

# Plans d'expériences numériques — (2) avec modèle —

LUC PRONZATO

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

Pec'num<sup>2015</sup>

# Plan I

- 1 Plans optimaux pour krigeage et processus Gaussiens
  - 1.1 Processus Gaussiens et krigeage
  - 1.2 Critères à base de MSE
  - 1.3 Échantillonnage à maximum d'entropie
- 2 Plans optimaux pour régression linéaire
  - 2.1 Régression linéaire
  - 2.2 Plans exacts
  - 2.3 Plans approximatifs
  - 2.4 Modèles produits tensoriels
  - 2.5 Conséquences pour les plans *space filling*
- 3 Plans optimaux pour prédiction Bayésienne
  - 3.1 Décomposition de Karhunen-Loève
  - 3.2 Prédiction Bayésienne
- 4 Au delà des plans *space filling*
- 5 Conclusions partie (2)

# Objectifs (idem partie (1))

## Expériences numériques: à base de simulations

- Habituellement,  $\mathbf{x} \in \mathbb{R}^d \mapsto$  observation  $Y(\mathbf{x})$  (phénomène physique)
- ici, simulation numérique:  $Y(\mathbf{x}) = f(\mathbf{x})$ , observation = évaluation d'une fonction  $f(\cdot)$  inconnue  
(pas de bruit de mesure)

# Objectifs (idem partie (1))

## Expériences numériques: à base de simulations

- Habituellement,  $\mathbf{x} \in \mathbb{R}^d \mapsto$  observation  $Y(\mathbf{x})$  (phénomène physique)
- ici, simulation numérique:  $Y(\mathbf{x}) = f(\mathbf{x})$ , observation = évaluation d'une fonction  $f(\cdot)$  inconnue  
(pas de bruit de mesure)

## A partir de paires $(\mathbf{X}_i, f(\mathbf{x}_i))$ , $i = 1, 2, \dots, n$

- optimisation: trouver  $\mathbf{x}^* = \arg \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x})$
- inversion: reconstruire  $\{\mathbf{x} \in \mathcal{X} : f(\mathbf{x}) = T\}$
- estimation d'une probabilité de défaillance:  $\text{Prob}\{f(\mathbf{x}) > C\}$  quand  $\mathbf{x}$  est distribué suivant une densité de probabilité  $\phi(\cdot)$
- analyse de sensibilité
- **approximation/interpolation de  $f(\cdot)$**  par une fonction  $\eta_n(\cdot)$ , à construire

# 1 Plans optimaux pour krigeage et processus Gaussiens

## 1.1 Processus Gaussiens et krigeage

Modèle pour  $f(\cdot)$  : processus Gaussien

$$f(\mathbf{x}) = \mathbf{r}^\top(\mathbf{x})\boldsymbol{\beta} + Z(\mathbf{x}), \text{ avec}$$

$\mathbf{r}(\mathbf{x})$  un vecteur de fonctions connues de  $\mathbf{x}$  (la tendance)

$Z(\mathbf{x})$  = réalisation d'un processus aléatoire (stationnaire au second ordre, typiquement Gaussien)

$$E\{Z(\mathbf{x})\} = 0, E\{Z(\mathbf{x})Z(\mathbf{x}')\} = \sigma^2 C(\mathbf{x} - \mathbf{x}'; \theta)$$

# 1 Plans optimaux pour krigeage et processus Gaussiens

## 1.1 Processus Gaussiens et krigeage

Modèle pour  $f(\cdot)$  : processus Gaussien

$$f(\mathbf{x}) = \mathbf{r}^\top(\mathbf{x})\boldsymbol{\beta} + Z(\mathbf{x}), \text{ avec}$$

$\mathbf{r}(\mathbf{x})$  un vecteur de fonctions connues de  $\mathbf{x}$  (la tendance)

$Z(\mathbf{x})$  = réalisation d'un processus aléatoire (stationnaire au second ordre, typiquement Gaussien)

$$E\{Z(\mathbf{x})\} = 0, E\{Z(\mathbf{x})Z(\mathbf{x}')\} = \sigma^2 C(\mathbf{x} - \mathbf{x}'; \theta)$$

Expériences de simulation (*computer experiments*)

Depuis (Sacks et al., 1989) : Choisir  $C(\delta; \theta)$  continue en  $\delta = 0$ ,  $C(0; \theta) = 1$

- 2 répétitions au même  $\mathbf{x}$  donne la même valeur  $f(\mathbf{x})$   
(pas de bruit de mesure)

Objectif = interpolation (ou extrapolation) : construire une prédiction  $\eta_n(\mathbf{x})$  pour une réalisation particulière de  $Z(\cdot)$

très différent de  $\neq$  la prédiction d'autres réalisations

( $\Rightarrow$  "simplement" estimer  $\beta$ )

Objectif = interpolation (ou extrapolation) : construire une prédiction  $\eta_n(\mathbf{x})$  pour une réalisation particulière de  $Z(\cdot)$

très différent de  $\neq$  la prédiction d'autres réalisations

( $\Rightarrow$  "simplement" estimer  $\beta$ )

Krigeage ordinaire

(les formules du krigage universel avec tendance  $\mathbf{r}^\top(\mathbf{x})\beta$ ,  $\beta \in \mathbb{R}^p$ ,  $p > 1$ , sont un peu plus compliquées) :

$$f(\mathbf{x}) = \beta + Z(\mathbf{x}) \rightarrow \eta_n(\mathbf{x}) = \eta_n[f](\mathbf{x})$$

BLUP (*Best Linear Unbiased Predictor*) en  $\mathbf{x}$ :  $\eta_n(\mathbf{x}) = \mathbf{v}_n^\top(\mathbf{x})\mathbf{y}_n$  avec

- $\mathbf{y}_n = (f(\mathbf{x}_1), \dots, f(\mathbf{x}_n))^\top$
- $\mathbf{v}_n(\mathbf{x})$  minimise  $E\{(\mathbf{v}_n^\top \mathbf{y}_n - [\beta + Z(\mathbf{x})])^2\}$
- sous la contrainte  $E\{\mathbf{v}_n^\top \mathbf{y}_n\} = \beta \sum_{i=1}^n \{\mathbf{v}_n\}_i = E\{f(\mathbf{x})\} = \beta$ , i.e.,  
 $\sum_{i=1}^n \{\mathbf{v}_n\}_i = 1$

Prédiction :  $\eta_n(\mathbf{x}) = \hat{\beta}^n + \mathbf{c}_n^\top(\mathbf{x})\mathbf{C}_n^{-1}(\mathbf{y}_n - \hat{\beta}^n\mathbf{1})$

**MSE** (*Mean-Squared Error*) proportionnelle à

$$\rho_n(\mathbf{x}) = \left( 1 - [\mathbf{c}_n^\top(\mathbf{x}) \mathbf{1}] \begin{bmatrix} \mathbf{C}_n & \mathbf{1} \\ \mathbf{1} & 0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{c}_n(\mathbf{x}) \\ 1 \end{bmatrix} \right)$$

[avec  $\{\mathbf{C}_n\}_{i,j} = C((X_i - X_j); \theta)$ ,  $\{\mathbf{c}_n(\mathbf{x})\}_i = C((X_i - \mathbf{x}); \theta)$ ,  $\hat{\beta}^n = (\mathbf{1}^\top \mathbf{C}_n^{-1} \mathbf{y}_n) / (\mathbf{1}^\top \mathbf{C}_n^{-1} \mathbf{1})$  (WLS) et  $\mathbf{1} = (1, \dots, 1)^\top$ ]

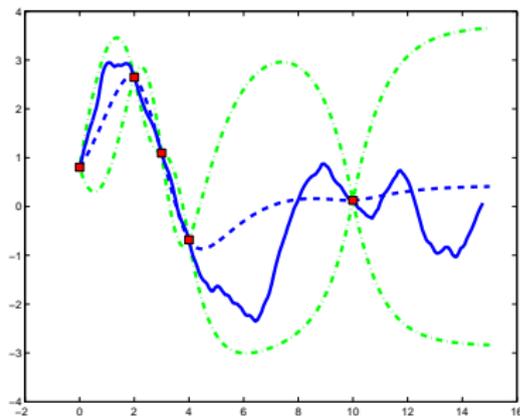
Prédiction :  $\eta_n(\mathbf{x}) = \hat{\beta}^n + \mathbf{c}_n^\top(\mathbf{x})\mathbf{C}_n^{-1}(\mathbf{y}_n - \hat{\beta}^n\mathbf{1})$

**MSE** (*Mean-Squared Error*) proportionnelle à

$$\rho_n(\mathbf{x}) = \left( 1 - [\mathbf{c}_n^\top(\mathbf{x}) \ 1] \begin{bmatrix} \mathbf{C}_n & \mathbf{1} \\ \mathbf{1} & 0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{c}_n(\mathbf{x}) \\ 1 \end{bmatrix} \right)$$

[avec  $\{\mathbf{C}_n\}_{i,j} = C((X_i - X_j); \theta)$ ,  $\{\mathbf{c}_n(\mathbf{x})\}_i = C((X_i - \mathbf{x}); \theta)$ ,  $\hat{\beta}^n = (\mathbf{1}^\top \mathbf{C}_n^{-1} \mathbf{y}_n) / (\mathbf{1}^\top \mathbf{C}_n^{-1} \mathbf{1})$  (WLS) et  $\mathbf{1} = (1, \dots, 1)^\top$ ]

Ex. avec  $d = 1$   
(noter que  $\rho_n(\mathbf{x}_i) = 0$ ,  $i = 1, \dots, n$  — pas de bruit de mesure)



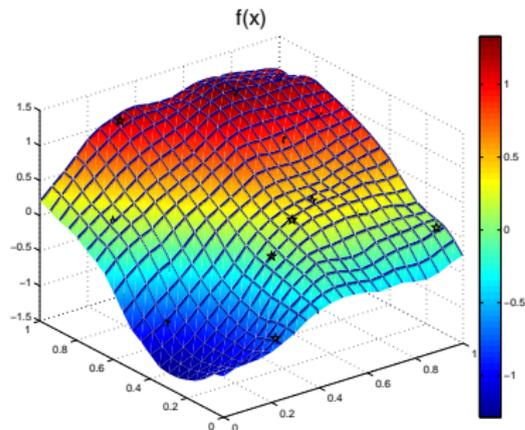
Prédiction :  $\eta_n(\mathbf{x}) = \hat{\beta}^n + \mathbf{c}_n^\top(\mathbf{x})\mathbf{C}_n^{-1}(\mathbf{y}_n - \hat{\beta}^n\mathbf{1})$

**MSE** (*Mean-Squared Error*) proportionnelle à

$$\rho_n(\mathbf{x}) = \left( 1 - [\mathbf{c}_n^\top(\mathbf{x}) \mathbf{1}] \begin{bmatrix} \mathbf{C}_n & \mathbf{1} \\ \mathbf{1} & 0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{c}_n(\mathbf{x}) \\ 1 \end{bmatrix} \right)$$

[avec  $\{\mathbf{C}_n\}_{i,j} = C((X_i - X_j); \theta)$ ,  $\{\mathbf{c}_n(\mathbf{x})\}_i = C((X_i - \mathbf{x}); \theta)$ ,  $\hat{\beta}^n = (\mathbf{1}^\top \mathbf{C}_n^{-1} \mathbf{y}_n) / (\mathbf{1}^\top \mathbf{C}_n^{-1} \mathbf{1})$  (WLS) et  $\mathbf{1} = (1, \dots, 1)^\top$ ]

Ex. avec  $d = 2$   
 ( $n = 20$ ,  $\mathbf{X}_n =$  hypercube latin aléatoire)



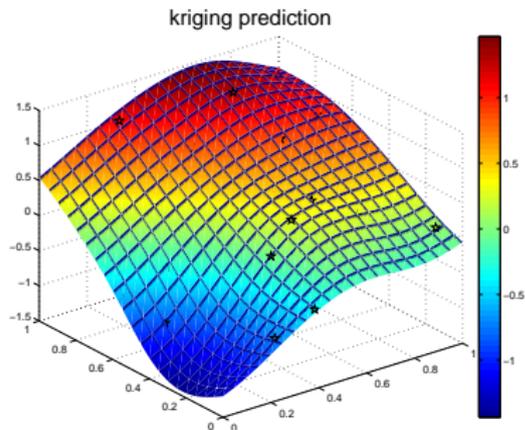
Prédiction :  $\eta_n(\mathbf{x}) = \hat{\beta}^n + \mathbf{c}_n^\top(\mathbf{x})\mathbf{C}_n^{-1}(\mathbf{y}_n - \hat{\beta}^n\mathbf{1})$

**MSE** (*Mean-Squared Error*) proportionnelle à

$$\rho_n(\mathbf{x}) = \left( 1 - [\mathbf{c}_n^\top(\mathbf{x}) \mathbf{1}] \left[ \begin{array}{cc} \mathbf{C}_n & \mathbf{1} \\ \mathbf{1} & 0 \end{array} \right]^{-1} \left[ \begin{array}{c} \mathbf{c}_n(\mathbf{x}) \\ 1 \end{array} \right] \right)$$

[avec  $\{\mathbf{C}_n\}_{i,j} = C((X_i - X_j); \theta)$ ,  $\{\mathbf{c}_n(\mathbf{x})\}_i = C((X_i - \mathbf{x}); \theta)$ ,  $\hat{\beta}^n = (\mathbf{1}^\top \mathbf{C}_n^{-1} \mathbf{y}_n) / (\mathbf{1}^\top \mathbf{C}_n^{-1} \mathbf{1})$  (WLS) et  $\mathbf{1} = (1, \dots, 1)^\top$ ]

Ex. avec  $d = 2$   
 ( $n = 20$ ,  $\mathbf{X}_n =$  hypercube latin aléatoire)



Prédiction :  $\eta_n(\mathbf{x}) = \hat{\beta}^n + \mathbf{c}_n^\top(\mathbf{x})\mathbf{C}_n^{-1}(\mathbf{y}_n - \hat{\beta}^n\mathbf{1})$

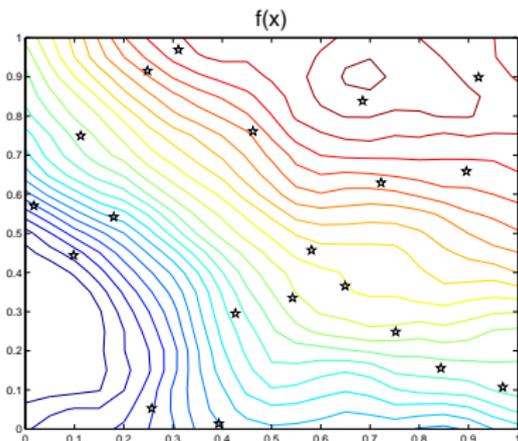
**MSE** (*Mean-Squared Error*) proportionnelle à

$$\rho_n(\mathbf{x}) = \left( 1 - [\mathbf{c}_n^\top(\mathbf{x}) \mathbf{1}] \begin{bmatrix} \mathbf{C}_n & \mathbf{1} \\ \mathbf{1} & 0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{c}_n(\mathbf{x}) \\ 1 \end{bmatrix} \right)$$

[avec  $\{\mathbf{C}_n\}_{i,j} = C((X_i - X_j); \theta)$ ,  $\{\mathbf{c}_n(\mathbf{x})\}_i = C((X_i - \mathbf{x}); \theta)$ ,  $\hat{\beta}^n = (\mathbf{1}^\top \mathbf{C}_n^{-1} \mathbf{y}_n) / (\mathbf{1}^\top \mathbf{C}_n^{-1} \mathbf{1})$  (WLS) et  $\mathbf{1} = (1, \dots, 1)^\top$ ]

Ex. avec  $d = 2$

( $n = 20$ ,  $\mathbf{X}_n =$  hypercube latin aléatoire)



Prédiction :  $\eta_n(\mathbf{x}) = \hat{\beta}^n + \mathbf{c}_n^\top(\mathbf{x})\mathbf{C}_n^{-1}(\mathbf{y}_n - \hat{\beta}^n\mathbf{1})$

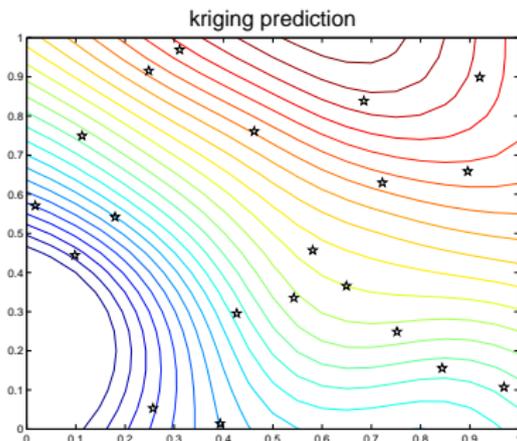
**MSE** (*Mean-Squared Error*) proportionnelle à

$$\rho_n(\mathbf{x}) = \left( 1 - [\mathbf{c}_n^\top(\mathbf{x}) \mathbf{1}] \left[ \begin{array}{cc} \mathbf{C}_n & \mathbf{1} \\ \mathbf{1} & 0 \end{array} \right]^{-1} \left[ \begin{array}{c} \mathbf{c}_n(\mathbf{x}) \\ 1 \end{array} \right] \right)$$

[avec  $\{\mathbf{C}_n\}_{i,j} = C((X_i - X_j); \theta)$ ,  $\{\mathbf{c}_n(\mathbf{x})\}_i = C((X_i - \mathbf{x}); \theta)$ ,  $\hat{\beta}^n = (\mathbf{1}^\top \mathbf{C}_n^{-1} \mathbf{y}_n) / (\mathbf{1}^\top \mathbf{C}_n^{-1} \mathbf{1})$  (WLS) et  $\mathbf{1} = (1, \dots, 1)^\top$ ]

Ex. avec  $d = 2$

( $n = 20$ ,  $\mathbf{X}_n =$  hypercube latin aléatoire)



## 1.2 Critères à base de MSE

**Idée naturelle : “minimiser  $\rho_n(\mathbf{x})$  pour tout  $\mathbf{x}$ ”**

En pratique :

- minimiser  $\text{MMSE}(\mathbf{X}_n) = \max_{\mathbf{x} \in \mathcal{X}} \rho_n(\mathbf{x})$
- minimiser  $\text{IMSE}(\mathbf{X}_n) = \int_{\mathcal{X}} \rho_n(\mathbf{x}) d\mu(\mathbf{x})$ , avec  $\mu(\cdot)$  une mesure d'intérêt sur  $\mathcal{X}$

## 1.2 Critères à base de MSE

**Idee naturelle : “minimiser  $\rho_n(\mathbf{x})$  pour tout  $\mathbf{x}$ ”**

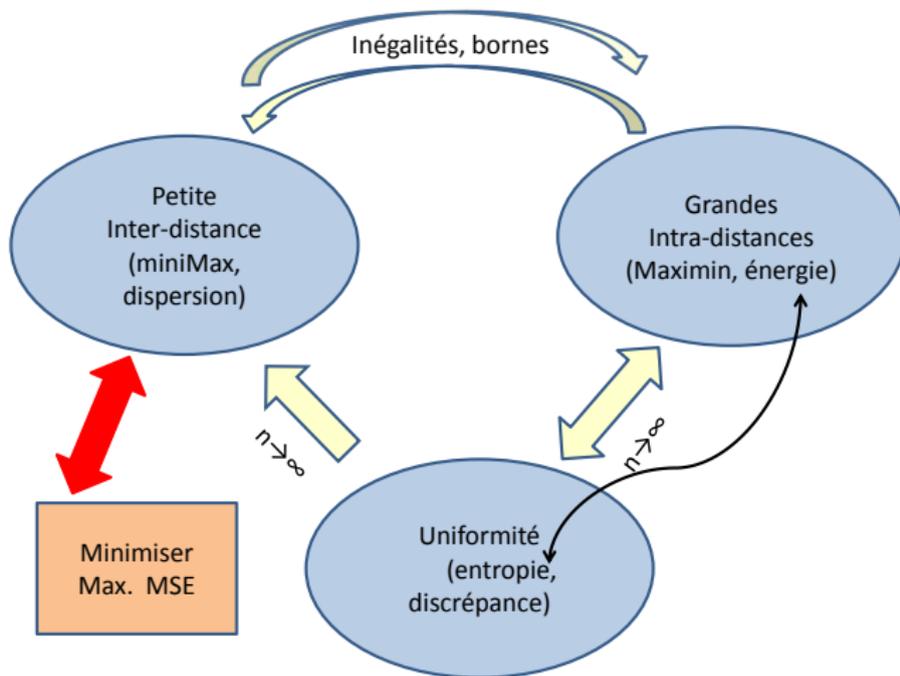
En pratique :

- minimiser  $\text{MMSE}(\mathbf{X}_n) = \max_{\mathbf{x} \in \mathcal{X}} \rho_n(\mathbf{x})$
- minimiser  $\text{IMSE}(\mathbf{X}_n) = \int_{\mathcal{X}} \rho_n(\mathbf{x}) d\mu(\mathbf{x})$ , avec  $\mu(\cdot)$  une mesure d'intérêt sur  $\mathcal{X}$

Les plans obtenus sont typiquement *space-filling* :

**Johnson et al. (1990)** : si  $C(\mathbf{x} - \mathbf{x}') = c(\|\mathbf{x} - \mathbf{x}'\|)$  avec  $c(\cdot)$  décroissante, alors  $\mathbf{X}_n^*$  optimal pour  $\Phi_{mM}(\cdot)$  (miniMax optimal) tend à être optimal pour  $\text{MMSE}(\mathbf{X}_n)$  avec la covariance  $C_a(\mathbf{x} - \mathbf{x}') = [C(\mathbf{x} - \mathbf{x}')]^a$  quand  $a \rightarrow \infty$

⇒ pas de points  $\mathbf{x}_i$  sur les bords de  $\mathcal{X}$



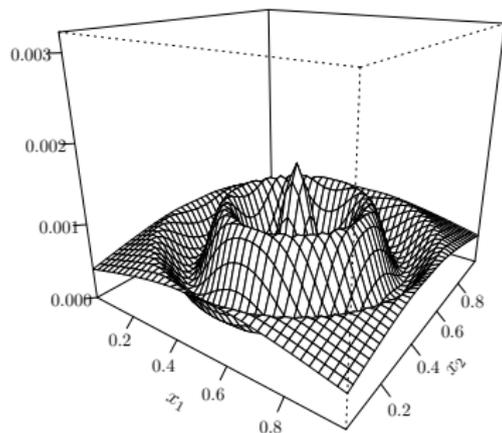
Ex. de plan IMSE-optimal (Gauthier and Pronzato, 2014a) :

$\mathcal{X}$  = grille de  $37^2 = 1369$  points

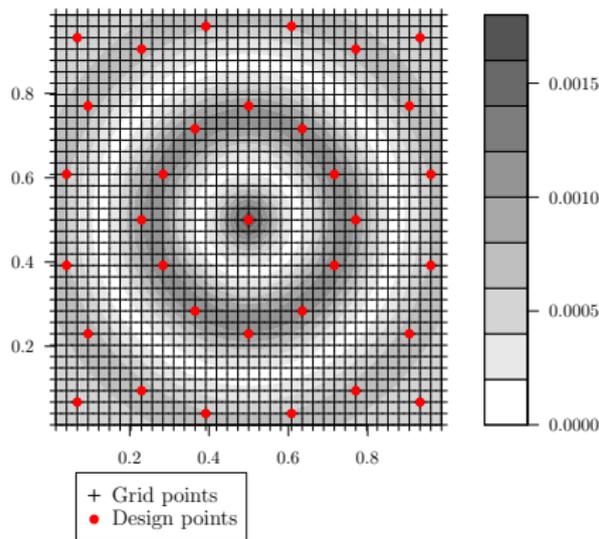
$$C(\mathbf{x} - \mathbf{x}') = C_1(\{\mathbf{x}\}_1 - \{\mathbf{x}'\}_1) \times C_2(\{\mathbf{x}\}_2 - \{\mathbf{x}'\}_2),$$

$$C_i(x - x') = (1 + 25/\sqrt{3}|x - x'|) \exp[-25/\sqrt{3}|x - x'|] \text{ (Matérn } 3/2)$$

Grid weight  $\omega_k$



33-point optimal design



### Calcul de $\text{MMSE}(\mathbf{X}_n)$ :

Calculer  $\rho_n(\mathbf{x}^{(k)})$  pour un ensemble fini  $\mathcal{X}_Q$  de  $Q$  valeurs  $\mathbf{x}_k$ ,  $1 \leq k \leq Q$  (par exemple, les  $Q$  premiers éléments d'une suite à faible discrédance sur  $\mathcal{X}$ ),

puis  $\text{MMSE}(\mathbf{X}_n) \simeq \max_k \rho_n(\mathbf{x}^{(k)})$ ,

à minimiser par exemple par recuit simulé

Calcul de MMSE( $\mathbf{X}_n$ ) :

