

# Journées MASCOT NUM

22-23 mars 07



## Introduction aux RKHS

L. Carraro, X. Bay, O. Roustant

Ecole des mines de Saint-Etienne

(nom@emse.fr)

# PLAN



- **Partie 1 (22 mars) :**
  - ◆ Motivation : régularisation
  - ◆ Noyaux de Bergmann
  - ◆ RKHS, définition et propriétés
  - ◆ Exemples
  - ◆ Retour sur la régularisation

# PLAN



- **Partie 2 (23 mars) :**
  - ◆ RKHS et régularisation, exemples.
  - ◆ Le noyau reproduisant et ses interprétations.
  - ◆ Discrédance et RKHS
  - ◆ Propagation d'incertitudes et RKHS
  - ◆ ~~Semi-RKHS~~

# PLAN



- **Partie 2 (23 mars) :**
  - ◆ RKHS et régularisation, exemples.
  - ◆ Le noyau reproduisant et ses interprétations.
  - ◆ Discrédance et RKHS
  - ◆ Propagation d'incertitudes et RKHS
  - ◆ Semi-RKHS

# RKHS et régularisation



- Théorème (Kimerldorf et Wahba 1970 ?)

Soit  $H$  un RKHS de noyau  $K$ . Le minimum du critère  $C$ , défini par :

$$C(f) = \mathbf{L}(\mathbf{y}, f(\mathbf{x})) = \sum_{i=1}^n D(y_i, f(\mathbf{x}_i)) + \lambda \|f\|_H^2$$

est obtenu pour une fonction  $f$  de la forme :

$$f = \sum_{i=1}^n \beta_i K_{\mathbf{x}_i}$$

# Exemples



- Splines linéaires en dim 1 :

- ◆  $D(y, z) = (y - z)^2 \quad \|f\|_H^2 = \int_0^1 f'(x)^2 dx$

- ◆ Solution :  $f(x) = \alpha + \sum_{i=1}^n \beta_i \min(x, x_i) = \alpha + \sum_{i=1}^n \beta_i (x - x_i)_+$

- Splines cubiques en dim 1 :

- ◆  $D(y, z) = (y - z)^2 \quad \|f\|_H^2 = \int_0^1 f''(x)^2 dx$

- ◆ Solution :  $f(x) = \alpha_0 + \alpha_1 x + \sum_{i=1}^n \beta_i (x - x_i)_+^3$

# Exemples



- Splines additives en dim d :

- ◆  $D(y, z) = (y - z)^2$

- ◆ Fonctions de la forme :

$$f = \alpha + \sum_{j=1}^d \phi_j(\mathbf{x}_j)$$

- ◆  $\|f\|_{\mathbf{H}}^2 = \sum_{j=1}^d \int f_{jj}(\mathbf{x}_j)^2 dt_j = \sum_{j=1}^d \int \phi''_j(\mathbf{x}_j)^2 d\mathbf{x}_j$

# Exemples



- Splines plaques minces en dim 2 :

- ◆  $D(y, z) = (y - z)^2$

- ◆  $\|f\|_H^2 = \int \int [f_{11}(x_1, x_2)^2 + 2f_{12}(x_1, x_2)^2 + f_{22}(x_1, x_2)^2] dx_1 dx_2$

- ◆ Solution :  $\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \sum_{i=1}^n \beta_i K(x, x_i)$

$$K(x, y) = G(x-y), \text{ où } G(x) = \|y\|^2 \text{ Log } \|y\|$$

- ◆ Remarque : K n'est pas définie positive !

