Modélisation statistique pour données fonctionnelles : approches non-asymptotiques et méthodes adaptatives.

Angelina Roche
thèse effectuée sous la direction d’Elodie Brunel et André Mas

7 juillet 2014
**Aim:** study the link between two random variables.

- \( Y \in \mathbb{R} \) a variable of interest.
- \( X \in \mathbb{H} \) an explanatory (functional) variable, with \( (\mathbb{H}, \langle ., . \rangle, \| . \|) \) a separable Hilbert space.

Typically \( \mathbb{H} = L^2([a, b]) \), \( \mathbb{H} \) = a Sobolev space...

**Observations:** \( (X_i, Y_i)_{i \in \{1, \ldots, n\}} \) a sample following the same distribution as \((X, Y)\).
Models and problems considered

- **Functional linear model**: \( Y = \langle \beta, X \rangle + \varepsilon \),
  with \( \beta \in \mathbb{H} \) and \( \varepsilon \) a noise term, centred, independent of \( X \), with finite variance.

- **Model without structural constraint**

- **Nonparametric regression**: \( Y = m(X) + \varepsilon \),
  with \( m : \mathbb{H} \rightarrow \mathbb{R} \) a function and \( \varepsilon \) a noise term.
Models and problems considered

- **Functional linear model**: $Y = \langle \beta, X \rangle + \varepsilon$,
  with $\beta \in \mathbb{H}$ and $\varepsilon$ a noise term, centred, independent of $X$, with finite variance.

  *Estimation of the slope function $\beta$.*
  *Goal: prediction of a new value of $Y$ given a new curve $X$.*

- **Model without structural constraint**

- **Nonparametric regression**: $Y = m(X) + \varepsilon$,
  with $m : \mathbb{H} \to \mathbb{R}$ a function and $\varepsilon$ a noise term.
Models and problems considered

- **Functional linear model**: \( Y = \langle \beta, X \rangle + \varepsilon \),
  with \( \beta \in \mathbb{H} \) and \( \varepsilon \) a noise term, centred, independent of \( X \), with finite variance.

  \[ \text{Estimation of the slope function } \beta. \]
  \[ \text{Goal: prediction of a new value of } Y \text{ given a new curve } X. \]

- **Model without structural constraint**

  \[ \text{Estimation of the conditional cumulative distribution function} \]
  \[ F : \mathbb{H} \times \mathbb{R} \rightarrow \mathbb{R} \]
  \[ (x, y) \mapsto F_x(y) = \mathbb{P}(Y \leq y | X = x). \]

- **Nonparametric regression**: \( Y = m(X) + \varepsilon \),
  with \( m : \mathbb{H} \rightarrow \mathbb{R} \) a function and \( \varepsilon \) a noise term.
Models and problems considered

- **Functional linear model**: \( Y = \langle \beta, X \rangle + \varepsilon \),
  with \( \beta \in \mathbb{H} \) and \( \varepsilon \) a noise term, centred, independent of \( X \), with finite variance.

  *Estimation of the slope function \( \beta \).*

  *Goal: prediction of a new value of \( Y \) given a new curve \( X \).*

- **Model without structural constraint**

  *Estimation of the conditional cumulative distribution function*

  \[ F : \mathbb{H} \times \mathbb{R} \rightarrow \mathbb{R} \]
  \[
  (x, y) \mapsto F^x(y) = \mathbb{P}(Y \leq y | X = x).
  \]

- **Nonparametric regression**: \( Y = m(X) + \varepsilon \),
  with \( m : \mathbb{H} \rightarrow \mathbb{R} \) a function and \( \varepsilon \) a noise term.

  *Minimisation of the conditional expectation :*

  \[
  x^* = \arg \min_{x \in C} \{ m(x) \}.
  \]
Outline

1. Prediction in the functional linear model
   - Estimation procedure
   - Theoretical results
   - Simulation results

2. Adaptive estimation of the conditional c.d.f
   - Bias-variance decomposition of the risk
   - Bandwidth selection device
   - Optimal estimation in the minimax sense
   - Simulation study

3. Response surface methodology for functional data
   - Response surface methodology
   - Extension to the functional setting
Outline

1. Prediction in the functional linear model
   - Estimation procedure
   - Theoretical results
   - Simulation results

2. Adaptive estimation of the conditional c.d.f
   - Bias-variance decomposition of the risk
   - Bandwidth selection device
   - Optimal estimation in the minimax sense
   - Simulation study

3. Response surface methodology for functional data
   - Response surface methodology
   - Extension to the functional setting
We suppose that

$$Y = \langle \beta, X \rangle + \varepsilon,$$  \hspace{1cm} (1)

with

- $X$ a centred random variable with values in a separable Hilbert space $(\mathcal{H}, \langle \cdot, \cdot \rangle, \| \cdot \|)$ with infinite dimension;
- $\beta$, the slope function: an unknown element of $\mathcal{H}$;
- $\varepsilon$ a noise term, centred, independent of $X$ and with unknown variance $\sigma^2$.

**Aim:** estimate the slope function $\beta$ using the information of the sample $\{(X_i, Y_i), i = 1, \ldots, n\}$ following (1).
Covariance operator

Multiplying the model equation $Y = \langle \beta, X \rangle + \varepsilon$ by $X(s)$ and taking expectation we obtain

$$E[XY] = E[\langle \beta, X \rangle X]$$

where

$$g \in \mathbb{H} = \Gamma \beta$$

is the covariance operator associated to $X$.

- $\Gamma$ positive compact self-adjoint
  - $\Rightarrow$ basis $(\psi_j)_{j \geq 1}$ of eigenfunctions
  - $(\lambda_j)_{j \geq 1}$ associated eigenvalues, non-increasing sequence.
- $\lambda_j \downarrow 0 \Rightarrow$ ill-posed inverse problem.
- For identifiability, we suppose that
  $$\text{Ker}(\Gamma) = \{0\} \iff \lambda_j > 0 \text{ for all } j.$$
Covariance operator

Multiplying the model equation $Y = \langle \beta, X \rangle + \varepsilon$ by $X(s)$ and taking expectation we obtain

$$
\mathbb{E}[YX] = \mathbb{E}[\langle \beta, X \rangle X]
$$

where

$$
\Gamma : f \in \mathbb{H} \mapsto \mathbb{E}[\langle X, f \rangle X]
$$

is the covariance operator associated to $X$.

- $\Gamma$ positive compact self-adjoint
  - $\Rightarrow$ basis $(\psi_j)_{j \geq 1}$ of eigenfunctions
    - $(\lambda_j)_{j \geq 1}$ associated eigenvalues, non-increasing sequence.
  - $\lambda_j \searrow 0 \Rightarrow$ ill-posed inverse problem.
- For identifiability, we suppose that
  $$
  \ker(\Gamma) = \{0\} \iff \lambda_j > 0 \text{ for all } j.
  $$
Covariance operator

Multiplying the model equation $Y = \langle \beta, X \rangle + \varepsilon$ by $X(s)$ and taking expectation we obtain

$$
\mathbb{E}[YX] = \mathbb{E}[\langle \beta, X \rangle X]
$$

\[ g \in H = \Gamma \beta \]

where

\[ \Gamma : f \in H \mapsto \mathbb{E}[\langle X, f \rangle X] \]

is the covariance operator associated to $X$.

- $\Gamma$ positive compact self-adjoint

  \[ \Rightarrow \text{basis } (\psi_j)_{j \geq 1} \text{ of eigenfunctions} \]

  \[ (\lambda_j)_{j \geq 1} \text{ associated eigenvalues, non-increasing sequence.} \]

- $\lambda_j \downarrow 0 \Rightarrow \text{ill-posed inverse problem.}$

- For identifiability, we suppose that

  \[ \text{Ker}(\Gamma) = \{0\} \Leftrightarrow \lambda_j > 0 \text{ for all } j. \]
Covariance operator

Multiplying the model equation $Y = \langle \beta, X \rangle + \varepsilon$ by $X(s)$ and taking expectation we obtain

$$E[YX] = E[\langle \beta, X \rangle X]$$

where $g \in \mathbb{H} = \Gamma \beta$

where

$$\Gamma : f \in \mathbb{H} \mapsto E[\langle X, f \rangle X]$$

is the covariance operator associated to $X$.