Calculer  $\rho_n(\mathbf{x}^{(k)})$  pour un ensemble fini  $\mathcal{X}_Q$  de  $Q$  valeurs  $\mathbf{x}_k$ ,  $1 \leq k \leq Q$  (par exemple, les  $Q$  premiers éléments d'une suite à faible discrédance sur  $\mathcal{X}$ ),  
 puis  $\text{MMSE}(\mathbf{X}_n) \simeq \max_k \rho_n(\mathbf{x}^{(k)})$ ,  
 à minimiser par exemple par recuit simulé

Calcul de IMSE( $\mathbf{X}_n$ ) =  $\int_{\mathcal{X}} \rho_n(\mathbf{x}) d\mu(\mathbf{x})$  (Gauthier and Pronzato, 2014a,b) :

Sans tendance ( $\mathbf{r}(\mathbf{x}) = \mathbf{0} \forall \mathbf{x}$ )  $\implies \rho_n(\mathbf{x}) = 1 - \mathbf{c}_n^\top(\mathbf{x}) \mathbf{C}_n^{-1} \mathbf{c}_n(\mathbf{x})$ ,  
 avec  $\{\mathbf{c}_n(\mathbf{x})\}_i = C(\mathbf{x} - \mathbf{x}_i)$ ,  $\{\mathbf{C}_n\}_{ij} = C(\mathbf{x}_i - \mathbf{x}_j)$

$$\begin{aligned} \text{IMSE}(\mathbf{X}_n) &= 1 - \text{trace} \left[ \mathbf{C}_n^{-1} \int_{\mathcal{X}} \mathbf{c}_n(\mathbf{x}) \mathbf{c}_n^\top(\mathbf{x}) d\mu(\mathbf{x}) \right] \\ &= 1 - \text{trace} \left[ \mathbf{C}_n^{-1} \boldsymbol{\Sigma}_n \right] \end{aligned}$$

Calcul pour un ensemble fini  $\mathcal{X}_Q$  de  $Q$  valeurs  $\mathbf{x}_k$ ,  $1 \leq k \leq Q$  :

$$\begin{aligned} \text{IMSE}(\mathbf{X}_n) &\simeq \widehat{\text{IMSE}}(\mathbf{X}_n) = \sum_{k=1}^Q w_k \rho_n(\mathbf{x}^{(k)}) \\ &= 1 - \text{trace} \left[ \mathbf{C}_n^{-1} \widehat{\boldsymbol{\Sigma}}_n \right] \end{aligned}$$

avec  $\sum_{k=1}^Q w_k = 1$  ( $w_k = 1/Q$  pour  $\mu$  uniforme)

et  $\widehat{\boldsymbol{\Sigma}}_n = \sum_{k=1}^Q w_k \mathbf{c}_n(\mathbf{x}^{(k)}) \mathbf{c}_n^\top(\mathbf{x}^{(k)})$

Calcul pour un ensemble fini  $\mathcal{X}_Q$  de  $Q$  valeurs  $\mathbf{x}_k$ ,  $1 \leq k \leq Q$  :

$$\begin{aligned} \text{IMSE}(\mathbf{X}_n) &\simeq \widehat{\text{IMSE}}(\mathbf{X}_n) = \sum_{k=1}^Q w_k \rho_n(\mathbf{x}^{(k)}) \\ &= 1 - \text{trace} \left[ \mathbf{C}_n^{-1} \widehat{\boldsymbol{\Sigma}}_n \right] \end{aligned}$$

avec  $\sum_{k=1}^Q w_k = 1$  ( $w_k = 1/Q$  pour  $\mu$  uniforme)

et  $\widehat{\boldsymbol{\Sigma}}_n = \sum_{k=1}^Q w_k \mathbf{c}_n(\mathbf{x}^{(k)}) \mathbf{c}_n^\top(\mathbf{x}^{(k)})$

Si en plus  $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathcal{X}_Q$ , avec  $\mathbf{x}_i = \mathbf{x}^{(k_i)}$ ,  $i = 1, \dots, n$ ,

alors  $\widehat{\boldsymbol{\Sigma}}_n = \{\mathbf{Q}\mathbf{W}\mathbf{Q}\}_{\mathbb{J}_n\mathbb{J}_n}$

avec  $\{\mathbf{Q}\}_{kl} = \mathbf{C}(\mathbf{x}^{(k)} - \mathbf{x}^{(\ell)})$ ,  $\mathbf{W} = \text{diag}\{w_1, \dots, w_Q\}$

et  $\mathbb{J}_n = \{k_1, \dots, k_n\}$

⇒  $\text{IMSE}(\mathbf{X}_n) \simeq 1 - \text{trace} \left[ \mathbf{Q}_{\mathbb{J}_n\mathbb{J}_n}^{-1} \{\mathbf{Q}\mathbf{W}\mathbf{Q}\}_{\mathbb{J}_n\mathbb{J}_n} \right]$

pas très coûteux à évaluer quand on a calculé  $\mathbf{Q}$  et  $\mathbf{Q}\mathbf{W}\mathbf{Q}$

(un peu plus compliqué avec tendance  $\mathbf{r}^\top(\mathbf{x})\boldsymbol{\beta}$ )

## Construction séquentielle d'un plan optimal :

Pour IMSE : rien de particulier, à l'étape  $n + 1$ ,  $\mathbf{X}_{n+1} = \{\mathbf{X}_n, \mathbf{x}_{n+1}\}$  avec

$$\mathbf{x}_{n+1}^* = \arg \min_{\mathbf{x} \in \mathcal{X}} \text{IMSE}(\{\mathbf{X}_n, \mathbf{x}\})$$

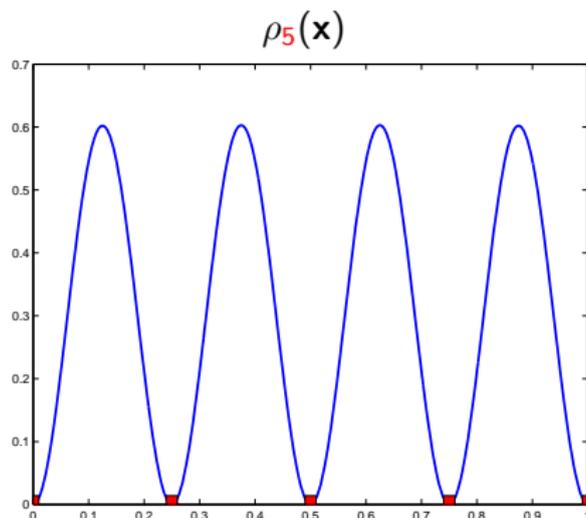
## Construction séquentielle d'un plan optimal :

Pour IMSE : rien de particulier, à l'étape  $n + 1$ ,  $\mathbf{X}_{n+1} = \{\mathbf{X}_n, \mathbf{x}_{n+1}\}$  avec

$$\mathbf{x}_{n+1}^* = \arg \min_{\mathbf{x} \in \mathcal{X}} \text{IMSE}(\{\mathbf{X}_n, \mathbf{x}\})$$

Pour MMSE : ne surtout pas choisir  $\mathbf{x}_{n+1}^* = \arg \min_{\mathbf{x} \in \mathcal{X}} \text{MMSE}(\{\mathbf{X}_n, \mathbf{x}\})$  !

⇒ prendre  $\mathbf{x}_{n+1}^* = \arg \max_{\mathbf{x} \in \mathcal{X}} \rho_n(\mathbf{x})$



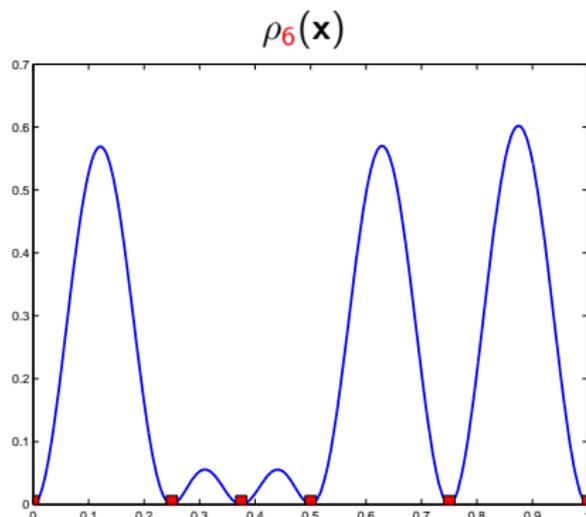
## Construction séquentielle d'un plan optimal :

Pour IMSE : rien de particulier, à l'étape  $n + 1$ ,  $\mathbf{X}_{n+1} = \{\mathbf{X}_n, \mathbf{x}_{n+1}\}$  avec

$$\mathbf{x}_{n+1}^* = \arg \min_{\mathbf{x} \in \mathcal{X}} \text{IMSE}(\{\mathbf{X}_n, \mathbf{x}\})$$

Pour MMSE : ne surtout pas choisir  $\mathbf{x}_{n+1}^* = \arg \min_{\mathbf{x} \in \mathcal{X}} \text{MMSE}(\{\mathbf{X}_n, \mathbf{x}\})$  !

⇒ prendre  $\mathbf{x}_{n+1}^* = \arg \max_{\mathbf{x} \in \mathcal{X}} \rho_n(\mathbf{x})$



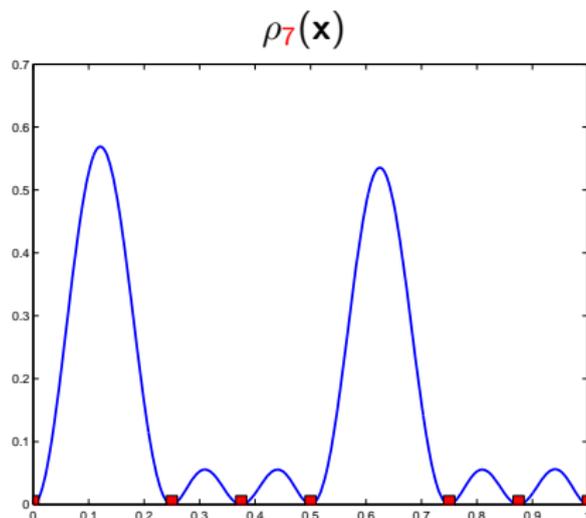
## Construction séquentielle d'un plan optimal :

Pour IMSE : rien de particulier, à l'étape  $n + 1$ ,  $\mathbf{X}_{n+1} = \{\mathbf{X}_n, \mathbf{x}_{n+1}\}$  avec

$$\mathbf{x}_{n+1}^* = \arg \min_{\mathbf{x} \in \mathcal{X}} \text{IMSE}(\{\mathbf{X}_n, \mathbf{x}\})$$

Pour MMSE : ne surtout pas choisir  $\mathbf{x}_{n+1}^* = \arg \min_{\mathbf{x} \in \mathcal{X}} \text{MMSE}(\{\mathbf{X}_n, \mathbf{x}\})$  !

⇒ prendre  $\mathbf{x}_{n+1}^* = \arg \max_{\mathbf{x} \in \mathcal{X}} \rho_n(\mathbf{x})$



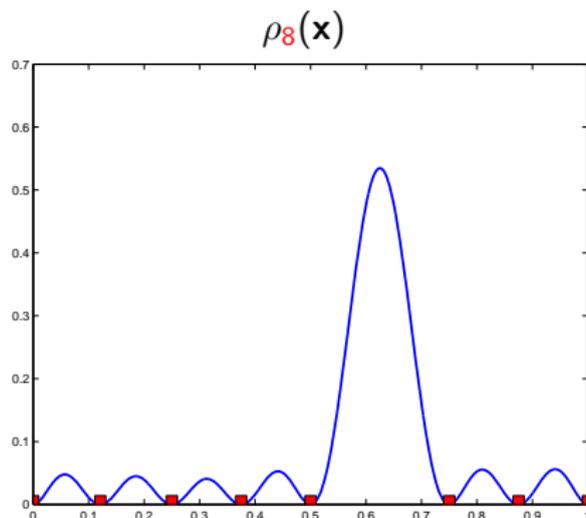
## Construction séquentielle d'un plan optimal :

Pour IMSE : rien de particulier, à l'étape  $n + 1$ ,  $\mathbf{X}_{n+1} = \{\mathbf{X}_n, \mathbf{x}_{n+1}\}$  avec

$$\mathbf{x}_{n+1}^* = \arg \min_{\mathbf{x} \in \mathcal{X}} \text{IMSE}(\{\mathbf{X}_n, \mathbf{x}\})$$

Pour MMSE : ne surtout pas choisir  $\mathbf{x}_{n+1}^* = \arg \min_{\mathbf{x} \in \mathcal{X}} \text{MMSE}(\{\mathbf{X}_n, \mathbf{x}\})$  !

⇒ prendre  $\mathbf{x}_{n+1}^* = \arg \max_{\mathbf{x} \in \mathcal{X}} \rho_n(\mathbf{x})$



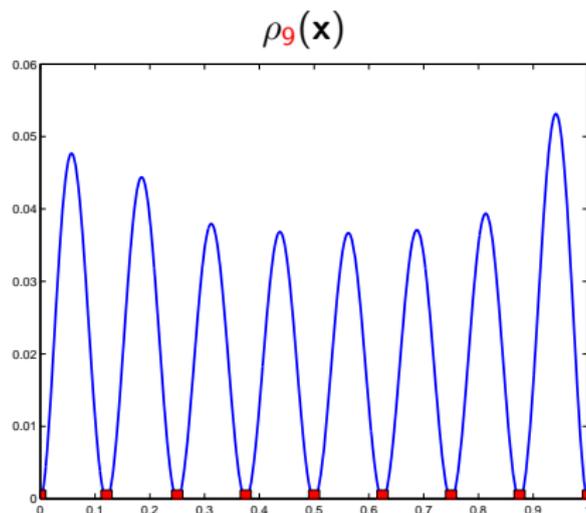
## Construction séquentielle d'un plan optimal :

Pour IMSE : rien de particulier, à l'étape  $n + 1$ ,  $\mathbf{X}_{n+1} = \{\mathbf{X}_n, \mathbf{x}_{n+1}\}$  avec

$$\mathbf{x}_{n+1}^* = \arg \min_{\mathbf{x} \in \mathcal{X}} \text{IMSE}(\{\mathbf{X}_n, \mathbf{x}\})$$

Pour MMSE : ne surtout pas choisir  $\mathbf{x}_{n+1}^* = \arg \min_{\mathbf{x} \in \mathcal{X}} \text{MMSE}(\{\mathbf{X}_n, \mathbf{x}\})$  !

⇒ prendre  $\mathbf{x}_{n+1}^* = \arg \max_{\mathbf{x} \in \mathcal{X}} \rho_n(\mathbf{x})$



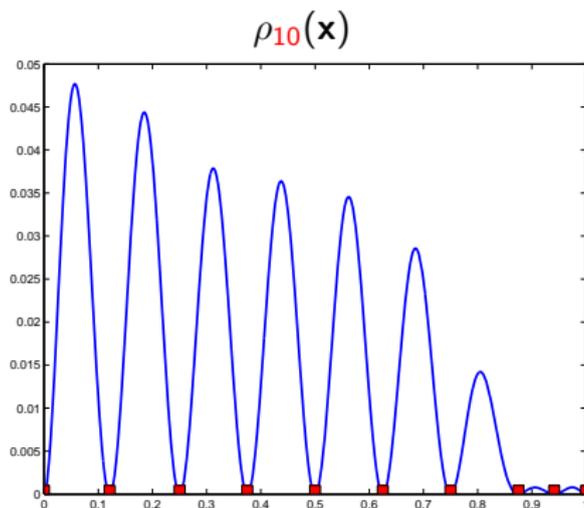
## Construction séquentielle d'un plan optimal :

Pour IMSE : rien de particulier, à l'étape  $n + 1$ ,  $\mathbf{X}_{n+1} = \{\mathbf{X}_n, \mathbf{x}_{n+1}\}$  avec

$$\mathbf{x}_{n+1}^* = \arg \min_{\mathbf{x} \in \mathcal{X}} \text{IMSE}(\{\mathbf{X}_n, \mathbf{x}\})$$

Pour MMSE : ne surtout pas choisir  $\mathbf{x}_{n+1}^* = \arg \min_{\mathbf{x} \in \mathcal{X}} \text{MMSE}(\{\mathbf{X}_n, \mathbf{x}\})$  !

⇒ prendre  $\mathbf{x}_{n+1}^* = \arg \max_{\mathbf{x} \in \mathcal{X}} \rho_n(\mathbf{x})$



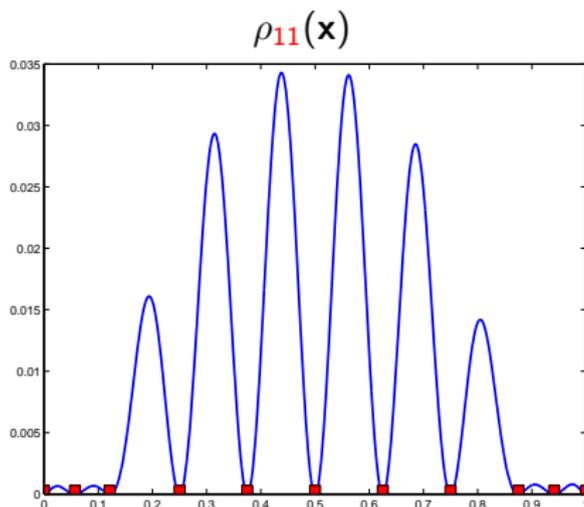
## Construction séquentielle d'un plan optimal :

Pour IMSE : rien de particulier, à l'étape  $n + 1$ ,  $\mathbf{X}_{n+1} = \{\mathbf{X}_n, \mathbf{x}_{n+1}\}$  avec

$$\mathbf{x}_{n+1}^* = \arg \min_{\mathbf{x} \in \mathcal{X}} \text{IMSE}(\{\mathbf{X}_n, \mathbf{x}\})$$

Pour MMSE : ne surtout pas choisir  $\mathbf{x}_{n+1}^* = \arg \min_{\mathbf{x} \in \mathcal{X}} \text{MMSE}(\{\mathbf{X}_n, \mathbf{x}\})$  !

⇒ prendre  $\mathbf{x}_{n+1}^* = \arg \max_{\mathbf{x} \in \mathcal{X}} \rho_n(\mathbf{x})$



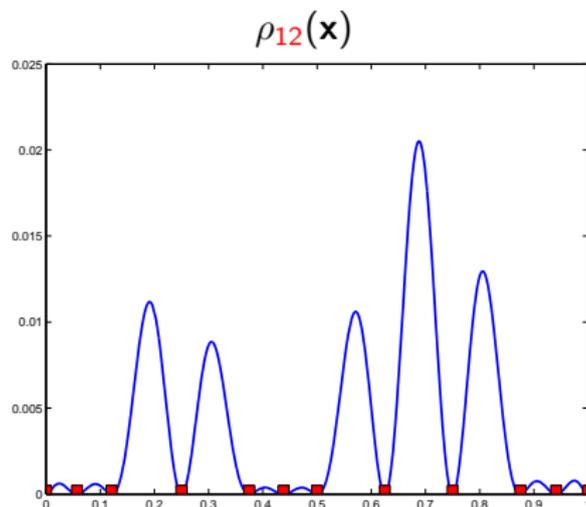
## Construction séquentielle d'un plan optimal :

Pour IMSE : rien de particulier, à l'étape  $n + 1$ ,  $\mathbf{X}_{n+1} = \{\mathbf{X}_n, \mathbf{x}_{n+1}\}$  avec

$$\mathbf{x}_{n+1}^* = \arg \min_{\mathbf{x} \in \mathcal{X}} \text{IMSE}(\{\mathbf{X}_n, \mathbf{x}\})$$

Pour MMSE : ne surtout pas choisir  $\mathbf{x}_{n+1}^* = \arg \min_{\mathbf{x} \in \mathcal{X}} \text{MMSE}(\{\mathbf{X}_n, \mathbf{x}\})$  !

⇒ prendre  $\mathbf{x}_{n+1}^* = \arg \max_{\mathbf{x} \in \mathcal{X}} \rho_n(\mathbf{x})$



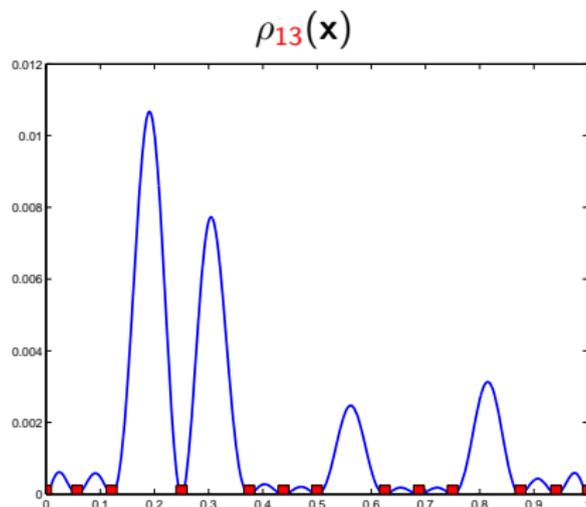
## Construction séquentielle d'un plan optimal :

Pour IMSE : rien de particulier, à l'étape  $n + 1$ ,  $\mathbf{X}_{n+1} = \{\mathbf{X}_n, \mathbf{x}_{n+1}\}$  avec

$$\mathbf{x}_{n+1}^* = \arg \min_{\mathbf{x} \in \mathcal{X}} \text{IMSE}(\{\mathbf{X}_n, \mathbf{x}\})$$

Pour MMSE : ne surtout pas choisir  $\mathbf{x}_{n+1}^* = \arg \min_{\mathbf{x} \in \mathcal{X}} \text{MMSE}(\{\mathbf{X}_n, \mathbf{x}\})$  !

⇒ prendre  $\mathbf{x}_{n+1}^* = \arg \max_{\mathbf{x} \in \mathcal{X}} \rho_n(\mathbf{x})$



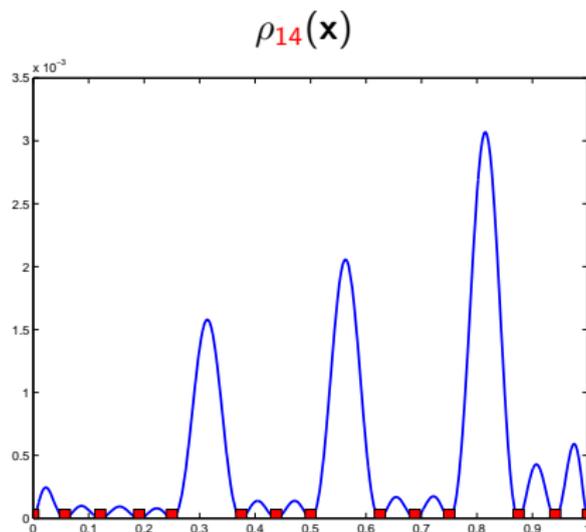
## Construction séquentielle d'un plan optimal :

Pour IMSE : rien de particulier, à l'étape  $n + 1$ ,  $\mathbf{X}_{n+1} = \{\mathbf{X}_n, \mathbf{x}_{n+1}\}$  avec

$$\mathbf{x}_{n+1}^* = \arg \min_{\mathbf{x} \in \mathcal{X}} \text{IMSE}(\{\mathbf{X}_n, \mathbf{x}\})$$

Pour MMSE : ne surtout pas choisir  $\mathbf{x}_{n+1}^* = \arg \min_{\mathbf{x} \in \mathcal{X}} \text{MMSE}(\{\mathbf{X}_n, \mathbf{x}\})$  !