# Exemples



- Ridge regression :

- ◆  $D(y, z) = (y - z)^2$

- ◆  $f = X(x) \beta$

- $X(x) = (1 \ x_1 \ \dots \ x_d)$  ou autre polynôme

- $X(x) = \left[ 1 \ R\left(\left\|\frac{x - \lambda_1}{\sigma}\right\|\right) \ \dots \ R\left(\left\|\frac{x - \lambda_p}{\sigma}\right\|\right) \right]$  : RBF

- ◆  $\|f\|_H^2 = \sum_{j=1}^p \beta_j^2$

# Exemples



- SVM en régression :
  - ◆  $D(y, z) = V_\epsilon(y - z)$  avec  $V_\epsilon(u) = (|u| - \epsilon) \mathbf{1}_{|u| > \epsilon}$
  - ◆ Critère de régularité = un des critères précédents...
- SVM en classification :
  - ◆ cas de deux classes codées par -1 et 1
  - ◆  $L(y, z) = (1 - yz)_+$
  - ◆ Critère de régularité = idem SVM en régression

# PLAN



- **Partie 2 (23 mars) :**
  - ◆ RKHS et régularisation, exemples.
  - ◆ Le noyau reproduisant et ses interprétations.
  - ◆ Discrédance et RKHS
  - ◆ Propagation d'incertitudes et RKHS
  - ◆ Semi-RKHS

# Les « normes » sur H



- «norme» la plus courante :

$$\|f\|_H^2 = \int_{\mathbb{R}^d} [P(x, D) f](x)^2 dx$$

$P(x, \cdot)$  est un polynôme en  $d$  variables (à coefficients non constants en général)

$D$  est l'opérateur de différentiation

$P(\cdot, D)$  est un opérateur différentiel

- Splines plaques minces :

$$P(x, u) = P(u) = u_1^2 + 2u_1 u_2 + u_2^2$$

$$P(D) = \frac{\partial^2}{\partial x_1^2} + 2 \frac{\partial^2}{\partial x_1 \partial x_2} + \frac{\partial^2}{\partial x_2^2}$$

# Le noyau K comme fonction de Green



Si  $P(D)^*$  est l'adjoint dans  $L^2$  de  $P(D)$  :

$$\langle f, K_x \rangle_H = f(x) = \langle f, P(D)^* P(D) K_x \rangle_{L^2}$$

Par suite,

$$P(D)^* P(D) K_x = \delta_x$$

K est donc la fonction de Green de l'opérateur différentiel  $P(D)^* P(D)$ , c'est-à-dire :

La solution de  $P(D)^* P(D) f = u$  est donnée par

$$f(x) = \langle u, K_x \rangle_{L^2} = \int K(x, y) u(y) dy$$

# Le noyau K comme fonction de Green



## ● Splines en dimension 1

### ◆ Splines linéaires

$$P(D)^* P(D) f = - f''$$

$$K_x(y) = (x-y)_+$$

### ◆ Splines cubiques

$$P(D)^* P(D) f = - f^{(4)}$$

$$K_x(y) = (x-y)_+^3$$

# le noyau K – vision fréquentielle



- Rappel sur Fourier

$$\hat{f}(\xi) = \int_{\mathbb{R}^d} e^{-i\langle t, \xi \rangle} f(t) dt$$

$$\widehat{P(D)f}(\xi) = P(i\xi)\hat{f}(\xi)$$

$$\int_{\mathbb{R}^d} |P(i\xi)|^2 |\hat{f}(\xi)|^2 d\xi = (2\pi)^d \int_{\mathbb{R}^d} |f(x)|^2 dx, \text{ Plancherel}$$

- Norme

$$\|f\|_{\mathbb{H}}^2 = 1/(2\pi)^d \int_{\mathbb{R}^d} |P(i\xi)|^2 |\hat{f}(\xi)|^2 d\xi$$

# le noyau K – vision fréquentielle



- Le noyau K

$$\text{Posons } \hat{R}(\xi) = 1/|P(i\xi)|^2$$

$$\text{Alors } K(x,y) = R(x-y)$$

- Splines en dim 1 :

linéaires :  $P(D) f = f'$  et  $P(i\xi) = i\xi$

$$\hat{R}(\xi) = 1/\xi^2$$

cubiques :  $P(D) f = f''$  et  $P(i\xi) = (i\xi)^2$

$$\hat{R}(\xi) = 1/\xi^4$$

# Le noyau K comme covariance



Soit  $Z = (Z(x))_x$  un PG centré de covariance K

On note H de RKHS associé et  $H_Z$  l'e.v. fermé (dans  $L^2(P)$ ) engendré par  $\{Z(x), x\}$ .