- $\Gamma$ positive compact self-adjoint
  - $\Rightarrow$ basis $(\psi_j)_{j \geq 1}$ of eigenfunctions
    - $(\lambda_j)_{j \geq 1}$ associated eigenvalues, non-increasing sequence.
  - $\lambda_j \downarrow 0 \Rightarrow$ ill-posed inverse problem.
- For identifiability, we suppose that

$$\text{Ker}(\Gamma) = \{0\} \iff \lambda_j > 0 \text{ for all } j.$$
Risk considered

Definition

The prediction error of an estimator $\hat{\beta}$ is the quantity

$$\mathbb{E} \left[ \left( \hat{Y}_{n+1} - \mathbb{E} \left[ Y_{n+1} \mid X_{n+1} \right] \right)^2 \mid (X_1, Y_1), \ldots, (X_n, Y_n) \right]$$

$$= \mathbb{E} \left[ \langle \hat{\beta} - \beta, X_{n+1} \rangle^2 \mid (X_1, Y_1), \ldots, (X_n, Y_n) \right]$$

$$= \langle \Gamma (\hat{\beta} - \beta), \hat{\beta} - \beta \rangle =: \| \hat{\beta} - \beta \|^2_{\Gamma}$$

with

- $(X_{n+1}, Y_{n+1})$ a copy of $(X, Y)$ independent of the sample;
- $\hat{Y}_{n+1}$ the prediction of $Y_{n+1}$ with the estimator $\hat{\beta}$:

$$\hat{Y}_{n+1} = \langle \hat{\beta}, X_{n+1} \rangle.$$
Risk considered

Definition

The prediction error of an estimator \( \hat{\beta} \) is the quantity

\[
\mathbb{E} \left[ \left( \hat{Y}_{n+1} - \mathbb{E} [Y_{n+1} | X_{n+1}] \right)^2 \right] | (X_1, Y_1), \ldots, (X_n, Y_n)
\]

\[
= \mathbb{E} \left[ \langle \hat{\beta} - \beta, X_{n+1} \rangle^2 \right] | (X_1, Y_1), \ldots, (X_n, Y_n)
\]

\[
= \langle \Gamma(\hat{\beta} - \beta), \hat{\beta} - \beta \rangle =: \| \hat{\beta} - \beta \|^2.
\]

with

- \((X_{n+1}, Y_{n+1})\) a copy of \((X, Y)\) independent of the sample;
- \(\hat{Y}_{n+1}\) the prediction of \(Y_{n+1}\) with the estimator \(\hat{\beta}\):

\[
\hat{Y}_{n+1} = \langle \hat{\beta}, X_{n+1} \rangle.
\]
Risk considered

Definition

The prediction error of an estimator $\hat{\beta}$ is the quantity

$$\mathbb{E} \left[ \left( \hat{Y}_{n+1} - \mathbb{E} [Y_{n+1} | X_{n+1}] \right)^2 | (X_1, Y_1), \ldots, (X_n, Y_n) \right]$$

$$= \mathbb{E} \left[ \langle \hat{\beta} - \beta, X_{n+1} \rangle^2 | (X_1, Y_1), \ldots, (X_n, Y_n) \right]$$

$$= \langle \Gamma (\hat{\beta} - \beta), \hat{\beta} - \beta \rangle =: \|\hat{\beta} - \beta\|_{\Gamma}^2$$

with

- $(X_{n+1}, Y_{n+1})$ a copy of $(X, Y)$ independent of the sample;
- $\hat{Y}_{n+1}$ the prediction of $Y_{n+1}$ with the estimator $\hat{\beta}$:

$$\hat{Y}_{n+1} = \langle \hat{\beta}, X_{n+1} \rangle.$$
Risk considered

Definition

The prediction error of an estimator $\hat{\beta}$ is the quantity

$$
\mathbb{E} \left[ \left( \hat{Y}_{n+1} - \mathbb{E} [Y_{n+1} | X_{n+1}] \right)^2 \right] | (X_1, Y_1), \ldots, (X_n, Y_n)
$$

$$
= \mathbb{E} \left[ (\hat{\beta} - \beta, X_{n+1})^2 | (X_1, Y_1), \ldots, (X_n, Y_n) \right]
$$

$$
= \langle \Gamma(\hat{\beta} - \beta), \hat{\beta} - \beta \rangle =: \|\hat{\beta} - \beta\|_{\Gamma}^2
$$

with

- $(X_{n+1}, Y_{n+1})$ a copy of $(X, Y)$ independent of the sample;
- $\hat{Y}_{n+1}$ the prediction of $Y_{n+1}$ with the estimator $\hat{\beta}$:

$$
\hat{Y}_{n+1} = \langle \hat{\beta}, X_{n+1} \rangle.
$$
Short overview of existing work

- **Estimation by projection or by roughness regularization.**
  - On fixed basis: Fourier, $B$-splines, general o.n.b...
  - On data-driven basis: functional PCA.

- **Numerous results with asymptotic point of view:** Cardot, Ferraty and Sarda (1999), Cai and Hall (2006), Hall and Horowitz (2007),...
  - ... but very few non-asymptotic results: Cardot and Johannes (2010, lower bounds on general $L^2$-risks), Comte and Johannes (2010, 2012; adaptive estimators).

- **Comte and Johannes (2010, 2012):**
  - → projection estimators on fixed basis;
  - → oracle-type inequalities for general weighted $L^2$ norms without including the prediction error;
  - → minimax convergence rates.

**Goal:** define an adaptive estimator by projection on the PCA basis.
Short overview of existing work

- **Estimation by projection or by roughness regularization.**
  - On fixed basis: Fourier, B-splines, general o.n.b...
  - On data-driven basis: functional PCA.

- **Numerous results with asymptotic point of view:** Cardot, Ferraty and Sarda (1999), Cai and Hall (2006), Hall and Horowitz (2007), ...
  - ... but very few non-asymptotic results: Cardot and Johannes (2010, lower bounds on general $L^2$-risks), Comte and Johannes (2010, 2012; adaptive estimators).

- **Comte and Johannes (2010, 2012):**
  - projection estimators on fixed basis;
  - oracle-type inequalities for general weighted $L^2$ norms without including the prediction error;
  - minimax convergence rates.

**Goal:** define an adaptive estimator by projection on the PCA basis.
Short overview of existing work

- **Estimation by projection or by roughness regularization.**
  
  On fixed basis: Fourier, B-splines, general o.n.b...
  On data-driven basis: functional PCA.

- **Numerous results with asymptotic point of view:** Cardot, Ferraty and Sarda (1999), Cai and Hall (2006), Hall and Horowitz (2007),...
  
  ... but very few non-asymptotic results: Cardot and Johannes (2010, lower bounds on general $\mathbb{L}^2$-risks), Comte and Johannes (2010, 2012; adaptive estimators).

- **Comte and Johannes (2010, 2012):**
  
  → projection estimators on fixed basis;
  
  → oracle-type inequalities for general weighted $\mathbb{L}^2$ norms without including the prediction error;
  
  → minimax convergence rates.

**Goal:** define an adaptive estimator by projection on the PCA basis.
Short overview of existing work

- **Estimation by projection or by roughness regularization.**
  - On fixed basis: Fourier, B-splines, general o.n.b...
  - On data-driven basis: functional PCA.

- **Numerous results with asymptotic point of view:** Cardot, Ferraty and Sarda (1999), Cai and Hall (2006), Hall and Horowitz (2007),...
  - ... but very few non-asymptotic results: Cardot and Johannes (2010, lower bounds on general $L^2$-risks), Comte and Johannes (2010, 2012; adaptive estimators).

- **Comte and Johannes (2010, 2012):**
  - projection estimators on fixed basis;
  - oracle-type inequalities for general weighted $L^2$ norms without including the prediction error;
  - minimax convergence rates.

