Plans d'expériences et codes numériques

Luc Pronzato

Laboratoire I3S,
CNRS/Univ. Nice Sophia Antipolis, France

Plan

- I) Quel objectif ?

Plan

- I) Quel objectif ?
- II) Non séquentiel ou séquentiel/adaptatif ?

Plan

- I) Quel objectif ?
- II) Non séquentiel ou séquentiel/adaptatif ?
- III) Quel modèle ?
  (paramétrique/non paramétrique – structure fixe ou variable)

Plan

- I) Quel objectif ?
- II) Non séquentiel ou séquentiel/adaptatif ?
- III) Quel modèle ?
  (paramétrique/non paramétrique – structure fixe ou variable)
- IV) Quelques problèmes. . .

Université Nice Sophia Antipolis

CNRS Centre National de la Recherche Scientifique

I) Quel objectif ?

$f(x)$ une fonction (inconnue) définie pour $x \in \mathcal{X} \subset \mathbb{R}^d$

Á partir de couples $X_i, f(X_i)$, $i = 1, 2 \ldots$ on va construire un modèle $\eta(\cdot)$ de $f(\cdot)$ dans un but

- d'exploration : $\eta(\cdot)$ doit approximer $f(\cdot)$ le mieux possible sur $\mathcal{X}$

I) Quel objectif ?

$f(x)$ une fonction (inconnue) définie pour $x \in \mathcal{X} \subset \mathbb{R}^d$

Á partir de couples $X_i, f(X_i)$, $i = 1, 2 \ldots$ on va construire un modèle $\eta(\cdot)$ de $f(\cdot)$ dans un but

- d'exploration : $\eta(\cdot)$ doit approximer $f(\cdot)$ le mieux possible sur $\mathcal{X}$
- d'optimisation :
  $\rightarrow$ utiliser $\eta(\cdot)$ pour déterminer $\arg_{x \in \mathcal{X}} f(x)$

I) Quel objectif ?

$f(x)$ une fonction (inconnue) définie pour $x \in \mathcal{X} \subset \mathbb{R}^d$

Á partir de couples $X_i, f(X_i)$, $i = 1, 2 \ldots$ on va construire un modèle $\eta(\cdot)$ de $f(\cdot)$ dans un but

- d'exploration : $\eta(\cdot)$ doit approximer $f(\cdot)$ le mieux possible sur $\mathcal{X}$
- d'optimisation :
  $\rightarrow$ utiliser $\eta(\cdot)$ pour déterminer $\arg_{x \in \mathcal{X}} f(x)$
- de maximisation de la diversité des réponses :
  $\rightarrow$ trouver des X_i tels que les $f(X_i)$ soient bien dispersés dans $\{f(x)\ : x \in \mathcal{X}\}$

I) Quel objectif ?

$f(x)$ une fonction (inconnue) définie pour $x \in \mathcal{X} \subset \mathbb{R}^d$

Á partir de couples $X_i, f(X_i)$, $i = 1, 2 \ldots$ on va construire un modèle $\eta(\cdot)$ de $f(\cdot)$ dans un but

- d'exploration : $\eta(\cdot)$ doit approximer $f(\cdot)$ le mieux possible sur $\mathcal{X}$
- d'optimisation :
  $\rightarrow$ utiliser $\eta(\cdot)$ pour déterminer $\arg_{x \in \mathcal{X}} f(x)$
- de maximisation de la diversité des réponses :
  $\rightarrow$ trouver des X_i tels que les $f(X_i)$ soient bien dispersés dans $\{f(x) \ : x \in \mathcal{X}\}$
- on peut avoir m fonctions $f_i(\cdot)$, $i = 1, \ldots, m$:
  $\rightarrow$ optimisation multicritères, diversité dans $\mathbb{R}^m \ldots$

