

# Plans exploratoires pour l'interpolation par noyaux sur des domaines non hypercubiques

Yves AUFFRAY<sup>1</sup> , Pierre BARBILLON<sup>2</sup> , Jean-Michel MARIN<sup>3</sup>

Université Paris-Sud 11  
Projet Select (Inria Saclay -île-de-France)

[pierre.barbillon@math.u-psud.fr](mailto:pierre.barbillon@math.u-psud.fr)

---

<sup>1</sup>Dassault Aviation, Université Paris-sud 11

<sup>2</sup>Université Paris-Sud 11, INRIA Saclay

<sup>3</sup>Université Montpellier 2

# Problématique

La modélisation mathématique d'un phénomène physique peut s'avérer très couteuse en terme de temps de calcul.

$$y = f(x^1, \dots, x^d) = f(\mathbf{x})$$

où  $\mathbf{x} = (x^1, \dots, x^d) \in E$  et  $y \in \mathbb{R}$ .

- $\mathbf{x} = (x^1, \dots, x^d)$  est le vecteur des variables d'entrées ;
- $y$  est une variable de sortie ;
- $E$  est l'espace de départ supposé compact inclus dans  $\mathbb{R}^d$ .

# Objectifs

**But :** Proposer un métamodèle à partir d'un plan d'expérience (ou Design)

$$\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$$

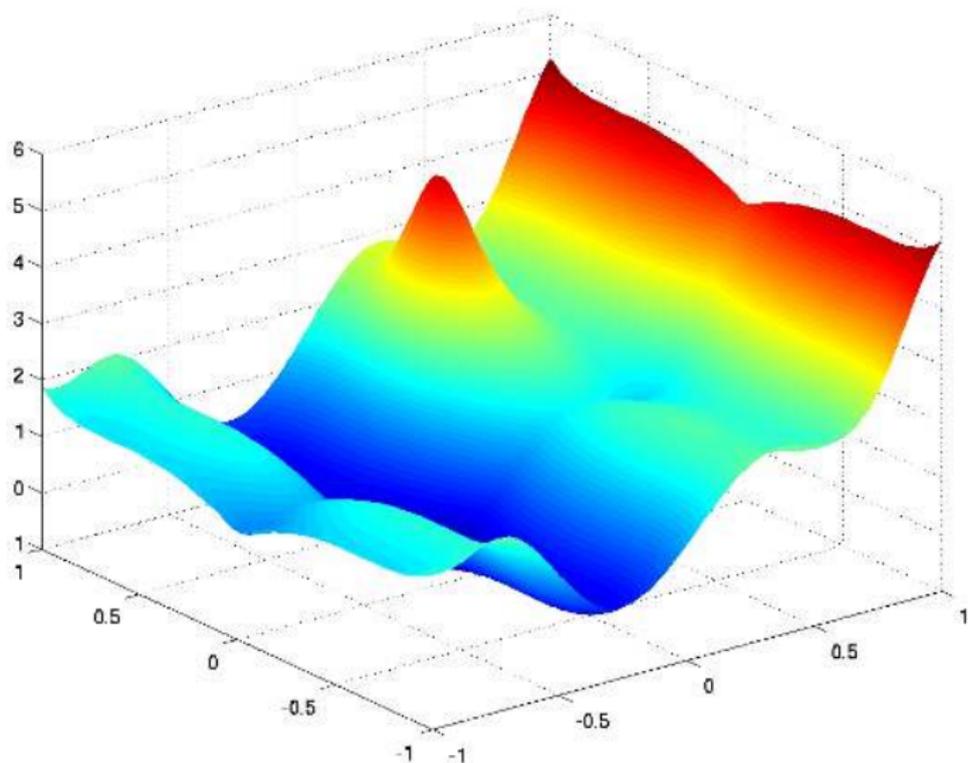
en lequel on calcule les

$$\{f(\mathbf{x}_1), \dots, f(\mathbf{x}_N)\}.$$

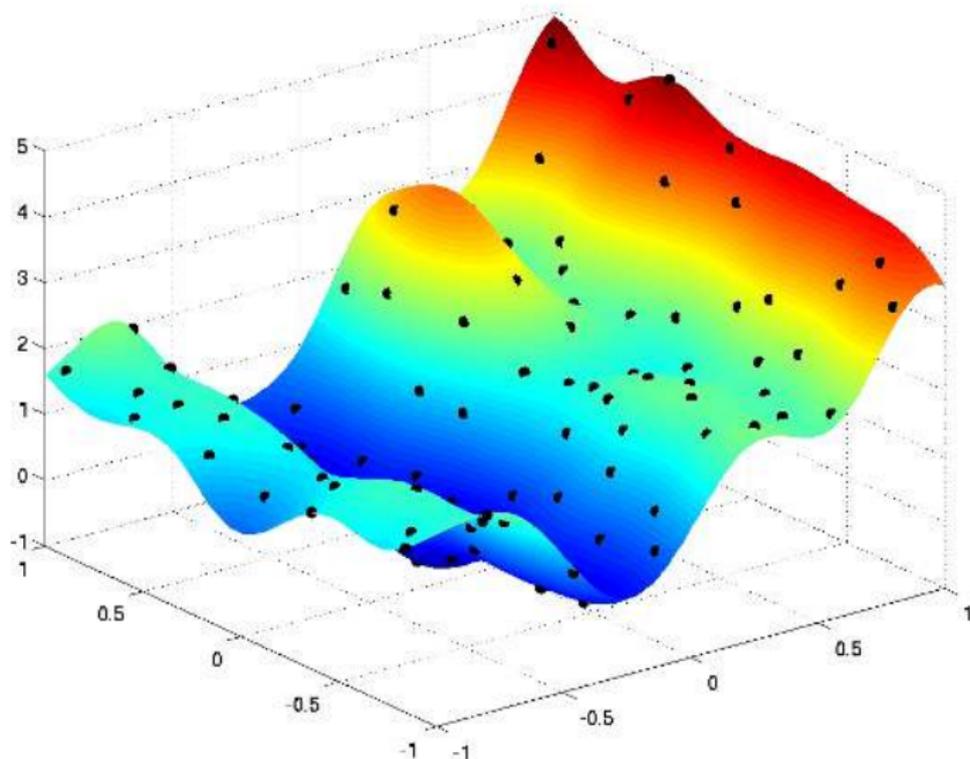
Ceci, dans pour répondre à des besoins de :

- pré-visualisation et étude préliminaire ;
- prédiction et optimisation ;
- analyse de sensibilité ;
- analyse probabiliste.

## Exemple : la vraie fonction



## Exemple : la fonction reconstituée



# Plan

- 1 Interpolation par noyaux
  - Noyaux définis positifs
  - Erreur en un point
  - Noyaux conditionnellement définis positifs
  
