 3.50 & 8.00 \\ 17.00 & 17.00 & 1.70 & 0.05 \\ 3.50 & 0.10 & 10.00 & 10.00 \\ 1.70 & 8.00 & 17.00 & 0.10 \\ 8.00 & 14.00 & 8.00 & 14.00 \end{pmatrix} \quad p = \begin{pmatrix} 0.1312 & 0.2329 & 0.2348 & 0.4047 \\ 0.1696 & 0.4135 & 0.1451 & 0.8828 \\ 0.5569 & 0.8307 & 0.3522 & 0.8732 \\ 0.0124 & 0.3736 & 0.2883 & 0.5743 \\ 0.8283 & 0.1004 & 0.3047 & 0.1091 \\ 0.5886 & 0.9991 & 0.6650 & 0.0381 \end{pmatrix} \quad c = \begin{pmatrix} 1.0 \\ 1.2 \\ 3.0 \\ 3.2 \end{pmatrix}$$

minimum global = **-3.32**

minimiseur global = [0.202, 0.150, 0.477, 0.275, 0.312, 0.657]

20 itérations en partant d'un plan à 50 points

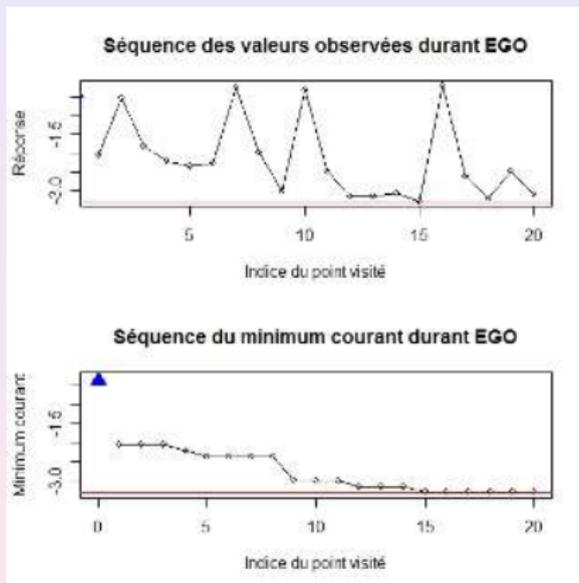


Figure: En démarrant avec un plan à 50 points, l'optimum global est atteint en 15 itérations.

20 itérations en partant d'un plan à 50 points

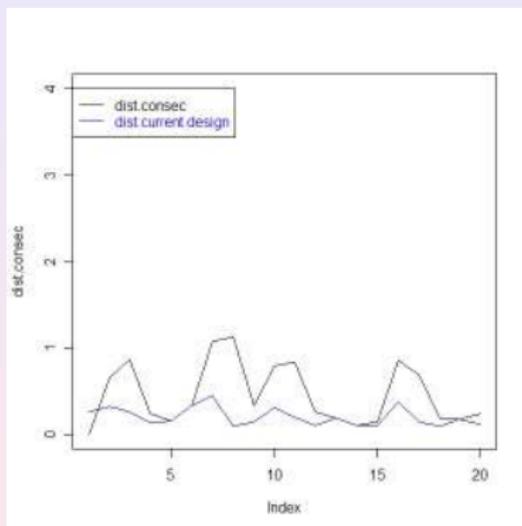
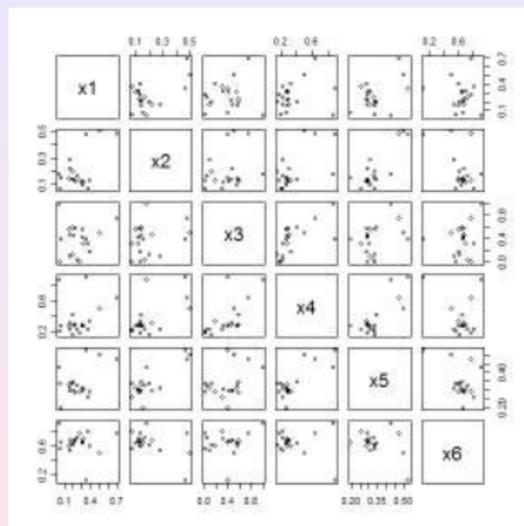


Figure: EGO visite séquentiellement la zone du minimum global de Hartmann. Il exploite l'information donnée par le plan initial.

90 itérations en partant d'un plan à 10 points

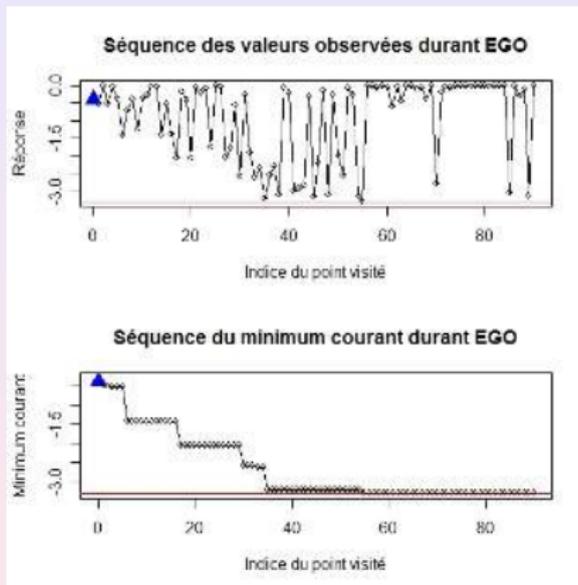


Figure: EGO trouve le minimum en 36 itérations. C'est peu si l'on considère que le plan initial ne compte que 10 points.