Notons que :

$$E(Z(x) Z(y)) := \langle Z(x), Z(y) \rangle_{L^2(P)} = K(x, y) = \langle K_x, K_y \rangle_H$$

L'application :

$$\begin{aligned} H_Z &\rightarrow H \\ Z_x &\rightarrow K_x \end{aligned}$$

se prolonge en une isométrie de  $H_Z$  sur H.

# RKHS et krigage



## ● Théorème (Duchon, Matheron...)

Soit  $z_i = f(x_i) + \varepsilon_i$ , où  $\varepsilon_i$  v.a.i.i.d.  $N(0, \sigma^2)$

où  $f$  est la réalisation d'un PG stationnaire centré de covariance  $K$ .

Alors, si  $\lambda = \sigma^2 / K(0)$ , la solution  $f_\lambda$  du problème :

$$\underset{f}{\text{Min}} C(f), \text{ où } C(f) = \mathbf{L}(\mathbf{z}, f(\mathbf{x})) = \sum_{i=1}^n (z_i - f(x_i))^2 + \lambda \|f\|_H^2$$

vérifie :  $f_\lambda(x) = E(Z(x) | Z(x_i) = z_i)$

### Remarque :

Cas d'une tendance avec loi a priori impropre

# PLAN



- **Partie 2 (23 mars) :**
  - ◆ RKHS et régularisation, exemples.
  - ◆ Le noyau reproduisant et ses interprétations.
  - ◆ **Discrédance et RKHS**
  - ◆ Propagation d'incertitudes et RKHS
  - ◆ Semi-RKHS

# Rappel sur discr pance



## ● Cas de la dimension 1

On approxime l'int grale

$$\mu = \int_0^1 f(x) dx$$

par une moyenne sur un plan  $P = \{x_1, \dots, x_n\}$  :

$$\hat{\mu} = \frac{1}{n} \sum_{i=0}^n f(x_i)$$

In galit s pr cisant la convergence :

$$|\mu - \hat{\mu}| \leq D(P) \phi(f)$$

exemple : discr pance infinie et  $\Phi(f) =$  variation de  $f$

# Discrépance et erreur de modèle (Hickernell)



On part du modèle :

$$z_i = g(x_i) + \varepsilon_i, \text{ où } \varepsilon_i \text{ v.a.i.i.d. } N(0, \sigma^2)$$

$g$  est dans un RKHS et vérifie :

$$g(x) = \mu + h(x), \text{ avec } E_x(h(x)) = \int_X h(x) \lambda(dx) = 0$$

On estime  $\mu$  sur le plan  $P$  par :

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n z_i$$

$$\text{Biais}(\hat{\mu}) := L_P(g) = \frac{1}{n} \sum_{i=1}^n g(x_i) - \int_X g(x) d\lambda(x)$$

$$L_P(g) = \int_X g(x) (\lambda_P - \lambda)(dx)$$

# discrépance et erreur de modèle



Ecart quadratique :

$$E\left[(\hat{\mu} - \mu)^2\right] = L_P(g)^2 + \frac{\sigma^2}{n}$$

$$L_P(g) = L_P(h) = \langle h, \eta \rangle, \text{ où } \eta(x) = \int_X K_x(y)(\lambda - \lambda_P)(dy)$$

Théorème (Hickernell) :

$$E\left[(\hat{\mu} - \mu)^2\right] \leq D(P) \|h\|_H^2 + \frac{\sigma^2}{n}$$

$$\text{où } D(P) = \int_X \int_X K(x, y)(\lambda_P - \lambda)(dx)(\lambda_P - \lambda)(dy)$$