**Goal:** define an adaptive estimator by projection on the PCA basis.
Outline

1. Prediction in the functional linear model
   - Estimation procedure
   - Theoretical results
   - Simulation results

2. Adaptive estimation of the conditional c.d.f
   - Bias-variance decomposition of the risk
   - Bandwidth selection device
   - Optimal estimation in the minimax sense
   - Simulation study

3. Response surface methodology for functional data
   - Response surface methodology
   - Extension to the functional setting
**fPCA**

functional Principal Components Regression

**Aim:**

Define an approximation space $S_m$ of dimension $D_m$ minimising the mean distance between $X$ and its projection on $S_m$.

$$S_m = \text{Vect}\{\psi_1, \ldots, \psi_{D_m}\}$$

By induction:

$$\psi_{k+1} \in \arg \min_{f \in \mathcal{H}} \mathbb{E} \left[ \|X - \Pi_k X - \langle X, f \rangle f \|^2 \right],$$

under the constraint $\langle \psi_{k+1}, \psi_j \rangle = 0$, for all $j \leq k$ et $\|\psi_{k+1}\| = 1$ ($\Pi_k$: projectorVect{$\psi_1, \ldots, \psi_k$}).

The family $(\psi_j)_{j \geq 1}$ is a o.n.b of $\mathcal{H}$ of eigenfunctions of the covariance operator

$$\Gamma : f \in \mathcal{H} \mapsto \mathbb{E} \left[ \langle X, f \rangle X \right].$$
fPCA
functional Principal Components Regression

Aim:
Define an approximation space $S_m$ of dimension $D_m$ minimising the mean distance between $X$ and its projection on $S_m$.

$$S_m = \text{Vect}\{\psi_1, \ldots, \psi_{D_m}\}$$

By induction:
$$\psi_{k+1} \in \arg \min_{f \in \mathcal{H}} \mathbb{E} \left[ \|X - \Pi_k X - \langle X, f \rangle f \|^2 \right],$$
under the constraint $\langle \psi_{k+1}, \psi_j \rangle = 0$, for all $j \leq k$ et $\|\psi_{k+1}\| = 1$ ($\Pi_k$: projector $\text{Vect}\{\psi_1, \ldots, \psi_k\}$).

The family $(\psi_j)_{j \geq 1}$ is a o.n.b of $\mathcal{H}$ of eigenfunctions of the covariance operator

$$\Gamma : f \in \mathcal{H} \mapsto \mathbb{E} [\langle X, f \rangle X].$$
Least-squares estimators

**Case 1: the basis \((\psi_j)_{j \geq 1}\) is known**

\[
\hat{\beta}_m^{(KB)} = \arg \min_{f \in S_m} \gamma_n(f),
\]

with \(S_m = \text{span}\{\psi_1, \ldots, \psi_{D_m}\}\),

where \((\psi_j)_{j \geq 1}\) are the eigenfunctions of the covariance operator

\[
\Gamma : f \in \mathbb{H} \mapsto \mathbb{E} \left[ \langle f, X \rangle X \right].
\]

**Case 2: the basis \((\psi_j)_{j \geq 1}\) is unknown**

\[
\hat{\beta}_m^{(FPCR)} = \arg \min_{f \in \hat{S}_m} \gamma_n(f),
\]

with \(\hat{S}_m = \text{span}\{\hat{\psi}_1, \ldots, \hat{\psi}_{D_m}\}\),

where \((\hat{\psi}_j)_{j \geq 1}\) are the eigenfunctions of the empirical covariance operator

\[
\Gamma_n : f \in \mathbb{H} \mapsto \frac{1}{n} \sum_{i=1}^{n} \langle f, X_i \rangle X_i.
\]

- \(\gamma_n : f \mapsto \frac{1}{n} \sum_{i=1}^{n} (Y_i - \langle f, X_i \rangle)^2\) is the least-squares contrast.
- \((D_m)_{m \geq 1}\) is a strictly increasing sequence such that \(D_1 \geq 1\) (e.g. \(D_m = m\) or \(D_m = 2m + 1\)).
Least-squares estimators

Case 1: the basis $(ψ_j)_{j\geq 1}$ is known

\[ \hat{β}^{(KB)}_m = \arg \min_{f \in S_m} \gamma_n(f), \]

with $S_m = \text{span}\{ψ_1, \ldots, ψ_{D_m}\}$,

where $(ψ_j)_{j\geq 1}$ are the eigenfunctions of the covariance operator

\[ Γ : f \in \mathbb{H} \mapsto \mathbb{E}[\langle f, X \rangle X]. \]

- $\gamma_n : f \mapsto \frac{1}{n} \sum_{i=1}^{n} (Y_i - \langle f, X_i \rangle)^2$ is the least-squares contrast.
- $(D_m)_{m\geq 1}$ is a strictly increasing sequence such that $D_1 \geq 1$ (e.g. $D_m = m$ or $D_m = 2m + 1$).

Case 2: the basis $(ψ_j)_{j\geq 1}$ is unknown

\[ \hat{β}^{(FPCR)}_m = \arg \min_{f \in \hat{S}_m} \gamma_n(f), \]

with $\hat{S}_m = \text{span}\{\hat{ψ}_1, \ldots, \hat{ψ}_{D_m}\}$,

where $(\hat{ψ}_j)_{j\geq 1}$ are the eigenfunctions of the empirical covariance operator

\[ Γ_n : f \in \mathbb{H} \mapsto \frac{1}{n} \sum_{i=1}^{n} \langle f, X_i \rangle X_i. \]
Dimension selection (I)

Problem:
How to choose the dimension $D_m$?

Best dimension for prediction error:

$D_{m^*}$ with

$$m^* \in \arg \min_{m=1, \ldots, N_n} \mathbb{E} \left[ \left\| \hat{\beta}_m^{(FPCR)} - \beta \right\|_\Gamma^2 \right]$$

$\rightarrow$ unknown in practice !!!

$\hat{\beta}_{m^*}^{(FPCR)}$ is the best estimator it is possible to select in the family

$\{ \hat{\beta}_m, m = 1, \ldots, N_n \}$. We call it oracle.
Dimension selection (II)

Bias-variance decomposition of the risk

\[
E \left[ \left\| \hat{\beta}_m^{(FPCR)} - \beta \right\|_\Gamma^2 \right] = E \left[ \left\| \hat{\Pi}_m \beta - \beta \right\|_\Gamma^2 \right] + E \left[ \left\| \hat{\beta}_m^{(FPCR)} - \hat{\Pi}_m \beta \right\|_\Gamma^2 \right],
\]

where \( \hat{\Pi}_m \beta \) is the orthogonal projection on \( \text{span}\{\hat{\psi}_1, \ldots, \hat{\psi}_{D_m}\} \).

Approximation error \( \sim \) bias term:
- decreases with the dimension \( D_m \);
- order unknown in practice (depends on the regularity of \( \beta \)).

Estimation error \( \sim \) variance term: \( \sim \sigma^2 \frac{D_m}{n} \quad \sigma^2: \) noise variance
Bias-variance decomposition of the risk

\[
\mathbb{E} \left[ \left\| \hat{\beta}_m^{(FPCR)} - \beta \right\|^2_\Gamma \right] = \mathbb{E} \left[ \left\| \hat{\Pi}_m \beta - \beta \right\|^2_\Gamma \right] + \mathbb{E} \left[ \left\| \hat{\beta}_m^{(FPCR)} - \hat{\Pi}_m \beta \right\|^2_\Gamma \right],
\]

where \( \hat{\Pi}_m \beta \) is the orthogonal projection on \( \text{span}\{\hat{\psi}_1, \ldots, \hat{\psi}_{D_m}\} \).

Approximation error \( \rightsquigarrow \) bias term:

- decreases with the dimension \( D_m \);
- order unknown in practice (depends on the regularity of \( \beta \)).

Estimation error \( \rightsquigarrow \) variance term:

\[ \sim \sigma^2 \frac{D_m}{n} \]  
\( \sigma^2 \): noise variance
Dimension selection (II)

Bias-variance decomposition of the risk

\[
\mathbb{E} \left[ \| \hat{\beta}_m^{(FPCR)} - \beta \|^2_\Gamma \right] = \mathbb{E} \left[ \| \hat{\Pi}_m \beta - \beta \|^2_\Gamma \right] + \mathbb{E} \left[ \| \hat{\beta}_m^{(FPCR)} - \hat{\Pi}_m \beta \|^2_\Gamma \right],
\]

where \( \hat{\Pi}_m \beta \) is the orthogonal projection on \( \text{span}\{\hat{\psi}_1, \ldots, \hat{\psi}_{Dm}\} \).