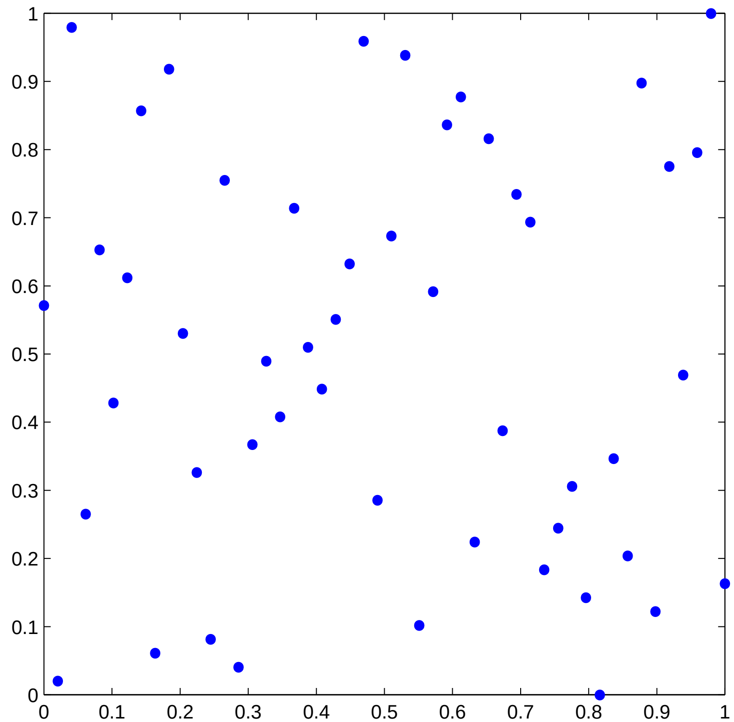
Préciser l'objectif est un préalable indispensable à la construction d'un plan "optimal" !

Préciser l'objectif est un préalable indispensable à la construction d'un plan "optimal" !

II) Non séquentiel ou séquentiel/adaptatif ?

Non-séquentiel, sans modèle :

propriété de remplissage d'espace ("space filling") → distance maximin ou minimax [Johnson *et al.*, 1990], hypercubes latins [Morris & Mitchell 1995]…



Université Nice SOPHIA ANTIPOLIS

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Non-séquentiel, avec modèle : si objectif = exploration, critère = maximum ou moyenne sur $\mathcal{X}$ de l'erreur quadratique moyenne de prédiction (EQM) ...

Non-séquentiel, avec modèle : si objectif = exploration, critère = maximum ou moyenne sur $\mathcal{X}$ de l'erreur quadratique moyenne de prédiction (EQM) ...

Disposer d'une EQM suppose un modèle $\eta(\cdot)$

... et si on dispose d'un modèle, on peut utiliser un plan adaptatif (construit pas à pas)

Problème central : comment prédire l'incertitude en un point x non encore exploré ?

Non-séquentiel, avec modèle : si objectif = exploration, critère = maximum ou moyenne sur $\mathcal{X}$ de l'erreur quadratique moyenne de prédiction (EQM) ...
Disposer d'une EQM suppose un modèle $\eta(\cdot)$
...et si on dispose d'un modèle, on peut utiliser un plan adaptatif (construit pas à pas)
Problème central : comment prédire l'incertitude en un point x non encore exploré ?

III) Quel modèle ?

Éviter une approche paramétrique dans un contexte non-paramétrique !

Modèle paramétrique :

$f(X_i) = \eta(X_i, \theta) + \varepsilon_i$, θ estimé par MC

$$\hat{\theta}^N = \arg\min \frac{1}{N} \sum_{i=1}^{N} [f(X_i) - \eta(X_i, \theta)]^2$$

Université Nice SOPHIA ANTIPOLIS

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Modèle paramétrique :

$f(X_i) = \eta(X_i, \theta) + \varepsilon_i$, θ estimé par MC

$$\hat{\theta}^N = \arg\min \frac{1}{N} \sum_{i=1}^{N} [f(X_i) - \eta(X_i, \theta)]^2$$