Remarque :

$$D(P) = \|\lambda_P - \lambda\|_H^2,$$

# Commentaires



- Interprétation probabiliste
- Cas  $X=[0,1]^d$  et  $\lambda$  uniforme :
  - ◆  $K$  = covariance du drap brownien
  - ◆  $D(P)$  = statistique de Cramer Von Mises :

$$D(P) = \int_{[0,1]^d} \left[ F_{\lambda_P}(x) - F_{\lambda}(x) \right]^2 dx$$

- Cas  $d=2$  :

Voir illustration

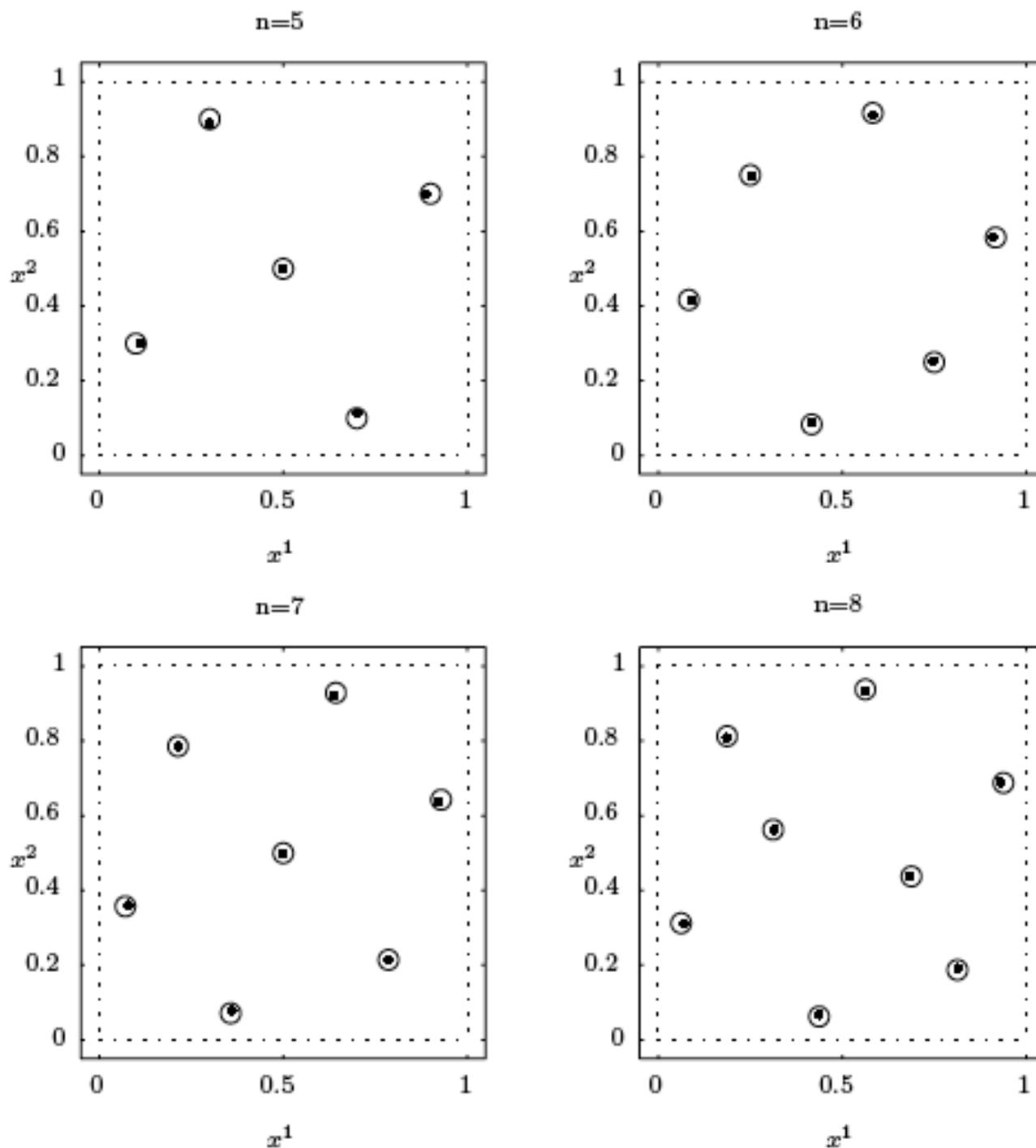


Figure 2. The designs  $\xi_n^b$  ( $\circ$ ) and  $\xi_{n,\text{equ}}^b$  ( $\cdot$ ) for the two-dimensional location model with bias given in Example 1.

# PLAN



- **Partie 2 (23 mars) :**
  - ◆ RKHS et régularisation, exemples.
  - ◆ Le noyau reproduisant et ses interprétations.
  - ◆ Discrédance et RKHS
  - ◆ **Propagation d'incertitudes et RKHS**
  - ◆ Semi-RKHS

# Propagation d'incertitudes (Carraro...)



$x$  est aléatoire, quelle est la loi de  $z(x)$  ?

- **MCV optimalité :**

modèle :  $z(x) = X(x)\beta + \varepsilon(x)$ , où  $\varepsilon(x)$  v.a.i.i.d.  $N(0, \sigma^2)$

Données : plan  $P$   $\{x_i, i=1\dots n\}$  et observations  $\{z_i, i=1\dots n\}$

modèle approché :  $z_{app}(x) = X(x)\hat{\beta} + \eta(x)$ , où  $\eta(x)$  v.a.i.i.d.  $N(0, \hat{\sigma}^2)$