⇒ prendre  $\mathbf{x}_{n+1}^* = \arg \max_{\mathbf{x} \in \mathcal{X}} \rho_n(\mathbf{x})$



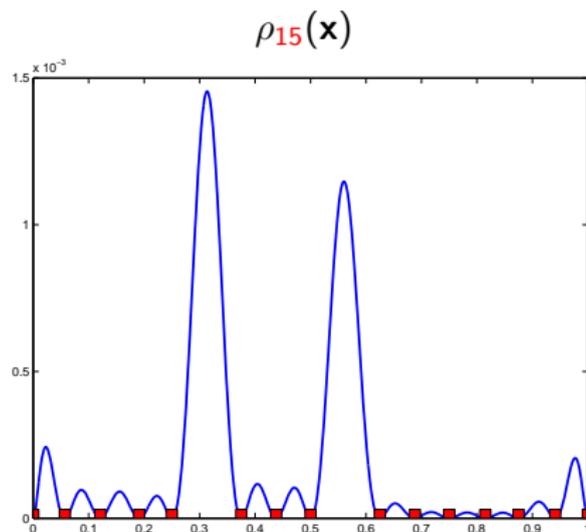
## Construction séquentielle d'un plan optimal :

Pour IMSE : rien de particulier, à l'étape  $n + 1$ ,  $\mathbf{X}_{n+1} = \{\mathbf{X}_n, \mathbf{x}_{n+1}\}$  avec

$$\mathbf{x}_{n+1}^* = \arg \min_{\mathbf{x} \in \mathcal{X}} \text{IMSE}(\{\mathbf{X}_n, \mathbf{x}\})$$

Pour MMSE : ne surtout pas choisir  $\mathbf{x}_{n+1}^* = \arg \min_{\mathbf{x} \in \mathcal{X}} \text{MMSE}(\{\mathbf{X}_n, \mathbf{x}\})$  !

⇒ prendre  $\mathbf{x}_{n+1}^* = \arg \max_{\mathbf{x} \in \mathcal{X}} \rho_n(\mathbf{x})$



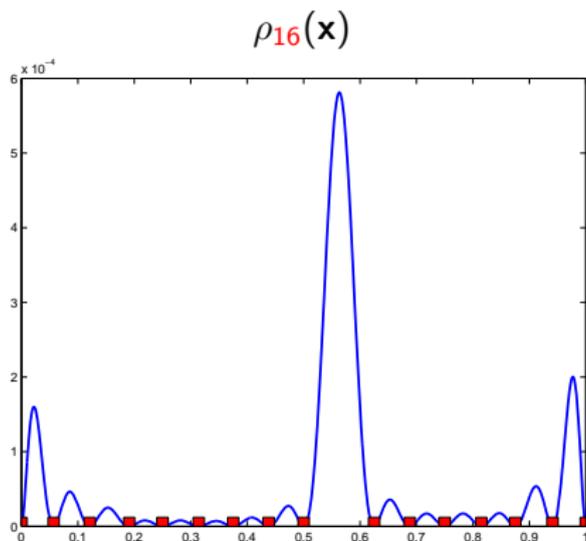
## Construction séquentielle d'un plan optimal :

Pour IMSE : rien de particulier, à l'étape  $n + 1$ ,  $\mathbf{X}_{n+1} = \{\mathbf{X}_n, \mathbf{x}_{n+1}\}$  avec

$$\mathbf{x}_{n+1}^* = \arg \min_{\mathbf{x} \in \mathcal{X}} \text{IMSE}(\{\mathbf{X}_n, \mathbf{x}\})$$

Pour MMSE : ne surtout pas choisir  $\mathbf{x}_{n+1}^* = \arg \min_{\mathbf{x} \in \mathcal{X}} \text{MMSE}(\{\mathbf{X}_n, \mathbf{x}\})$  !

⇒ prendre  $\mathbf{x}_{n+1}^* = \arg \max_{\mathbf{x} \in \mathcal{X}} \rho_n(\mathbf{x})$



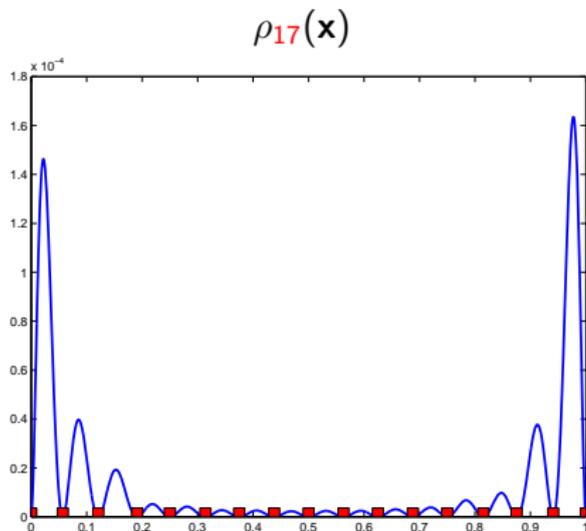
## Construction séquentielle d'un plan optimal :

Pour IMSE : rien de particulier, à l'étape  $n + 1$ ,  $\mathbf{X}_{n+1} = \{\mathbf{X}_n, \mathbf{x}_{n+1}\}$  avec

$$\mathbf{x}_{n+1}^* = \arg \min_{\mathbf{x} \in \mathcal{X}} \text{IMSE}(\{\mathbf{X}_n, \mathbf{x}\})$$

Pour MMSE : ne surtout pas choisir  $\mathbf{x}_{n+1}^* = \arg \min_{\mathbf{x} \in \mathcal{X}} \text{MMSE}(\{\mathbf{X}_n, \mathbf{x}\})$  !

⇒ prendre  $\mathbf{x}_{n+1}^* = \arg \max_{\mathbf{x} \in \mathcal{X}} \rho_n(\mathbf{x})$



### 1.3 Échantillonnage à maximum d'entropie (*Maximum Entropy Sampling*) (Shewry and Wynn, 1987)

Sans tendance ( $\mathbf{r}(\mathbf{x}) = \mathbf{0} \forall \mathbf{x}$ ) :

- $\mathbf{z}_Q \triangleq$  le vecteur des  $Z(\mathbf{x}^{(k)})$ ,  $\mathbf{x}^{(k)} \in \mathcal{X}_Q$
  - $\mathbf{z}_n \triangleq$  le vecteur des  $Z(\mathbf{x}_i)$ ,  $i = 1, \dots, n$  (les observations)
  - $H_1(\mathbf{z}) \triangleq - \int \varphi(\mathbf{z}) \log[\varphi(\mathbf{z})] d\mathbf{z}$  entropie (Shannon) de la loi  $\varphi(\mathbf{z})$   
= mesure de "dispersion"
- $H_1(\mathbf{z}_1|\mathbf{z}_2) \triangleq$  entropie de la loi conditionnelle  $\varphi_{1|2}(\mathbf{z}_1|\mathbf{z}_2)$

## 1.3 Échantillonnage à maximum d'entropie (*Maximum Entropy Sampling*) (Shewry and Wynn, 1987)

Sans tendance ( $\mathbf{r}(\mathbf{x}) = \mathbf{0} \forall \mathbf{x}$ ) :

- ▶  $\mathbf{z}_Q \triangleq$  le vecteur des  $Z(\mathbf{x}^{(k)})$ ,  $\mathbf{x}^{(k)} \in \mathcal{X}_Q$
- ▶  $\mathbf{z}_n \triangleq$  le vecteur des  $Z(\mathbf{x}_i)$ ,  $i = 1, \dots, n$  (les observations)
- ▶  $H_1(\mathbf{z}) \triangleq - \int \varphi(\mathbf{z}) \log[\varphi(\mathbf{z})] d\mathbf{z}$  entropie (Shannon) de la loi  $\varphi(\mathbf{z})$   
= mesure de "dispersion"
- $H_1(\mathbf{z}_1|\mathbf{z}_2) \triangleq$  entropie de la loi conditionnelle  $\varphi_{1|2}(\mathbf{z}_1|\mathbf{z}_2)$

Alors, 
$$H_1(\mathbf{y}_Q) = H_1(\mathbf{y}_n) + E\{H_1(\mathbf{y}_Q|\mathbf{y}_n)\}$$

### 1.3 Échantillonnage à maximum d'entropie (*Maximum Entropy Sampling*) (Shewry and Wynn, 1987)

Sans tendance ( $\mathbf{r}(\mathbf{x}) = \mathbf{0} \forall \mathbf{x}$ ) :

- $\mathbf{z}_Q \triangleq$  le vecteur des  $Z(\mathbf{x}^{(k)})$ ,  $\mathbf{x}^{(k)} \in \mathcal{X}_Q$
  - $\mathbf{z}_n \triangleq$  le vecteur des  $Z(\mathbf{x}_i)$ ,  $i = 1, \dots, n$  (les observations)
  - $H_1(\mathbf{z}) \triangleq - \int \varphi(\mathbf{z}) \log[\varphi(\mathbf{z})] d\mathbf{z}$  entropie (Shannon) de la loi  $\varphi(\mathbf{z})$   
= mesure de "dispersion"
- $H_1(\mathbf{z}_1|\mathbf{z}_2) \triangleq$  entropie de la loi conditionnelle  $\varphi_{1|2}(\mathbf{z}_1|\mathbf{z}_2)$

Alors, 
$$\boxed{\underbrace{H_1(\mathbf{y}_Q)}_{=\text{constante}} = H_1(\mathbf{y}_n) + \underbrace{E\{H_1(\mathbf{y}_Q|\mathbf{y}_n)\}}_{\text{à minimiser}}}$$

### 1.3 Échantillonnage à maximum d'entropie (*Maximum Entropy Sampling*) (Shewry and Wynn, 1987)

Sans tendance ( $\mathbf{r}(\mathbf{x}) = \mathbf{0} \forall \mathbf{x}$ ) :

- $\mathbf{z}_Q \triangleq$  le vecteur des  $Z(\mathbf{x}^{(k)})$ ,  $\mathbf{x}^{(k)} \in \mathcal{X}_Q$
- $\mathbf{z}_n \triangleq$  le vecteur des  $Z(\mathbf{x}_i)$ ,  $i = 1, \dots, n$  (les observations)
- $H_1(\mathbf{z}) \triangleq - \int \varphi(\mathbf{z}) \log[\varphi(\mathbf{z})] d\mathbf{z}$  entropie (Shannon) de la loi  $\varphi(\mathbf{z})$   
= mesure de "dispersion"

$H_1(\mathbf{z}_1|\mathbf{z}_2) \triangleq$  entropie de la loi conditionnelle  $\varphi_{1|2}(\mathbf{z}_1|\mathbf{z}_2)$

Alors, 
$$\underbrace{H_1(\mathbf{y}_Q)}_{=\text{constante}} = H_1(\mathbf{y}_n) + \underbrace{E\{H_1(\mathbf{y}_Q|\mathbf{y}_n)\}}_{\text{à minimiser}}$$

Minimiser  $E\{H_1(\mathbf{y}_Q|\mathbf{y}_n)\}$  par rapport à  $\mathbf{X}_n \Leftrightarrow$  maximiser  $H_1(\mathbf{y}_n)$

Z(x) supposé Gaussien  $\Rightarrow$  maximiser  $\det[\mathbf{C}_n]$

**= critère intra-distances**

### Construction séquentielle d'un plan optimal :

$$\mathbf{x}_{n+1} = \arg \max_{\mathbf{x} \in \mathcal{X}} \det[\mathbf{C}_{n+1}] = \arg \max_{\mathbf{x} \in \mathcal{X}} \underbrace{\det \begin{bmatrix} \mathbf{C}_n & \mathbf{c}_n(\mathbf{x}) \\ \mathbf{c}_n^\top(\mathbf{x}) & 1 \end{bmatrix}}_{= \det[\mathbf{C}_n] (1 - \underbrace{\mathbf{c}_n^\top(\mathbf{x}) \mathbf{C}_n^{-1} \mathbf{c}_n(\mathbf{x})}_{=\rho_n(\mathbf{x})})}$$

$$\Rightarrow \boxed{\mathbf{x}_{n+1} = \arg \max_{\mathbf{x} \in \mathcal{X}} \rho_n(\mathbf{x})}$$

### Construction séquentielle d'un plan optimal :

$$\mathbf{x}_{n+1} = \arg \max_{\mathbf{x} \in \mathcal{X}} \det[\mathbf{C}_{n+1}] = \arg \max_{\mathbf{x} \in \mathcal{X}} \underbrace{\det \begin{bmatrix} \mathbf{C}_n & \mathbf{c}_n(\mathbf{x}) \\ \mathbf{c}_n^\top(\mathbf{x}) & 1 \end{bmatrix}}_{= \det[\mathbf{C}_n] (1 - \underbrace{\mathbf{c}_n^\top(\mathbf{x}) \mathbf{C}_n^{-1} \mathbf{c}_n(\mathbf{x})}_{=\rho_n(\mathbf{x})})}$$

$$\Rightarrow \boxed{\mathbf{x}_{n+1} = \arg \max_{\mathbf{x} \in \mathcal{X}} \rho_n(\mathbf{x})}$$

### Les plans obtenus sont typiquement *space-filling* :

Johnson et al. (1990) : si  $C(\mathbf{x} - \mathbf{x}') = c(\|\mathbf{x} - \mathbf{x}'\|)$  avec  $c(\cdot)$  décroissante, alors  $\mathbf{X}_n^*$  optimal pour  $\Phi_{Mm}(\cdot)$  (Maximin optimal) tend à être optimal pour  $\det[\mathbf{C}_n]$  avec la covariance  $C_a(\mathbf{x} - \mathbf{x}') = [C(\mathbf{x} - \mathbf{x}')]^a$  quand  $a \rightarrow \infty$

$$\Rightarrow \boxed{\text{il y a des points } \mathbf{x}_i \text{ sur les bords de } \mathcal{X}}$$

Prenons  $C_a(\mathbf{x} - \mathbf{x}') = \exp(-a\|\mathbf{x} - \mathbf{x}'\|^2)$  :

Pour  $a$  assez grand,  $\mathbf{C}_n$  est à diagonale strict. dominante :

$$\Delta_i \triangleq \sum_{j \neq i} |C_a(\mathbf{x}_i - \mathbf{x}_j)| < C_a(0) = 1$$

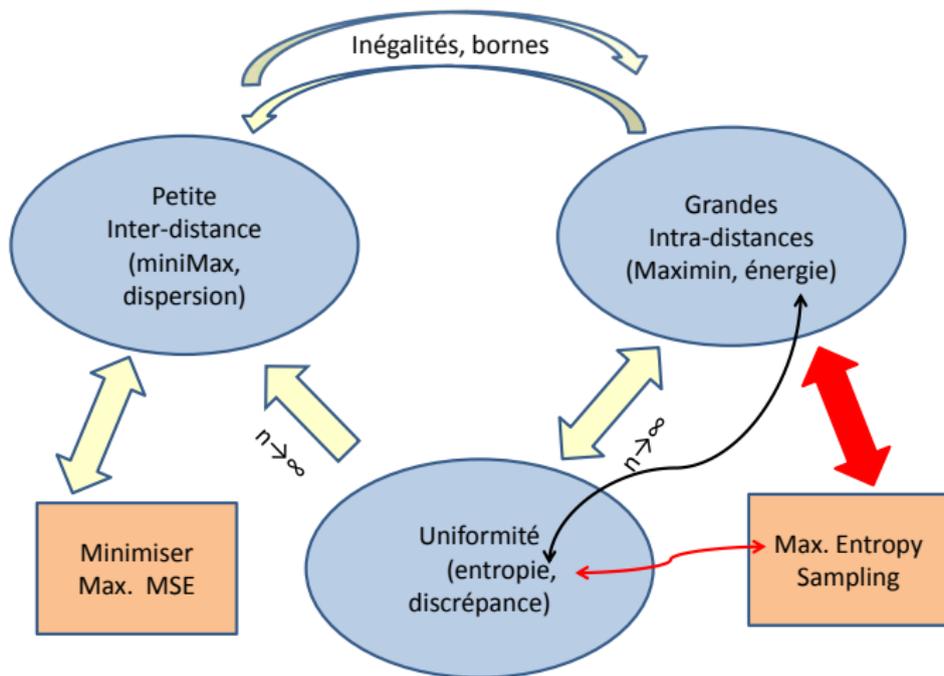
$$\begin{aligned} \Rightarrow 0 > \log \det(\mathbf{C}_n) &> \sum_{i=1}^n \log(1 - \Delta_i) \simeq - \sum_{i \neq j} C_a(\mathbf{x}_i - \mathbf{x}_j) \\ &= n + n^2(\pi/a)^{d/2} [H_2(\hat{\varphi}_n) - 1] \end{aligned}$$

(> : Th. Gershgorin, voir (Joseph et al., 2015))

avec  $H_2(\hat{\varphi}_n)$  entropie de Tsallis d'ordre 2 de l'estimateur à noyau

$\hat{\varphi}_n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n K_{1/(4a)}(\|\mathbf{x} - \mathbf{x}_i\|)$  et  $K_{\sigma^2}(\cdot) =$  densité de  $\mathcal{N}(0, \sigma^2)$ , voir § I-2.1

⇒ maximiser  $\det[\mathbf{C}_n] \approx$  maximiser  $H_2(\hat{\varphi}_n)$



Voir (Ko et al., 1995) pour calcul plan optimal  $\mathbf{X}_n^*$

Avec tendance ( $f(\mathbf{x}) = \mathbf{r}^\top(\mathbf{x})\boldsymbol{\beta} + Z(\mathbf{x})$ ) :

maximiser  $\det[\mathbf{C}_n] \det[\mathbf{R}_n^\top \mathbf{C}_n^{-1} \mathbf{R}_n]$  avec

$$\mathbf{R}_n = \begin{pmatrix} \mathbf{r}^\top(\mathbf{x}_1) \\ \vdots \\ \mathbf{r}^\top(\mathbf{x}_n) \end{pmatrix}$$

voir (Santner et al., 2003, Chap. 6)

## 2 Plans optimaux pour régression linéaire

### 2.1 Régression linéaire

Observations  $y_i = y(\mathbf{x}_i) = \mathbf{r}^\top(\mathbf{x}_i)\boldsymbol{\gamma} + \varepsilon_i$ ,  $\boldsymbol{\gamma} \in \mathbb{R}^p$   
avec  $(\varepsilon_i)$  i.i.d.,  $E\{\varepsilon_i\} = 0$ ,  $\text{var}\{\varepsilon_i\} = \sigma^2 \forall i$

Estimation de  $\boldsymbol{\gamma}$  par moindres carrés (MC) (*Least-Squares (LS)*)

$$\hat{\boldsymbol{\gamma}}_n = (\mathbf{R}_n^\top \mathbf{R}_n)^{-1} \mathbf{R}_n^\top \mathbf{y}_n, \text{ avec } \mathbf{y}_n = (y_1, \dots, y_n)^\top \text{ et } \mathbf{R}_n = \begin{pmatrix} \mathbf{r}^\top(\mathbf{x}_1) \\ \vdots \\ \mathbf{r}^\top(\mathbf{x}_n) \end{pmatrix}$$

$E\{\hat{\boldsymbol{\gamma}}_n\} = \boldsymbol{\gamma}$  (estimateur non biaisé)

## 2 Plans optimaux pour régression linéaire

### 2.1 Régression linéaire

Observations  $y_i = y(\mathbf{x}_i) = \mathbf{r}^\top(\mathbf{x}_i)\boldsymbol{\gamma} + \varepsilon_i$ ,  $\boldsymbol{\gamma} \in \mathbb{R}^p$   
 avec  $(\varepsilon_i)$  i.i.d.,  $E\{\varepsilon_i\} = 0$ ,  $\text{var}\{\varepsilon_i\} = \sigma^2 \forall i$

Estimation de  $\boldsymbol{\gamma}$  par moindres carrés (MC) (*Least-Squares (LS)*)

$$\hat{\boldsymbol{\gamma}}_n = (\mathbf{R}_n^\top \mathbf{R}_n)^{-1} \mathbf{R}_n^\top \mathbf{y}_n, \text{ avec } \mathbf{y}_n = (y_1, \dots, y_n)^\top \text{ et } \mathbf{R}_n = \begin{pmatrix} \mathbf{r}^\top(\mathbf{x}_1) \\ \vdots \\ \mathbf{r}^\top(\mathbf{x}_n) \end{pmatrix}$$

$E\{\hat{\boldsymbol{\gamma}}_n\} = \boldsymbol{\gamma}$  (estimateur non biaisé)

$$\text{Covariance} = \text{cov}(\hat{\boldsymbol{\gamma}}_n) = \sigma^2 (\mathbf{R}_n^\top \mathbf{R}_n)^{-1} = \frac{\sigma^2}{n} \underbrace{\left[ \sum_{i=1}^n \mathbf{r}(\mathbf{x}_i) \mathbf{r}^\top(\mathbf{x}_i) \right]}_{\mathbf{M}_n}^{-1}$$

$$\text{cov}(\hat{\boldsymbol{\gamma}}_n) = \frac{\sigma^2}{n} \mathbf{M}_n^{-1}, \text{ avec}$$

$$\mathbf{M}_n = \mathbf{M}(\mathbf{X}_n) = \frac{1}{n} \sum_{i=1}^n \mathbf{r}(\mathbf{x}_i) \mathbf{r}^\top(\mathbf{x}_i) \in \mathbb{R}^{p \times p}$$

= matrice d'information (moyenne par observation)

$$\text{cov}(\hat{\gamma}_n) = \frac{\sigma^2}{n} \mathbf{M}_n^{-1}, \text{ avec}$$

$$\mathbf{M}_n = \mathbf{M}(\mathbf{X}_n) = \frac{1}{n} \sum_{i=1}^n \mathbf{r}(\mathbf{x}_i) \mathbf{r}^\top(\mathbf{x}_i) \in \mathbb{R}^{p \times p}$$

= matrice d'information (moyenne par observation)

Plan d'expérience optimal  $\mathbf{X}_n^*$  : maximise une fonction scalaire  $\Phi(\cdot)$  de  $\mathbf{M}_n$  (avec  $\Phi(\cdot)$  croissante pour Loewner)

- **E-optimalité** : maximiser  $\lambda_{\min}(\mathbf{M}_n)$   
(minimiser le plus grand axe des ellipsoïdes de confiance sur  $\gamma$ )

$$\text{cov}(\hat{\gamma}_n) = \frac{\sigma^2}{n} \mathbf{M}_n^{-1}, \text{ avec}$$

$$\mathbf{M}_n = \mathbf{M}(\mathbf{X}_n) = \frac{1}{n} \sum_{i=1}^n \mathbf{r}(\mathbf{x}_i) \mathbf{r}^\top(\mathbf{x}_i) \in \mathbb{R}^{p \times p}$$

= matrice d'information (moyenne par observation)

Plan d'expérience optimal  $\mathbf{X}_n^*$  : maximise une fonction scalaire  $\Phi(\cdot)$  de  $\mathbf{M}_n$  (avec  $\Phi(\cdot)$  croissante pour Loewner)

- **E-optimalité** : maximiser  $\lambda_{\min}(\mathbf{M}_n)$   
(minimiser le plus grand axe des ellipsoïdes de confiance sur  $\gamma$ )
- **A-optimalité** : maximiser  $-\text{trace}[\mathbf{M}_n^{-1}] \Leftrightarrow$  maximiser  $1/\text{trace}[\mathbf{M}_n^{-1}]$   
(minimiser la somme du carré des longueur des axes des ellipsoïdes de confiance sur  $\gamma$ )

$$\text{cov}(\hat{\gamma}_n) = \frac{\sigma^2}{n} \mathbf{M}_n^{-1}, \text{ avec}$$

$$\mathbf{M}_n = \mathbf{M}(\mathbf{X}_n) = \frac{1}{n} \sum_{i=1}^n \mathbf{r}(\mathbf{x}_i) \mathbf{r}^\top(\mathbf{x}_i) \in \mathbb{R}^{p \times p}$$

= matrice d'information (moyenne par observation)

Plan d'expérience optimal  $\mathbf{X}_n^*$  : maximise une fonction scalaire  $\Phi(\cdot)$  de  $\mathbf{M}_n$  (avec  $\Phi(\cdot)$  croissante pour Loewner)

- **E-optimality** : maximiser  $\lambda_{\min}(\mathbf{M}_n)$   
(minimiser le plus grand axe des ellipsoïdes de confiance sur  $\gamma$ )
- **A-optimality** : maximiser  $-\text{trace}[\mathbf{M}_n^{-1}] \Leftrightarrow$  maximiser  $1/\text{trace}[\mathbf{M}_n^{-1}]$   
(minimiser la somme du carré des longueur des axes des ellipsoïdes de confiance sur  $\gamma$ )
- plus généralement, **L-optimality** : maximiser  $-\text{trace}[\mathbf{L}\mathbf{M}_n^{-1}]$   
(on considérera seulement le cas  $\mathbf{L}$  symétrique définie positive)

- **D-optimalité** : maximiser  $\log \det \mathbf{M}_n$   
(minimiser le volume des ellipsoïdes de confiance sur  $\gamma$ )  
Très utilisé :
  - un plan *D*-optimal est invariant par reparamétrisation

$$\det \mathbf{M}'_n(\beta(\theta)) = \det \mathbf{M}_n(\theta) \det^{-2} \left( \frac{\partial \beta}{\partial \theta^\top} \right)$$

- $\Rightarrow$  très souvent répétitions d'un petit nombre de conditions expérimentales différentes (on a supposé les  $\varepsilon_i$  i.i.d.  $\Rightarrow$  plusieurs observations au même  $\mathbf{x}_i$  apportent de l'information)

► Modèles produits tensoriels

## 2.2 Plans exacts

$n$  observations en  $\mathbf{x}_1, \dots, \mathbf{x}_n$ ,  $\mathbf{x}_i \in \mathbb{R}^d$

Maximiser  $\Phi(\mathbf{M}_n)$  par rapport à  $\mathbf{X}_n = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \in \mathbb{R}^{n \times d}$

avec  $\mathbf{M}_n = \mathbf{M}(\mathbf{X}_n) = \frac{1}{n} \sum_{i=1}^n \mathbf{r}(\mathbf{x}_i) \mathbf{r}^\top(\mathbf{x}_i)$

## 2.2 Plans exacts

$n$  observations en  $\mathbf{x}_1, \dots, \mathbf{x}_n$ ,  $\mathbf{x}_i \in \mathbb{R}^d$

Maximiser  $\Phi(\mathbf{M}_n)$  par rapport à  $\mathbf{X}_n = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \in \mathbb{R}^{n \times d}$

avec  $\mathbf{M}_n = \mathbf{M}(\mathbf{X}_n) = \frac{1}{n} \sum_{i=1}^n \mathbf{r}(\mathbf{x}_i) \mathbf{r}^\top(\mathbf{x}_i)$

- Si la dimension  $n \times d$  du problème n'est pas trop grande
  - ➡ algorithme "standard" (mais attention aux contraintes et aux optimas locaux !)