90 itérations en partant d'un plan à 10 points

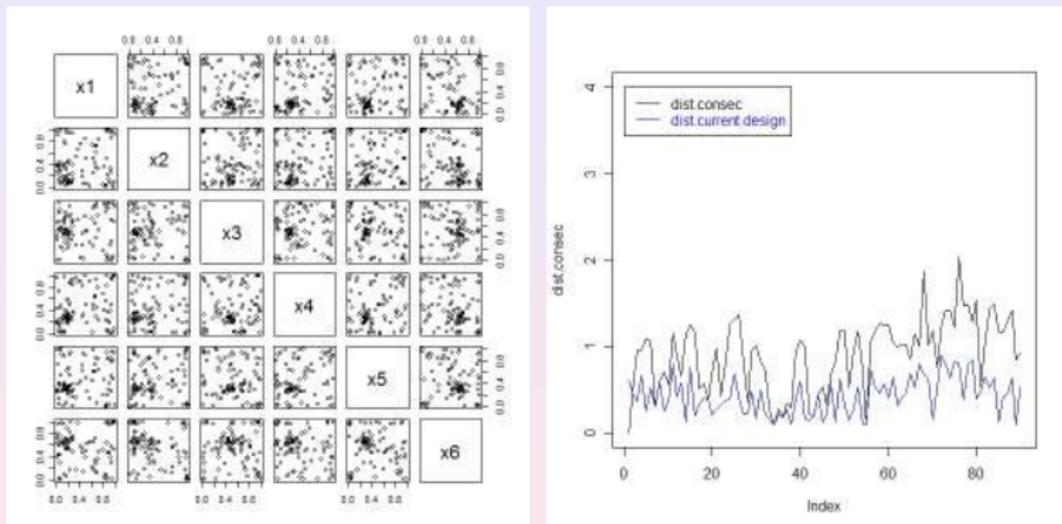


Figure: L'algorithme a un comportement nettement plus exploratoire qu'en partant de 50 points. On pourrait penser que Hartmann est facile à optimiser. *Essayons en tirant 500 points au hasard...*

Minimisation approchée par 500 tirages Monte-Carlo

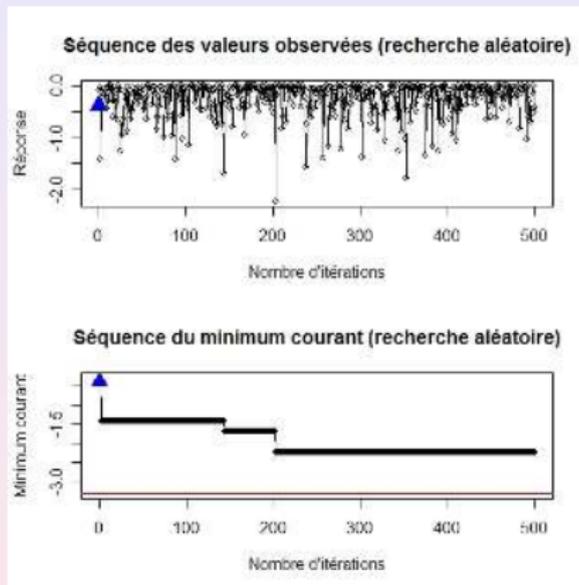


Figure: On ne trouve ici pas mieux que -2.2 en 500 itérations, alors que le minimum est à -3.32 .

Optimiser en alternant CL et évaluations parallèles

Idee: optimisation parallèle synchrone en utilisant CL à chaque itération pour obtenir les points d'explorations. Les mensonges du CL sont corrigés à la fin de chaque itération, après les évaluations parallèles du simulateur.

Constant Liar à plusieurs périodes

```
1: function CLMIN.STAGES(X, Y,  $y$ ,  $n_{proc}$ ,  $n_{init}$ )
2:   for  $i \leftarrow 1, n_{init}$  do
3:      $L = \min(\mathbf{Y})$ 
4:     for  $j \leftarrow 1, n_{proc}$  do
5:        $\mathbf{x}^{n+j} = \operatorname{argmax}_{\mathbf{x} \in D} \mathbb{E}I(\mathbf{x})$  ▷ avec  $\mathbf{X}$  and  $\mathbf{Y}_{CL}$ 
6:        $\mathbf{X} = \mathbf{X} \cup \{\mathbf{x}^{n+j}\}$ 
7:        $\mathbf{Y}_{CL} = \mathbf{Y} \cup \{L\}$ 
8:     end for
9:      $\mathbf{Y} = \mathbf{Y} \cup y(\{\mathbf{x}^{n+j}\}_{1 \leq j \leq n_{proc}})$  ▷ Evaluations du simulateur
10:    Ré-estimation du Krigage
11:  end for
12: end function
```

CL avec 10 proc. en parallèle, 50 points initiaux

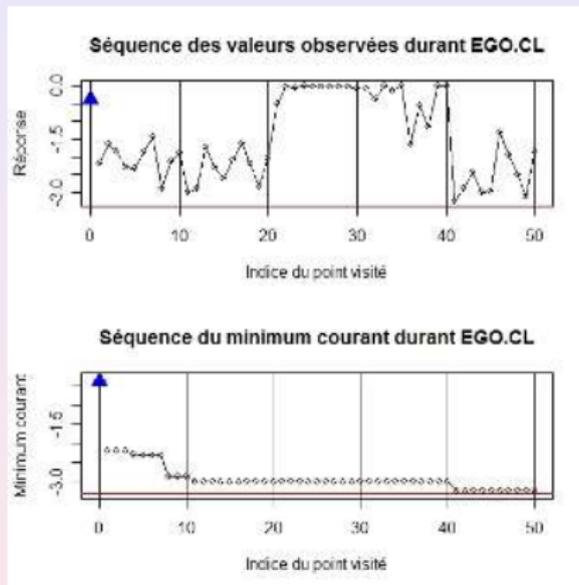


Figure: On peut encore réduire les délais: avec 10 processeurs en parallèle, le minimum est atteint en 5 unités de temps.