Pour connaître la loi de  $z(x)$ , on utilise la loi de  $z_{app}(x)$

## Propriétés :

$$E(z_{app}(x)) = E(z(x))$$

$$Var(z_{app}(x)) \geq Var(z(x))$$

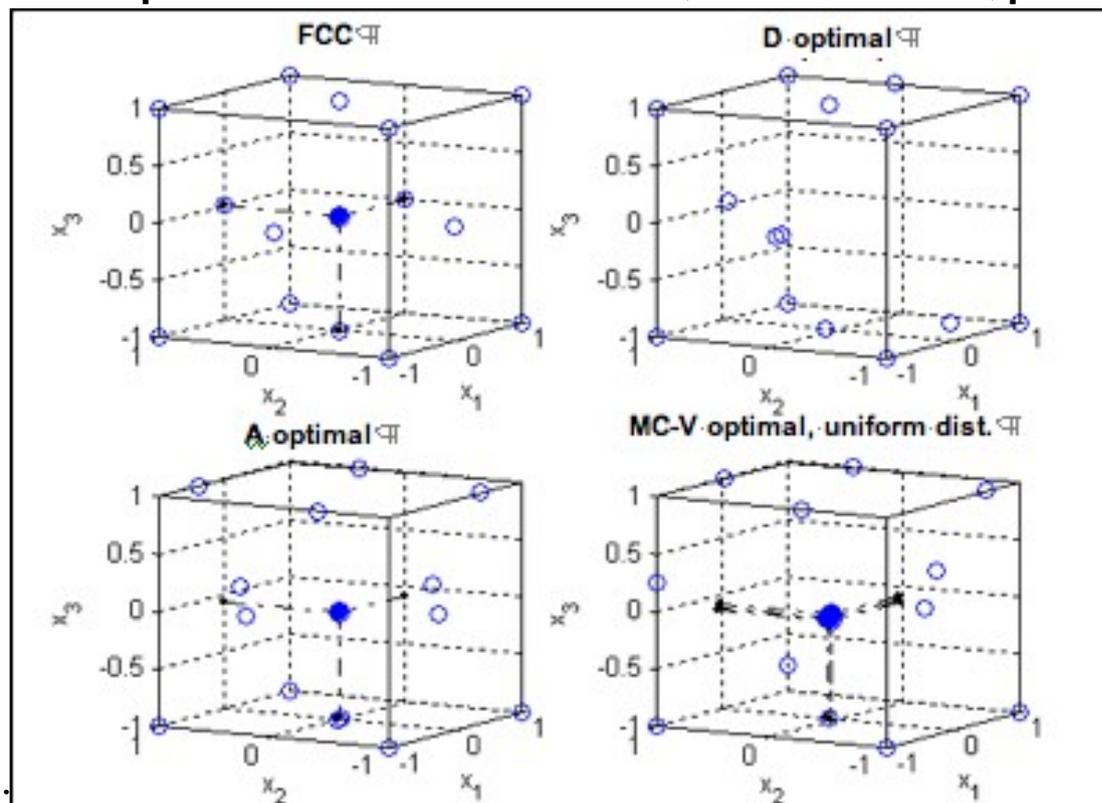
# Propagation d'incertitudes



Critère de choix du plan P :

$$D(P) = \text{Var}(z_{app}(x)) - \text{Var}(z(x))$$

$$D(P) = \sigma^2 \left[ m'(X'X)^{-1}m + \text{Tr}((X'X)^{-1}\Gamma) \right]$$