## 2.2 Plans exacts

$n$  observations en  $\mathbf{x}_1, \dots, \mathbf{x}_n$ ,  $\mathbf{x}_i \in \mathbb{R}^d$

Maximiser  $\Phi(\mathbf{M}_n)$  par rapport à  $\mathbf{X}_n = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \in \mathbb{R}^{n \times d}$

avec  $\mathbf{M}_n = \mathbf{M}(\mathbf{X}_n) = \frac{1}{n} \sum_{i=1}^n \mathbf{r}(\mathbf{x}_i) \mathbf{r}^\top(\mathbf{x}_i)$

- Si la dimension  $n \times d$  du problème n'est pas trop grande
  - ➡ algorithme "standard" (mais attention aux contraintes et aux optimas locaux !)
- Si  $d$  est grand ➡ algorithme spécifique

**Méthode d'échange** : à l'itération  $k$ , échanger **un** point de support  $\mathbf{x}_j$  par un "meilleur point"  $\mathbf{x}^*$  dans  $\mathcal{X}$  (meilleur au sens de  $\Phi(\mathbf{M}_n)$ )

$$\mathbf{X}_n^k = (\mathbf{x}_1, \dots, \boxed{\begin{array}{c} \mathbf{x}_j \\ \updownarrow \\ \mathbf{x}^* \end{array}}, \dots, \mathbf{x}_n)$$

## Algorithme de Fedorov (1972) :

A chaque itération  $k$ , considérer successivement les  $n$  échanges possible, en partant chaque fois de  $\mathbf{X}_n^k$   
retenir le “meilleur” d'entre eux  $\Rightarrow \mathbf{X}_n^{k+1}$

$$\mathbf{X}_n^k = ( \mathbf{x}_1, \dots, \mathbf{x}_j, \dots, \mathbf{x}_n )$$

$\begin{array}{ccc} \updownarrow & & \updownarrow \\ \mathbf{x}_1^* & & \mathbf{x}_j^* \end{array}$

## Algorithme de Fedorov (1972) :

A chaque itération  $k$ , considérer successivement les  $n$  échanges possible, en partant chaque fois de  $\mathbf{X}_n^k$   
 retenir le “meilleur” d'entre eux  $\Rightarrow \mathbf{X}_n^{k+1}$

$$\mathbf{X}_n^k = ( \mathbf{x}_1, \dots, \mathbf{x}_j, \dots, \mathbf{x}_n )$$

$$\begin{array}{ccc} \updownarrow & & \updownarrow \\ \mathbf{x}_1^* & & \mathbf{x}_j^* \end{array} \quad \begin{array}{ccc} \updownarrow & & \updownarrow \\ \mathbf{x}_n^* & & \end{array}$$

Une itération =  $n$  optimisations en dimension  $d$   
 suivies du classement de  $n$  valeurs de critère

## Algorithme DETMAX de Mitchell (1974) :

Si une observation supplémentaire était possible : choix optimal

$$\mathbf{x}_{n+1}^{k+} = (\mathbf{x}_1, \dots, \mathbf{x}_j, \dots, \mathbf{x}_n, \mathbf{x}_{n+1}^*)$$

Il faut ensuite supprimer un point pour revenir à un plan à  $n$  points : considérer les  $n + 1$  suppressions possibles, retenir la moins pénalisante en terme de  $\Phi(\mathbf{M}_n)$

## Algorithme DETMAX de Mitchell (1974) :

Si une observation supplémentaire était possible : choix optimal

$$\mathbf{x}_{n+1}^{k+} = (\mathbf{x}_1, \dots, \mathbf{x}_j, \dots, \mathbf{x}_n, \mathbf{x}_{n+1}^*)$$

Il faut ensuite supprimer un point pour revenir à un plan à  $n$  points : considérer les  $n + 1$  suppressions possibles, retenir la moins pénalisante en terme de  $\Phi(\mathbf{M}_n)$

▣ globalement, on a échangé un  $\mathbf{x}_j$  avec  $\mathbf{x}_{n+1}^*$

(ici, excursion de longueur 1, des excursions plus longues sont possibles...)

Une itération = une optimisation en dimension  $d$

suivie du classement de  $n + 1$  valeurs de critère

- Itérations moins coûteuses pour DETMAX que pour Fedorov, mais il en faut souvent davantage

- Itérations moins coûteuses pour DETMAX que pour Fedorov, mais il en faut souvent davantage
- Blocage possible :
  - DETMAX: le point à supprimer est  $\mathbf{x}_{n+1}^*$
  - Fedorov: pas d'amélioration en optimisant **un seul**  $\mathbf{x}_i$  à la fois

- Itérations moins coûteuses pour DETMAX que pour Fedorov, mais il en faut souvent davantage
- Blocage possible :
  - DETMAX: le point à supprimer est  $\mathbf{x}_{n+1}^*$
  - Fedorov: pas d'amélioration en optimisant **un seul**  $\mathbf{x}_i$  à la fois
- ▲ les deux méthodes convergent vers un optimum local ▲  
(on peut combiner avec un recuit simulé)

- Itérations moins coûteuses pour DETMAX que pour Fedorov, mais il en faut souvent davantage
- Blocage possible :
  - DETMAX: le point à supprimer est  $\mathbf{x}_{n+1}^*$
  - Fedorov: pas d'amélioration en optimisant **un seul**  $\mathbf{x}_i$  à la fois
- ▲ les deux méthodes convergent vers un optimum local ▲  
(on peut combiner avec un recuit simulé)
- Autres méthodes :
  - *Branch and bound* : convergence garantie vers l'optimum, mais plutôt compliqué (et lent) (Welch, 1982)
  - Arrondi d'un **plan approximatif** (défini par ses points de support  $\mathbf{x}_i$  et les masses associées  $w_i^*$ ,  $i = 1, \dots, m$ ) :  
choisir  $n$  entiers  $r_i$  ( $r_i =$  nb. de répétitions d'observations en  $\mathbf{x}_i$ ) tels que  $\sum_{i=1}^m r_i = n$  et  $r_i/n \approx w_i^*$  (voir par ex. (Pukelsheim and Reider, 1992))

## 2.3 Plans approximatifs (*approximate design theory*)

(Chernoff, 1953; Kiefer and Wolfowitz, 1960; Fedorov, 1972; Silvey, 1980; Pukelsheim, 1993) ...

$$\mathbf{M}_n = \mathbf{M}(\mathbf{X}_n) = \frac{1}{n} \sum_{i=1}^n \mathbf{r}(\mathbf{x}_i) \mathbf{r}^\top(\mathbf{x}_i)$$

la forme additive est ici essentielle (liée à l'indépendance des erreurs  $\varepsilon_i$ )

## 2.3 Plans approximatifs (*approximate design theory*)

(Chernoff, 1953; Kiefer and Wolfowitz, 1960; Fedorov, 1972; Silvey, 1980; Pukelsheim, 1993) ...

$$\mathbf{M}_n = \mathbf{M}(\mathbf{X}_n) = \frac{1}{n} \sum_{i=1}^n \mathbf{r}(\mathbf{x}_i) \mathbf{r}^\top(\mathbf{x}_i)$$

la forme additive est ici essentielle (liée à l'indépendance des erreurs  $\varepsilon_i$ )

Si plusieurs  $\mathbf{x}_i$  coïncident (répétitions), avec seulement  $m < n$  différents  $\mathbf{x}_i$

$$\mathbf{M}(\mathbf{X}_n) = \sum_{i=1}^m \frac{r_i}{n} \mathbf{r}(\mathbf{x}_i) \mathbf{r}^\top(\mathbf{x}_i)$$

- $\frac{r_i}{n}$  = proportion d'observations réalisées en  $\mathbf{x}_i$
- = "pourcentage d'effort expérimental" en  $\mathbf{x}_i$
- = poids  $w_i$  du point de support  $\mathbf{x}_i$

$$\mathbf{M}(\mathbf{X}_n) = \sum_{i=1}^m w_i \mathbf{r}(\mathbf{x}_i) \mathbf{r}^\top(\mathbf{x}_i)$$

$$\Rightarrow \text{plan d'expérience } \mathbf{X}_n \Leftrightarrow \left\{ \begin{array}{ccc} \mathbf{x}_1 & \cdots & \mathbf{x}_m \\ w_1 & \cdots & w_m \end{array} \right\} \text{ avec } \sum_{i=1}^m w_i = 1$$

$\Rightarrow$  mesure de probabilité discrète sur les  $\mathbf{x}_i$ ,

avec les contraintes  $w_i = r_i/n$

$$\mathbf{M}(\mathbf{X}_n) = \sum_{i=1}^m w_i \mathbf{r}(\mathbf{x}_i) \mathbf{r}^\top(\mathbf{x}_i)$$

$$\Rightarrow \text{plan d'expérience } \mathbf{X}_n \Leftrightarrow \left\{ \begin{array}{ccc} \mathbf{x}_1 & \cdots & \mathbf{x}_m \\ w_1 & \cdots & w_m \end{array} \right\} \text{ avec } \sum_{i=1}^m w_i = 1$$

$\Rightarrow$  mesure de probabilité discrète sur les  $\mathbf{x}_i$ ,

avec les contraintes  $w_i = r_i/n$

$\Rightarrow$  relaxons les contraintes : utilisons seulement  $w_i \geq 0$  et  $\sum_{i=1}^m w_i = 1$

$\Rightarrow \xi =$  mesure de probabilité discrète sur  $\mathcal{X}$

points de support  $\mathbf{x}_i$  et masses associées  $w_i$

= "plan approximatif"

$$\mathbf{M}(\mathbf{X}_n) = \sum_{i=1}^m w_i \mathbf{r}(\mathbf{x}_i) \mathbf{r}^\top(\mathbf{x}_i)$$

⇒ plan d'expérience  $\mathbf{X}_n \Leftrightarrow \left\{ \begin{array}{ccc} \mathbf{x}_1 & \cdots & \mathbf{x}_m \\ w_1 & \cdots & w_m \end{array} \right\}$  avec  $\sum_{i=1}^m w_i = 1$

⇒ mesure de probabilité discrète sur les  $\mathbf{x}_i$ ,

avec les contraintes  $w_i = r_i/n$

⇒ relaxons les contraintes : utilisons seulement  $w_i \geq 0$  et  $\sum_{i=1}^m w_i = 1$

⇒  $\xi =$  mesure de probabilité discrète sur  $\mathcal{X}$

points de support  $\mathbf{x}_i$  et masses associées  $w_i$

= "plan approximatif"

Expression + générale :  $\xi =$  mesure de probabilité quelconque sur  $\mathcal{X}$

$$\mathbf{M}(\xi) = \int_{\mathcal{X}} \mathbf{r}(\mathbf{x}) \mathbf{r}^\top(\mathbf{x}) \xi(d\mathbf{x}) \text{ avec } \int_{\mathcal{X}} \xi(d\mathbf{x}) = 1$$

$\mathbf{M}(\xi) \in$  enveloppe convexe de  $\mathcal{M}$  = ensemble des matrices de rang 1

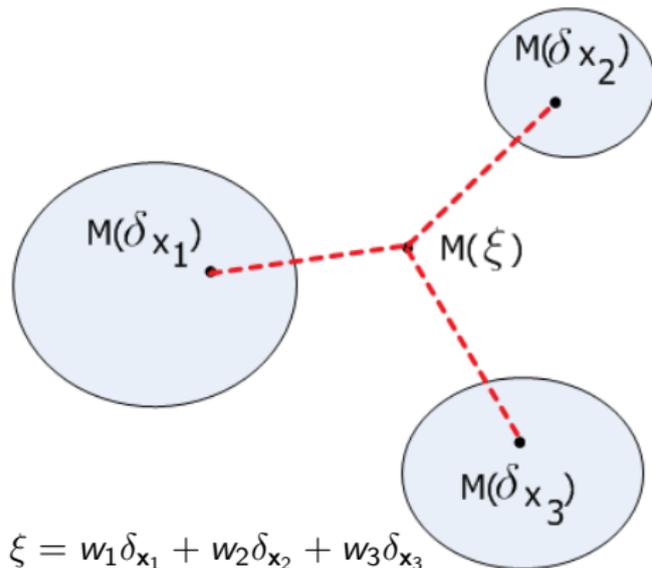
$$\mathbf{M}(\delta_x) = \mathbf{r}(\mathbf{x})\mathbf{r}^T(\mathbf{x})$$

$\mathbf{M}(\xi)$  est symétrique  $p \times p \Rightarrow$  dans un espace de dimension  $q = \frac{p(p+1)}{2}$

$\mathbf{M}(\xi) \in$  enveloppe convexe de  $\mathcal{M} =$  ensemble des matrices de rang 1

$$\mathbf{M}(\delta_x) = \mathbf{r}(\mathbf{x})\mathbf{r}^T(\mathbf{x})$$

$\mathbf{M}(\xi)$  est symétrique  $p \times p \Rightarrow$  dans un espace de dimension  $q = \frac{p(p+1)}{2}$



$$\xi = w_1\delta_{x_1} + w_2\delta_{x_2} + w_3\delta_{x_3}$$

## Théorème de Caratheodory :

$\mathbf{M}(\xi)$  peut s'écrire comme la combinaison linéaire d'au plus  $q + 1$  éléments de  $\mathcal{M}$ :

$$\mathbf{M}(\xi) = \sum_{i=1}^m w_i \mathbf{r}(\mathbf{x}_i) \mathbf{r}^\top(\mathbf{x}_i), \quad m \leq \frac{p(p+1)}{2} + 1$$

$\Rightarrow \xi =$  mesure de prob. discrète avec au plus  $\frac{p(p+1)}{2} + 1$  points de support

▣ vrai en particulier pour le plan optimal !

[Même mieux : pour beaucoup de critères  $\Phi(\cdot)$ , si  $\xi^*$  est optimal (maximise  $\Phi[\mathbf{M}(\xi)]$ ) alors  $\mathbf{M}(\xi^*)$  est sur un bord de l'enveloppe convexe de  $\mathcal{M}$  et  $\frac{p(p+1)}{2}$  points de support suffisent]

## Théorème de Caratheodory :

$\mathbf{M}(\xi)$  peut s'écrire comme la combinaison linéaire d'au plus  $q + 1$  éléments de  $\mathcal{M}$ :

$$\mathbf{M}(\xi) = \sum_{i=1}^m w_i \mathbf{r}(\mathbf{x}_i) \mathbf{r}^\top(\mathbf{x}_i), \quad m \leq \frac{p(p+1)}{2} + 1$$

$\Rightarrow \xi =$  mesure de prob. discrète avec au plus  $\frac{p(p+1)}{2} + 1$  points de support

▣ vrai en particulier pour le plan optimal !

[Même mieux : pour beaucoup de critères  $\Phi(\cdot)$ , si  $\xi^*$  est optimal (maximise  $\Phi[\mathbf{M}(\xi)]$ ) alors  $\mathbf{M}(\xi^*)$  est sur un bord de l'enveloppe convexe de  $\mathcal{M}$  et  $\frac{p(p+1)}{2}$  points de support suffisent]

Supposons que l'on ait trouvé  $\xi^* = \sum_{i=1}^m w_i^* \delta_{\mathbf{x}_i}$  optimal

► pour  $n$  fixé, choisir les  $r_i$  pour que  $\frac{r_i}{n} \simeq w_i^*$  optimal

▣ arrondi d'un plan approximatif

## Théorème de Caratheodory :

$\mathbf{M}(\xi)$  peut s'écrire comme la combinaison linéaire d'au plus  $q + 1$  éléments de  $\mathcal{M}$ :

$$\mathbf{M}(\xi) = \sum_{i=1}^m w_i \mathbf{r}(\mathbf{x}_i) \mathbf{r}^\top(\mathbf{x}_i), \quad m \leq \frac{p(p+1)}{2} + 1$$

$\Rightarrow \xi =$  mesure de prob. discrète avec au plus  $\frac{p(p+1)}{2} + 1$  points de support

▣ vrai en particulier pour le plan optimal !

[Même mieux : pour beaucoup de critères  $\Phi(\cdot)$ , si  $\xi^*$  est optimal (maximise  $\Phi[\mathbf{M}(\xi)]$ ) alors  $\mathbf{M}(\xi^*)$  est sur un bord de l'enveloppe convexe de  $\mathcal{M}$  et  $\frac{p(p+1)}{2}$  points de support suffisent]

Supposons que l'on ait trouvé  $\xi^* = \sum_{i=1}^m w_i^* \delta_{\mathbf{x}_i}$  optimal

► pour  $n$  fixé, choisir les  $r_i$  pour que  $\frac{r_i}{n} \simeq w_i^*$  optimal

▣ arrondi d'un plan approximatif

**Pourquoi les plans approximatifs sont intéressants ?**

**En quoi cela simplifie le problème d'optimisation ?**

⇒ Maximiser  $\Phi(\mathbf{M}(\xi))$  **concave** avec  $\mathbf{M}(\xi)$  dans un **ensemble convexe**

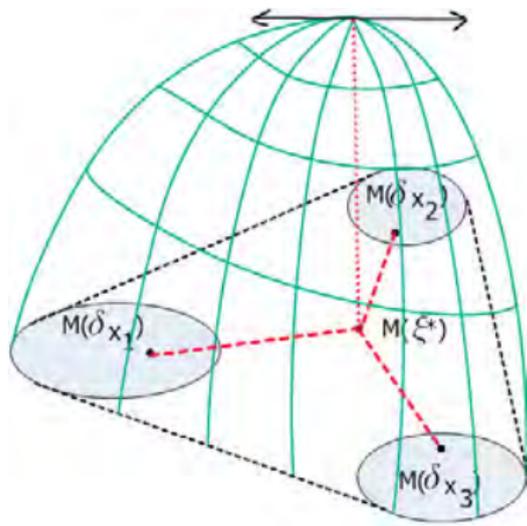
Ex.  $D$ -optimalité :  $\forall \mathbf{M}_1 \succeq \mathbf{O}, \mathbf{M}_2 \succeq \mathbf{O}$ , avec  $\mathbf{M}_1 \not\prec \mathbf{M}_2, \forall \alpha, 0 < \alpha < 1$ ,  
 $\log \det[(1 - \alpha)\mathbf{M}_1 + \alpha\mathbf{M}_2] > (1 - \alpha) \log \det \mathbf{M}_1 + \alpha \log \det \mathbf{M}_2$   
 $\Rightarrow \log \det[\cdot]$  (strictement) concave

ensemble convexe + critère concave  $\Rightarrow$  un maximum unique !

⇒ Maximiser  $\Phi(\mathbf{M}(\xi))$  **concave** avec  $\mathbf{M}(\xi)$  dans un **ensemble convexe**

Ex.  $D$ -optimalité :  $\forall \mathbf{M}_1 \succeq \mathbf{O}, \mathbf{M}_2 \succeq \mathbf{O}$ , avec  $\mathbf{M}_1 \not\propto \mathbf{M}_2, \forall \alpha, 0 < \alpha < 1$ ,  
 $\log \det[(1 - \alpha)\mathbf{M}_1 + \alpha\mathbf{M}_2] > (1 - \alpha) \log \det \mathbf{M}_1 + \alpha \log \det \mathbf{M}_2$   
 $\Rightarrow \log \det[\cdot]$  (strictement) concave

ensemble convexe + critère concave  $\Rightarrow$  un maximum unique !



$\xi^*$  est optimal  $\Leftrightarrow$  dérivée directionnelle  
 $\leq 0$  dans toutes les directions

⇒ “Théorème d'Équivalence” de Kiefer and Wolfowitz (1960)

$\Xi$  = ensemble des mesures de prob. sur  $\mathcal{X}$

$\Phi(\cdot)$  concave,  $\phi(\xi) = \Phi[\mathbf{M}(\xi)]$

$$F_{\phi}(\xi; \nu) = \lim_{\alpha \rightarrow 0^+} \frac{\phi[(1-\alpha)\xi + \alpha\nu] - \phi(\xi)}{\alpha}$$

= dérivée directionnelle de  $\phi(\cdot)$  en  $\xi$  dans la direction  $\nu$

**Théorème d'Équivalence :**

$$\xi^* \text{ maximise } \phi(\xi) \Leftrightarrow \max_{\nu \in \Xi} F_{\phi}(\xi^*; \nu) \leq 0$$

⇒ “Théorème d'Équivalence” de Kiefer and Wolfowitz (1960)

$\Xi$  = ensemble des mesures de prob. sur  $\mathcal{X}$

$\Phi(\cdot)$  concave,  $\phi(\xi) = \Phi[\mathbf{M}(\xi)]$

$F_\phi(\xi; \nu) = \lim_{\alpha \rightarrow 0^+} \frac{\phi[(1-\alpha)\xi + \alpha\nu] - \phi(\xi)}{\alpha}$   
 = dérivée directionnelle de  $\phi(\cdot)$  en  $\xi$  dans la direction  $\nu$

**Théorème d'Équivalence :**

$\xi^*$  maximise  $\phi(\xi) \Leftrightarrow \max_{\nu \in \Xi} F_\phi(\xi^*; \nu) \leq 0$

⇒ Prend une forme très simple quand  $\Phi(\cdot)$  est différentiable

$$\xi^* \text{ maximise } \phi(\xi) \Leftrightarrow \max_{\mathbf{x} \in \mathcal{X}} F_\phi(\xi^*; \delta_{\mathbf{x}}) \leq 0$$

⇒ On peut vérifier l'optimalité de  $\xi^*$  en traçant  $F_\phi(\xi^*; \delta_{\mathbf{x}})$

### Ex. : $D$ -optimalité

- $\xi_D^*$  maximise  $\log \det[\mathbf{M}(\xi)]$  par rapport à  $\xi \in \Xi$
- $\Leftrightarrow \max_{\mathbf{x} \in \mathcal{X}} d(\xi_D^*, \mathbf{x}) \leq p = \dim(\gamma)$
- $\Leftrightarrow \xi_D^*$  minimise  $\max_{\mathbf{x} \in \mathcal{X}} d(\xi, \mathbf{x})$  par rapport à  $\xi \in \Xi$

avec  $d(\xi, \mathbf{x}) = \mathbf{r}^\top(\mathbf{x})\mathbf{M}^{-1}(\xi)\mathbf{r}(\mathbf{x})$

De plus,  $d(\xi_D^*, \mathbf{x}_i) = p$  pour tout point de support  $\mathbf{x}_i$  de  $\xi_D^*$

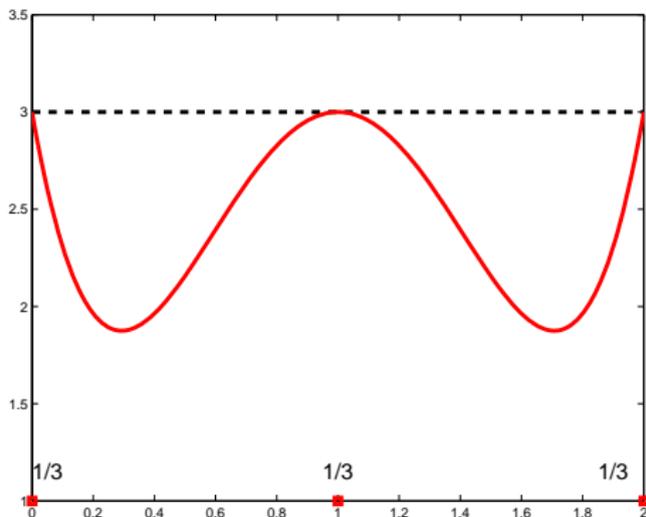
Ex. :  $\mathbf{r}(x) = (1 \ x \ x^2)^\top$  ( $p = 3$ ) erreurs i.i.d.,  $\mathcal{X} = [0, 2]$