Dans un contexte stochastique, avec (ε_i) i.i.d. de moyenne 0 et variance σ^2 et (X_i) i.i.d. avec mesure de probabilité ξ sur $\mathcal{X}$, alors $\hat{\theta}^N \stackrel{\text{p.s.}}{\to} \bar{\theta}$ et

$$\sqrt{N}(\hat{\theta}^N - \bar{\theta}) \stackrel{\text{d}}{\to} \zeta \sim \mathcal{N}(0, \sigma^2 \mathbf{M}^{-1}(\xi, \bar{\theta}))$$

avec $\mathbf{M}(\xi, \theta) = \int_{\mathcal{X}} \frac{\partial \eta(x,\theta)}{\partial \theta} \frac{\partial \eta(x,\theta)}{\partial \theta^\top} \, \xi(dx)$

Soient θ fixé et $X_1, \ldots, X_n$ fixés (tels que $\mathbf{M}(\xi_n, \theta)$ de rang plein). Choisir

$$X_{k+1} = \mathbf{arg} \max_{x \in \mathcal{X}} \frac{\partial \eta(x,\theta)}{\partial \theta^\top} \mathbf{M}^{-1}(\xi_k, \theta) \frac{\partial \eta(x,\theta)}{\partial \theta}, \; k \geq n$$

assure $\xi_k \to \xi^*(\theta) = \arg\max_\xi \det \mathbf{M}(\xi, \theta)$ (D-optimalité) quand $k \to \infty$, avec ξ_k = mesure empirique de $X_1, \ldots, X_k$

Soient θ fixé et $X_1, \ldots, X_n$ fixés (tels que $\mathbf{M}(\xi_n, \theta)$ de rang plein). Choisir

$$X_{k+1} = \arg\max_{x \in \mathcal{X}} \frac{\partial \eta(x,\theta)}{\partial \theta^\top} \mathbf{M}^{-1}(\xi_k, \theta) \frac{\partial \eta(x,\theta)}{\partial \theta}\,,\ k \geq n$$

assure $\xi_k \to \xi^*(\theta) = \arg\max_\xi \det \mathbf{M}(\xi, \theta)$ (D-optimalité) quand $k \to \infty$, avec ξ_k = mesure empirique de $X_1, \ldots, X_k$

Problème : $\xi^*(\theta)$ a $p(p+1)/2$ points de support au plus (et souvent seulement p) avec $p = \dim(\theta)$

$\Rightarrow$ ξ_k n'est pas "space filling"

Quand on ré-estime θ à chaque étape, soit

$$X_{k+1} = \mathbf{arg}\max_{x \in \mathcal{X}} \frac{\partial \eta(x,\theta)}{\partial \theta^\top}_{|\hat{\theta}^k} \mathbf{M}^{-1}(\xi_k, \hat{\theta}^k) \frac{\partial \eta(x,\theta)}{\partial \theta}_{|\hat{\theta}^k}, \quad k \geq n \quad (1)$$

la convergence de $\hat{\theta}^k$ vers $\bar{\theta}$, et de ξ_k vers $\xi^*(\bar{\theta})$, n'est pas prouvée en général, mais jamais mise en défaut. . .

Quand on ré-estime θ à chaque étape, soit

$$X_{k+1} = \arg\max_{x\in\mathcal{X}} \frac{\partial\eta(x,\theta)}{\partial\theta^\top}_{|\hat\theta^k} \mathbf{M}^{-1}(\xi_k,\hat\theta^k)\frac{\partial\eta(x,\theta)}{\partial\theta}_{|\hat\theta^k}, \quad k\geq n \qquad (1)$$

la convergence de $\hat\theta^k$ vers $\bar\theta$, et de ξ_k vers $\xi^*(\bar\theta)$, n'est pas prouvée en général, mais jamais mise en défaut. . .

On pourrait penser utiliser (1) dans un cadre déterministe avec une fonction d'approximation pour modèle $\eta(x,\theta)$ (réseau de neurones, RBF. . .)