CL avec 10 proc. en parallèle, 50 points initiaux

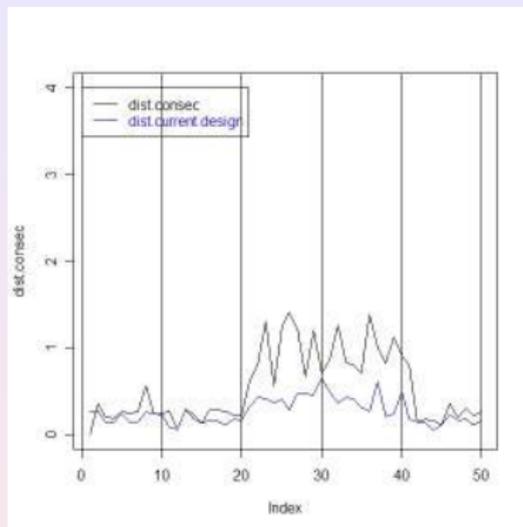
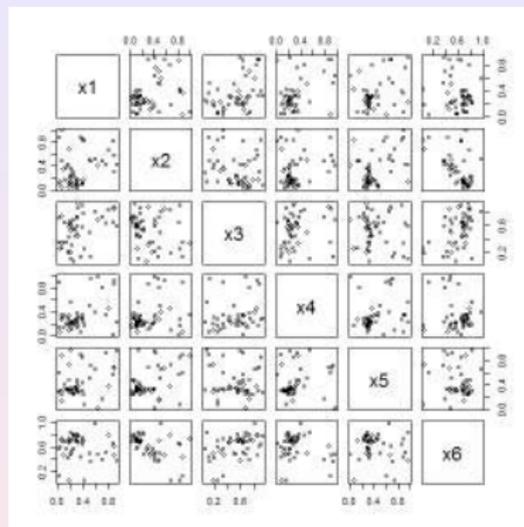


Figure: L'algorithme alterne ici entre une première phase d'exploitation (deux premières unités de temps), une phase plus exploratoire (3^{ème} et 4^{ème} unités de temps), puis une phase finale d'exploitation durant laquelle il trouve le minimum.

CLmin avec 10 proc. en parallèle, 10 points initiaux

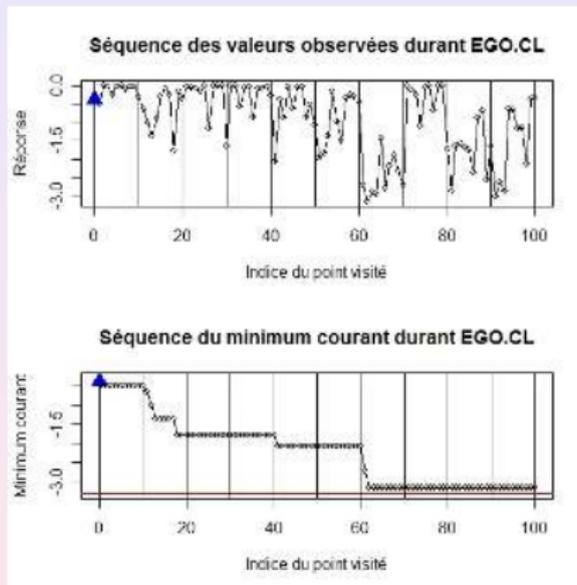
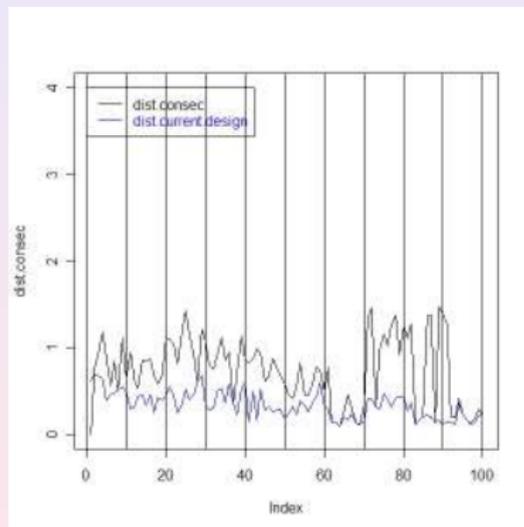
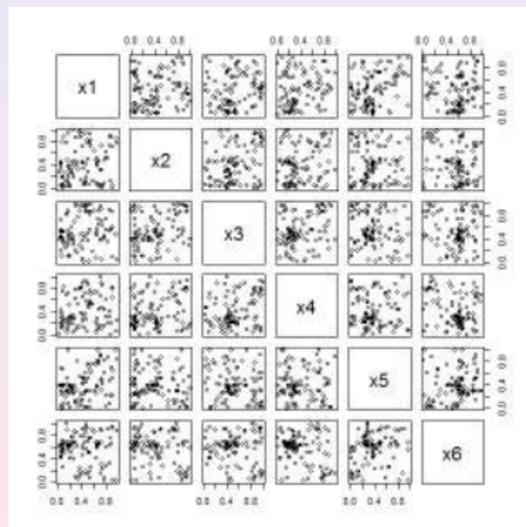


Figure: En partant d'un plan à 10 points, CL_{min} avec 10 processeurs permet de trouver le minimum en 7 unités de temps.