- 2 Plans exploratoires
  - Critères de dispersion
  - Recuit simulé pour le critère maximin

# Noyau défini positif

Nous appelons noyau une fonction  $K : E \times E \rightarrow \mathbb{R}$  symétrique.

## Définition : Défini positif

$K : E \times E \rightarrow \mathbb{R}$  est un noyau défini positif si  
 $\forall N \in \mathbb{N}, \forall \mathbf{x}_1, \dots, \mathbf{x}_N \in E, \forall \lambda_1, \dots, \lambda_N \in \mathbb{R}$

$$\sum_{i,j} \lambda_i \lambda_j K(\mathbf{x}_i, \mathbf{x}_j) \geq 0.$$

## Exemples :

- Noyaux gaussiens  $K_\theta(\mathbf{x}, \mathbf{y}) = e^{-\theta \|\mathbf{x} - \mathbf{y}\|_2^2}$
- Noyaux exponentiels  $K_\theta(\mathbf{x}, \mathbf{y}) = e^{-\theta \|\mathbf{x} - \mathbf{y}\|_\nu^\nu}$  ( $0 < \nu \leq 2$ )
- Noyaux multiquadratiques inverses  $K_\theta(\mathbf{x}, \mathbf{y}) = \frac{1}{\sqrt{1 + \theta \|\mathbf{x} - \mathbf{y}\|^2}}$

## R.K.H.S.

On notera :

- $K_{\mathbf{x}}$  la fonction partielle  $\mathbf{x}' \mapsto K(\mathbf{x}, \mathbf{x}')$  ;
- $\mathcal{F}_K$  l'espace de fonctions engendré par  $\{K_{\mathbf{x}} : \mathbf{x} \in E\}$ .

### Théorème (Aronszajn)

Soit  $K$  un noyau positif. Il existe un unique espace de Hilbert de fonctions  $(\mathcal{H}_K, \langle \cdot, \cdot \rangle_{\mathcal{H}_K})$  tel que :

- 1  $\mathcal{F}_K \subset \mathcal{H}_K$
- 2 La propriété de reproduction est satisfaite :

$$\forall f \in \mathcal{H}_K, \mathbf{x} \in E, f(\mathbf{x}) = \langle f, K_{\mathbf{x}} \rangle_{\mathcal{H}_K} .$$

$\mathcal{H}_K$  est un espace de Hilbert à noyau reproduisant (R.K.H.S.) de noyau reproduisant  $K$ .