Approximation error \( \rightsquigarrow \) bias term:

- decreases with the dimension \( D_m \);
- order unknown in practice (depends on the regularity of \( \beta \)).

Estimation error \( \rightsquigarrow \) variance term:

\[ \sim \sigma^2 \frac{D_m}{n} \quad \sigma^2: \text{noise variance} \]
Dimension selection (III)

Dimension selection criterion

We select

\[ \hat{m} \in \arg \min_{m=1,\ldots,N} \left\{ \gamma_n(\hat{\beta}_m^{(FPCR)}) + \kappa \hat{\sigma}_m^2 \frac{D_m}{n} \right\} \]

with

\[ \hat{\sigma}_m^2 := \frac{1}{n} \sum_{i=1}^{n} \left( Y_i - \langle \hat{\beta}_m^{(FPCR)} , X_i \rangle \right)^2 = \gamma_n(\hat{\beta}_m^{(FPCR)}) \]

an estimator of the noise variance \( \sigma^2 \).
Outline

1. Prediction in the functional linear model
   - Estimation procedure
   - Theoretical results
   - Simulation results

2. Adaptive estimation of the conditional c.d.f
   - Bias-variance decomposition of the risk
   - Bandwidth selection device
   - Optimal estimation in the minimax sense
   - Simulation study

3. Response surface methodology for functional data
   - Response surface methodology
   - Extension to the functional setting
Assumptions

- **Assumption on the noise:** there exists $p > 4$, such that $\mathbb{E}[\varepsilon^p] < +\infty$.

- **Assumption on the target function $\beta$:** there exists $r, R > 0$ such that

$$
\beta \in \mathcal{W}^R_r := \left\{ f \in \mathcal{H}, \sum_{j \geq 1} j^r < f, \psi_j >^2 \leq R^2 \right\}
$$

- **Assumptions on the process $X$:**
  - on the principal components scores:
    - $\sup_{j \geq 1} \mathbb{E} \left[ \frac{\langle X, \psi_j \rangle^{2\ell}}{\lambda_j^\ell} \right] \leq \ell! b^{\ell-1}$, for all $\ell \geq 1$  
    - For all $j \neq k$, $\langle X, \psi_j \rangle$ is independent of $\langle X, \psi_k \rangle$.
  
- on the eigenvalues of $\Gamma$:
  - $\lambda_1 > \lambda_2 > \ldots$
  - $cj^{-a} \leq \lambda_j \leq Cj^{-a}$ with $a > 1, c, C > 0$ (polynomial decrease) or $ce^{-ja} \leq \lambda_j \leq Ce^{-ja}, a, c, C > 0$ (exponential decrease).
  - There exists a constant $\gamma > 0$ such that $(j\lambda_j \max \{\ln^{1+\gamma(j)}, 1\})_{j \geq 1}$ is decreasing.

\[ \rightarrow \text{Brownian motion: } \lambda_j = \pi^{-2}(j - 0.5)^{-2}, \text{ Brownian bridge: } \lambda_j = \pi^{-2}j^{-2} \]
Assumptions

- **Assumption on the noise:** there exists $p > 4$, such that $\mathbb{E}[\varepsilon^p] < +\infty$.

- **Assumption on the target function $\beta$:** there exists $r, R > 0$ such that

$$\beta \in \mathcal{W}_r^R := \left\{ f \in \mathcal{H}, \sum_{j \geq 1} j^r < f, \psi_j >^2 \leq R^2 \right\}$$

- **Assumptions on the process $X$:**

  - on the principal components scores:
    - $\sup_{j \geq 1} \mathbb{E} \left[ \frac{\langle X, \psi_j \rangle^{2\ell}}{\lambda_j^{\ell}} \right] \leq \ell!b^{\ell-1}$, for all $\ell \geq 1$ → Verified for all Gaussian processes
    - For all $j \neq k$, $\langle X, \psi_j \rangle$ is independent of $\langle X, \psi_k \rangle$.

  - on the eigenvalues of $\Gamma$:
    - $\lambda_1 > \lambda_2 > \ldots$
    - $cj^{-a} \leq \lambda_j \leq Cj^{-a}$ with $a > 1, c, C > 0$ (polynomial decrease) or $ce^{-ja} \leq \lambda_j \leq Ce^{-ja}, a, c, C > 0$ (exponential decrease).
    - There exists a constant $\gamma > 0$ such that $(j\lambda_j \max\{\ln^{1+\gamma}(j), 1\})_{j \geq 1}$ is decreasing.

→ Brownian motion: $\lambda_j = \pi^{-2}(j - 0.5)^{-2}$, Brownian bridge: $\lambda_j = \pi^{-2}j^{-2}$
Assumptions

- **Assumption on the noise:** there exists $p > 4$, such that $\mathbb{E}[\varepsilon^p] < +\infty$.
- **Assumption on the target function $\beta$:** there exists $r, R > 0$ such that

$$
\beta \in \mathcal{W}^R_r := \left\{ f \in \mathbb{H}, \sum_{j \geq 1} j^r \leq f, \psi_j \right\}^2 \leq R^2
$$

- **Assumptions on the process $X$:**
  - on the principal components scores:
    - $\sup_{j \geq 1} \mathbb{E} \left[ \frac{\langle X, \psi_j \rangle^{2\ell}}{\lambda_j^{\ell}} \right] \leq \ell! b^{\ell-1}$, for all $\ell \geq 1$ → Verified for all Gaussian processes
    - For all $j \neq k$, $\langle X, \psi_j \rangle$ is independent of $\langle X, \psi_k \rangle$.
  - on the eigenvalues of $\Gamma$:
    - $\lambda_1 > \lambda_2 > \ldots$
    - $c j^{-a} \leq \lambda_j \leq C j^{-a}$ with $a > 1$, $c, C > 0$ (polynomial decrease) or $c e^{-j^a} \leq \lambda_j \leq C e^{-j^a}$, $a, c, C > 0$ (exponential decrease).
    - There exists a constant $\gamma > 0$ such that $(j \lambda_j \max\{\ln^{1+\gamma(j)}, 1\})_{j \geq 1}$ is decreasing.

→ Brownian motion: $\lambda_j = \pi^2 (j - 0.5)^{-2}$, Brownian bridge: $\lambda_j = \pi^{-2} j^{-2}$
Theorem
Under the previous assumptions and if $a + r/2 > 2$ (for the polynomial decrease),

$$
\mathbb{E} \left[ \left\| \hat{\beta}_m^{(FPCR)} - \beta \right\|^2_\Gamma \right] \leq C_1 \min_{m=1,...,N_n} \left\{ \mathbb{E} \left[ \left\| \hat{\Pi}_m \beta - \beta \right\|^2_\Gamma \right] + \kappa \sigma^2 \frac{D_m}{n} \right\} + \frac{C_2}{n},
$$

where $C_1, C_2 > 0$ are independent of $n$ and $\beta$ and $\hat{\Pi}_m$ is the orthogonal projector onto $\hat{S}_m$.

Rates of convergence

<table>
<thead>
<tr>
<th></th>
<th>Polynomial decrease</th>
<th>Exponential decrease</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sup_{\beta \in \mathcal{W}_r} \mathbb{E} \left[ \left| \hat{\beta}<em>m^{(FPCR)} - \beta \right|^2</em>\Gamma \right]$</td>
<td>$c j^{-a} \leq \lambda_j \leq C j^{-a}$</td>
<td>$ce^{-j^a} \leq \lambda_j \leq Ce^{-j^a}$</td>
</tr>
<tr>
<td></td>
<td>$\leq Cn^{-(a+r)/(a+r+1)}$</td>
<td>$\leq Cn^{-1}(\ln(n))^{1/a}$</td>
</tr>
</tbody>
</table>

→ coincides with the lower-bounds established by Cardot and Johannes (2010).