# Extension avec erreur de modèle



modèle :  $z(x) = X(x)\beta + h(x) + \varepsilon(x)$ , où  $\varepsilon(x)$  v.a.i.i.d.  $N(0, \sigma^2)$

Avec  $h$  appartenant à un RKHS formé de fonctions « orthogonales » aux  $X_j(x)$ .

Critère contenant :

un terme de biais

un terme de variance

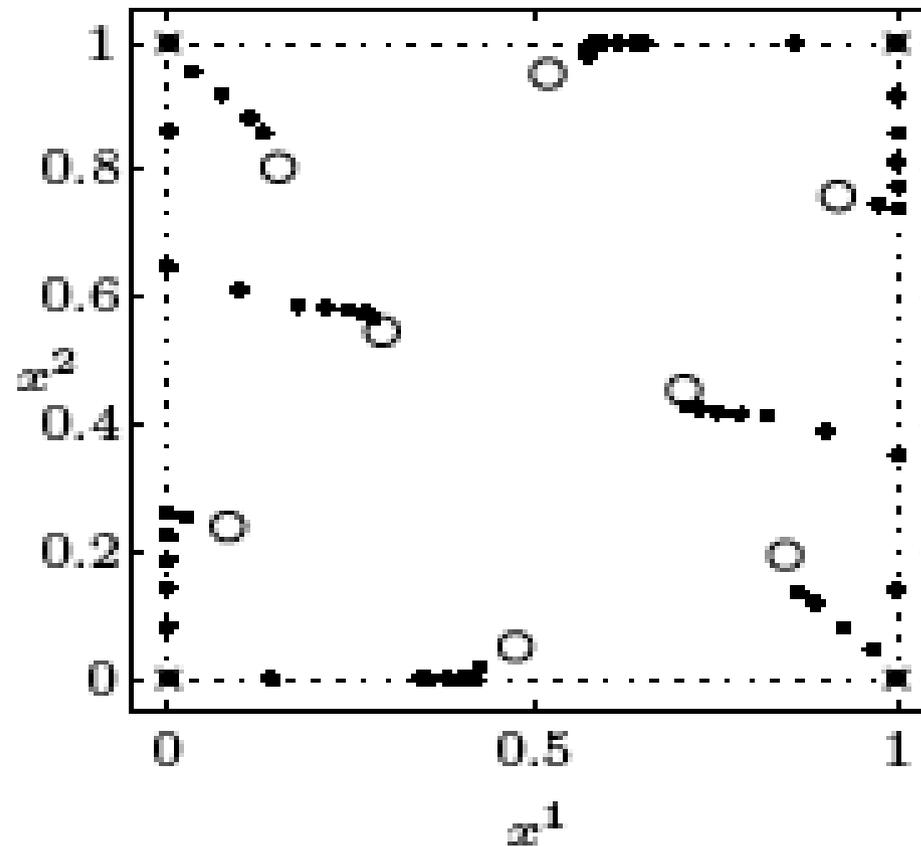
une pondération entre les deux dépendant de :