# Interpolation

Nous fixons  $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \subset E$  et nous notons :

- $\mathcal{H}_K(\mathbf{X}) = \text{Vect}(K_{\mathbf{x}_1}, \dots, K_{\mathbf{x}_N})$  le sous espace vectoriel de  $\mathcal{H}_K$  ;
- Pour  $f \in \mathcal{H}_K$ , on note  $S_{K,\mathbf{X}}(f)$  le projeté orthogonal de  $f$  sur  $\mathcal{H}_K(\mathbf{X})$ .

## Proposition

$S_{K,\mathbf{X}}(f)$  est l'interpolateur de  $f$ , de norme minimale : i.e.  $S_{K,\mathbf{X}}(f)$  est solution du problème

$$\begin{cases} \min_{g \in \mathcal{H}_K} \|g\|_{\mathcal{H}_K} \\ g(\mathbf{x}_k) = f(\mathbf{x}_k), \quad k = 1, \dots, N \end{cases} .$$

# Formulation lagrangienne

- $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \subset E$ .
- $K$  est un noyau défini positif.

## Proposition

Il existe des uniques  $u_1, \dots, u_N$  appartenant à  $\mathcal{F}_K(\mathbf{X})$ , tels que

$$u_k(\mathbf{x}_l) = \delta_{k,l}, \forall k, l \in \{1, \dots, N\}. \quad (1)$$

Ainsi, pour tout  $f \in \mathcal{H}_K$ ,  $S_{K,\mathbf{X}}(f) = \sum_{k=1}^N f(\mathbf{x}_k)u_k$ .

Soit  $\mathbf{x} \in E$ ,

$$\begin{aligned} |f(\mathbf{x}) - S_{K,\mathbf{x}}(f)(\mathbf{x})| &= \left| \langle K_{\mathbf{x}} - \sum_{k=1}^N u_k(\mathbf{x}) K_{\mathbf{x}_k}, f \rangle_{\mathcal{H}_K} \right| \\ &\leq \|f\|_{\mathcal{H}_K} \|K_{\mathbf{x}} - \sum_{k=1}^N u_k(\mathbf{x}) K_{\mathbf{x}_k}\|_{\mathcal{H}_K}. \end{aligned}$$

On note  $P_{\mathbf{x}}(\mathbf{x}) = \|K_{\mathbf{x}} - \sum_{k=1}^N u_k(\mathbf{x}) K_{\mathbf{x}_k}\|_{\mathcal{H}_K}$ .

## Majoration pour un noyau gaussien

Dans le cas d'un noyau gaussien  $K_\theta(\mathbf{x}, \mathbf{y}) = e^{-\theta \|\mathbf{x} - \mathbf{y}\|_2^2}$ , la majoration suivante est donnée dans [Schaback \(1994\)](#) :

$$P_{\mathbf{x}}(\mathbf{x}) \leq \exp\left(-\frac{\delta}{2h_{\mathbf{x}}^2}\right),$$

où  $\delta$  dépend du  $\theta$  du noyau et

$$h_{\mathbf{x}} = \sup_{\mathbf{y}} \min_{1 \leq j \leq N} \|\mathbf{y} - \mathbf{x}_j\|.$$

d'où la majoration sur l'interpolateur :

$$\sup_{\mathbf{x}} |f(\mathbf{x}) - S_{K, \mathbb{P}, \mathbf{x}}(f)(\mathbf{x})| \leq \|f\|_{\mathcal{H}_K} \exp\left(-\frac{\delta}{2h_{\mathbf{x}}^2}\right).$$

Soit  $\mathbb{P} \subset \mathbb{R}^E$  un espace vectoriel de dimension  $n$ .

Dans [Auffray & Barbillon \(2009\)](#), nous donnons une définition plus générale que celle usuellement adoptée ([Schaback \(2007\)](#)) du concept de noyau  $\mathbb{P}$ -conditionnellement défini positif.