→ The estimator is optimal in the minimax sense
Oracle inequality and rates

Theorem

Under the previous assumptions and if $a + r/2 > 2$ (for the polynomial decrease),

$$
\mathbb{E} \left[ \left\| \hat{\beta}^{(FPCR)}_m - \beta \right\|_\Gamma^2 \right] \leq C_1 \min_{m=1, \ldots, N_n} \left\{ \mathbb{E} \left[ \left\| \hat{\Pi}_m \beta - \beta \right\|_\Gamma^2 \right] + \kappa \sigma^2 \frac{D_m}{n} \right\} + \frac{C_2}{n},
$$

where $C_1, C_2 > 0$ are independent of $n$ and $\beta$ and $\hat{\Pi}_m$ is the orthogonal projector onto $\hat{S}_m$.

Rates of convergence

<table>
<thead>
<tr>
<th>Polynomial decrease</th>
<th>Exponential decrease</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c j^{-a} \leq \lambda_j \leq C j^{-a}$</td>
<td>$c e^{-j^a} \leq \lambda_j \leq C e^{-j^a}$</td>
</tr>
</tbody>
</table>

$$
\sup_{\beta \in \mathcal{N}_r} \mathbb{E} \left[ \left\| \hat{\beta}^{(FPCR)}_m - \beta \right\|_\Gamma^2 \right] \leq C n^{- (a+r)/(a+r+1)} \leq C n^{-1} (\ln(n))^{1/a}
$$

→ coincides with the lower-bounds established by Cardot and Johannes (2010).

→ The estimator is optimal in the minimax sense.
Outline

1 Prediction in the functional linear model
   - Estimation procedure
   - Theoretical results
   - Simulation results

2 Adaptive estimation of the conditional c.d.f
   - Bias-variance decomposition of the risk
   - Bandwidth selection device
   - Optimal estimation in the minimax sense
   - Simulation study

3 Response surface methodology for functional data
   - Response surface methodology
   - Extension to the functional setting
Simulation of $X$

$$X = \sum_{j=1}^{100} \sqrt{\lambda_j} \xi_j \psi_j,$$

with $\xi_1, \ldots, \xi_{100}$ independent realizations of $\mathcal{N}(0, 1)$ and $\psi_j(x) = \sqrt{2} \sin(\pi(j - 0.5)x)$.

$$\lambda_j = j^{-2}$$  

$$\lambda_j = j^{-3}$$  

$$\lambda_j = e^{-j}$$

**Figure:** Sample of 5 random curves
\[ \beta_1(t) = \exp\left(-\left(t - 0.3\right)^2 / 0.05\right) \cos(4\pi t), \quad n = 1000 \]

\[ \lambda_j = j^{-2} \]

\[ \lambda_j = j^{-3} \]

\[ \lambda_j = \exp(-j) \]

\[ \beta_2(t) = \ln(15t^2 + 10) + \cos(4\pi t), \quad n = 1000 \]

\[ \lambda_j = j^{-2} \]

\[ \lambda_j = j^{-3} \]

\[ \lambda_j = \exp(-j) \]
Comparison with cross-validation

We compare our selection criterion with other methods:

- **Cross validation:**

\[ \hat{m}^{CV} := \arg \min_{m=1,...,N} \frac{1}{n} \sum_{i=1}^{n} \left( Y_i - \hat{Y}_i^{(-i)} \right)^2, \]

where \( \hat{Y}_i^{(-i)} \) is the prediction of \( Y \) made from the sample \( \{(X_j, Y_j), j \neq i\} \).

- **Generalized cross-validation:**

\[ \hat{m}^{GCV} := \arg \min_{m=1,...,N} \frac{\gamma_n(\hat{\beta}_m)}{ \left( 1 - \frac{\text{tr}(H_m)}{n} \right)^2}, \]

where \( \hat{Y}_i^{(m)} := \langle \hat{\beta}_m, X_i \rangle \) (prediction of \( Y \)) and \( H_m \) is the Hat matrix defined by \( \hat{Y}^{(m)} = H_m Y \).
Comparison with cross-validation

Estimation of $\beta_1$

Estimation of $\beta_2$

Figure: Left: comparison of estimators $\hat{\beta}_m$ when $m$ is selected by minimization of the penalized criterion or the CV criterion. Right: comparison with the GCV criterion. $n = 2000$, $\lambda_j = j^{-3}$.  

23 / 54
Comparison with cross-validation

Comparison of risks

Figure: Boxplot of prediction errors calculated from 500 independent samples. Estimation of $\beta_1, \lambda_j = j^{-3}$. 
Comparison with cross-validation

Ratio to the oracle

Figure: Ratio $\frac{\|\hat{\beta}_m - \beta\|_1^2}{\|\hat{\beta}_m^* - \beta\|_1^2}$ where $\|\hat{\beta}_m^* - \beta\|_1^2 = \min_{1, \ldots, N_n} \left\{ \|\hat{\beta}_m - \beta\|_1^2 \right\}$. Estimation of $\beta_1$, $\lambda_j = j^{-3}$. 
Comparison with cross-validation

Ratio to the oracle

Figure: Ratio $\frac{\|\hat{\beta}_m - \beta\|_1^2}{\|\hat{\beta}_m^* - \beta\|_1^2}$ where $\|\hat{\beta}_m^* - \beta\|_1^2 = \min_{1, \ldots, N_n} \left\{ \|\hat{\beta}_m - \beta\|_1^2 \right\}$. Estimation of $\beta_1$, $\lambda_j = j^{-3}$.
Outline

1 Prediction in the functional linear model
   - Estimation procedure
   - Theoretical results
   - Simulation results

2 Adaptive estimation of the conditional c.d.f
   - Bias-variance decomposition of the risk
   - Bandwidth selection device
   - Optimal estimation in the minimax sense
   - Simulation study

3 Response surface methodology for functional data
   - Response surface methodology
   - Extension to the functional setting
Aim: estimate the conditional distribution function

\[ F^x(y) = \mathbb{P}(Y \leq y | X = x) \]

using the information of the sample \( \{(X_i, Y_i), i = 1, \ldots, n\} \) following the same distribution as \((X, Y)\).
Estimation method

- **Kernel estimation**

\[
\hat{F}_{h,d}^x(y) = \frac{\sum_{i=1}^n K_h \left( d(X_i, x) \right) \mathbf{1}_{\{Y_i \leq y\}}}{\sum_{i=1}^n K_h \left( d(X_i, x) \right)}
\]

where

- \( K : \mathbb{R} \to \mathbb{R}_+ \) is a kernel function. It verifies \( \int \mathbb{R} K(t) dt = 1 \).
- \( h > 0 \) is a bandwidth.
- \( d : \mathbb{H}^2 \to \mathbb{R}_+ \) is a general pseudometric.

**Reference:** Ferraty et al. (2006, 2010):

- Almost complete and uniform almost complete convergence (with bias-variance decomposition).
- Rates of convergence on some examples of processes.

**Purposes**

- provide a data-driven choice for the bandwidth \( h \) with nonasymptotic theoretical results;
- discuss the choice of the semi-metric \( d \) in the kernel;
- compute optimal rates of convergence under various regularity assumptions.
Estimation method

- **Kernel estimation**

\[
\hat{F}^x_{h,d}(y) = \frac{\sum_{i=1}^n K_h \left( d(X_i, x) \right) 1\{Y_i \leq y\}}{\sum_{i=1}^n K_h \left( d(X_i, x) \right)}
\]

where

- \( K : \mathbb{R} \rightarrow \mathbb{R}_+ \) is a kernel function. It verifies \( \int_{\mathbb{R}} K(t) dt = 1 \).
- \( h > 0 \) is a bandwidth.
- \( d : \mathbb{H}^2 \rightarrow \mathbb{R}_+ \) is a general pseudometric.

**Reference:** Ferraty et al. (2006, 2010):

- Almost complete and uniform almost complete convergence (with bias-variance decomposition).
- Rates of convergence on some examples of processes.

**Purposes**

- provide a data-driven choice for the bandwidth \( h \) with nonasymptotic theoretical results;
- discuss the choice of the semi-metric \( d \) in the kernel;
- compute optimal rates of convergence under various regularity assumptions.
Estimation method

- **Kernel estimation**

\[ \hat{F}_{h,d}^x(y) = \frac{\sum_{i=1}^{n} K_h(d(X_i, x)) 1\{Y_i \leq y\}}{\sum_{i=1}^{n} K_h(d(X_i, x))} \]

where
- \( K : \mathbb{R} \rightarrow \mathbb{R}_+ \) is a kernel function. It verifies \( \int_{\mathbb{R}} K(t)dt = 1 \).
- \( h > 0 \) is a bandwidth.
- \( d : \mathbb{H}^2 \rightarrow \mathbb{R}_+ \) is a general pseudometric.