Université Nice SOPHIA ANTIPOLIS

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Ex. 1 : RBF $\eta(x) = \sum_{j=1}^{p} \theta_j K(x - N_j)$ (par ex. $K(\cdot)$ gaussien)
$\longrightarrow$ modèle linéaire à p paramètres
$p = 5$ fixé, $N_1, \ldots, N_5$ fixés, $(X_1, \ldots, X_5)$ = plan D-optimal
puis X_i là où la variance de la prédiction est maximale
(ce qui suppose $f(X_i) = \eta(X_i) + \varepsilon_i \ldots$)

Ex. 1 : RBF $\eta(x) = \sum_{j=1}^{p} \theta_j K(x - N_j)$ (par ex. $K(\cdot)$ gaussien)

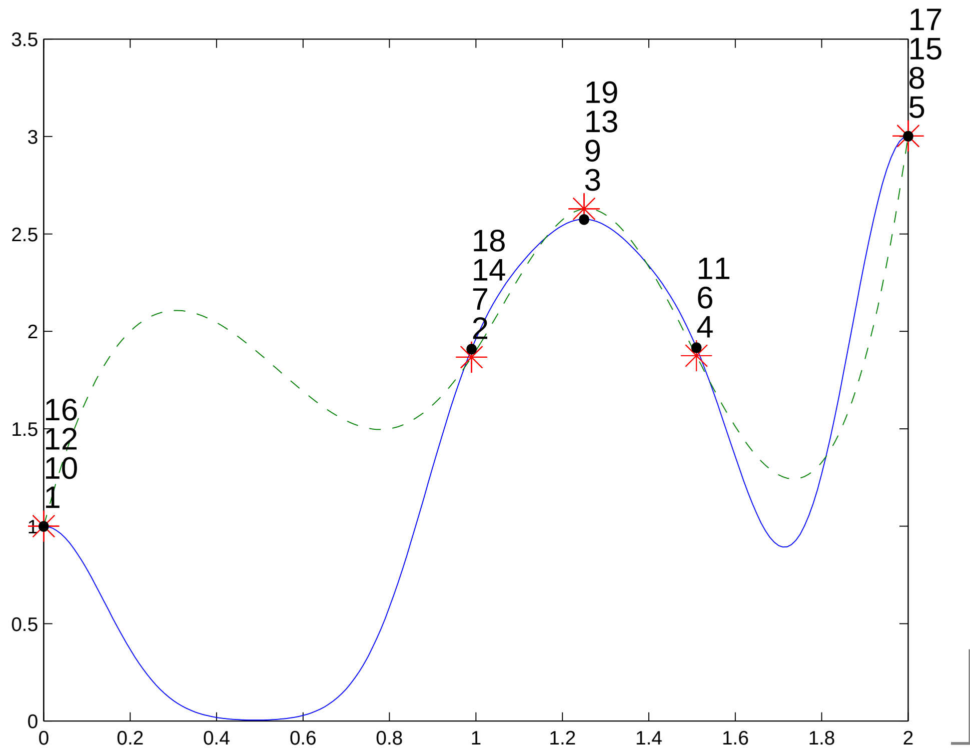
$\longrightarrow$ modèle linéaire à p paramètres

$p = 5$ fixé, $N_1, \ldots, N_5$ fixés, $(X_1, \ldots, X_5)$ = plan D-optimal

puis X_i là où la variance de la prédiction est maximale

(ce qui suppose $f(X_i) = \eta(X_i) + \varepsilon_i$ ...)

$\longrightarrow$ pas d'exploration :
le plan reste concentré en
p points



Université Nice SOPHIA ANTIPOLIS

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Prenons p croissant (structure variable) :