Nous étendons les résultats précédents à ce nouveau paradigme. Plus précisément,

- Il existe un unique espace semi-hilbertien associé au couple  $(K, \mathbb{P})$ , noté  $\mathcal{H}_{K, \mathbb{P}}$  qui généralise le R.K.H.S. associé à un noyau défini positif.
- Pour tout  $f \in \mathcal{H}_{K, \mathbb{P}}$  et un design  $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$  fixé, nous donnons  $S_{K, \mathbb{P}, \mathbf{X}}(f)$  qui est l'interpolateur de semi-norme minimale de  $f$  sur  $\mathbf{X}$  et qui généralise  $S_{K, \mathbf{X}}(f)$ .

Nous pouvons ainsi traiter les cas :

- où  $K$  est un noyau défini positif associé à un espace  $\mathbb{P}$  de polynôme de degré au plus  $n - 1$ .
- des différents types de krigeage :  $S_{K, \mathbb{P}, \mathbf{x}}(f)$  correspond alors aux estimateurs BLUP.
- des noyaux de splines de plaques minces ([Wahba \(1990\)](#)) associés à un espace de polynôme.

# Minimax

Un plan d'expérience  $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$  est dit minimax s'il **minimise** :

$$h_{\mathbf{X}} = \sup_{\mathbf{y}} \min_{1 \leq j \leq N} \|\mathbf{y} - \mathbf{x}_j\|.$$

- Ce critère est adapté à la convergence de la méthode d'interpolation par noyau.
- Mais, il est très difficile de construire un plan minimax en pratique !

# Maximin

Un plan  $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$  est dit maximin s'il **maximise** :

$$\zeta_{\mathbf{X}} = \min_{1 \leq i, j \leq N} \|\mathbf{x}_i - \mathbf{x}_j\|,$$

et si le nombre de paires réalisant la distance  $\zeta_{\mathbf{X}}$  est minimal.

# Convergence pour un plan maximin

## Proposition

Soit un plan maximin  $\mathbf{X}$  avec  $\zeta_{\mathbf{X}} = \min_{1 \leq i, j \leq N} \|\mathbf{x}_i - \mathbf{x}_j\|$ . La réunion des boules  $\mathcal{B}(\mathbf{x}_i, \zeta_{\mathbf{X}})$  recouvre le domaine.

Ainsi pour ce plan  $\mathbf{X}$ ,

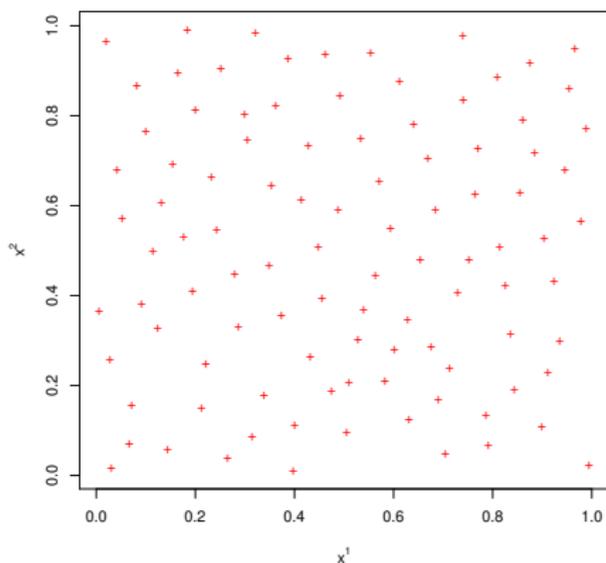
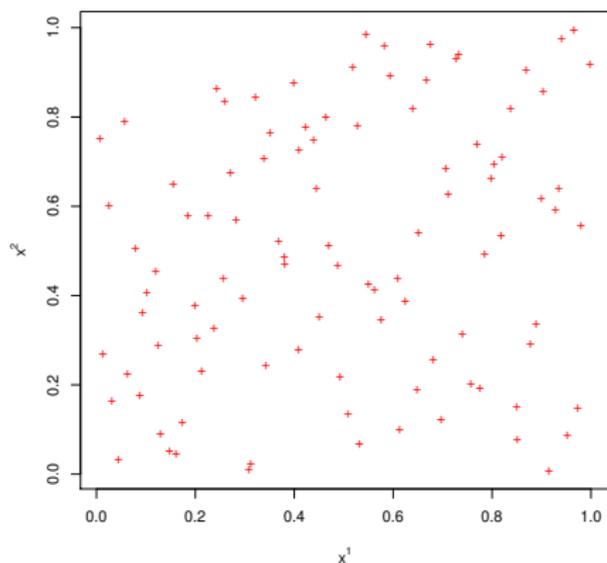
$$h_{\mathbf{X}} = \sup_{\mathbf{y}} \min_{1 \leq j \leq N} \|\mathbf{y} - \mathbf{x}_j\| \leq \zeta_{\mathbf{X}}.$$