**Reference:** Ferraty et al. (2006, 2010):
- Almost complete and uniform almost complete convergence (with bias-variance decomposition).
- Rates of convergence on some examples of processes.

**Purposes**
- **provide a data-driven choice for the bandwidth \( h \) with nonasymptotic theoretical results;**
- discuss the choice of the semi-metric \( d \) in the kernel;
- compute optimal rates of convergence under various regularity assumptions.
Estimation method

- **Kernel estimation**

\[
\hat{F}_{h,d}^x(y) = \frac{\sum_{i=1}^n K_h \left(d(X_i, x)\right) \mathbf{1}_{\{Y_i \leq y\}}}{\sum_{i=1}^n K_h \left(d(X_i, x)\right)}
\]

where

- \( K : \mathbb{R} \to \mathbb{R}_+ \) is a kernel function. It verifies \( \int_{\mathbb{R}} K(t)dt = 1 \).
- \( h > 0 \) is a bandwidth.
- \( d : \mathbb{H}^2 \to \mathbb{R}_+ \) is a general pseudometric.

**Reference:** Ferraty *et al.* (2006, 2010):

- Almost complete and uniform almost complete convergence (with bias-variance decomposition).
- Rates of convergence on some examples of processes.

**Purposes**

- provide a data-driven choice for the bandwidth \( h \) with nonasymptotic theoretical results;
- discuss the choice of the semi-metric \( d \) in the kernel;
- compute optimal rates of convergence under various regularity assumptions.
Estimation method

- **Kernel estimation**

\[
\hat{F}_{h,d}^x(y) = \frac{\sum_{i=1}^{n} K_h(d(X_i, x)) \mathbf{1}_{\{Y_i \leq y\}}}{\sum_{i=1}^{n} K_h(d(X_i, x))}
\]

where
- \( K : \mathbb{R} \to \mathbb{R}_+ \) is a kernel function. It verifies \( \int_{\mathbb{R}} K(t)dt = 1 \).
- \( h > 0 \) is a bandwidth.
- \( d : \mathbb{H}^2 \to \mathbb{R}_+ \) is a general pseudometric.

**Reference:** Ferraty et al. (2006, 2010):
- Almost complete and uniform almost complete convergence (with bias-variance decomposition).
- Rates of convergence on some examples of processes.

**Purposes**
- provide a data-driven choice for the bandwidth \( h \) with nonasymptotic theoretical results;
- discuss the choice of the semi-metric \( d \) in the kernel;
- compute optimal rates of convergence under various regularity assumptions.
Outline

1 Prediction in the functional linear model
   - Estimation procedure
   - Theoretical results
   - Simulation results

2 Adaptive estimation of the conditional c.d.f
   - Bias-variance decomposition of the risk
   - Bandwidth selection device
   - Optimal estimation in the minimax sense
   - Simulation study

3 Response surface methodology for functional data
   - Response surface methodology
   - Extension to the functional setting
Considered risk

- For the main part of the talk: \( d(x, x') = \|x - x'\|, (x, x') \in \mathbb{H}. \)

\[
\begin{align*}
\hat{F}_h^x(y) & := \hat{F}_{h,d}^x(y) = \sum_{i=1}^n \frac{K_h(\|X_i - x\|)1_{\{Y_i \leq y\}}}{\sum_{i=1}^n K_h(\|X_i - x\|)} \\
\end{align*}
\]

- Integrated risk

\[
\mathcal{R}(\hat{F}_h, F) := \mathbb{E} \left[ \int_B \left( \int_D (\hat{F}_h^x(y) - F^x(y))^2 \, dy \right) \, d\mathbb{P}_X(x) \right] = \mathbb{E} \left[ \|\hat{F}_h^{X'} - F^{X'}\|_D^2 1_B(X') \right]
\]

with

- \( X' \) is a copy of \( X \), independent of the data-sample.
- \( D \) is a compact subset of \( \mathbb{R} \);
- \( B \) is a bounded subset of \( \mathbb{H} \).
Assumptions to control the risk

- **Assumptions on the kernel**
  - \( \text{supp}(K) \subset [0; 1] \)
  - \( 0 < c_K \leq K(t) \leq C_K < +\infty, \ t \in [0; 1] \)

- **Assumption on the target function** \( F \):
  \[
  \exists \beta \in (0; 1), \ \exists C_D > 0, \ \forall x, x' \in \mathbb{H}, \ \|F^x - F^{x'}\|_D \leq C_D \|x - x'\|^\beta
  \]

  \(\rightarrow\) \( F \) belongs to a Hölder space with smoothness index \( \beta \).

- **Assumption on the process** \( X \):
  - through the small ball probabilities
    \[
    \varphi(h) := \mathbb{P}(|X| \leq h) \text{ and } \varphi^{x_0}(h) := \mathbb{P}(|X - x_0| \leq h), \ x_0 \in \mathbb{H}.
    \]
  - \( \exists c_\varphi, C_\varphi > 0, \) such that
    \[
    \forall h > 0, \ \forall x_0 \in B, \ c_\varphi \varphi(h) \leq \varphi^{x_0}(h) \leq C_\varphi \varphi(h).
    \]
Upper-bound for the risk

Proposition

Under the previous assumptions, there exists $C > 0$, such that, for any $h > 0$,

$$\mathcal{R}(\hat{F}_h, F) \leq C \left( h^{2\beta} + \frac{1}{n\varphi(h)} \right),$$
Upper-bound for the risk

Proposition
Under the previous assumptions, there exists $C > 0$, such that, for any $h > 0$,

$$\mathcal{R}(\hat{F}_h, F) \leq C \left( h^{2\beta} + \frac{1}{n\varphi(h)} \right),$$

Unknown oracle choice

$$h^* = \arg\min_{h \in \mathcal{H}_n} \mathcal{R}(\hat{F}_h, F)$$
Upper-bound for the risk

Proposition

Under the previous assumptions, there exists $C > 0$, such that, for any $h > 0$,

$$
\mathcal{R}(\hat{F}_h, F) \leq C \left( h^{2\beta} + \frac{1}{n\varphi(h)} \right),
$$

Unknown oracle choice

$$
\hat{h}^* = \arg \min_{h \in \mathcal{H}_n} \mathcal{R}(\hat{F}_h, F) \leq C \left( h^{2\beta} + \frac{1}{n\varphi(h)} \right)
$$
Upper-bound for the risk

Proposition

Under the previous assumptions, there exists $C > 0$, such that, for any $h > 0$,

$$
\mathcal{R}(\hat{F}_h, F) \leq C \left( h^{2\beta} + \frac{1}{n\varphi(h)} \right),
$$

Unknown oracle choice

$$
h^\ast = \arg\min_{h \in \mathcal{H}_n} \mathcal{R}(\hat{F}_h, F) \leq C \left( h^{2\beta} + \frac{1}{n\varphi(h)} \right)
$$

**Question:** How to choose $h$ without the knowledge of $\beta$ and $\varphi(h)$?
Outline

1 Prediction in the functional linear model
   - Estimation procedure
   - Theoretical results
   - Simulation results

2 Adaptive estimation of the conditional c.d.f
   - Bias-variance decomposition of the risk
   - Bandwidth selection device
   - Optimal estimation in the minimax sense
   - Simulation study

3 Response surface methodology for functional data
   - Response surface methodology
   - Extension to the functional setting
Bandwidth selection device

Inspired from the work of Goldenshluger and Lepski (2011)

Bias-variance decomposition of the risk

\[ R \left( \hat{F}_h, F \right) = \mathbb{E} \left[ \| F^{X'} - \mathbb{E} \left[ \hat{F}_h^{X'} | X' \right] \|_D^2 \mathbf{1}_B(X') \right] + \mathbb{E} \left[ \| \mathbb{E} \left[ \hat{F}_h^{X'} | X' \right] - \hat{F}_h^{X'} \|_D^2 \mathbf{1}_B(X') \right]. \]