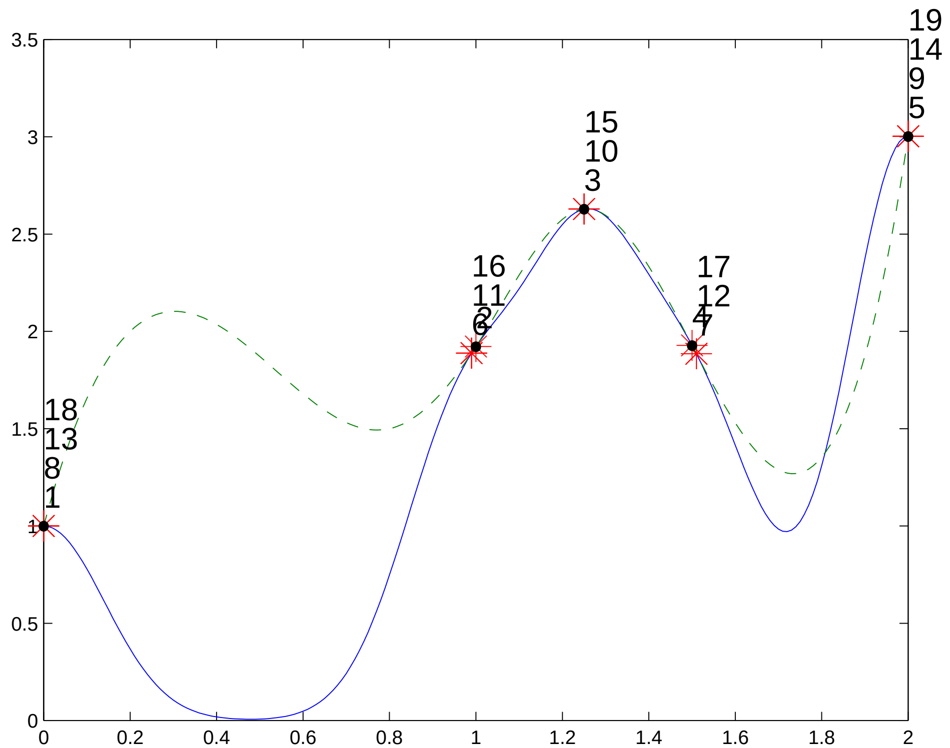
$N_1, \ldots, N_5$ et $(X_1, \ldots, X_5)$ comme précédemment

puis N_i et X_i au point de variance de prédiction maximale

Prenons p croissant (structure variable) :

$N_1, \ldots, N_5$ et $(X_1, \ldots, X_5)$ comme précédemment
puis N_i et X_i au point de variance de prédiction maximale

$\longrightarrow$ pas beaucoup plus d'exploration : le plan reste concentré



Le problème n'est pas lié au choix de $\eta(x, \theta)$: mêmes difficultés pour polynômes, ou "fonction universelle"

Le problème n'est pas lié au choix de $\eta(x,\theta)$: mêmes difficultés pour polynômes, ou "fonction universelle"

Ex. 2 : $\eta(x,\theta) = \int_0^{x+\theta_1} \frac{\theta_2\theta_4}{1+\theta_4^2-\cos(\theta_2\, t)} \cos[\exp(t)]\, dt + \theta_3$

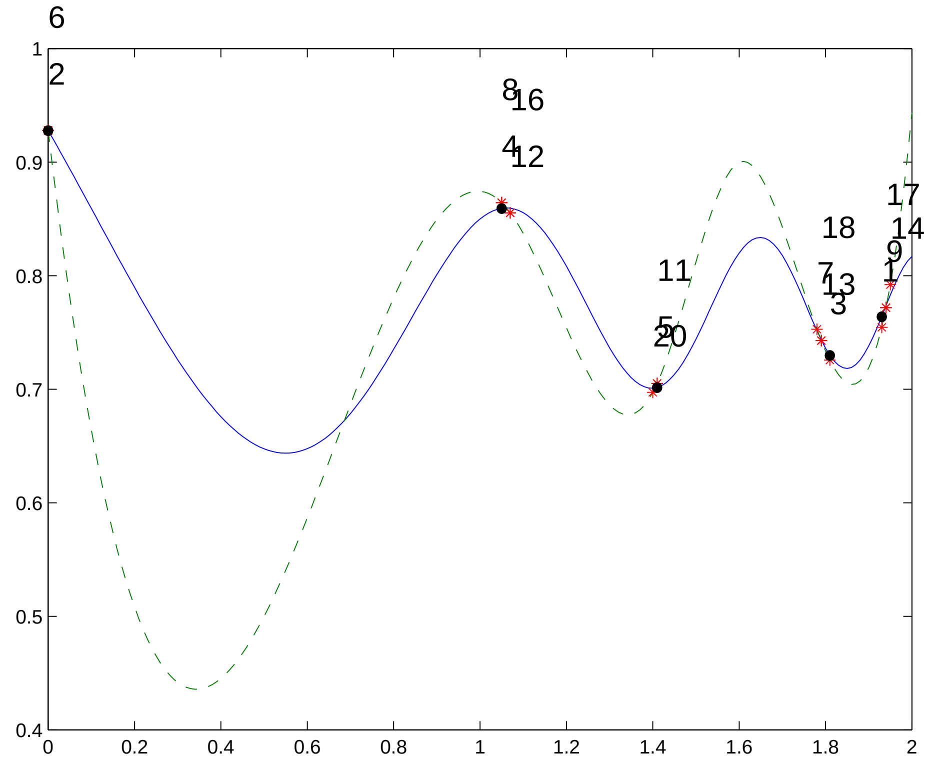
(famille dense dans $\mathcal{C}(I)$ pout tout $I=[a,b]\subset\mathbb{R}$)