D'où la majoration,

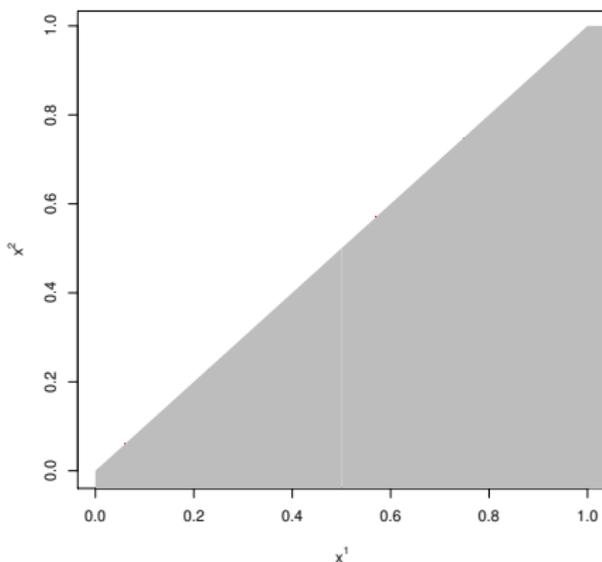
$$\sup_{\mathbf{x}} |f(\mathbf{x}) - S_{K, \mathbf{X}}(f)(\mathbf{x})| \leq \|f\|_{\mathcal{H}_K} \exp\left(-\frac{\delta}{2\zeta_{\mathbf{X}}^2}\right).$$

## Technique standard : hypercube latin maximin

**Morris & Mitchell (1995)** ont proposé un algorithme de type recuit simulé afin d'obtenir un plan maximin dans la classe des hypercubes latins.



Un plan en hypercube latin n'a plus de sens dans un domaine non hypercubique et dans certains cas, il est même impossible à obtenir.



# Algorithme

On propose un algorithme basé sur une méthode de recuit simulé dans un espace d'état  $E^N \subset \mathbb{R}^{d \times N}$  afin de maximiser la fonction  $\mathbf{X} \mapsto \zeta_{\mathbf{x}}$ .

## Initialisation

Tirage uniforme d'un grand nombre de points dans  $E$ .

Estimation la covariance empirique des points tirés, notée  $\hat{\Sigma}$ .

$N$  points sont gardés et forme le design initial  $\mathbf{X}^{(0)} = \{\mathbf{x}_1^{(0)}, \dots, \mathbf{x}_N^{(0)}\}$ , point de départ de l'algorithme.

Choix d'une température et d'une variance initiales.

Choix de schémas de diminution de la température et de la variance.

- 1 Choix d'une paire de points  $(\mathbf{x}_i^{(t)}, \mathbf{x}_j^{(t)})$  dans  $\mathbf{X}^{(t)}$  suivant une loi multinomiale dont les probabilités sont proportionnelles à  $\frac{1}{\|\mathbf{x}_i^{(t)} - \mathbf{x}_j^{(t)}\|}$  ;
- 2 Choix avec probabilité  $(\frac{1}{2})$  de l'un des deux points, notons  $\mathbf{x}_k^{(t)}$  le point retenu ;
- 3 Utilisation d'une marche aléatoire gaussienne contrainte pour proposer un nouveau point :

$$\mathbf{x}_k^{prop} \sim \mathcal{N}_d(\mathbf{x}_k^{(t)}, \min(\tau_{min}, \tau_t) \hat{\Sigma}) \mathbf{1}(\cdot \in E),$$