CL avec 10 proc. en parallèle, 10 points initiaux



Conclusions et perspectives sur le q -EI

Retour experimental

- Le calcul de critères multipoints sur base de Krigeage par Monte-Carlo est abordable **quelque soit la dimension** de l'espace des variables
- La stratégie du Constant Liar a donné des résultats très prometteurs sur les cas test en dimensions 1 à 6.

Conclusions et perspectives sur le q -EI

Retour experimental

- Le calcul de critères multipoints sur base de Krigeage par Monte-Carlo est abordable **quelque soit la dimension** de l'espace des variables
- La stratégie du Constant Liar a donné des résultats très prometteurs sur les cas test en dimensions 1 à 6.

Pistes de travaux futurs

- Appliquer ces outils à des problèmes industriels de plus grande dimension (à l'étude sur le simulateur MORET)
- Optimisation du q -EI (mutation de bons plans d'expériences candidats?)
- Applications en optimisation en temps fini.

Sommaire

- 1 Métamodèles de Krigeage pour l'approximation et l'optimisation
 - Bases du Krigeage et package DiceKriging
 - Expected Improvement et optimisation sur base de Krigeage
- 2 Parallélisation(s) de l'optimisation sur base de Krigeage
 - *EI* multipoints et heuristiques parallèles
 - DiceOptim. Application en dimension 6.
- 3 Derniers développements en optimisation et approximation
 - Optimisation séquentielle en temps fini
 - Quelques extensions du modèle de Krigeage

Optimisation en temps fini (avec R. Le Riche)

Intuition: il n'y a pas de raison que maximiser l'*expected improvement* à chaque itération d'une stratégie séquentielle soit la meilleure chose à faire ...

Optimisation en temps fini (avec R. Le Riche)

Intuition: il n'y a pas de raison que maximiser l'*expected improvement* à chaque itération d'une stratégie séquentielle soit la meilleure chose à faire . . .

But: montrer que le comportement optimal dépend du crédit d'itérations, et que la "valeur" d'un point ne se limite pas au gain immédiat qu'il occasionne.

Optimisation en temps fini (avec R. Le Riche)

Intuition: il n'y a pas de raison que maximiser l'*expected improvement* à chaque itération d'une stratégie séquentielle soit la meilleure chose à faire . . .

But: montrer que le comportement optimal dépend du crédit d'itérations, et que la "valeur" d'un point ne se limite pas au gain immédiat qu'il occasionne.

Moyen: utiliser le q -*EI* comme critère de référence en traitant les points et les réponses à venir comme des variables aléatoires.

Optimisation en temps fini: notations

n_0 désigne le nb d'expériences initiales, $(\mathbf{X}^0, \mathbf{Y}^0)$ le plan et les observations associées, et \mathbf{x}^{n_0+j} ($j \in [1, r]$) les points choisis au fil des itérations.

- \mathbf{X}^{n_0+j} et \mathbf{Y}^{n_0+j} désignent respectivement le plan \mathbf{X}^0 et les observations \mathbf{Y}^0 augmentés des \mathbf{x}^{n_0+i} et $y(\mathbf{x}^{n_0+i})$ associés ($1 \leq i \leq j$, $j \in [1, r]$),
- Les \mathcal{X}^{n_0+j} sont les versions aléatoires des \mathbf{x}^{n_0+j} ($j \in [1, r]$)
- Les \mathbb{X}^{n_0+j} et \mathbb{Y}^{n_0+j} sont les versions aléatoires des \mathbf{X}^{n_0+j} et \mathbf{Y}^{n_0+j} :
 - $\mathbb{X}^{n_0+j} = \mathbf{X}^0 \cup \{\mathcal{X}^{n_0+1}, \dots, \mathcal{X}^{n_0+j}\}$
 - $\mathbb{Y}^{n_0+j} = \mathbf{Y}^0 \cup \{Y(\mathcal{X}^{n_0+1}), \dots, Y(\mathcal{X}^{n_0+j})\}$

Optimisation en temps fini: stratégies déterministes

Definition

On appelle *stratégie déterministe à horizon r* ($r \in \mathbb{N}^*$) toute suite finie $\mathcal{S} = (s_j)_{j \in [1, r]}$ de fonctions mesurables

$$s_j : (D \times \mathbb{R})^{n_0 + j - 1} \longrightarrow D$$

Definition

L'amélioration donnée (*in fine*) par la stratégie $\mathcal{S} = (s_j)_{j \in [1, r]}$ est

$$\begin{aligned} i(\mathcal{S}) &= (\min(y(\mathbf{X}^0)) - \min(y(\mathbf{x}^{n_0+1}), \dots, y(\mathbf{x}^{n_0+r})))^+ \\ &= i(s_1(\mathbf{X}^0, \mathbf{Y}^0), \dots, s_r(\mathbf{X}^{r-1}, \mathbf{Y}^{r-1})) \end{aligned}$$

Optimisation en temps fini: stratégies déterministes

Definition

On appelle *stratégie déterministe à horizon r* ($r \in \mathbb{N}^*$) toute suite finie $\mathcal{S} = (s_j)_{j \in [1, r]}$ de fonctions mesurables

$$s_j : (D \times \mathbb{R})^{n_0+j-1} \longrightarrow D$$

Definition

L'amélioration donnée (*in fine*) par la stratégie $\mathcal{S} = (s_j)_{j \in [1, r]}$ est

$$\begin{aligned} i(\mathcal{S}) &= (\min(y(\mathbf{X}^0)) - \min(y(\mathbf{x}^{n_0+1}), \dots, y(\mathbf{x}^{n_0+r})))^+ \\ &= i(s_1(\mathbf{X}^0, \mathbf{Y}^0), \dots, s_r(\mathbf{X}^{r-1}, \mathbf{Y}^{r-1})) \end{aligned}$$

Rq: on aimerait maximiser i ... mais on ne peut l'évaluer qu'*a posteriori*!