Le problème n'est pas lié au choix de $\eta(x,\theta)$: mêmes difficultés pour polynômes, ou “fonction universelle”

Ex. 2 : $\eta(x,\theta) = \int_0^{x+\theta_1} \frac{\theta_2\theta_4}{1+\theta_4^2-\cos(\theta_2\,t)}\,\cos[\exp(t)]\,dt + \theta_3$

(famille dense dans $\mathcal{C}(I)$ pout tout $I=[a,b]\subset\mathbb{R}$)

$\longrightarrow$ pas vraiment d'exploration : le plan reste concentré



Université Nice SOPHIA ANTIPOLIS

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

$\longrightarrow$ Changeons la façon de prédire l'incertitude en x !

$\rightarrow$ Changeons la façon de prédire l'incertitude en x !

Ex. 1 (suite) : RBF avec Bootstrap et Leave-One-Out [Gazut, Martinez, Issanchou 2006] :

$\eta_i(x, \theta^{(i)}) = \sum_{j=1,\, j\neq i}^{n} \theta_j K(x - N_j)$

$\rightarrow$ variance $V(x)$ des n prédictions (MC) en x

$\rightarrow$ observation en $X_{n+1} = \arg\max_x V(x)$ et $N_{n+1} = X_{n+1}$

$\rightarrow$ Changeons la façon de prédire l'incertitude en x !

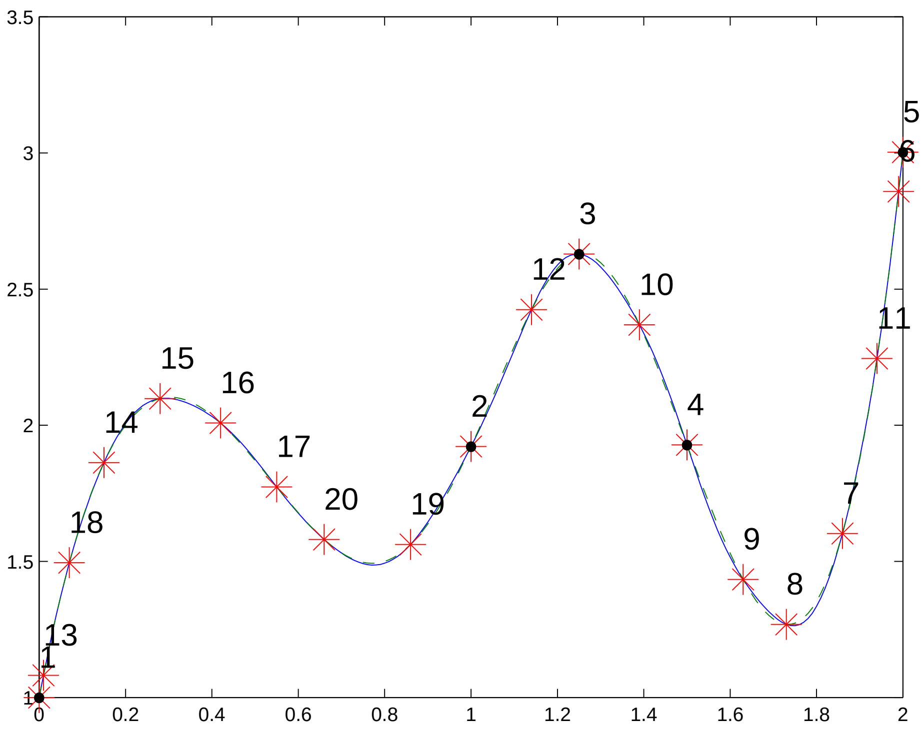
Ex. 1 (suite) : RBF avec Bootstrap et Leave-One-Out [Gazut, Martinez, Issanchou 2006] :