la configuration proposée est notée

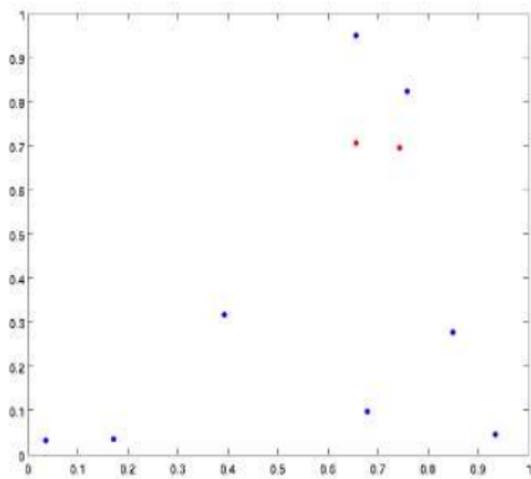
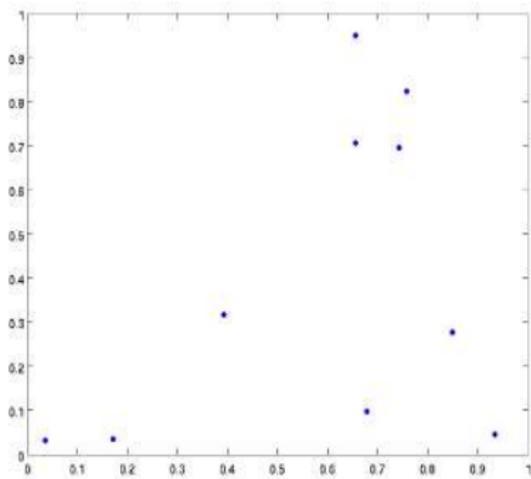
$$\mathbf{X}^{prop} = \{\mathbf{x}_1^{(t)}, \dots, \mathbf{x}_{k-1}^{(t)}, \mathbf{x}_k^{prop}, \mathbf{x}_{k+1}^{(t)}, \dots, \mathbf{x}_N^{(t)}\};$$

- 4 Accepter  $\mathbf{X}^{(t+1)} = \mathbf{X}^{prop}$  avec probabilité

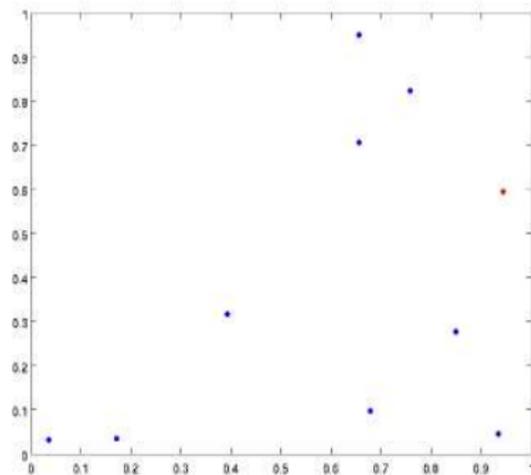
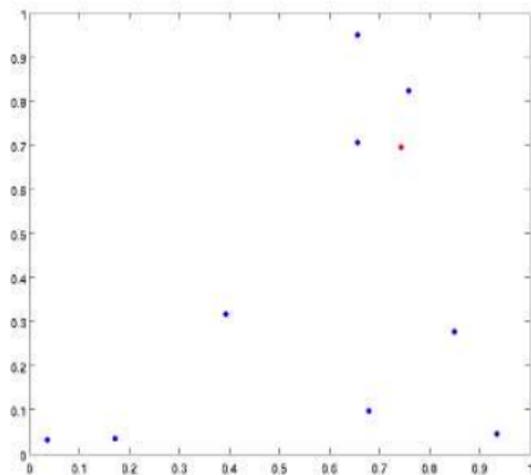
$$\min \left( 1, \exp \left( \frac{\zeta_{\mathbf{X}^{prop}} - \zeta_{\mathbf{X}^{(t)}}}{T_t} \right) \right),$$

sinon  $\mathbf{X}^{(t+1)} = \mathbf{X}^{(t)}$ .

# Choix d'une paire de points

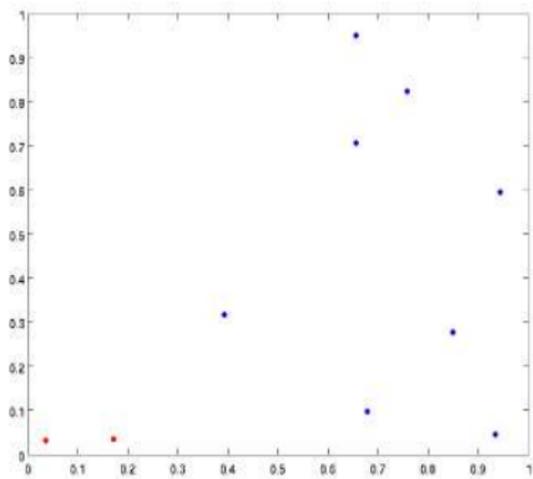
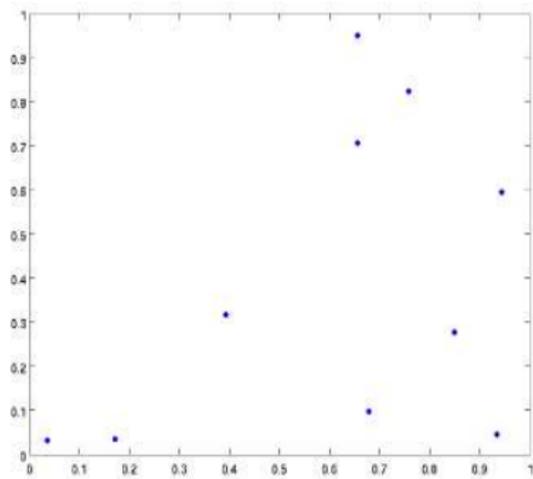