Optimisation en temps fini: EI d'une stratégie

Definition

L'amélioration espérée d'une stratégie $\mathcal{S} = (s_j)_{j \in [1, r]}$ (vue de 0) est donné par

$$\begin{aligned} EI(\mathcal{S}) &= \mathbb{E} \left[(\min(Y(\mathbf{X}^0)) - \min(Y(s_1(\mathbf{X}, \mathbf{Y})), \dots, Y(s_r(\mathbb{X}^{r-1}, \mathbb{Y}^{r-1}))))^+ | A_0 \right] \\ &= \mathbb{E} \left[I(s_1(\mathbf{X}^0, \mathbf{Y}^0), s_2(\mathbb{X}^1, \mathbb{Y}^1), \dots, s_r(\mathbb{X}^{r-1}, \mathbb{Y}^{r-1})) | A_0 \right], \end{aligned} \tag{1}$$

où A_0 est l'évènement " $Y(\mathbf{X}^0) = \mathbf{Y}^0$ ".

Definition

On note \mathcal{P}_r le problème: trouver $\mathcal{S}^* = (s_j^*)_{j \in [1, r]}$ maximisant $EI(\cdot)$

Optimisation en temps fini: stratégie optimale

Theorem

Dans \mathcal{P}_r , choisir \mathbf{x}^{n_0+r} après $r - 1$ itérations revient à maximiser $\mathbb{E}[I(\cdot)|A_{r-1}]$

Optimisation en temps fini: stratégie optimale

Theorem

Dans \mathcal{P}_r , choisir \mathbf{x}^{n_0+r} après $r - 1$ itérations revient à maximiser $\mathbb{E}[I(\cdot)|A_{r-1}]$

Corollary

La solution $S^* = (s_1^*, \dots, s_r^*)$ de \mathcal{P}_r est donnée par la récursion:

$$\begin{cases} \mathbf{x}^{n_0+r} = s_r^*(\mathbf{X}^{r-1}, \mathbf{Y}^{r-1}) = \arg \max_{\mathbf{x} \in D} \mathbb{E}[I(\mathbf{x})|A_{r-1}] \\ \mathbf{x}^{n_0+r-1} = s_{r-1}^*(\mathbf{X}^{r-2}, \mathbf{Y}^{r-2}) = \arg \max_{\mathbf{x} \in D} \mathbb{E}[I(\mathbf{x}, s_r^*(\mathbf{X}^{r-1}(\mathbf{x}), \mathbf{Y}^{r-1}(\mathbf{x})))|A_{r-2}] \\ \dots \\ \mathbf{x}^{n_0+1} = s_1^*(\mathbf{X}^0, \mathbf{Y}^0) = \arg \max_{\mathbf{x} \in D} \mathbb{E}[I(\mathbf{x}, s_1^*(\mathbf{X}^1(\mathbf{x}), \mathbf{Y}^1(\mathbf{x})), \dots, s_r^*(\mathbf{X}^{r-1}(\mathbf{x}), \mathbf{Y}^{r-1}(\mathbf{x})))|A_0] \end{cases}$$

Optimisation en temps fini: exemple à deux points

Stratégie proposée

$\mathcal{S}(\mathbf{a})$ = choisir \mathbf{a} à la 1^{ère} itération, puis maximiser le "1-point EL " à la 2^{nde}:

$$\mathcal{X}^{n_0+2} = \mathcal{S}_2(\mathbb{X}^1, \mathbb{Y}^1) = \arg \max_{\mathbf{x} \in D} \mathbb{E} \left[\left(\min(\mathbb{Y}^1) - Y(\mathbf{x}) \right)^+ \mid A_0, Y(\mathbf{a}) \right]$$

Optimisation en temps fini: exemple à deux points

Stratégie proposée

$\mathcal{S}(\mathbf{a})$ = choisir \mathbf{a} à la 1^{ère} itération, puis maximiser le "1-point EI " à la 2^{nde}:

$$\mathcal{X}^{n_0+2} = \mathcal{s}_2(\mathbb{X}^1, \mathbb{Y}^1) = \arg \max_{\mathbf{x} \in D} \mathbb{E} \left[\left(\min(\mathbb{Y}^1) - Y(\mathbf{x}) \right)^+ \mid A_0, Y(\mathbf{a}) \right]$$

On peut alors développer l' EI de $\mathcal{S}(\mathbf{a})$ comme suit:

$$\begin{aligned} EI(\mathcal{S}(\mathbf{a})) &= \mathbb{E} \left[\left(\min(\mathbb{Y}^0) - \min(Y(\mathbf{a}), Y(\mathcal{X}^{n_0+2})) \right)^+ \mid A_0 \right] \\ &= \underbrace{\mathbb{E} \left[\left(\min(\mathbb{Y}^0) - Y(\mathbf{a}) \right)^+ \mid A_0 \right]}_{EI(\mathbf{a}) = EI_{0,1}^0(\mathbf{a})} + \underbrace{\mathbb{E} \left[\left(\min(\mathbb{Y}^1) - Y(\mathcal{X}^{n_0+2}) \right)^+ \mid A_0 \right]}_{EI_{1,2}^0(\mathbf{a})} \end{aligned}$$

Optimisation en temps fini: exemple à deux points

Stratégie proposée

$\mathcal{S}(\mathbf{a})$ = choisir \mathbf{a} à la 1^{ère} itération, puis maximiser le "1-point El " à la 2^{nde}:

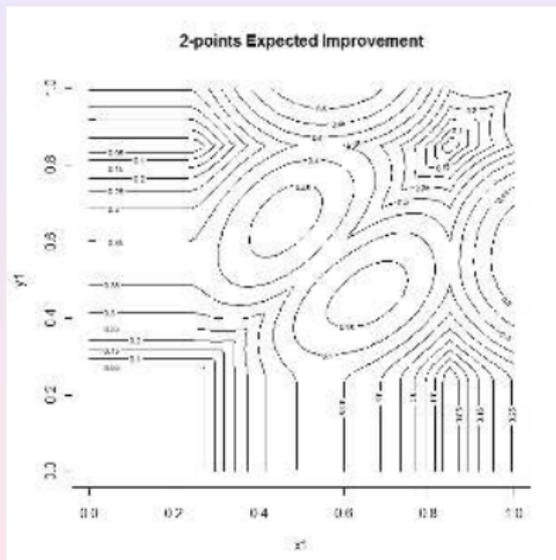
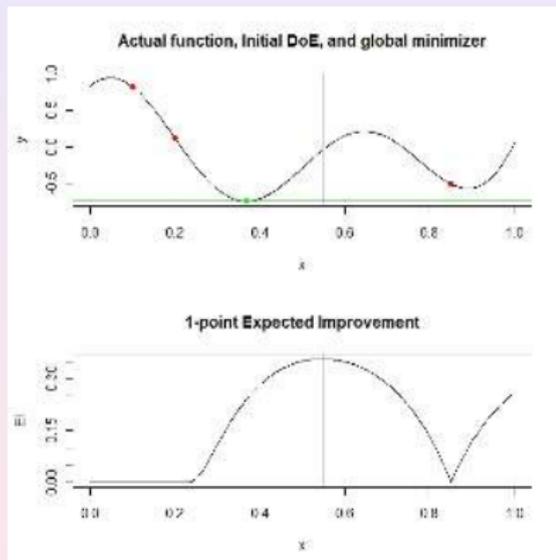
$$\mathcal{X}^{n_0+2} = \mathbf{s}_2(\mathbb{X}^1, \mathbb{Y}^1) = \arg \max_{\mathbf{x} \in D} \mathbb{E} \left[\left(\min(\mathbb{Y}^1) - Y(\mathbf{x}) \right)^+ \mid A_0, Y(\mathbf{a}) \right]$$

On peut alors développer l' El de $\mathcal{S}(\mathbf{a})$ comme suit:

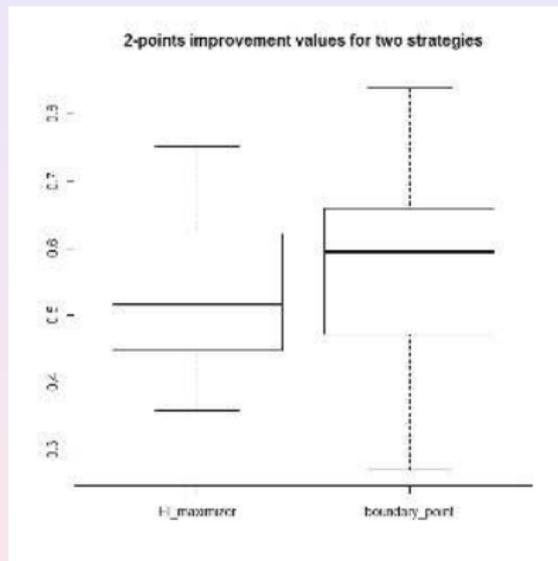
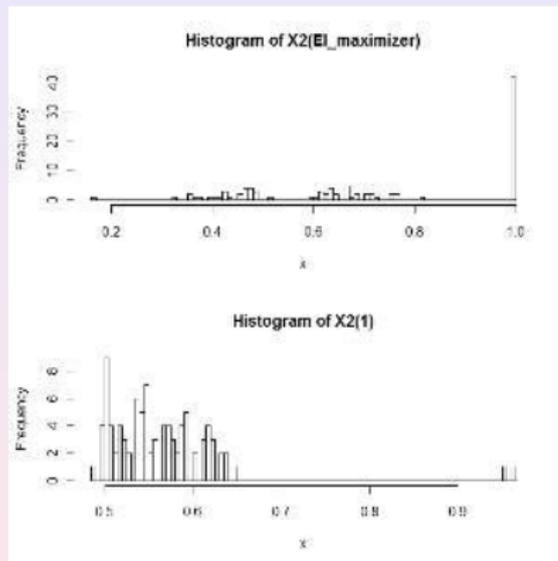
$$\begin{aligned} El(\mathcal{S}(\mathbf{a})) &= \mathbb{E} \left[\left(\min(\mathbb{Y}^0) - \min(Y(\mathbf{a}), Y(\mathcal{X}^{n_0+2})) \right)^+ \mid A_0 \right] \\ &= \underbrace{\mathbb{E} \left[\left(\min(\mathbb{Y}^0) - Y(\mathbf{a}) \right)^+ \mid A_0 \right]}_{El(\mathbf{a}) = El_{0,1}^0(\mathbf{a})} + \underbrace{\mathbb{E} \left[\left(\min(\mathbb{Y}^1) - Y(\mathcal{X}^{n_0+2}) \right)^+ \mid A_0 \right]}_{El_{1,2}^0(\mathbf{a})} \end{aligned}$$

$El_{1,2}^0(\mathbf{a}) \approx \frac{1}{k} \sum_{i=1}^k \mathbb{E} \left[\left(\min(\mathbb{Y}^1(\mathbf{a})) - Y(\mathcal{X}^{n_0+2}(\mathbf{a})) \right)^+ \mid A_0, Y(\mathbf{a}) = y_a^i \right]$ avec
 $y_a^i \sim \mathcal{N}(m_0(\mathbf{a}), \mathbf{s}_0^2(\mathbf{a}))$ indépendamment.