$$\lambda = \frac{\sigma^2}{\|h\|_H^2 + \sigma^2}$$

# Illustration (Hickernell)



$$X(x) \beta = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$



# Bibliographie



- Aronszajn N. (1950), Theory of reproducing kernels, Trans. AMS, 68, p. 337-404
- Amato U., Antoniadis A., Pensky M. (2006), Wavelet kernel penalized estimation for non-equispaced design regression, Statistics and Computing, 16, 1, 37-56, 2006
- Bergmann S. (1970), The Kernel Function and Conformal Mappings. AMS, Providence.
- Carraro L., Corre B., Helbert C., Roustant O., Josserand S. (2007), Optimal designs for the propagation of uncertainty in computer experiments, to appear in Chemometrics and Intelligent Laboratory Systems.
- Duchon J. (1980), Fonctions splines homogènes à plusieurs variables, Thèse d'Etat, Université de Grenoble.
- Duchon J. (2007), Modèles intrinsèques et probabilités conditionnelles, en préparation.
- Feuillard V. (2007), Analyse d'une base de données pour la calibration d'un code de calcul, Thèse de doctorat, Univ. Paris VI. Soutenance prévue : mai.
- Girosi F., Jones M., Poggò T. (1995), Regularization theory and neural network architectures, Neural Computation, 7, p. 219-269.
- Hickernell F.J. (1999), Goodness-of-fit statistics, discrepancies and robust designs, Stat. & Prob. Letters, 44, p. 73-78.

# Bibliographie

Aronszajn N. (1950), Theory of reproducing kernels, Trans. AMS, 68, p. 337-404

Bergmann S. (1970), The Kernel Function and Conformal Mappings. AMS, Providence.

Carraro L., Corre B., Helbert C., Roustant O., Josserand S. (2007), Optimal designs for the propagation of uncertainty in computer experiments, to appear in Chemometrics and Intelligent Laboratory Systems.

Duchon J. (1980), Fonctions splines homogènes à plusieurs variables, Thèse d'Etat, Université de Grenoble.

Duchon J. (2007), Modèles intrinsèques et probabilités conditionnelles, en préparation.

Feuillard V. (2007), Analyse d'une base de données pour la calibration d'un code de calcul, Thèse de doctorat, Univ. Paris VI. Soutenance prévue : mai.

Girosi F., Jones M., Poggio T. (1995), Regularization theory and neural network architectures, Neural Computation, 7, p. 219-269.

Hickernell F.J. (1999), Goodness-of-fit statistics, discrepancies and robust designs, Stat. & Prob. Letters, 44, p. 73-78.

# Bibliographie



- Lu F., Keles S., Lin Y., Wright G., Wahba G. (2006), Kernel Regularization and Dimension Reduction, Tech. Report n°1119, U. of Wisconsin.
- Schwartz L. (1964), Sous-espaces hilbertiens d'espaces vectoriels topologiques et noyaux associés (noyaux reproduisants), J. Anal. Math., 13, p. 115-256.
- Vazquez E. (2005), Modélisation comportementale de systèmes non linéaires multivariables par méthodes à noyaux et applications, thèse de doctorat de l'U d'Orsay.
- Wahba G. (1990), Spline models for observational data, SIAM.
- Wahba G. (2000), An Introduction to Model Building with Reproducing Kernel Hilbert Spaces, Tech. Report n°1020, U. of Wisconsin.
- Yue R., Hickernell F.J. (1999), Robust designs for fitting linear models with misspecification, Stat. Sinica, 9, p. 1053-1069.