# Marche aléatoire

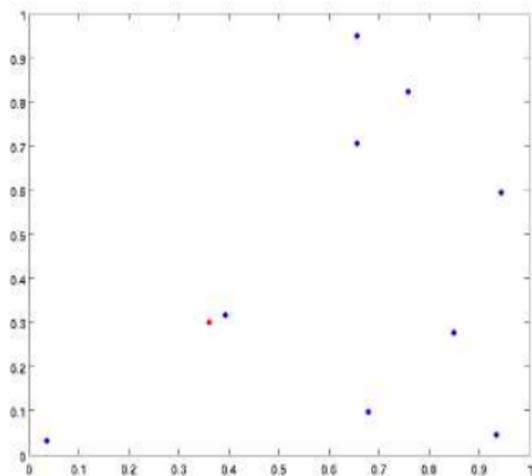
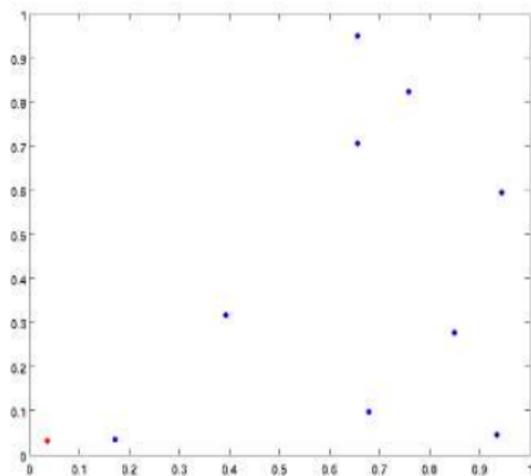


On accepte la transition.

# Choix d'une paire de points



## Marche aléatoire

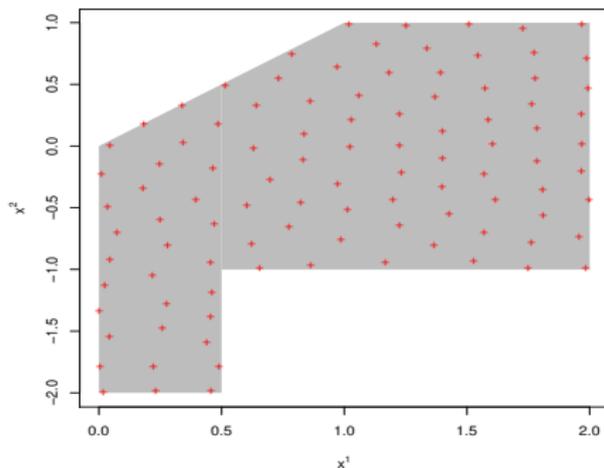
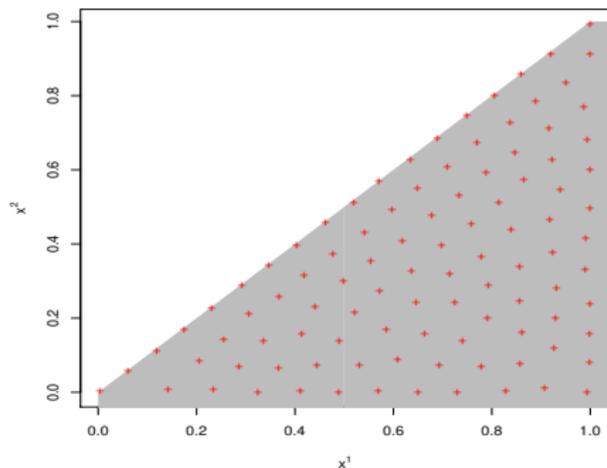


On accepte avec probabilité

$$\exp\left(\frac{\zeta_{\mathbf{x}^{prop}} - \zeta_{\mathbf{x}^{(t)}}}{T_t}\right).$$

# Exemples

Ces plans de  $N = 100$  points sont obtenus pour  $10^6$  itérations de l'algorithme.