▸  $d(\xi, x)$  en fonction de  $x$

Ex. :  $\mathbf{r}(x) = (1 \ x \ x^2)^\top$  ( $p = 3$ ) erreurs i.i.d.,  $\mathcal{X} = [0, 2]$

▸  $d(\xi, x)$  en fonction de  $x$

$$\xi_D^* = \left\{ \begin{array}{ccc} 0 & 1 & 2 \\ 1/3 & 1/3 & 1/3 \end{array} \right\}$$

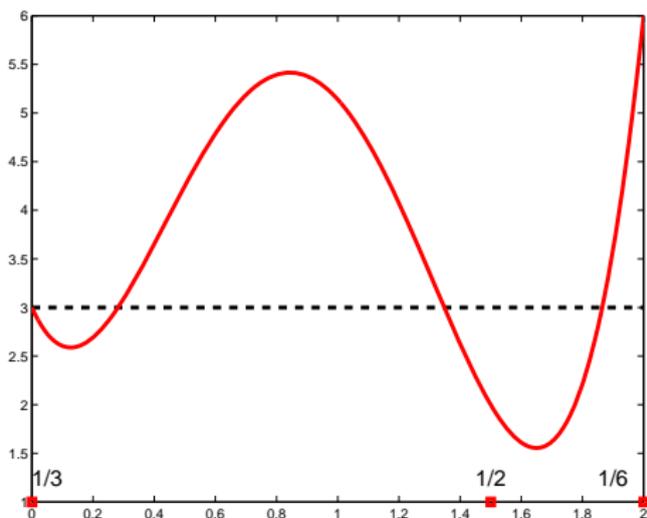
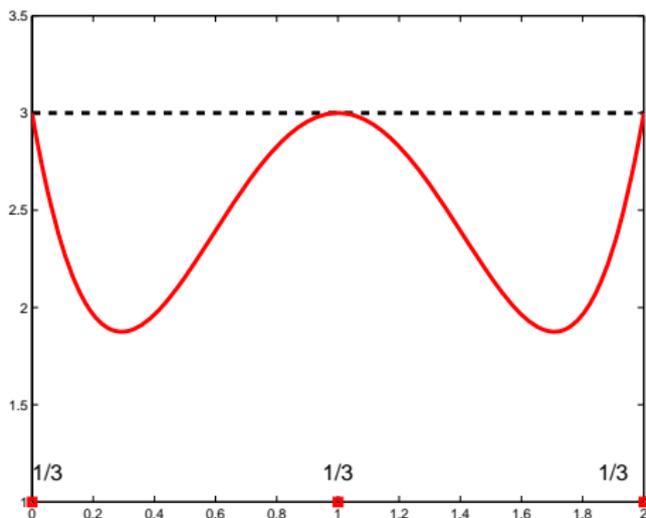


Ex. :  $\mathbf{r}(x) = (1 \ x \ x^2)^\top$  ( $p = 3$ ) erreurs i.i.d.,  $\mathcal{X} = [0, 2]$

⇒  $d(\xi, x)$  en fonction de  $x$

$$\xi_D^* = \begin{Bmatrix} 0 & 1 & 2 \\ 1/3 & 1/3 & 1/3 \end{Bmatrix}$$

$$\xi = \begin{Bmatrix} 0 & 1.5 & 2 \\ 1/3 & 1/2 & 1/6 \end{Bmatrix}$$



Le Th. d'Équivalence de KW relie l'optimalité en estimation (précision sur  $\gamma$ ) à l'optimalité en prédiction :

$$\text{nvar}[\mathbf{r}^\top(\mathbf{x})\hat{\gamma}^n] = \sigma^2 \mathbf{r}^\top(\mathbf{x})\mathbf{M}^{-1}(\xi)\mathbf{r}(\mathbf{x}) = \sigma^2 d(\xi, \mathbf{x})$$

$D$ -optimalité  $\Leftrightarrow G$ -optimalité

⇒  $\xi_D^*$  minimise le maximum de la variance de prédiction sur  $\mathcal{X}$

Le Th. d'Équivalence de KW relie l'optimalité en estimation (précision sur  $\gamma$ ) à l'optimalité en prédiction :

$$\text{nvar}[\mathbf{r}^\top(\mathbf{x})\hat{\gamma}^n] = \sigma^2 \mathbf{r}^\top(\mathbf{x})\mathbf{M}^{-1}(\xi)\mathbf{r}(\mathbf{x}) = \sigma^2 d(\xi, \mathbf{x})$$

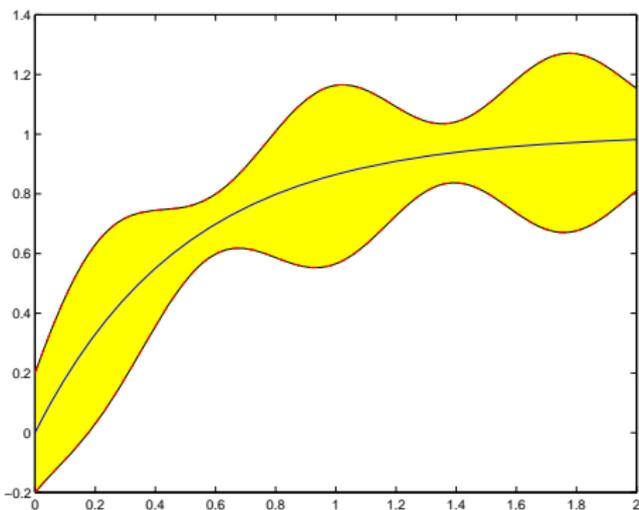
$D$ -optimalité  $\Leftrightarrow$   $G$ -optimalité

⇒  $\xi_D^*$  minimise le maximum de la variance de prédiction sur  $\mathcal{X}$

$\eta(\mathbf{x}, \hat{\gamma}^n)$

$\eta(\mathbf{x}, \hat{\gamma}^n) \pm 2$  écarts-types

⇒ placer l'observation suivante là où  $d(\xi, \mathbf{x})$  est grand



► Modèles produits tensoriels

# Construction d'un plan approximatif optimal

Idée centrale (**▲** pour  $\Phi(\cdot)$  différentiable **▲**) : plus grande pente

## Fedorov–Wynn :

- 1 : Choisir  $\xi^1$  non dégénéré ( $\det \mathbf{M}(\xi^1) > 0$ )
- 2 : Calculer  $\mathbf{x}_k^* = \arg \max_{\mathcal{X}} F_\phi(\xi^k; \delta_{\mathbf{x}})$   
Si  $F_\phi(\xi^k; \delta_{\mathbf{x}_k^*}) < \epsilon$ , stop:  $\xi^k$  est  $\epsilon$ -optimal
- 3 :  $\xi^{k+1} = (1 - \alpha_k)\xi^k + \alpha_k\delta_{\mathbf{x}_k^*}$  (mesure de Dirac en  $\mathbf{x}_k^*$ )  
[Vertex Direction]  
 $k \rightarrow k + 1$ , retourner au pas 2

# Construction d'un plan approximatif optimal

Idee centrale (**▲** pour  $\Phi(\cdot)$  différentiable **▲**) : plus grande pente

**Fedorov–Wynn :**

- 1 : Choisir  $\xi^1$  non dégénéré ( $\det \mathbf{M}(\xi^1) > 0$ )
- 2 : Calculer  $\mathbf{x}_k^* = \arg \max_{\mathcal{X}} F_\phi(\xi^k; \delta_{\mathbf{x}})$   
Si  $F_\phi(\xi^k; \delta_{\mathbf{x}_k^*}) < \epsilon$ , stop:  $\xi^k$  est  $\epsilon$ -optimal
- 3 :  $\xi^{k+1} = (1 - \alpha_k)\xi^k + \alpha_k \delta_{\mathbf{x}_k^*}$  (mesure de Dirac en  $\mathbf{x}_k^*$ )  
[Vertex Direction]  
 $k \rightarrow k + 1$ , retourner au pas 2

Longueur de pas  $\alpha_k$ ?

$$\begin{aligned} \Rightarrow \alpha_k &= \arg \max \phi(\xi^{k+1}) \\ &= \frac{d(\xi^k, \mathbf{x}_k^*) - p}{p[d(\xi^k, \mathbf{x}_k^*) - 1]} \text{ pour } D\text{-optimalité (Fedorov, 1972)} \end{aligned}$$

→ convergence monotone

$$\Rightarrow \alpha_k > 0, \quad \lim_{k \rightarrow \infty} \alpha_k = 0, \quad \sum_{i=1}^{\infty} \alpha_k = \infty$$

((Wynn, 1970) pour  $D$ -optimalité)

## Remarques :

- Planification séquentielle, un  $\mathbf{x}_i$  à la fois qui entre dans  $\mathbf{M}(\mathbf{X})$  :

$$\mathbf{M}(\mathbf{X}_{k+1}) = \frac{k}{k+1} \mathbf{M}(\mathbf{X}_k) + \frac{1}{k+1} \mathbf{r}(\mathbf{x}_{k+1}) \mathbf{r}^\top(\mathbf{x}_{k+1})$$

avec  $\mathbf{x}_{k+1} = \arg \max_{\mathbf{x} \in \mathcal{X}} F_\phi(\xi^k; \delta_{\mathbf{x}})$

$\Leftrightarrow$  algorithm de Wynn avec  $\alpha_k = \frac{1}{k+1}$

## Remarques :

- Planification séquentielle, un  $\mathbf{x}_i$  à la fois qui entre dans  $\mathbf{M}(\mathbf{X})$  :

$$\mathbf{M}(\mathbf{X}_{k+1}) = \frac{k}{k+1} \mathbf{M}(\mathbf{X}_k) + \frac{1}{k+1} \mathbf{r}(\mathbf{x}_{k+1}) \mathbf{r}^\top(\mathbf{x}_{k+1})$$

avec  $\mathbf{x}_{k+1} = \arg \max_{\mathbf{x} \in \mathcal{X}} F_\phi(\xi^k; \delta_{\mathbf{x}})$

$\Leftrightarrow$  algorithm de Wynn avec  $\alpha_k = \frac{1}{k+1}$

- **Convergence assurée vers l'optimum**

## Remarques :

- Planification séquentielle, un  $\mathbf{x}_i$  à la fois qui entre dans  $\mathbf{M}(\mathbf{X})$  :

$$\mathbf{M}(\mathbf{X}_{k+1}) = \frac{k}{k+1} \mathbf{M}(\mathbf{X}_k) + \frac{1}{k+1} \mathbf{r}(\mathbf{x}_{k+1}) \mathbf{r}^\top(\mathbf{x}_{k+1})$$

avec  $\mathbf{x}_{k+1} = \arg \max_{\mathbf{x} \in \mathcal{X}} F_\phi(\xi^k; \delta_{\mathbf{x}})$

$\Leftrightarrow$  algorithm de Wynn avec  $\alpha_k = \frac{1}{k+1}$

- **Convergence assurée vers l'optimum**
- Il existe des méthodes plus rapides :
  - supprimer des points de support de  $\xi^k$  ( $\approx$  autoriser  $\alpha_k$  à être  $< 0$ ) (Atwood, 1973; Böhning, 1985, 1986)
  - combiner avec un gradient projeté (ou une méthode du 2nd ordre) (Wu, 1978)
  - algorithm multiplicatif (Titterton, 1976; Torsney, 1983, 2009; Yu, 2010) (pour  $A$  et  $D$  optimalité, loin de l'optimum)
  - combiner différentes méthodes (Yu, 2011)
  - sujet toujours actif — surtout pour  $\Phi(\cdot)$  non différentiable...

## 2.4 Modèles produits tensoriels

D-optimalité (également vrai pour A-optimalité sous certaines conditions  
(Schwabe, 1996))

$[\mathbf{r}^{(k)}(\mathbf{x})]^\top \boldsymbol{\theta}^{(k)} \triangleq \sum_{i=1}^{d_k} \theta_i^{(k)} x^i$  polynôme de degré  $d_k$ ,  $\dim(\boldsymbol{\theta}^{(k)}) = p_k = 1 + d_k$

Modèle global pour  $\mathbf{x} = (\{\mathbf{x}\}_1, \{\mathbf{x}\}_2, \dots, \{\mathbf{x}\}_d)^\top$  :

$$\mathbf{r}^\top(\mathbf{x})\boldsymbol{\gamma} = \prod_{k=1}^d [\mathbf{r}^{(k)}(\mathbf{x})]^\top \boldsymbol{\theta}^{(k)},$$

$$\text{degré total } \sum_{k=1}^d d_k, \dim(\boldsymbol{\gamma}) = \prod_{k=1}^d p_k$$

## 2.4 Modèles produits tensoriels

D-optimalité (également vrai pour A-optimalité sous certaines conditions (Schwabe, 1996))

$[\mathbf{r}^{(k)}(\mathbf{x})]^\top \boldsymbol{\theta}^{(k)} \triangleq \sum_{i=1}^{d_k} \theta_i^{(k)} x^i$  polynôme de degré  $d_k$ ,  $\dim(\boldsymbol{\theta}^{(k)}) = p_k = 1 + d_k$

Modèle global pour  $\mathbf{x} = (\{\mathbf{x}\}_1, \{\mathbf{x}\}_2, \dots, \{\mathbf{x}\}_d)^\top$  :

$$\mathbf{r}^\top(\mathbf{x})\boldsymbol{\gamma} = \prod_{k=1}^d [\mathbf{r}^{(k)}(\mathbf{x})]^\top \boldsymbol{\theta}^{(k)},$$

$$\text{degré total } \sum_{k=1}^d d_k, \dim(\boldsymbol{\gamma}) = \prod_{k=1}^d p_k$$

Exemple :

$$\begin{aligned} \mathbf{r}^\top(\mathbf{x})\boldsymbol{\gamma} &= (\theta_0^{(1)} + \theta_1^{(1)}\{\mathbf{x}\}_1 + \theta_2^{(1)}\{\mathbf{x}\}_1^2) \times (\theta_0^{(2)} + \theta_1^{(2)}\{\mathbf{x}\}_2 + \theta_2^{(2)}\{\mathbf{x}\}_2^2) \\ &= \gamma_0 + \gamma_1\{\mathbf{x}\}_1 + \gamma_2\{\mathbf{x}\}_2 + \gamma_{12}\{\mathbf{x}\}_1\{\mathbf{x}\}_2 + \gamma_{11}\{\mathbf{x}\}_1^2 + \gamma_{22}\{\mathbf{x}\}_2^2 \\ &\quad + \gamma_{112}\{\mathbf{x}\}_1^2\{\mathbf{x}\}_2 + \gamma_{122}\{\mathbf{x}\}_1\{\mathbf{x}\}_2^2 + \gamma_{1122}\{\mathbf{x}\}_1^2\{\mathbf{x}\}_2^2 \end{aligned}$$

## 2.4 Modèles produits tensoriels

D-optimalité (également vrai pour A-optimalité sous certaines conditions (Schwabe, 1996))

$[\mathbf{r}^{(k)}(\mathbf{x})]^\top \boldsymbol{\theta}^{(k)} \triangleq \sum_{i=1}^{d_k} \theta_i^{(k)} x^i$  polynôme de degré  $d_k$ ,  $\dim(\boldsymbol{\theta}^{(k)}) = p_k = 1 + d_k$

Modèle global pour  $\mathbf{x} = (\{\mathbf{x}\}_1, \{\mathbf{x}\}_2, \dots, \{\mathbf{x}\}_d)^\top$  :

$$\mathbf{r}^\top(\mathbf{x})\boldsymbol{\gamma} = \prod_{k=1}^d [\mathbf{r}^{(k)}(\mathbf{x})]^\top \boldsymbol{\theta}^{(k)},$$

$$\text{degré total } \sum_{k=1}^d d_k, \dim(\boldsymbol{\gamma}) = \prod_{k=1}^d p_k$$

Exemple :

$$\begin{aligned} \mathbf{r}^\top(\mathbf{x})\boldsymbol{\gamma} &= (\theta_0^{(1)} + \theta_1^{(1)}\{\mathbf{x}\}_1 + \theta_2^{(1)}\{\mathbf{x}\}_1^2) \times (\theta_0^{(2)} + \theta_1^{(2)}\{\mathbf{x}\}_2 + \theta_2^{(2)}\{\mathbf{x}\}_2^2) \\ &= \gamma_0 + \gamma_1\{\mathbf{x}\}_1 + \gamma_2\{\mathbf{x}\}_2 + \gamma_{12}\{\mathbf{x}\}_1\{\mathbf{x}\}_2 + \gamma_{11}\{\mathbf{x}\}_1^2 + \gamma_{22}\{\mathbf{x}\}_2^2 \\ &\quad + \gamma_{112}\{\mathbf{x}\}_1^2\{\mathbf{x}\}_2 + \gamma_{122}\{\mathbf{x}\}_1\{\mathbf{x}\}_2^2 + \gamma_{1122}\{\mathbf{x}\}_1^2\{\mathbf{x}\}_2^2 \end{aligned}$$

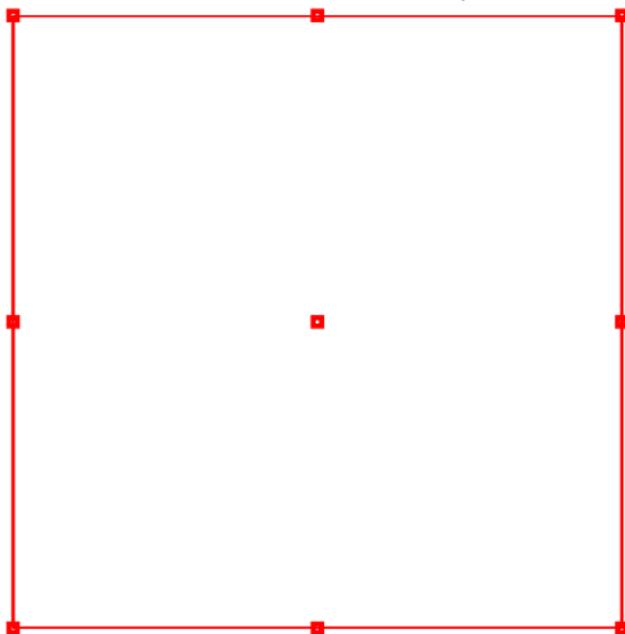
Plan *D*-optimal (approximatif) = produit tensoriel des *d* plans *D*-optimaux (vrai pour tout type de modèle, pas uniquement polynômes)

Polynôme de degré  $k$ : plan  $D$ -optimal supporté en  $k + 1$  points (sur  $[-1, 1]$ , les racines de  $(1 - t^2)P'_k(t)$ , avec  $P_k(t) \triangleq k$ -ème polynôme de Legendre) recevant la même masse  $1/(k + 1)$

Polynôme de degré  $k$ : plan  $D$ -optimal supporté en  $k + 1$  points (sur  $[-1, 1]$ , les racines de  $(1 - t^2)P'_k(t)$ , avec  $P_k(t) \triangleq k$ -ème polynôme de Legendre) recevant la même masse  $1/(k + 1)$

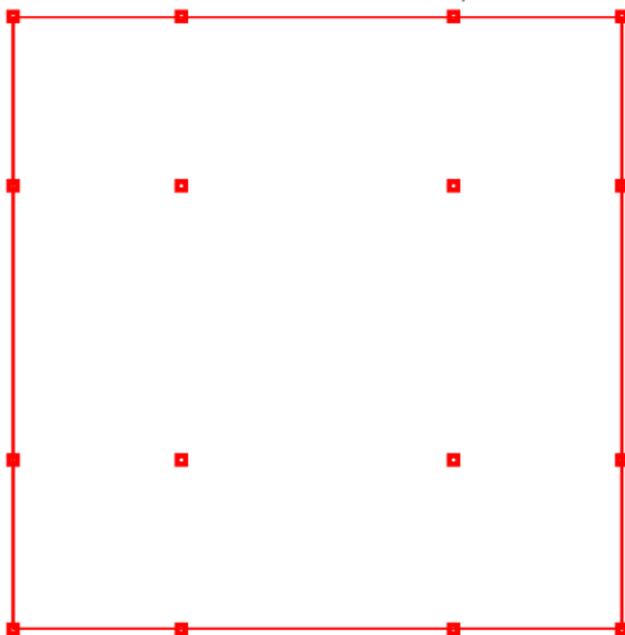
dimension 2,  $d_1 = d_2 = 2$

9 points, masses =  $1/9$



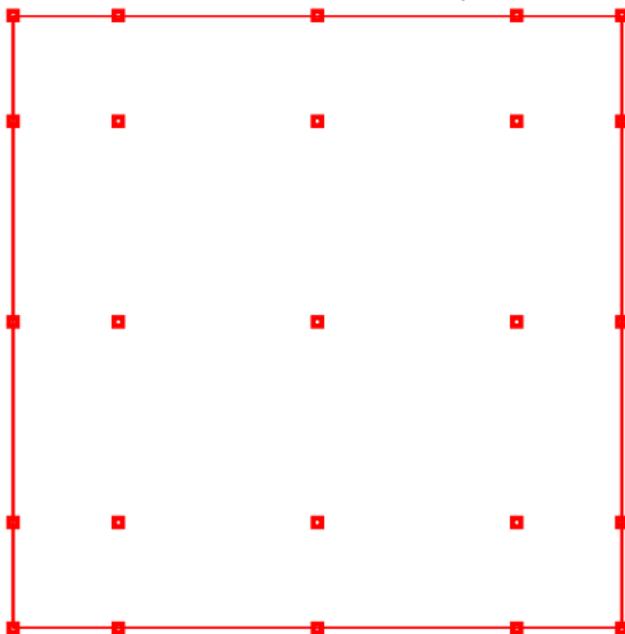
Polynôme de degré  $k$ : plan  $D$ -optimal supporté en  $k + 1$  points (sur  $[-1, 1]$ , les racines de  $(1 - t^2)P'_k(t)$ , avec  $P_k(t) \triangleq k$ -ème polynôme de Legendre) recevant la même masse  $1/(k + 1)$

dimension 2,  $d_1 = d_2 = 3$   
16 points, masses =  $1/16$



Polynôme de degré  $k$ : plan  $D$ -optimal supporté en  $k + 1$  points (sur  $[-1, 1]$ , les racines de  $(1 - t^2)P'_k(t)$ , avec  $P_k(t) \triangleq k$ -ème polynôme de Legendre) recevant la même masse  $1/(k + 1)$

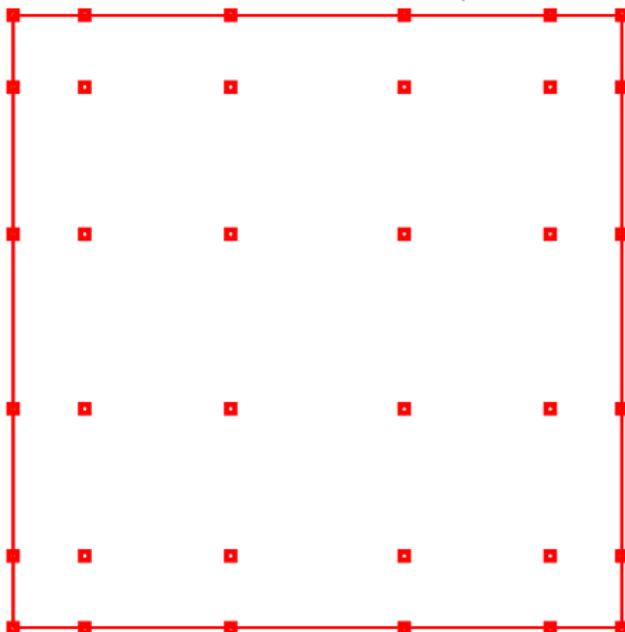
dimension 2,  $d_1 = d_2 = 4$   
25 points, masses =  $1/25$



Polynôme de degré  $k$ : plan  $D$ -optimal supporté en  $k + 1$  points (sur  $[-1, 1]$ , les racines de  $(1 - t^2)P'_k(t)$ , avec  $P_k(t) \triangleq k$ -ème polynôme de Legendre) recevant la même masse  $1/(k + 1)$

dimension 2,  $d_1 = d_2 = 5$

36 points, masses =  $1/36$



Somme de polynômes ?