$\eta_i(x, \theta^{(i)}) = \sum_{j=1,\, j\neq i}^{n} \theta_j K(x - N_j)$

$\rightarrow$ variance $V(x)$ des n prédictions (MC) en x

$\rightarrow$ observation en $X_{n+1} = \arg\max_x V(x)$ et $N_{n+1} = X_{n+1}$

$\rightarrow$ exploration



Université Nice SOPHIA ANTIPOLIS

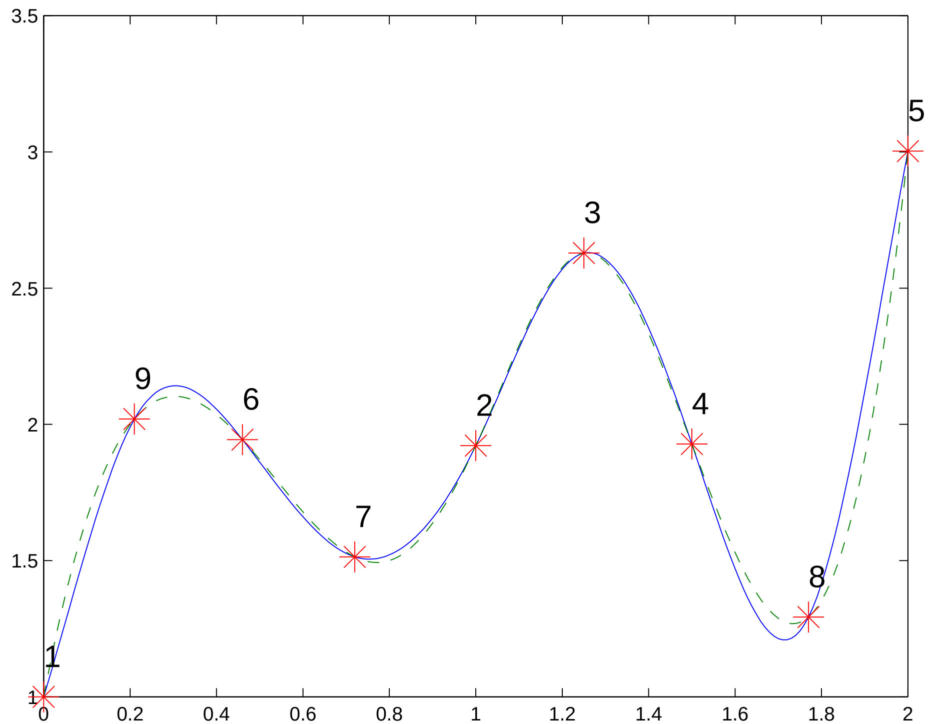
CNRS CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Ex. 3 : Krigeage