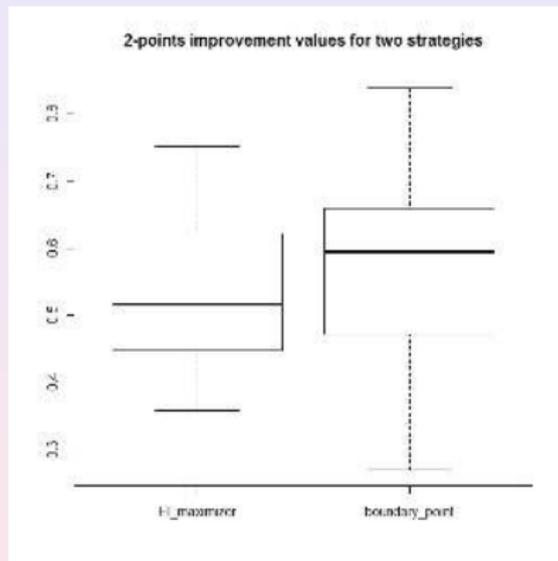
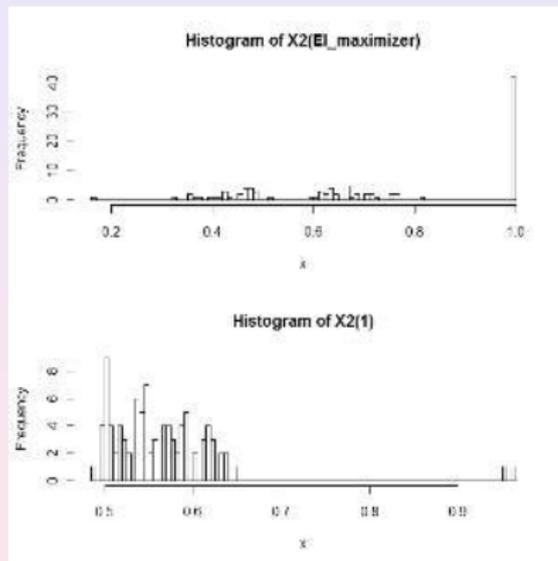
Optimisation en temps fini: exemple à deux points



Optimisation en temps fini: exemple à deux points



Optimisation en temps fini: exemple à deux points



D. Ginsbourger and R. Le Riche.

Towards GP-based optimization with finite time horizon.

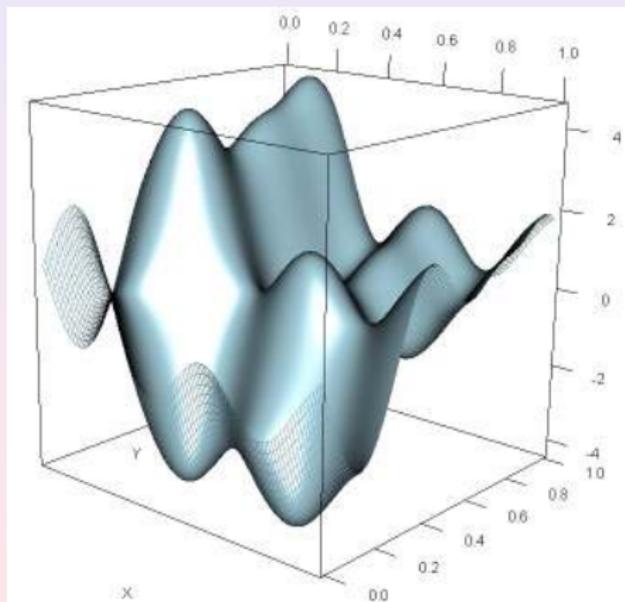
Proceedings of the 9th International Workshop in Model-Oriented Design and Analysis, Submitted 2010.

Sommaire

- 1 Métamodèles de Krigeage pour l'approximation et l'optimisation
 - Bases du Krigeage et package DiceKriging
 - Expected Improvement et optimisation sur base de Krigeage
- 2 Parallélisation(s) de l'optimisation sur base de Krigeage
 - *EI* multipoints et heuristiques parallèles
 - DiceOptim. Application en dimension 6.
- 3 Derniers développements en optimisation et approximation
 - Optimisation séquentielle en temps fini
 - Quelques extensions du modèle de Krigeage

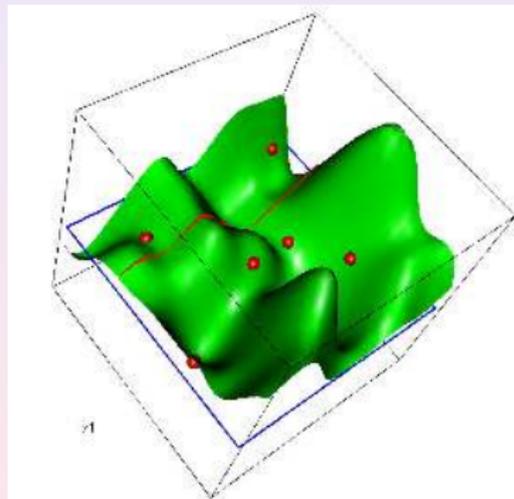
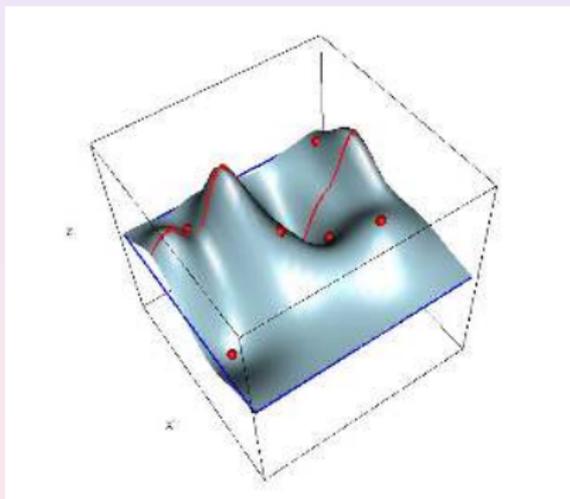
Krigeage additif