$$\mathbf{r}^\top(\mathbf{x})\boldsymbol{\gamma} = \sum_{k=1}^d [\mathbf{r}^{(k)}(x)]^\top \boldsymbol{\theta}^{(k)},$$

$$\text{degré total } \max_{k=1}^d d_k, \dim(\boldsymbol{\gamma}) = (\sum_{k=1}^d p_k) - 1 = \sum_{k=1}^d d_k + d - 1$$

Somme de polynômes ?

$$\mathbf{r}^\top(\mathbf{x})\boldsymbol{\gamma} = \sum_{k=1}^d [\mathbf{r}^{(k)}(\mathbf{x})]^\top \boldsymbol{\theta}^{(k)},$$

$$\text{degré total } \max_{k=1}^d d_k, \dim(\boldsymbol{\gamma}) = (\sum_{k=1}^d p_k) - 1 = \sum_{k=1}^d d_k + d - 1$$

Exemple :

$$\begin{aligned} \mathbf{r}^\top(\mathbf{x})\boldsymbol{\gamma} &= (\theta_0^{(1)} + \theta_1^{(1)}\{\mathbf{x}\}_1 + \theta_2^{(1)}\{\mathbf{x}\}_1^2) + (\theta_0^{(2)} + \theta_1^{(2)}\{\mathbf{x}\}_2 + \theta_2^{(2)}\{\mathbf{x}\}_2^2) \\ &= \gamma_0 + \gamma_1\{\mathbf{x}\}_1 + \gamma_2\{\mathbf{x}\}_2 + \gamma_{11}\{\mathbf{x}\}_1^2 + \gamma_{22}\{\mathbf{x}\}_2^2 \end{aligned}$$

(pas de termes croisés)

Ici encore, plan  $D$ -optimal (approximatif) = produit tensoriel des  $d$  plans  $D$ -optimaux (Schwabe, 1996)

Somme de polynômes ?

$$\mathbf{r}^\top(\mathbf{x})\boldsymbol{\gamma} = \sum_{k=1}^d [\mathbf{r}^{(k)}(\mathbf{x})]^\top \boldsymbol{\theta}^{(k)},$$

$$\text{degré total } \max_{k=1}^d d_k, \dim(\boldsymbol{\gamma}) = (\sum_{k=1}^d p_k) - 1 = \sum_{k=1}^d d_k + d - 1$$

Exemple :

$$\begin{aligned} \mathbf{r}^\top(\mathbf{x})\boldsymbol{\gamma} &= (\theta_0^{(1)} + \theta_1^{(1)}\{\mathbf{x}\}_1 + \theta_2^{(1)}\{\mathbf{x}\}_1^2) + (\theta_0^{(2)} + \theta_1^{(2)}\{\mathbf{x}\}_2 + \theta_2^{(2)}\{\mathbf{x}\}_2^2) \\ &= \gamma_0 + \gamma_1\{\mathbf{x}\}_1 + \gamma_2\{\mathbf{x}\}_2 + \gamma_{11}\{\mathbf{x}\}_1^2 + \gamma_{22}\{\mathbf{x}\}_2^2 \end{aligned}$$

(pas de termes croisés)

Ici encore, plan  $D$ -optimal (approximatif) = produit tensoriel des  $d$  plans  $D$ -optimaux (Schwabe, 1996)

Difficilement utilisable en grande dimension :

$d$  polynômes de degré  $k \rightsquigarrow (k+1)^d$  points de support

mais une leçon générale à tirer, et une extension possible vers les processus Gaussiens et le krigeage

## 2.5 Conséquences pour les plans *space filling*

$D$ -optimalité + polynômes  $\Rightarrow$  davantage de points vers les bords quand le degré augmente

Théorème d'Erdős-Turan : les racines  $r$  de polynômes orthogonaux sur  $[0, 1]$  sont asymptotiquement distribuée suivant la loi de l'arcsinus, de densité

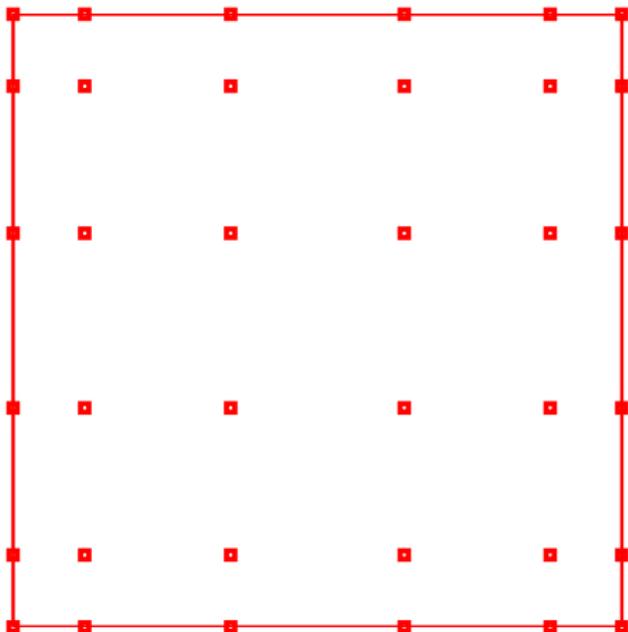
$$\varphi_0(r) = \frac{1}{\pi \sqrt{r(1-r)}}$$

$\Rightarrow$  Remarque générale : **il faut mettre + de points vers les bords !**

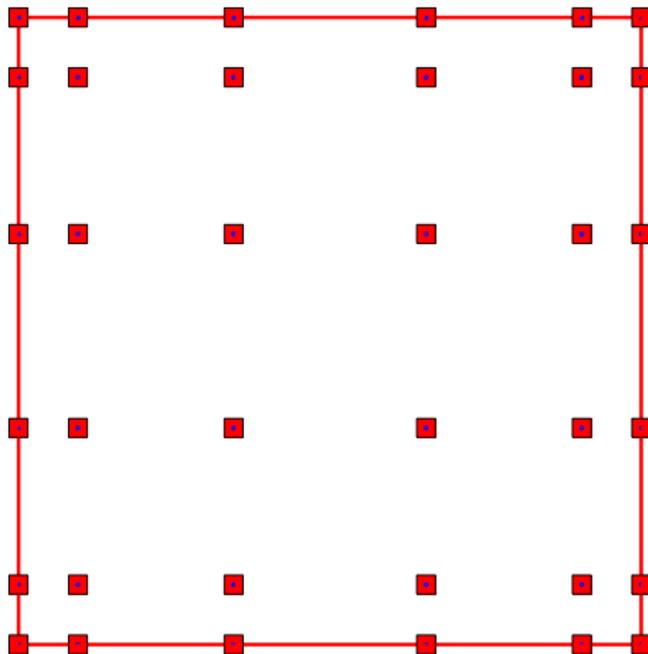
Pour contrer l'effet de bord Dette and Pepelyshev (2010) :

- choisir un plan *space filling* (par ex., Maximin, Lh Maximin, faible discrédance),
- pour tout  $j = 1, \dots, d$ , transformer les  $j$ -èmes coordonnées  $\{\mathbf{x}_i\}_j$  suivant  $T : x \mapsto z = T(x) = \frac{1 + \cos(\pi x)}{2}$   
( $x \sim$  uniforme  $\rightarrow z \sim$  arcsinus),
- utiliser le plan transformé  $\mathbf{Z}_n = (\mathbf{z}_1, \dots, \mathbf{z}_n)$

dimension 2, polynômes de degré 5 :  
plan  $D$ -optimal 36 points, masses  $=1/36$

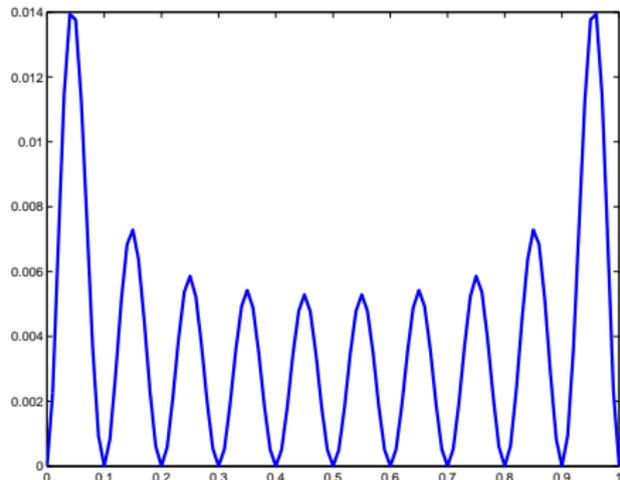


dimension 2,  $n = 36$   
plan Maximin optimal transformé (arcsinus)



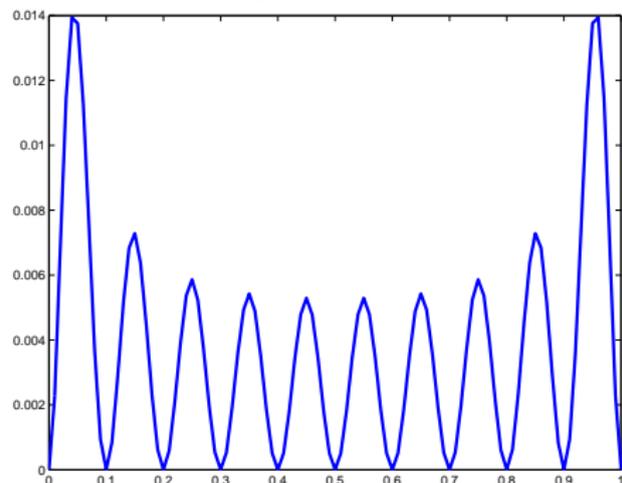
Exemple effet de bord :  $d = 1$   $n = 11$  observations dans  $[0, 1]$ , krigeage ordinaire pour covariance  $C(t) = \exp(-50 t^2) \Rightarrow$  tracé de  $\rho_n(x)$

$x_n$  Maximin

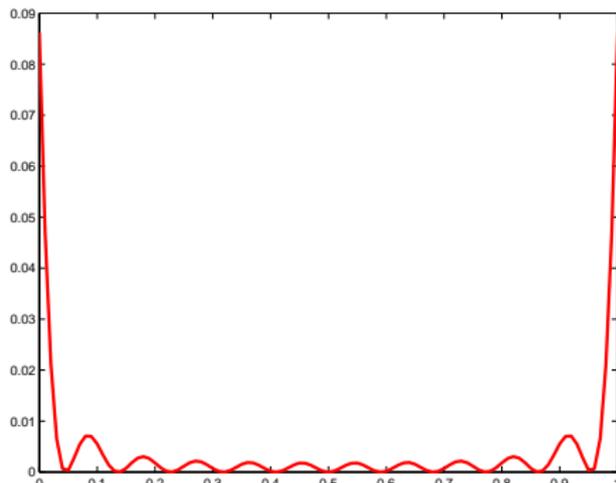


Exemple effet de bord :  $d = 1$   $n = 11$  observations dans  $[0, 1]$ , krigeage ordinaire pour covariance  $C(t) = \exp(-50 t^2) \Rightarrow$  tracé de  $\rho_n(x)$

$X_n$  Maximin



$X_n$  miniMax

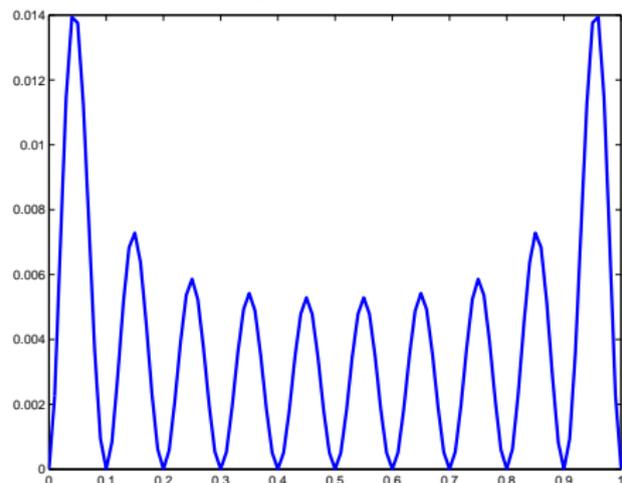


Si distribution uniforme des points

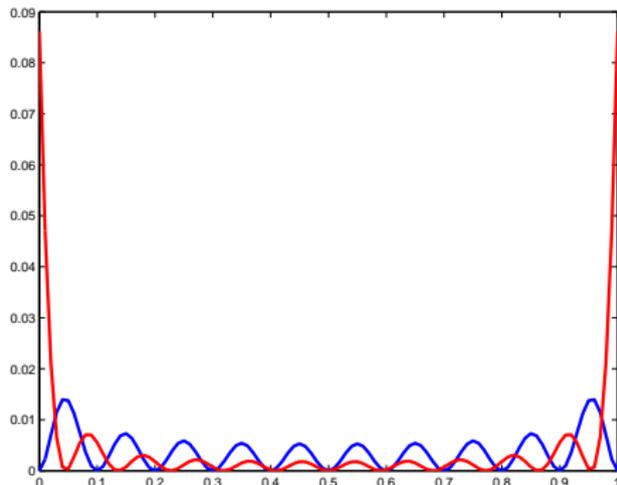
- $\Rightarrow$  moins de points disponibles près des bords
- $\Rightarrow$  moins de précision près des bords

Exemple effet de bord :  $d = 1$   $n = 11$  observations dans  $[0, 1]$ , krigage ordinaire pour covariance  $C(t) = \exp(-50 t^2) \Rightarrow$  tracé de  $\rho_n(x)$

$X_n$  Maximin



Maximin et miniMax



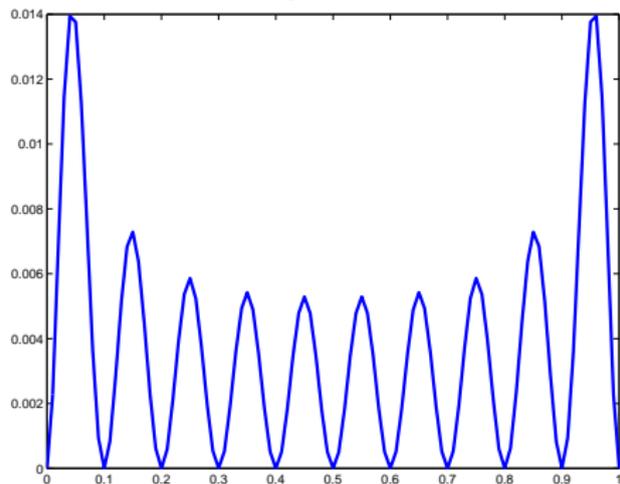
Si distribution uniforme des points

$\Rightarrow$  moins de points disponibles près des bords

$\Rightarrow$  moins de précision près des bords

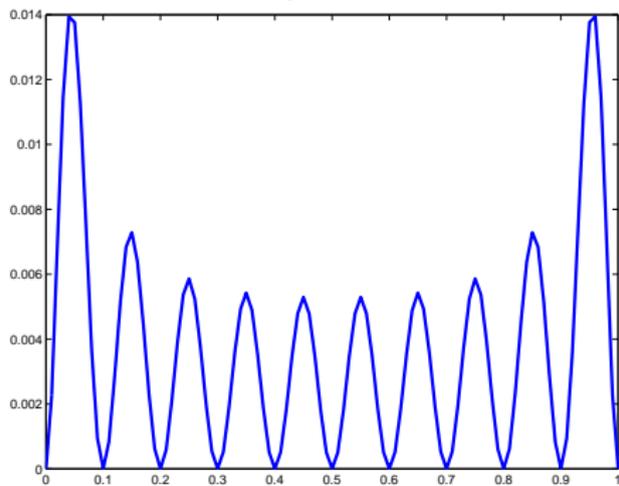
... Mais la transformation  $T : x \mapsto z = T(x) = \frac{1 + \cos(\pi x)}{2}$  peut être trop sévère

Maximin

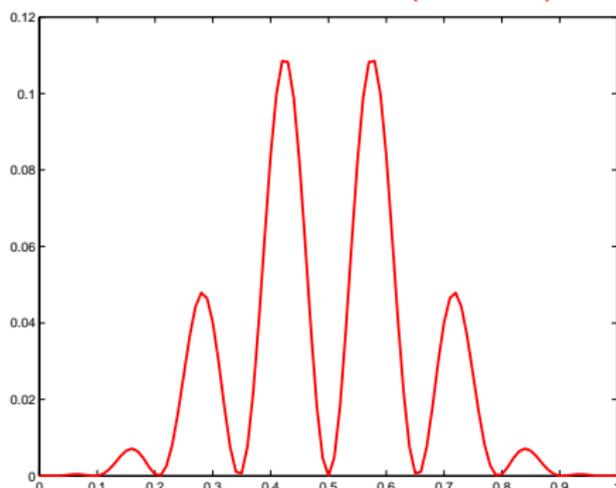


... Mais la transformation  $T : x \mapsto z = T(x) = \frac{1 + \cos(\pi x)}{2}$  peut être trop sévère

Maximin

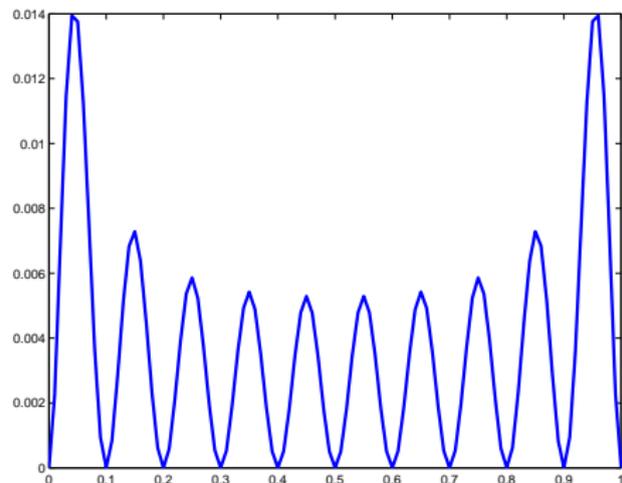


Maximin transformé (arcsinus)

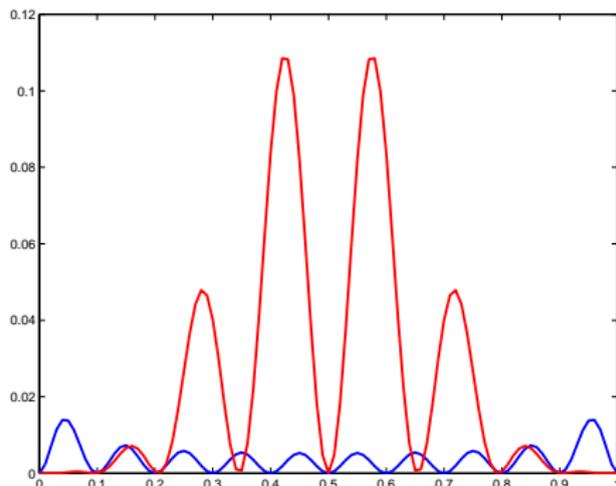


... Mais la transformation  $T : x \mapsto z = T(x) = \frac{1 + \cos(\pi x)}{2}$  peut être trop sévère

Maximin



Maximin et Maximin transformé



Distribution de l'arcsinus : maximise  $\tilde{\Phi}_{[0]}(\xi) = \exp \left[ \int_0^1 \int_0^1 \log \|x - y\| \xi(dx) \xi(dy) \right]$   
(version continue de  $\bar{\phi}_{[0]}(\mathbf{X}) = \exp \left[ \sum_{i < j} \mu_{ij} \log(d_{ij}) \right]$ , voir § I-1.5)

Distribution de l'arcsinus : maximise  $\tilde{\Phi}_{[0]}(\xi) = \exp \left[ \int_0^1 \int_0^1 \log \|x - y\| \xi(dx) \xi(dy) \right]$   
 (version continue de  $\bar{\phi}_{[0]}(\mathbf{X}) = \exp \left[ \sum_{i < j} \mu_{ij} \log(d_{ij}) \right]$ , voir § I-1.5)

► La maximisation de

$$\tilde{\Phi}_{[q]}(\xi) = \left[ \int_0^1 \int_0^1 \|x - y\|^{-q} \xi(dx) \xi(dy) \right]^{-1/q}, \quad 0 < q < 1$$

(version continue de  $\bar{\phi}_{[q]}(\mathbf{X}) = \left[ \sum_{i < j} \mu_{ij} d_{ij}^{-q} \right]^{-1/q}$ , voir § I-1.5) donne une  
 mesure  $\xi$  de densité  $\varphi_q(x) = \frac{x^{(q-1)/2}(1-x)^{(q-1)/2}}{B(\frac{q+1}{2}, \frac{q+1}{2})}$  (distribution Beta) (Zhigljavsky  
 et al., 2010)

(tend vers l'arcsinus pour  $q \rightarrow 0$ , vers l'uniforme pour  $q \rightarrow 1$ )

Pour contrer l'effet de bord (Dette and Pepelyshev, 2010) :

- choisir un plan *space filling*
- pour tout  $j = 1, \dots, d$ , transformer les  $j$ -èmes coordonnées  $\{\mathbf{x}_i\}_j$  suivant  $T : x \mapsto z = T(x)$  tel que  $x = \int_0^z \varphi_q(t) dt$   
( $x \sim \text{uniforme} \rightarrow z \sim \varphi_q$ ),
- utiliser le plan transformé  $\mathbf{Z}_n = (\mathbf{z}_1, \dots, \mathbf{z}_n)$

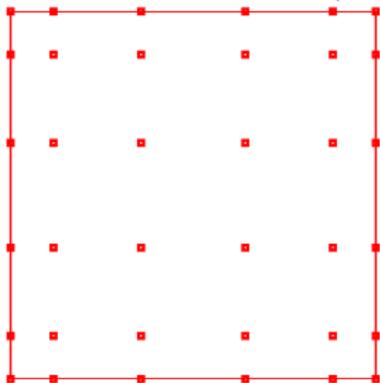
Pour contrer l'effet de bord (Dette and Pepelyshev, 2010) :

- choisir un plan *space filling*
- pour tout  $j = 1, \dots, d$ , transformer les  $j$ -èmes coordonnées  $\{\mathbf{x}_i\}_j$  suivant  $T : x \mapsto z = T(x)$  tel que  $x = \int_0^z \varphi_q(t) dt$   
( $x \sim \text{uniforme} \rightarrow z \sim \varphi_q$ ),
- utiliser le plan transformé  $\mathbf{Z}_n = (\mathbf{z}_1, \dots, \mathbf{z}_n)$

dimension 2, polynômes de degré 5

plan  $D$ -optimal :

36 points, masses = 1/36

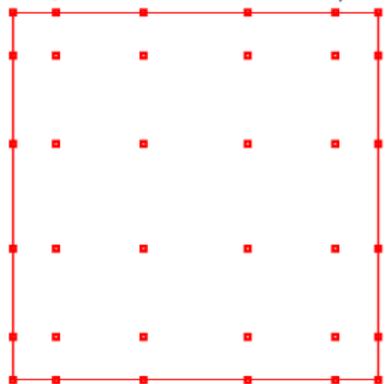


Pour contrer l'effet de bord (Dette and Pepelyshev, 2010) :

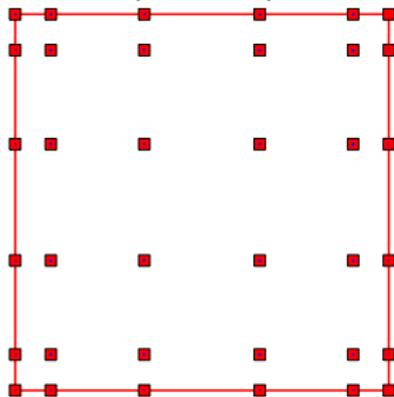
- choisir un plan *space filling*
- pour tout  $j = 1, \dots, d$ , transformer les  $j$ -èmes coordonnées  $\{\mathbf{x}_i\}_j$  suivant  $T : x \mapsto z = T(x)$  tel que  $x = \int_0^z \varphi_q(t) dt$   
( $x \sim \text{uniforme} \rightarrow z \sim \varphi_q$ ),
- utiliser le plan transformé  $\mathbf{Z}_n = (\mathbf{z}_1, \dots, \mathbf{z}_n)$

dimension 2, polynômes de degré 5  
plan  $D$ -optimal :

36 points, masses = 1/36



dimension 2,  $n = 36$   
plan Maximin optimal transformé  
(arcsinus)