- Variance term of order \( \frac{1}{n \varphi(h)} \) → can be estimated:
  \[ \hat{V}(h) = \kappa \frac{\ln n}{n \hat{\varphi}(h)} \text{ where } \hat{\varphi}(h) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{\|X\| \leq h\}. \]

- How to approximate the bias term ?
  \[ \hat{A}(h) = \max_{h' \in \mathcal{H}_n} \left( \| \hat{F}_h^{X'} - \hat{F}_{h' \vee h}^{X'} \|_D^2 - \hat{V}(h') \right) \]

- Finally \( \hat{h} = \arg \min_{h \in \mathcal{H}_n} \left\{ \hat{A}(h) + \hat{V}(h) \right\} \Rightarrow \hat{F}_{\hat{h}}^{X'}. \)
Bandwidth selection device

Inspired from the work of Goldenshluger and Lepski (2011)

Bias-variance decomposition of the risk

\[ \mathcal{R} \left( \hat{F}_h, F \right) = \mathbb{E} \left[ \left\| F^{X'} - \mathbb{E} \left[ \hat{F}_h^{X'} | X' \right] \right\|_D^2 \mathbf{1}_B(X') \right] + \mathbb{E} \left[ \left\| \mathbb{E} \left[ \hat{F}_h^{X'} | X' \right] - \hat{F}_h^{X'} \right\|_D^2 \mathbf{1}_B(X') \right]. \]

- **Variance term** of order \( \frac{1}{n \varphi(h)} \) → can be estimated:

  \[ \hat{V}(h) = \kappa \frac{\ln n}{n \hat{\varphi}(h)} \text{ where } \hat{\varphi}(h) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{\|X\| \leq h\}. \]

- **How to approximate the bias term?**

  \[ \hat{A}(h) = \max_{h' \in \mathcal{H}_n} \left( \| \hat{F}_{h'}^{X'} - \hat{F}_h^{X'} \|_D^2 - \hat{V}(h') \right) \]

- **Finally** \( \hat{h} = \arg \min_{h \in \mathcal{H}_n} \left\{ \hat{A}(h) + \hat{V}(h) \right\} \Rightarrow \hat{F}_h^{X'}. \)
Bandwidth selection device

Inspired from the work of Goldenshluger and Lepski (2011)

Bias-variance decomposition of the risk

\[ \mathcal{R} \left( \hat{F}_h, F \right) = \mathbb{E} \left[ \left\| F^{X'} - \mathbb{E} \left[ \hat{F}^{X'}_h | X' \right] \right\|_D^2 \mathbf{1}_B(X') \right] + \mathbb{E} \left[ \left\| \mathbb{E} \left[ \hat{F}^{X'}_h | X' \right] - \hat{F}^{X'}_h \right\|_D^2 \mathbf{1}_B(X') \right]. \]

- **Variance term** of order \( \frac{1}{n \varphi(h)} \) → can be estimated:

  \[ \hat{V}(h) = \kappa \frac{\ln n}{n \hat{\varphi}(h)} \text{ where } \hat{\varphi}(h) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{\|X\| \leq h\}. \]

- How to approximate the **bias term**?

  \[ \hat{A}(h) = \max_{h' \in \mathcal{H}_n} \left( \| \hat{F}^{X'}_{h'} - \hat{F}^{X'}_{h \vee h} \|_D^2 - \hat{V}(h') \right) + \]

- Finally \( \hat{h} = \arg \min_{h \in \mathcal{H}_n} \left\{ \hat{A}(h) + \hat{V}(h) \right\} \Rightarrow \hat{F}^{X'}_h. \)
Bandwidth selection device

Inspired from the work of Goldenshluger and Lepski (2011)

Bias-variance decomposition of the risk

\[
R\left(\hat{F}_h, F\right) = \mathbb{E}\left[\|F^{X'} - \mathbb{E}\left[\hat{F}_h^{X'} | X'\right]\|_D^2 \mathbf{1}_B(X')\right] + \mathbb{E}\left[\mathbb{E}\left[\hat{F}_h^{X'} | X'\right] - \hat{F}_h^{X'}\|_D^2 \mathbf{1}_B(X')\right].
\]

- **Variance term** of order \(\frac{1}{n\varphi(h)}\) → can be estimated:

\[
\hat{V}(h) = \kappa \frac{\ln n}{n\hat{\varphi}(h)} \text{ where } \hat{\varphi}(h) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{\|X\| \leq h\}.
\]

- How to approximate the **bias term**?

\[
\hat{A}(h) = \max_{h' \in \mathcal{H}_n} \left(\|\hat{F}_h^{X'} - \hat{F}_{h'\vee h}^{X'}\|_D^2 - \hat{V}(h')\right) +
\]

- Finally, \(\hat{h} = \arg \min_{h \in \mathcal{H}_n} \left\{\hat{A}(h) + \hat{V}(h)\right\} \Rightarrow \hat{F}_h^{X'}\).
Bandwidth selection device
Inspired from the work of Goldenshluger and Lepski (2011)

Bias-variance decomposition of the risk

\[
\mathcal{R}(\hat{F}_h, F) = \mathbb{E} \left[ \|F^{X'} - \mathbb{E} \left[ \hat{F}_h^{X'} | X' \right] \|_D^2 1_B(X') \right] + \mathbb{E} \left[ \| \mathbb{E} \left[ \hat{F}_h^{X'} | X' \right] - \hat{F}_h^{X'} \|_D^2 1_B(X') \right].
\]

- **Variance term** of order \( \frac{1}{n\varphi(h)} \) → can be estimated:

  \[
  \hat{V}(h) = \kappa \frac{\ln n}{n\hat{\varphi}(h)} \quad \text{where} \quad \hat{\varphi}(h) = \frac{1}{n} \sum_{i=1}^{n} 1\{\|X\| \leq h\}. 
  \]

- How to approximate the **bias term**?

  \[
  \hat{A}(h) = \max_{h' \in \mathcal{H}_n} \left( \| \hat{F}_h^{X'} - \hat{F}_{h' \vee h}^{X'} \|_D^2 - \hat{V}(h') \right) +
  \]

- Finally \( \hat{h} = \arg \min_{h \in \mathcal{H}_n} \left\{ \hat{A}(h) + \hat{V}(h) \right\} \Rightarrow \hat{F}_{\hat{h}}^{X'} \).
Main result: nonasymptotic adaptive risk bound

Theorem

Under the previous assumptions, and if the collection $\mathcal{H}_n$ is not too large, there exist 2 constants $c, C > 0$ such that

$$
\mathcal{R}(\hat{F}_h, F) \leq c \min_{h \in \mathcal{H}_n} \left\{ h^{2\beta} + \frac{\ln(n)}{n\varphi(h)} \right\} + \frac{C}{n}.
$$
Outline

1. Prediction in the functional linear model
   - Estimation procedure
   - Theoretical results
   - Simulation results

2. Adaptive estimation of the conditional c.d.f
   - Bias-variance decomposition of the risk
   - Bandwidth selection device
   - Optimal estimation in the minimax sense
   - Simulation study

3. Response surface methodology for functional data
   - Response surface methodology
   - Extension to the functional setting
Additional assumption on the small ball probability

\[ \varphi(h) = \mathbb{P}(\|X\| \leq h), \ h > 0. \]

3 possible assumptions on the decay of the s.b.p.