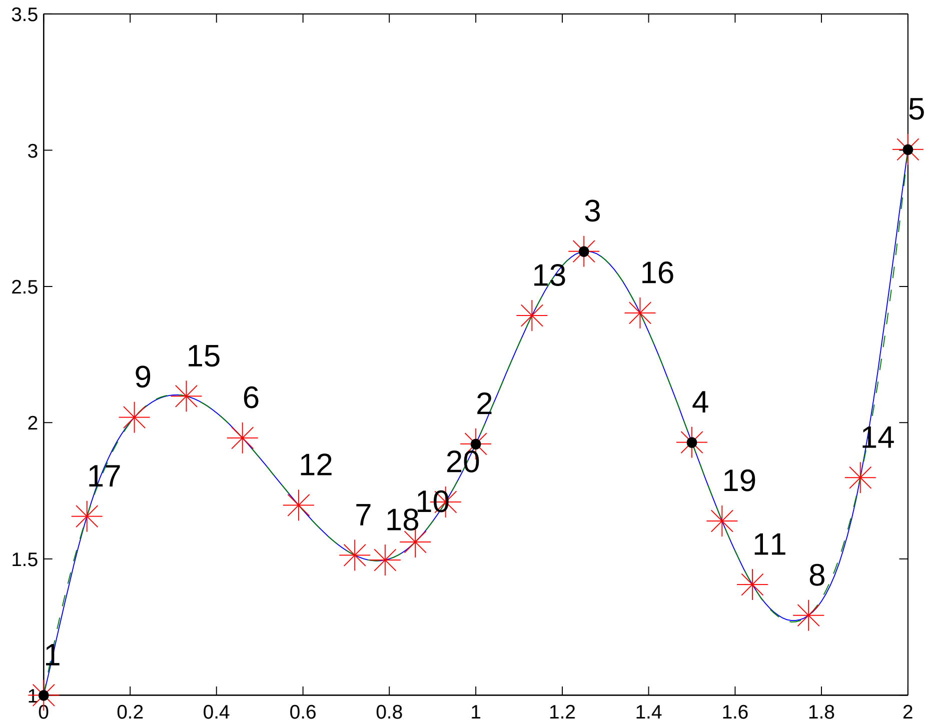
Même données que pour Ex. 1, modèle : $f(x) = \beta + Z(x,\omega)$ avec $Z(x,\omega)$ réalisation d'un processus gaussien, $\mathbb{E}\{Z(x,\omega)\} = 0$, $\mathbb{E}\{Z(x,\omega)Z(u,\omega)\} = \sigma^2\, C(|x-u|,\theta)$ (ici, β et σ^2 inconnus et θ fixé)

X_{n+1} au point de variance de krigeage maximale

$\longrightarrow$ exploration
(9 points seulement)



→ exploration (20 points)



III) Quelques problèmes...

- algorithmes pour plans non-séquentiels sans modèle ("space-filling")

III) Quelques problèmes…

- algorithmes pour plans non-séquentiels sans modèle (“space-filling”)
- plans séquentiels, méthodes d’exploration, apprentissage actif, modèles (Bootstrap et Leave-One-Out, krigeage…)

III) Quelques problèmes...

- algorithmes pour plans non-séquentiels sans modèle (“space-filling”)
- plans séquentiels, méthodes d’exploration, apprentissage actif, modèles (Bootstrap et Leave-One-Out, krigeage. . .)
- propriétés théoriques : asymptotique, prise en compte de l’estimation de θ dans $C(\cdot, \theta)$ pour prédiction par krigeage —aspect dual de la construction du plan—

III) Quelques problèmes...

- algorithmes pour plans non-séquentiels sans modèle ("space-filling")
- plans séquentiels, méthodes d'exploration, apprentissage actif, modèles (Bootstrap et Leave-One-Out, krigeage...)
- propriétés théoriques : asymptotique, prise en compte de l'estimation de θ dans $C(\cdot, \theta)$ pour prédiction par krigeage —aspect dual de la construction du plan—
- autres problématiques que l'exploration (optimisation...)

III) Quelques problèmes...

- algorithmes pour plans non-séquentiels sans modèle ("space-filling")
- plans séquentiels, méthodes d'exploration, apprentissage actif, modèles (Bootstrap et Leave-One-Out, krigeage...)
- propriétés théoriques : asymptotique, prise en compte de l'estimation de θ dans $C(\cdot, \theta)$ pour prédiction par krigeage —aspect dual de la construction du plan—
- autres problématiques que l'exploration (optimisation...)
- problèmes réels, benchmarks...