# Heuristique pour le choix des paramètres initiaux

- 10000 plans  $(\mathbf{X}_1, \dots, \mathbf{X}_{10000})$  sont tirés de manière uniforme, et on fixe

$$T_0 = \text{mediane}(\zeta_{\mathbf{x}_i})_{i=1 \dots 10000} .$$

- On règle  $\tau_0$ , la variance de la transition, sur un nombre réduit d'itérations de l'algorithme afin d'avoir un taux d'acceptation d'environ 30%.

# Schéma de température et de variance

Pour la température, ce schéma de diminution est choisi :

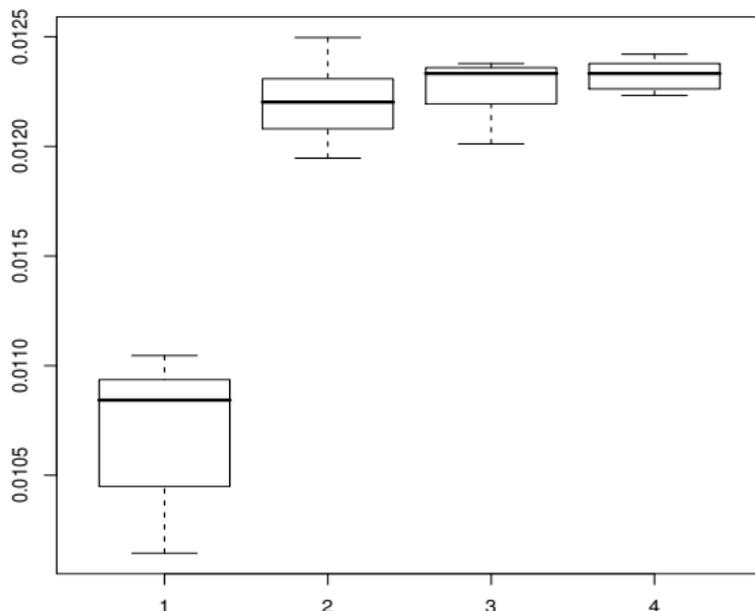
$$T_t = T_0 / \log(t).$$

Pour la variance, ces quatre schémas sont testés :

- 1  $\tau_t = \tau_0 / t$
- 2  $\tau_t = \tau_0 / \sqrt{t}$
- 3  $\tau_t = \tau_0 / \log(t)$
- 4  $\tau_t = \tau_0 / \log(\log(t))$

## Comparaison des schémas de diminution de la variance

Nous testons ces quatre schémas sur 10 appels à l'algorithme avec  $10^6$  itérations pour trouver un plan de 100 points dans  $[0, 1]^2$ .



## Hypothèses de Convergence

En utilisant ce noyau de transition, nous sommes sous les hypothèses faites dans [Bélisle \(1992\)](#) à savoir :

- 1 L'espace d'état  $E^N$  est fermé borné dans  $\mathbb{R}^{d \times N}$ .  $\zeta : \mathbf{X} \mapsto \zeta_{\mathbf{X}}$  est continue. De plus, il existe  $\mathbf{X}^* \in E^N$  tel que  $\zeta$  y atteigne son minimum et tel que pour tout  $\epsilon > 0$ , l'ensemble  $\{\mathbf{X} \in E^N \text{ tel que } \|\mathbf{X} - \mathbf{X}^*\| \leq \epsilon\}$  a une mesure de Lebesgue strictement positive.
- 2 Le noyau de transition  $R$  est absolument continu par rapport à la mesure de Lebesgue et sa densité est minorée strictement supérieure à 0. C'est à dire

$$R(\mathbf{X}, B) = \int_B r(\mathbf{X}, \mathbf{Y}) d\mathbf{Y} \text{ et } \inf_{\mathbf{X}, \mathbf{Y} \in E^N} r(\mathbf{X}, \mathbf{Y}) > 0.$$

- 3 Pour tout ouvert  $G \subset E^N$ ,  $\mathbf{X} \mapsto R(\mathbf{X}, G)$  est continu.
- 4  $(T_t)_{t \geq 0}$  décroît vers 0.

Sous ces hypothèses 1 – 4,  $(\zeta_{\mathbf{X}^{(k)}})_{k \geq 0}$  converge en probabilité vers  $\zeta^*$  qui est le minimum global de la fonction  $\zeta$ .