Thèse de Nicolas Durrande (EMSE, soutenance prévue: Automne 2011)



Krigeage "sous contraintes fonctionnelles"

Thèse de Bertrand Gauthier (EMSE, soutenance prévue: Automne 2010)



Plans d'expériences pour simulateurs stochastiques

Thèse de Victor Picheny (EMSE & UFL, soutenance: 15 Octobre 2009)

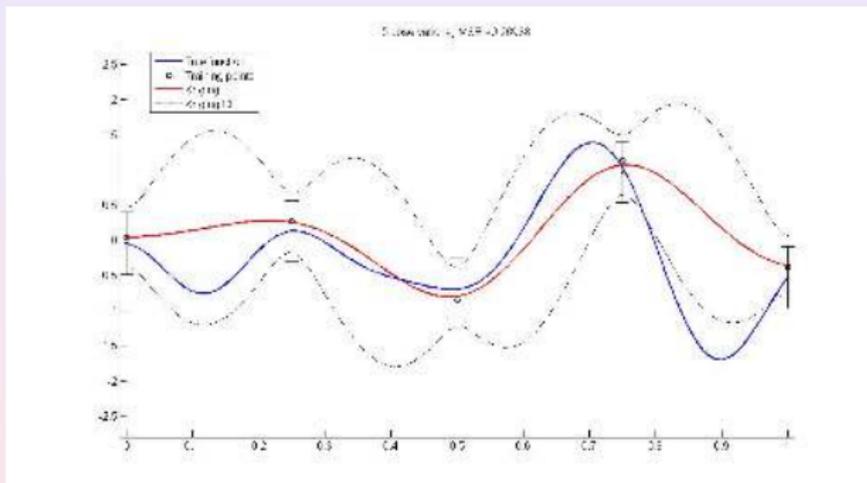


Figure: Un premier plan d'expériences "variables-précision".
Les barres indiquent la précision de chaque observation.

Plans d'expériences pour simulateurs stochastiques

Thèse de Victor Picheny (EMSE & UFL, soutenance: 15 Octobre 2009)

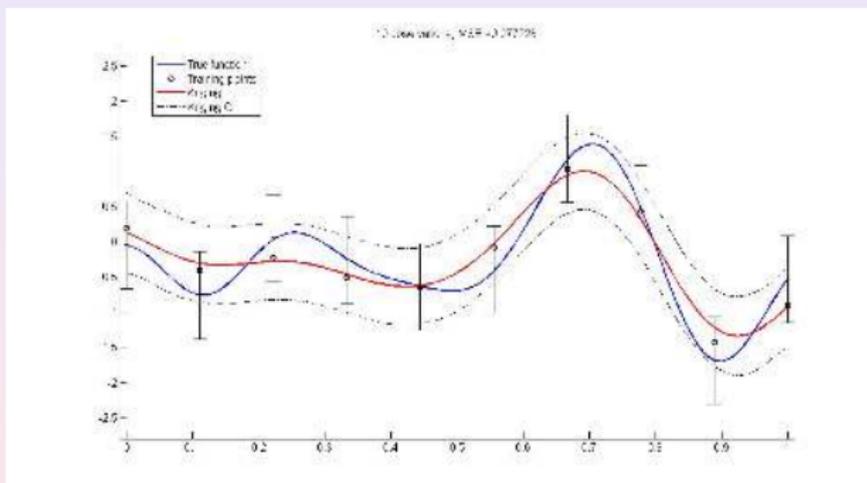


Figure: Un autre plan d'expériences dans l'espace "variables-précision", avec le même budget total ($\sum_1^n N_j = C$).

Plans d'expériences pour simulateurs stochastiques

Thèse de Victor Picheny (EMSE & UFL, soutenance: 15 Octobre 2009)

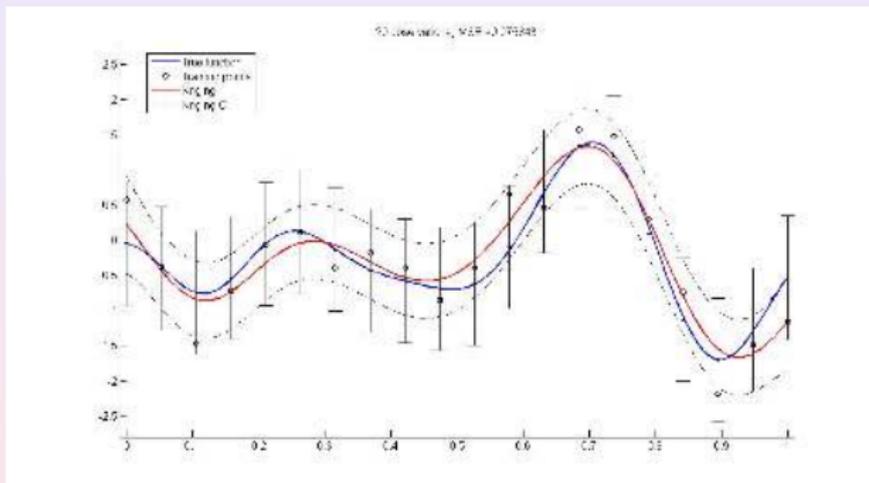


Figure: Un grand nombre d'observations très bruitées semble meilleur que quelques observations précises ...

Merci pour votre attention!



Des questions?