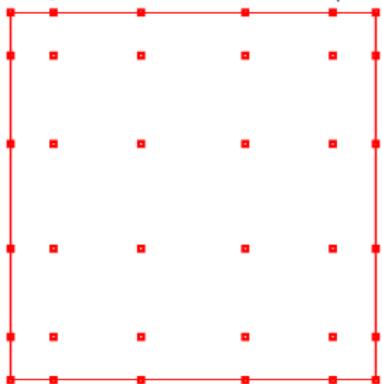


Pour contrer l'effet de bord (Dette and Pepelyshev, 2010) :

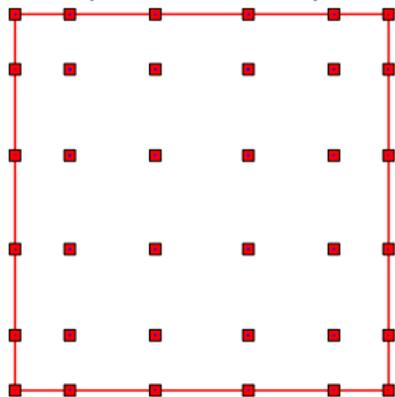
- choisir un plan *space filling*
- pour tout  $j = 1, \dots, d$ , transformer les  $j$ -èmes coordonnées  $\{x_i\}_j$  suivant  $T : x \mapsto z = T(x)$  tel que  $x = \int_0^z \varphi_q(t) dt$  ( $x \sim \text{uniforme} \rightarrow z \sim \varphi_q$ ),
- utiliser le plan transformé  $\mathbf{Z}_n = (z_1, \dots, z_n)$

dimension 2, polynômes de degré 5  
plan  $D$ -optimal :

36 points, masses = 1/36

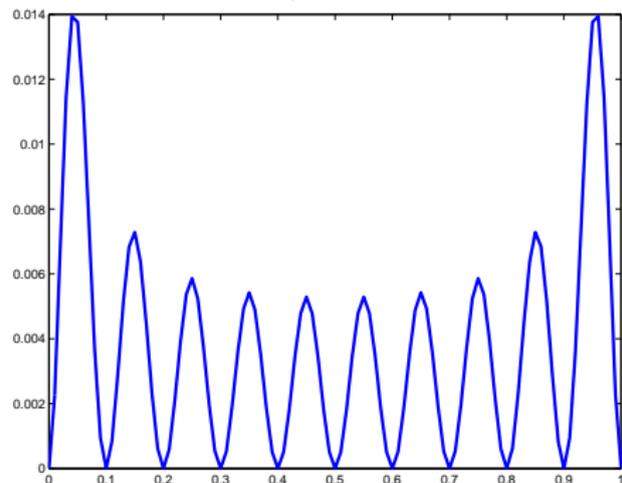


dimension 2,  $n = 36$   
plan Maximin optimal transformé  
(Beta,  $q = 0.4$ )



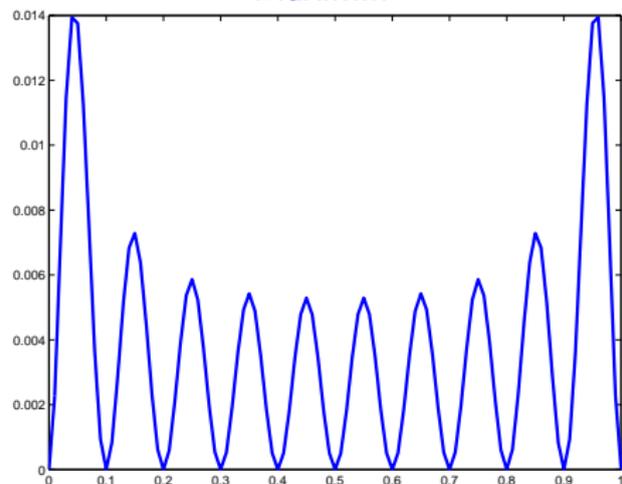
... Pour une transformation Beta bien choisie ( $q = 0.84$ )

Maximin

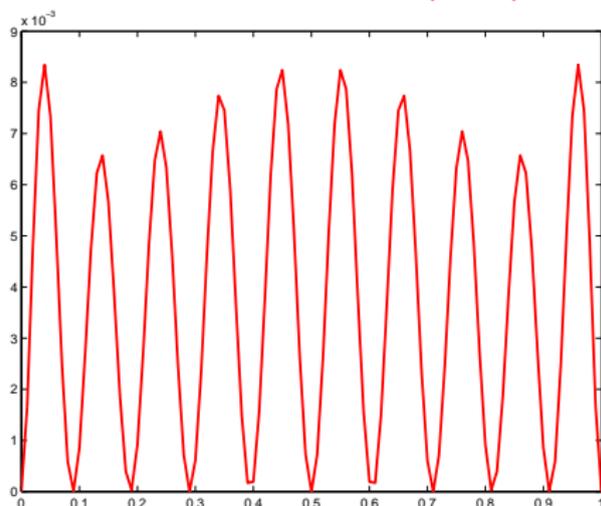


... Pour une transformation Beta bien choisie ( $q = 0.84$ )

Maximin

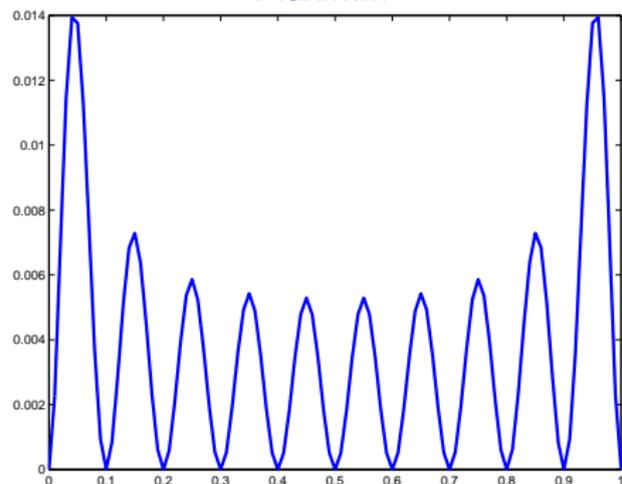


Maximin transformé (Beta)

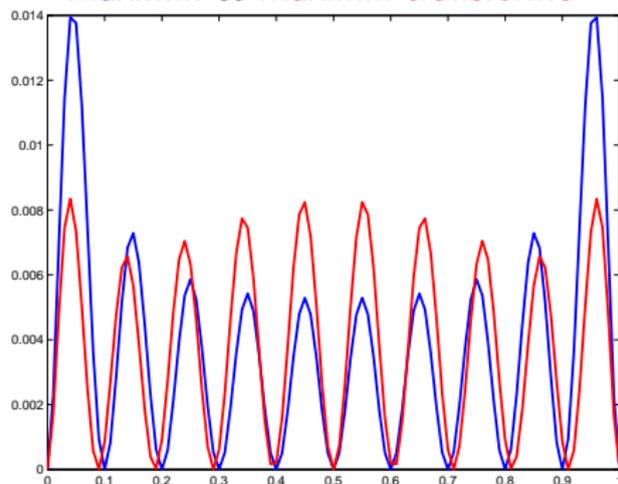


... Pour une transformation Beta bien choisie ( $q = 0.84$ )

Maximin



Maximin et Maximin transformé



Choix d'un  $q$  convenable ?

- Optimiser un critère de précision lié à  $\rho_n(\mathbf{x})$   
(dépend de la covariance  $C(\cdot)$ )

▲ utilisable seulement si  $\mathcal{X} = \text{hypercube}$  ▲

## 3 Plans optimaux pour prédiction Bayésienne

### 3.1 Décomposition de Karhunen-Loève d'un processus Gaussien

Modèle sans tendance :  $f(\mathbf{x}) = Z(\mathbf{x})$ , processus Gaussien

$E\{Z(\mathbf{x})\} = 0$ ,  $E\{Z(\mathbf{x})Z(\mathbf{x}')\} = C(\mathbf{x}, \mathbf{x}')$  ( $= C(\mathbf{x} - \mathbf{x}')$  si stationnaire)

$$\text{IMSE}_\mu(\mathbf{X}_n) \triangleq \int_{\mathcal{X}} E \left\{ [Z(\mathbf{x}) - E\{Z(\mathbf{x})|\mathbf{y}_n\}]^2 \right\} d\mu(\mathbf{x})$$

## 3 Plans optimaux pour prédiction Bayésienne

### 3.1 Décomposition de Karhunen-Loève d'un processus Gaussien

Modèle sans tendance :  $f(\mathbf{x}) = Z(\mathbf{x})$ , processus Gaussien

$E\{Z(\mathbf{x})\} = 0$ ,  $E\{Z(\mathbf{x})Z(\mathbf{x}')\} = C(\mathbf{x}, \mathbf{x}')$  ( $= C(\mathbf{x} - \mathbf{x}')$  si stationnaire)

$$\text{IMSE}_\mu(\mathbf{X}_n) \triangleq \int_{\mathcal{X}} E \left\{ [Z(\mathbf{x}) - E\{Z(\mathbf{x})|\mathbf{y}_n\}]^2 \right\} d\mu(\mathbf{x})$$

L'opérateur intégral  $T_\mu$  défini par

$$\forall f \in L^2(\mathcal{X}, \mu), \forall \mathbf{x} \in \mathcal{X}, \quad T_\mu[f](\mathbf{x}) = \int_{\mathcal{X}} f(\mathbf{x}')K(\mathbf{x}, \mathbf{x}')d\mu(\mathbf{x}')$$

est diagonalisable :

valeurs propres  $\lambda_i$ ,  $i = 1, 2, 3 \dots$  (rangées par ordre  $\searrow$ )

fonctions propres associées  $\varphi_i(\cdot)$  (étendues sur  $\mathcal{X}$ ), avec

$$\int_{\mathcal{X}} \varphi_i(\mathbf{x})\varphi_j(\mathbf{x})d\mu(\mathbf{x}) = \delta_{ij}$$

$$Z'(\mathbf{x}) \triangleq P_{\mathbb{H}_\mu} [Z_{\mathbf{x}}] = \sum_i \zeta_i \sqrt{\lambda_i} \varphi_i(\mathbf{x})$$

avec les  $\zeta_i$  i.i.d.  $\mathcal{N}(0, 1)$

$P_{\mathbb{H}_\mu}$  = projection  $\perp$  sur l'espace "de ce qui a une contribution à  $\text{IMSE}_\mu$ "

$Z'(\mathbf{x}) = \sum_i \gamma_i \varphi_i(\mathbf{x})$  où les v.a.  $\gamma_i$  sont indépendantes  $\mathcal{N}(0, \lambda_i)$

$$Z'(\mathbf{x}) \triangleq P_{\mathbb{H}_\mu} [Z_{\mathbf{x}}] = \sum_i \zeta_i \sqrt{\lambda_i} \varphi_i(\mathbf{x})$$

avec les  $\zeta_i$  i.i.d.  $\mathcal{N}(0, 1)$

$P_{\mathbb{H}_\mu}$  = projection  $\perp$  sur l'espace "de ce qui a une contribution à  $\text{IMSE}_\mu$ "

$Z'(\mathbf{x}) = \sum_i \gamma_i \varphi_i(\mathbf{x})$  où les v.a.  $\gamma_i$  sont indépendantes  $\mathcal{N}(0, \lambda_i)$

Pour  $m$  fixé,

$$\begin{aligned} Z'(\mathbf{x}) &= \sum_{i=1}^m \gamma_i \varphi_i(\mathbf{x}) + \sum_{i>m} \gamma_i \varphi_i(\mathbf{x}) \\ &\simeq \boxed{Z''(\mathbf{x}) = \sum_{i=1}^m \gamma_i \varphi_i(\mathbf{x}) + \varepsilon(\mathbf{x})} \end{aligned}$$

avec  $E\{\varepsilon(\mathbf{x}_i)\} = 0$ ,  $E\{\varepsilon(\mathbf{x}_i)\varepsilon(\mathbf{x}_j)\} = \sigma^2 \delta_{ij}$  et  $\sigma^2 = \sum_{i>m} \lambda_i$

$$\boxed{Z''(\mathbf{x}_i) = \phi^\top(\mathbf{x}_i) \boldsymbol{\gamma} + \varepsilon_i}$$

= modèle de régression linéaire (comme au § 2.1)  
(fonction propres  $\varphi_i(\cdot)$  au lieu de polynômes)

## 3.2 Prédiction Bayésienne pour $Z''(\mathbf{x}_i) = \phi^\top(\mathbf{x}_i)\gamma + \varepsilon_i$

Estimation par MC :

$$\hat{\gamma}_n = (\Phi_n^\top \Phi_n)^{-1} \Phi_n^\top \mathbf{y}_n, \text{ avec } \mathbf{y}_n = (y_1, \dots, y_n)^\top \text{ et } \Phi_n = \begin{pmatrix} \phi^\top(\mathbf{x}_1) \\ \vdots \\ \phi^\top(\mathbf{x}_n) \end{pmatrix}$$

$$\text{cov}(\hat{\gamma}_n) = \sigma^2 (\Phi_n^\top \Phi_n)^{-1} = \frac{\sigma^2}{n} \underbrace{\left[ \sum_{i=1}^n \phi(\mathbf{x}_i) \phi^\top(\mathbf{x}_i) \right]}_{\mathbf{M}_n}^{-1}$$

### 3.2 Prédiction Bayésienne pour $Z''(\mathbf{x}_i) = \phi^\top(\mathbf{x}_i)\gamma + \varepsilon_i$

Estimation par MC :

$$\hat{\gamma}_n = (\Phi_n^\top \Phi_n)^{-1} \Phi_n^\top \mathbf{y}_n, \text{ avec } \mathbf{y}_n = (y_1, \dots, y_n)^\top \text{ et } \Phi_n = \begin{pmatrix} \phi^\top(\mathbf{x}_1) \\ \vdots \\ \phi^\top(\mathbf{x}_n) \end{pmatrix}$$

$$\text{cov}(\hat{\gamma}_n) = \sigma^2 (\Phi_n^\top \Phi_n)^{-1} = \frac{\sigma^2}{n} \underbrace{\left[ \sum_{i=1}^n \phi(\mathbf{x}_i) \phi^\top(\mathbf{x}_i) \right]}_{\mathbf{M}_n}^{-1}$$

Prédiction en  $\mathbf{x}$  :  $\eta_n(\mathbf{x}) = \phi^\top(\mathbf{x}) \hat{\gamma}_n$

$$\text{IMSE}(\mathbf{X}_n) = \int_{\mathcal{X}} \phi^\top(\mathbf{x}) \text{cov}(\hat{\gamma}_n) \phi(\mathbf{x}) d\mu(\mathbf{x}) = \frac{\sigma^2}{n} \text{trace}[\mathbf{M}_n^{-1}]$$

= critère de A-optimalité  
(il faut  $n \geq m$  pour avoir  $\mathbf{M}_n$  de rang plein)

Estimation Bayésienne : loi *a priori*  $\mathcal{N}(\mathbf{0}, \mathbf{\Lambda}_m)$  pour  $\gamma$ ,  
avec  $\mathbf{\Lambda}_m = \text{diag}\{\lambda_1, \dots, \lambda_m\}$

$$\hat{\gamma}_n = [\mathbf{\Phi}_n^\top \mathbf{\Phi}_n / \sigma^2 + \mathbf{\Lambda}_m^{-1}]^{-1} [\mathbf{\Phi}_n^\top \mathbf{y}_n / \sigma^2]$$

$$\text{cov}(\hat{\gamma}_n) = [\mathbf{\Phi}_n^\top \mathbf{\Phi}_n / \sigma^2 + \mathbf{\Lambda}_m^{-1}]^{-1} = \frac{\sigma^2}{n} \underbrace{\underbrace{\frac{1}{n} \sum_{i=1}^n \phi(\mathbf{x}_i) \phi^\top(\mathbf{x}_i)}_{\mathbf{M}_n}}_{\mathbf{M}_B(\mathbf{x}_n)} + \frac{\sigma^2}{n} \mathbf{\Lambda}_m^{-1} \Big]^{-1}$$

Estimation Bayésienne : loi *a priori*  $\mathcal{N}(\mathbf{0}, \mathbf{\Lambda}_m)$  pour  $\gamma$ ,  
avec  $\mathbf{\Lambda}_m = \text{diag}\{\lambda_1, \dots, \lambda_m\}$

$$\hat{\gamma}_n = [\mathbf{\Phi}_n^\top \mathbf{\Phi}_n / \sigma^2 + \mathbf{\Lambda}_m^{-1}]^{-1} [\mathbf{\Phi}_n^\top \mathbf{y}_n / \sigma^2]$$

$$\text{cov}(\hat{\gamma}_n) = [\mathbf{\Phi}_n^\top \mathbf{\Phi}_n / \sigma^2 + \mathbf{\Lambda}_m^{-1}]^{-1} = \frac{\sigma^2}{n} \underbrace{\underbrace{\frac{1}{n} \sum_{i=1}^n \phi(\mathbf{x}_i) \phi^\top(\mathbf{x}_i)}_{\mathbf{M}_n}}_{\mathbf{M}_B(\mathbf{x}_n)} + \frac{\sigma^2}{n} \mathbf{\Lambda}_m^{-1}]^{-1}$$

Prédiction en  $\mathbf{x}$  :  $\eta_n(\mathbf{x}) = \phi^\top(\mathbf{x}) \hat{\gamma}_n$

$$\text{IMSE}(\mathbf{X}_n) = \int_{\mathcal{X}} \phi^\top(\mathbf{x}) \text{cov}(\hat{\gamma}_n) \phi(\mathbf{x}) d\mu(\mathbf{x}) = \frac{\sigma^2}{n} \text{trace}[\mathbf{M}_B^{-1}(\mathbf{X}_n)] =$$

critère de *A*-optimalité appliqué à  $\mathbf{M}_B(\mathbf{X}_n)$   
( $\mathbf{M}_B(\mathbf{X}_n)$  de rang plein  $\forall m$ )

On dispose de tout l'arsenal de construction de plans optimaux (Pilz, 1983)

Plans exacts : Spöck and Pilz (2010)  $\Rightarrow$  prédiction de processus spatiaux  
(convergence non garantie vers l'optimum)

On dispose de tout l'arsenal de construction de plans optimaux (Pilz, 1983)

Plans exacts : Spöck and Pilz (2010)  $\Rightarrow$  prédiction de processus spatiaux  
(convergence non garantie vers l'optimum)

Plans approximatifs : convergence garantie vers l'optimum  $\xi^*$ ,  
mais il faut "arrondir" la solution (agrégation du support de  $\xi^*$ )

**En pratique** :

On ne contrôle pas le nb. de points  $N \Rightarrow$  des paramètres de réglage :

$m$  (niveau de troncature),

$$\alpha = \sigma^2/n$$

$N$  points  $\Rightarrow$  initialisation pour optimisation de la vraie  $\text{IMSE}(\mathbf{X}_N)$

(assez facile : optimum loin du bord de  $\mathcal{X}$ )

On dispose de tout l'arsenal de construction de plans optimaux (Pilz, 1983)

Plans exacts : Spöck and Pilz (2010)  $\Rightarrow$  prédiction de processus spatiaux  
(convergence non garantie vers l'optimum)

Plans approximatifs : convergence garantie vers l'optimum  $\xi^*$ ,  
mais il faut "arrondir" la solution (agrégation du support de  $\xi^*$ )

**En pratique** :

On ne contrôle pas le nb. de points  $N \Rightarrow$  des paramètres de réglage :

$m$  (niveau de troncature),

$$\alpha = \sigma^2/n$$

$N$  points  $\Rightarrow$  initialisation pour optimisation de la vraie IMSE( $\mathbf{X}_N$ )

(assez facile : optimum loin du bord de  $\mathcal{X}$ )

Valeurs propres et vecteurs propres :

$\blacktriangleright$  calcul sur un ensemble fini  $\mathcal{X}_Q$  de  $Q$  valeurs  $\mathbf{x}_k$ ,  $1 \leq k \leq Q$

$\blacktriangleright$  diagonalisation de  $\mathbf{Q}\mathbf{W}$  avec

$$\{\mathbf{Q}\}_{kl} = C(\mathbf{x}^{(k)}, \mathbf{x}^{(\ell)}), \mathbf{W} = \text{diag}\{w_1, \dots, w_Q\}$$

( $w_k = 1/Q$  pour  $\mu$  uniforme)

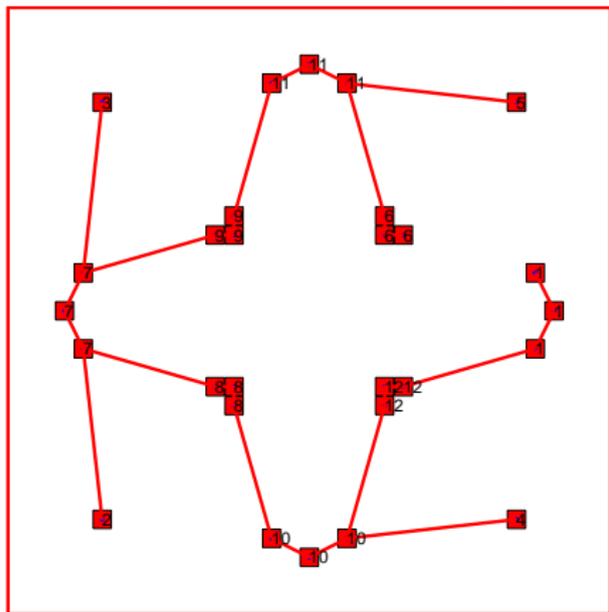
Exemple :

$$d = 2, C(\mathbf{x}, \mathbf{x}') = (1 + 10\|\mathbf{x} - \mathbf{x}'\|) \exp(-10\|\mathbf{x} - \mathbf{x}'\|) \text{ (Matérn } 3/2)$$

$$\mathcal{X}_Q = \text{grille régulière de } 33 \times 33 = 1089 \text{ points, } \sigma^2 = \sum_{i>m} \lambda_i$$

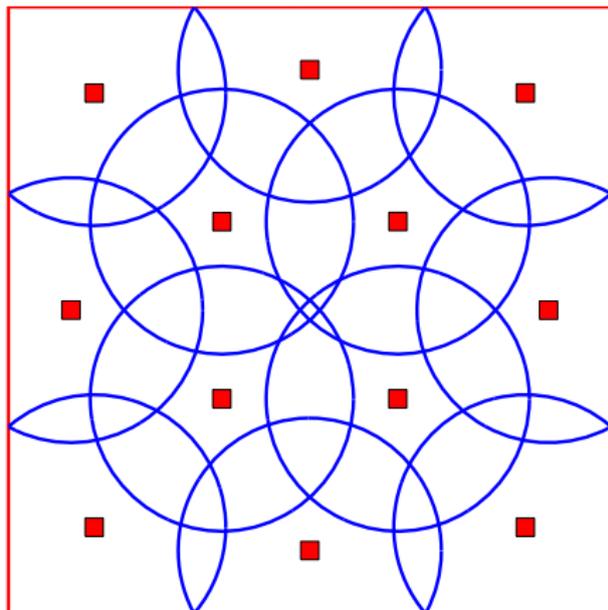
$$m = n = 10$$

agrégation du support de  $\xi^*$



$$m = n = 10$$

$$\Rightarrow N = 12$$

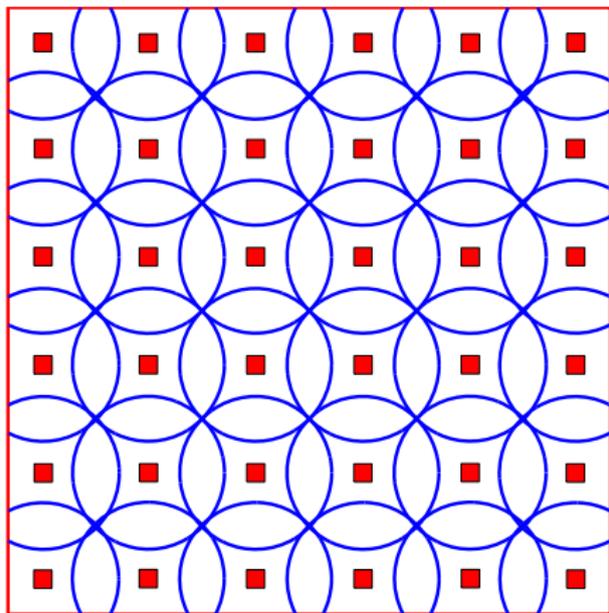


Exemple :

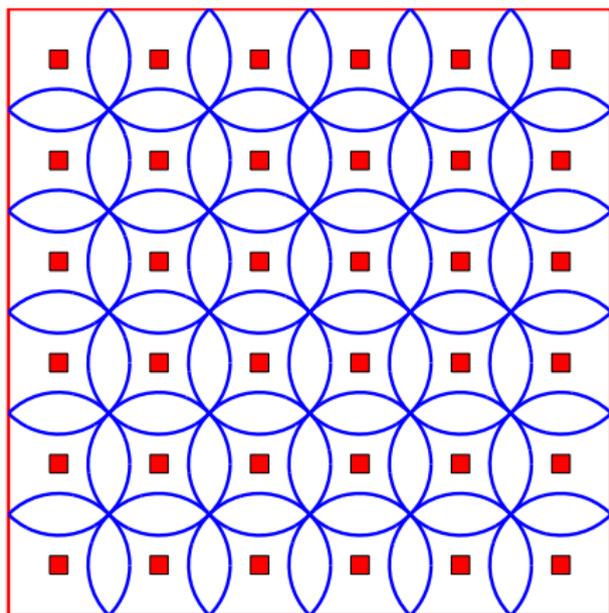
$d = 2$ ,  $C(\mathbf{x}, \mathbf{x}') = (1 + 10\|\mathbf{x} - \mathbf{x}'\|) \exp(-10\|\mathbf{x} - \mathbf{x}'\|)$  (Matérn 3/2)

$\mathcal{X}_Q =$  grille régulière de  $33 \times 33 = 1089$  points,  $\sigma^2 = \sum_{i>m} \lambda_i$

$m = 30, n = 10 \Rightarrow N = 36$



$\mathbf{X}_{36}$  miniMax optimal



Utilisable pour  $\mathcal{X}$  quelconque : travailler avec  $\mathcal{X}_Q$  obtenu à partir de  $Q$  points d'une suite à faible discrédance sur  $\mathcal{X}$

Utilisable pour  $\mathcal{X}$  quelconque : travailler avec  $\mathcal{X}_Q$  obtenu à partir de  $Q$  points d'une suite à faible discrédance sur  $\mathcal{X}$

Avec tendance,  $f(\mathbf{x}) = Z(\mathbf{x}) + \mathbf{r}^\top(\mathbf{x})\boldsymbol{\beta}$  ?

On peut faire la même chose (Spöck and Pilz, 2010), avec des formules un peu plus compliquées

Utilisable pour  $\mathcal{X}$  quelconque : travailler avec  $\mathcal{X}_Q$  obtenu à partir de  $Q$  points d'une suite à faible discrédance sur  $\mathcal{X}$

Avec tendance,  $f(\mathbf{x}) = Z(\mathbf{x}) + \mathbf{r}^\top(\mathbf{x})\beta$  ?

On peut faire la même chose (Spöck and Pilz, 2010), avec des formules un peu plus compliquées

Plus délicat : on voudrait ne pas mélanger les fonctions propres  $\varphi_i(\cdot)$  avec les composantes  $\{\mathbf{r}\}_j(\cdot)$  de la tendance

▣ travail en cours (LP + Bertrand Gauthier)

▶ Conclusions partie (2)

## 4 Au delà des plans *space filling*

**Plans optimaux pour le krigeage : une difficulté cachée :**  
 **$\theta$  dans la covariance  $C(\cdot; \theta)$  est inconnu**

⇒ estimer  $\theta$  à partir des mêmes observations que celles utilisées pour construire  $\eta_n(x)$

## 4 Au delà des plans *space filling*

**Plans optimaux pour le krigeage : une difficulté cachée :**  
 $\theta$  dans la covariance  $C(\cdot; \theta)$  est inconnu

⇒ estimer  $\theta$  à partir des mêmes observations que celles utilisées pour construire  $\eta_n(\mathbf{x})$

⇒ Maximum de vraisemblance  $\hat{\theta}^n$   
 ( $Z(\mathbf{x})$  est supposé Gaussien)

⇒ terme correctif (Harville and Jeske, 1992; Abt, 1999) :

$$\hat{\rho}_n(\mathbf{x}; \theta) = \rho_n(\mathbf{x}; \theta) + \text{trace}\left\{\mathbf{M}_\theta^{-1} \frac{\partial \mathbf{v}_n(\mathbf{x}; \theta)}{\partial \theta} \mathbf{C}_n \frac{\partial \mathbf{v}_n(\mathbf{x}; \theta)}{\partial \theta^\top}\right\}$$

avec :

$\mathbf{v}_n(\mathbf{x}; \theta)$  tel que  $\eta_n(\mathbf{x}) = \mathbf{v}_n^\top(\mathbf{x}; \theta) \mathbf{y}_n$

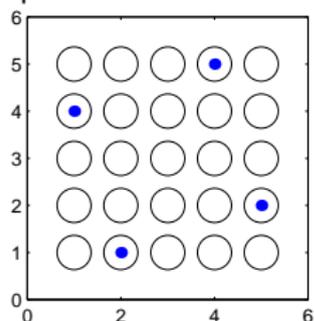
$\mathbf{M}_\theta$  = matrice d'information de Fisher (MIF) pour  $\theta$

Exemple (Zimmerman, 2006) :

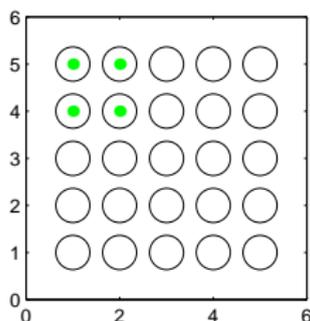
$$E\{Z(\mathbf{x})Z(\mathbf{x}')\} = \theta^{\|\mathbf{x}-\mathbf{x}'\|}, \theta = 0.3$$

$\mathcal{X}$  = grille régulière  $5 \times 5$

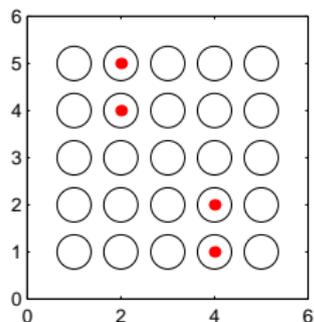
prediction for  $\theta$  known



estimation of  $\theta$



prediction with  $\theta$  estimated



prédiction pour  $\theta$  connu :

$\mathbf{X}_4$  minimise  $\max_{\mathbf{x} \in \mathcal{X}} \rho_4(\mathbf{x})$

estimation de  $\theta$ :

$\mathbf{X}_4$  maximise  $\det \mathbf{M}_\theta$

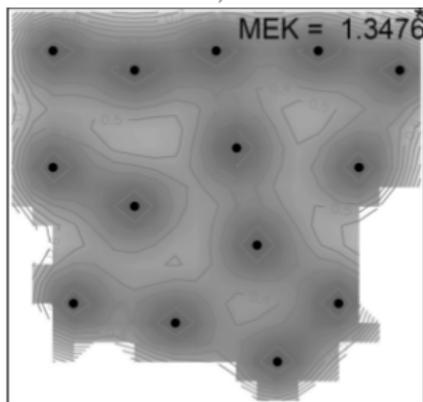
prédiction avec  $\theta$  estimé :

$\mathbf{X}_4$  minimise  $\max_{\mathbf{x} \in \mathcal{X}} \hat{\rho}_4(\mathbf{x}; \theta)$

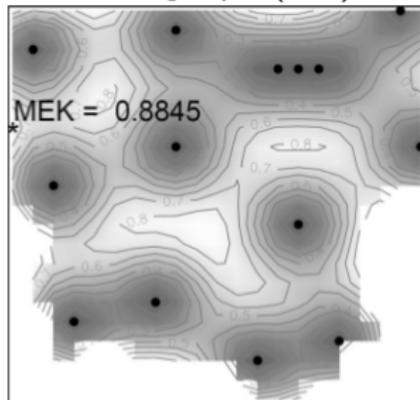
Exemple (Müller et al., 2014) :

Concentration en NH4 en mer du Nord (collaboration avec MUMM, Belgique) — données simulées, krigeage avec noyau Matérn 3/2

$\hat{\rho}_{14}(\mathbf{x}; \theta)$  pour plan miniMax-optimal  
 $\mathbf{X}_{mM, n=14}^*$



$\hat{\rho}_{14}(\mathbf{x}; \theta)$  pour  $\mathbf{X}_{14}^*$  qui minimise  
 $\max_{\mathbf{x} \in \mathcal{X}} \hat{\rho}_{14}(\mathbf{x}; \theta)$



Choisir  $\mathbf{X}_n$  qui minimise  $\max_{\mathbf{x} \in \mathcal{X}} \hat{\rho}_n(\mathbf{x}; \theta)$  est difficile

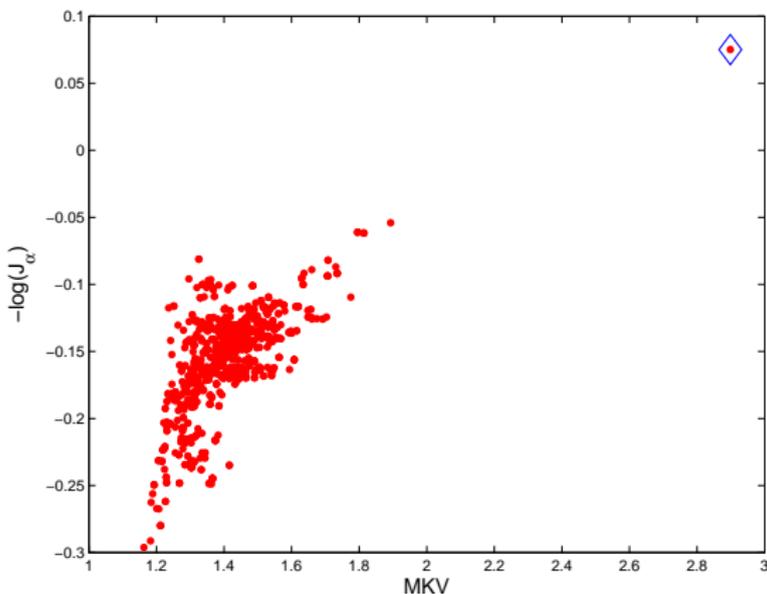
Choisir  $\mathbf{X}_n$  qui minimise  $\max_{\mathbf{x} \in \mathcal{X}} \hat{\rho}_n(\mathbf{x}; \theta)$  est difficile

⇒ utiliser un critère compromis entre *space filling* et agrégation de points

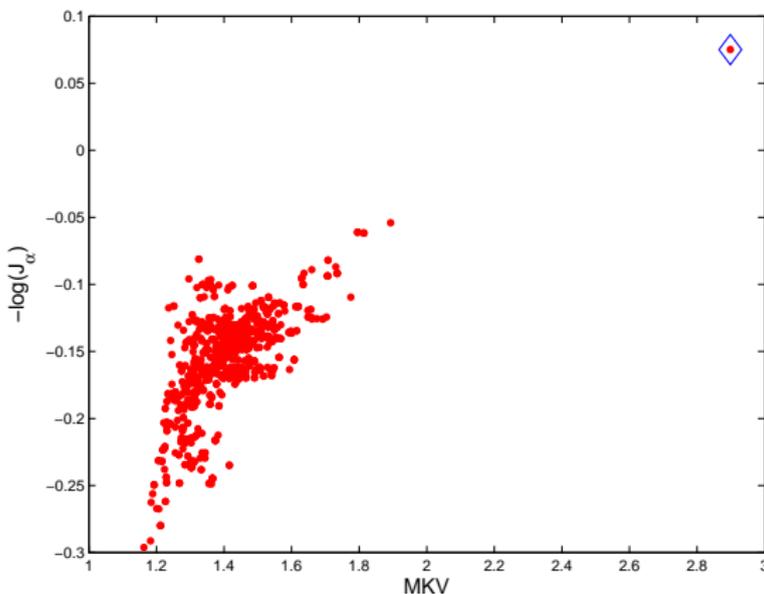
par ex., choisir  $\mathbf{X}_n$  qui maximise  $\gamma \log \det(\mathbf{M}_\beta) + (1 - \gamma) \log \det(\mathbf{M}_\theta)$  (Müller et al., 2011, 2014), avec

- $\mathbf{M}_\beta$  = MIF pour paramètres  $\beta$  de tendance (maximisation → *space filling*)
- $\mathbf{M}_\theta$  = MIF pour paramètres  $\theta$  de corrélation (maximization → *agrégation*)

Exemple :  $n = 7$ ,  $d = 2$ ,  $C(\mathbf{x} - \mathbf{x}'; \theta) = \exp(-\theta \|\mathbf{x} - \mathbf{x}'\|)$ ,  $\theta = 0.7$ ,  
 1000 Lh (999 aléatoires +  $\diamond$  pour plan Maximin optimal)  
 MKV =  $\max_{\mathbf{x} \in \mathcal{X}} \hat{\rho}_n(\mathbf{x}; \theta)$ ,  $J_\alpha = \det^\alpha(\mathbf{M}_\beta) + \det^{1-\alpha}(\mathbf{M}_\theta)$  ( $\alpha = 0.8$ )



Exemple :  $n = 7$ ,  $d = 2$ ,  $C(\mathbf{x} - \mathbf{x}'; \theta) = \exp(-\theta \|\mathbf{x} - \mathbf{x}'\|)$ ,  $\theta = 0.7$ ,  
 1000 Lh (999 aléatoires +  $\diamond$  pour plan Maximin optimal)  
 $MKV = \max_{\mathbf{x} \in \mathcal{X}} \hat{\rho}_n(\mathbf{x}; \theta)$ ,  $J_\alpha = \det^\alpha(\mathbf{M}_\beta) + \det^{1-\alpha}(\mathbf{M}_\theta)$  ( $\alpha = 0.8$ )



Cependant, l'effet du terme correctif dans

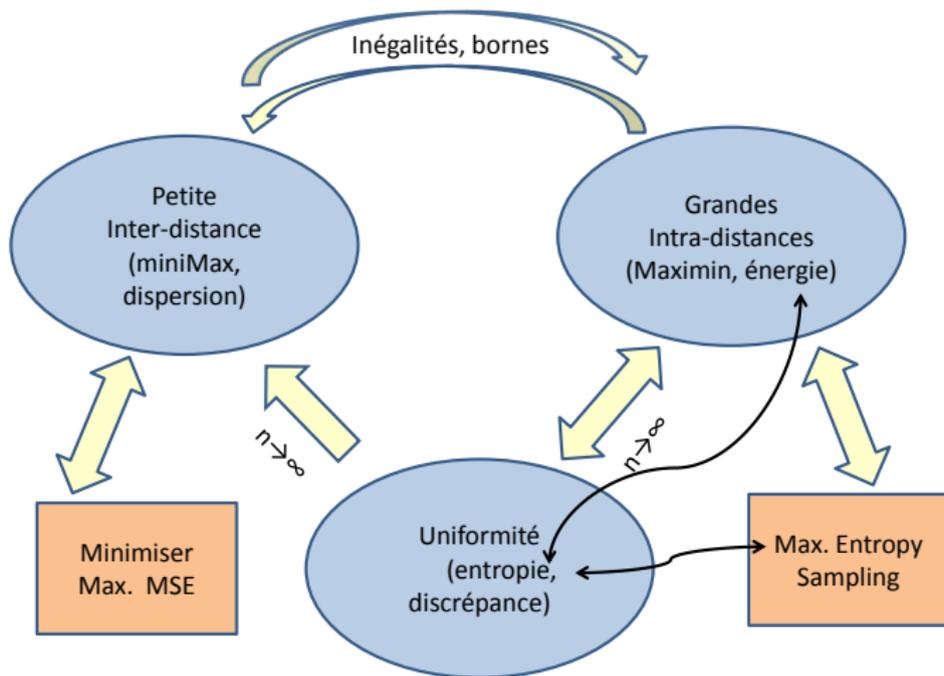
$$\hat{\rho}_n(\mathbf{x}; \theta) = \rho_n(\mathbf{x}; \theta) + \text{trace}\left\{ \mathbf{M}_\theta^{-1} \frac{\partial \mathbf{v}_n(\mathbf{x}; \theta)}{\partial \theta} \mathbf{C}_n \frac{\partial \mathbf{v}_n(\mathbf{x}; \theta)}{\partial \theta^\top} \right\}$$

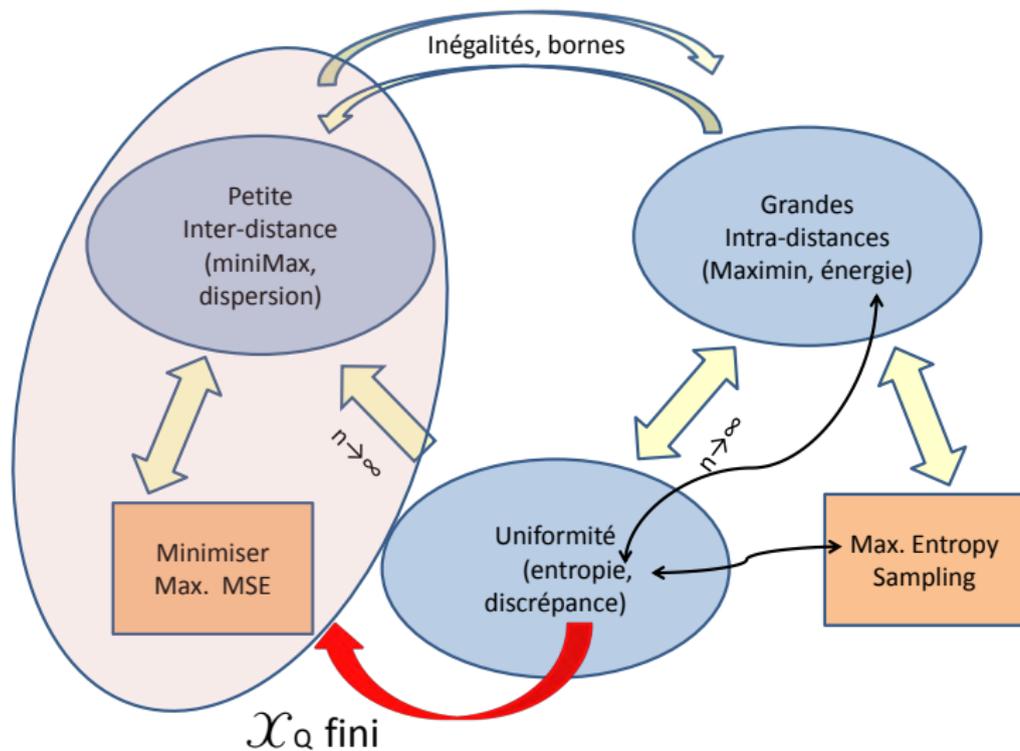
s'atténue rapidement quand  $n$  augmente

## 5 Conclusions partie (2) — avec modèle

- Les critères reposant sur un modèle de type processus Gaussien dépendent de la covariance choisie (de  $\theta$  dans  $C(\cdot, \theta)$ )
  - entropie, MMSE, IMSE
    - ➡ moyenne sur  $\theta$ , (Joseph et al., 2015) (entropie)
    - ➡ pire cas par rapport à  $\theta$ , (Spöck and Pilz, 2010) (IMSE)
- Mettre un peu plus de points vers les bords qu'au centre du domaine
- Mettre quelques points proches pour faciliter l'estimation de  $\theta$

Tester plusieurs approches (aucune n'est parfaite) en comparant les valeurs de différents critères





# Références I

- Abt, M., 1999. Estimating the prediction mean squared error in gaussian stochastic processes with exponential correlation structure. *Scandinavian Journal of Statistics* 26 (4), 563–578.
- Atwood, C., 1973. Sequences converging to  $D$ -optimal designs of experiments. *Annals of Statistics* 1 (2), 342–352.
- Böhning, D., 1985. Numerical estimation of a probability measure. *Journal of Statistical Planning and Inference* 11, 57–69.
- Böhning, D., 1986. A vertex-exchange-method in  $D$ -optimal design theory. *Metrika* 33, 337–347.
- Chernoff, H., 1953. Locally optimal designs for estimating parameters. *Annals of Math. Stat.* 24, 586–602.
- Dette, H., Pepelyshev, A., 2010. Generalized latin hypercube design for computer experiments. *Technometrics* 52 (4), 421–429.
- Fedorov, V., 1972. *Theory of Optimal Experiments*. Academic Press, New York.
- Gauthier, B., Pronzato, L., 2014a. Approximation of IMSE-optimal designs via quadrature rules and spectral decomposition. *Communications in Statistics – Simulation and Computation* (to appear).
- Gauthier, B., Pronzato, L., 2014b. Spectral approximation of the IMSE criterion for optimal designs in kernel-based interpolation models. *SIAM/ASA J. Uncertainty Quantification* 2, 805–825, doi 10.1137/130928534.
- Harville, D., Jeske, D., 1992. Mean squared error of estimation or prediction under a general linear model. *Journal of the American Statistical Association* 87 (419), 724–731.
- Johnson, M., Moore, L., Ylvisaker, D., 1990. Minimax and maximin distance designs. *Journal of Statistical Planning and Inference* 26, 131–148.

# Références II

- Joseph, V., Gul, E., Ba, S., 2015. Maximum projection designs for computer experiments. *Biometrika* To appear.
- Kiefer, J., Wolfowitz, J., 1960. The equivalence of two extremum problems. *Canadian Journal of Mathematics* 12, 363–366.
- Ko, C.-W., Lee, J., Queyranne, M., 1995. An exact algorithm for maximum entropy sampling. *Operations Research* 43 (4), 684–691.
- Mitchell, T., 1974. An algorithm for the construction of “*D*-optimal” experimental designs. *Technometrics* 16, 203–210.
- Müller, W., Pronzato, L., Rendas, J., Waldl, H., 2014. Efficient prediction designs for random fields. *Applied Stochastic Models in Business and Industry* To appear.
- Müller, W., Pronzato, L., Waldl, H., 2011. Beyond space-filling: An illustrative case. *Procedia Environmental Sciences* 7, 14–19, doi 10.1016/j.proenv.2011.07.004.
- Pilz, J., 1983. *Bayesian Estimation and Experimental Design in Linear Regression Models*. Vol. 55. Teubner-Texte zur Mathematik, Leipzig, (also Wiley, New York, 1991).
- Pukelsheim, F., 1993. *Optimal Experimental Design*. Wiley, New York.
- Pukelsheim, F., Reider, S., 1992. Efficient rounding of approximate designs. *Biometrika* 79 (4), 763–770.
- Sacks, J., Welch, W., Mitchell, T., Wynn, H., 1989. Design and analysis of computer experiments. *Statistical Science* 4 (4), 409–435.
- Santner, T., Williams, B., Notz, W., 2003. *The Design and Analysis of Computer Experiments*. Springer, Heidelberg.
- Schwabe, R., 1996. *Optimum Designs for Multi-Factor Models*. Springer, New York.

## Références III

- Shewry, M., Wynn, H., 1987. Maximum entropy sampling. *Applied Statistics* 14, 165–170.
- Silvey, S., 1980. *Optimal Design*. Chapman & Hall, London.
- Spöck, G., Pilz, J., 2010. Spatial sampling design and covariance-robust minimax prediction based on convex design ideas. *Stochastic Environmental Research and Risk Assessment* 24 (3), 463–482.
- Titterton, D., 1976. Algorithms for computing  $D$ -optimal designs on a finite design space. In: *Proc. of the 1976 Conference on Information Science and Systems*. Dept. of Electronic Engineering, John Hopkins University, Baltimore, pp. 213–216.
- Torsney, B., 1983. A moment inequality and monotonicity of an algorithm. In: Kortanek, K., Fiacco, A. (Eds.), *Proc. Int. Symp. on Semi-infinite Programming and Applications*. Springer, Heidelberg, pp. 249–260.
- Torsney, B., 2009.  $W$ -iterations and ripples therefrom. In: Pronzato, L., Zhigljavsky, A. (Eds.), *Optimal Design and Related Areas in Optimization and Statistics*. Springer, Ch. 1, pp. 1–12.
- Welch, W., 1982. Branch-and-bound search for experimental designs based on  $D$ -optimality and other criteria. *Technometrics* 24 (1), 41–28.
- Wu, C., 1978. Some algorithmic aspects of the theory of optimal designs. *Annals of Statistics* 6 (6), 1286–1301.
- Wynn, H., 1970. The sequential generation of  $D$ -optimum experimental designs. *Annals of Math. Stat.* 41, 1655–1664.
- Yu, Y., 2010. Strict monotonicity and convergence rate of Titterton's algorithm for computing  $D$ -optimal designs. *Comput. Statist. Data Anal.* 54, 1419–1425.
- Yu, Y., 2011.  $D$ -optimal designs via a cocktail algorithm. *Stat. Comput.* 21, 475–481.
- Zhigljavsky, A., Dette, H., Pepelyshev, A., 2010. A new approach to optimal design for linear models with correlated observations. *Journal of the American Statistical Association* 105 (491), 1093–1103.
- Zimmerman, D., 2006. Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. *Environmetrics* 17 (6), 635–652.