- **Fast decay**
  \[ \varphi(h) \asymp h^\gamma \exp\left(-ch^{-\alpha}\right), \ \gamma \in \mathbb{R}, \alpha > 0. \]

  **Ex:** if \( X \) is a brownian motion, assumption satisfied with \( \alpha = 2 \).

- **Intermediate decay**
  \[ \varphi(h) \asymp h^\gamma \exp\left(-c \ln^{-\alpha}(1/h)\right), \ \gamma \in \mathbb{R}, \alpha > 1. \]

- **Low decay**
  \[ \varphi(h) \asymp h^\gamma, \ \gamma > 0. \]

  **Ex:** if \( X \in \mathbb{R}^d \) (random vector), assumption satisfied with \( \gamma = d \).
### Rates of convergence

| Fast decay for $\varphi(h)$  
   (slow rates) | Intermediate decay for $\varphi(h)$  
   (intermediate) | Low decay for $\varphi(h)$  
   (fast rates) |
|------------------|------------------|------------------|
| $\mathcal{R}(\hat{F}_h, F) \lesssim \cdots$  
   (adaptive rate) | $(\ln(n))^{-2\beta/\alpha}$ | $\exp\left(-\frac{2\beta}{c_2^{1/\alpha}} \ln^{1/\alpha}(n)\right)$ | $\left(\frac{n}{\ln(n)}\right)^{-\frac{2\beta}{2\beta+\gamma}}$ |

→ similar rates to the ones obtained by Ferraty et al. (2006), but for an adaptive bandwidth.
## Rates of convergence

<table>
<thead>
<tr>
<th></th>
<th>Fast decay for $\varphi(h)$ (slow rates)</th>
<th>Intermediate decay $\varphi(h)$ (intermediate)</th>
<th>Low decay $\varphi(h)$ (fast rates)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) $\mathcal{R}(\hat{F}_h, F) \lesssim \cdots$ (adaptive rate)</td>
<td>$(\ln(n))^{-2\beta/\alpha}$</td>
<td>$\exp\left(-\frac{2\beta}{c_1/\alpha} \ln^{1/\alpha}(n)\right)$</td>
<td>$\left(\frac{n}{\ln(n)}\right)^{-\frac{2\beta}{2\beta+\gamma}}$</td>
</tr>
<tr>
<td>(b) Minimax rate $\inf_{\tilde{F}} \sup_{F, X} \mathcal{R}(\tilde{F}, F) \gtrsim \cdots$ (lower bound)</td>
<td>$(\ln(n))^{-2\beta/\alpha}$</td>
<td>$\exp\left(-\frac{2\beta}{c_1/\alpha} \ln^{1/\alpha}(n)\right)$</td>
<td>$n^{-\frac{2\beta}{2\beta+\gamma}}$</td>
</tr>
</tbody>
</table>

$\rightarrow$ similar rates to the ones obtained by Mas (2012) for regression estimation.

$\rightarrow$ the estimator is then optimal in the minimax sense, up to the extra $\ln(n)$ factor.
Outline

1 Prediction in the functional linear model
   - Estimation procedure
   - Theoretical results
   - Simulation results

2 Adaptive estimation of the conditional c.d.f
   - Bias-variance decomposition of the risk
   - Bandwidth selection device
   - Optimal estimation in the minimax sense
   - Simulation study

3 Response surface methodology for functional data
   - Response surface methodology
   - Extension to the functional setting
Implementation

- **Choice of** $K$: uniform kernel $K = 1_{[0,1]}$.

- **Choice of** $\mathcal{H}_n$: $\mathcal{H}_n = \{C/k, 1 \leq k \leq k_{\text{max}}\}$.

- **Simulation of** $X$:
  - $(W(t))_t$ a brownian motion,
  - $(\xi_j)_{j \geq 0}$ i.i.d. $N(0, 1)$.

\[
X(t) = W(t) + \xi_0
\]

\[
X(t) = \xi_0 + \sqrt{2} \sum_{j=1}^{150} \xi_j \frac{e^{-j}}{\sqrt{j}} \sin(\pi(j - 0.5)t)
\]

\[
X(t) = \xi_0 + \sqrt{2} \xi_1 \sin(\pi t/2) + \xi_2 \sin(3\pi t/2)/\sqrt{2}
\]
Estimators

Conditional c.d.f estimation in a regression model

Observations: \((X_i, Y_i)_{i \in \{1, \ldots, 500\}}\) such that \(Y_i = \left(\int_0^1 \beta(t)X_i(t)dt\right)^2 + \varepsilon_i\) with \(\beta(t) = \sin(4\pi t)\) and \(\varepsilon_i \sim \mathcal{N}(0, 0.1)\).

Fast decay for \(\varphi(h)\)

Intermediate decay for \(\varphi(h)\)

Low decay for \(\varphi(h)\)

.. true conditional c.d.f. | --- estimators \((\hat{F}_h)_{h \in \mathcal{H}_n}\),
oracle estimator \(\hat{F}_{h^*}\), | --- adaptive estimator \(\hat{F}_{\hat{h}}\).
Estimators
Conditional c.d.f estimation in a Gaussian mixture model

Observations: \((X_i, Y_i)_{i \in \{1, \ldots, 500\}}\) such that
\(Y_i | X_i = x \sim 0.5 \mathcal{N}(8 - 4 \|x\|, 1) + 0.5 \mathcal{N}(8 + 4 \|x\|, 1),\)

Fast decay for \(\varphi(h)\)
Intermediate decay for \(\varphi(h)\)
Low decay for \(\varphi(h)\)

. . true conditional c.d.f. \(\hat{F}_h\)
oracle estimator \(\hat{F}_{h^*}\),
estimators \((\hat{F}_h)_{h \in \mathcal{H}_n}\),
adaptive estimator \(\hat{F}_{\hat{h}}\).
Outline

1. Prediction in the functional linear model
   - Estimation procedure
   - Theoretical results
   - Simulation results

2. Adaptive estimation of the conditional c.d.f
   - Bias-variance decomposition of the risk
   - Bandwidth selection device
   - Optimal estimation in the minimax sense
   - Simulation study

3. Response surface methodology for functional data
   - Response surface methodology
   - Extension to the functional setting
Outline

1. Prediction in the functional linear model
   - Estimation procedure
   - Theoretical results
   - Simulation results

2. Adaptive estimation of the conditional c.d.f
   - Bias-variance decomposition of the risk
   - Bandwidth selection device
   - Optimal estimation in the minimax sense
   - Simulation study

3. Response surface methodology for functional data
   - Response surface methodology
   - Extension to the functional setting
Response surface methodology

Brief history

- **Box and Wilson (1950):** optimal conditions for chemical experimentation → widely used in industry.

- **Sacks *et al.* (1989):** Extension to numerical experiments
  - Lee and Hajela (1996): conception of rotor blades...

- **Recent advances:** Facer and Müller (2003), Khuri and Mukhopadhyay (2010), Georgiou, Stylianou and Aggarwal (2014).
Methodology

**Goal:** minimisation of \( (x_1, \ldots, x_n) \mapsto m(x_1, \ldots, x_d), \) **unknown.**

**Information available:**

\[
y_i = m(x_{1,i}, \ldots, x_{d,i}) + \varepsilon_i, \quad i = 1, \ldots, n,
\]

\((x_{1,i}, \ldots, x_{d,i})_{i=1}^n\) chosen by the user and \(n\) as small as possible.

**Example:**

- \(m(x_1, x_2) = x_1^2 + x_2^2;\)
- \(\varepsilon \sim \mathcal{N}(0, 1).\)
Methodology

Goal: minimisation of \((x_1, \ldots, x_n) \mapsto m(x_1, \ldots, x_d)\), unknown.

Information available:

\[ y_i = m(x_{1,i}, \ldots, x_{d,i}) + \varepsilon_i, \quad i = 1, \ldots, n, \]

\((x_{1,i}, \ldots, x_{d,i})_{i=1}^n\) chosen by the user and \(n\) as small as possible.

Example:

- \(m(x_1, x_2) = x_1^2 + x_2^2;\)
- \(\varepsilon \sim \mathcal{N}(0, 1).\)
Methodology

Goal: minimisation of \((x_1, \ldots, x_n) \mapsto m(x_1, \ldots, x_d), \) unknown.

Information available:

\[ y_i = m(x_{1,i}, \ldots, x_{d,i}) + \varepsilon_i, \ i = 1, \ldots, n, \]

\((x_{1,i}, \ldots, x_{d,i})_{i=1}^n\) chosen by the user and \(n\) as small as possible.

Example:

- \(m(x_1, x_2) = x_1^2 + x_2^2;\)
- \(\varepsilon \sim \mathcal{N}(0, 1).\)

Legend:

- Initial point
Methodology

Goal: minimisation of \((x_1, \ldots, x_n) \mapsto m(x_1, \ldots, x_d),\) unknown.

Information available:

\[ y_i = m(x_{1,i}, \ldots, x_{d,i}) + \varepsilon_i, \ i = 1, \ldots, n, \]
\[ (x_{1,i}, \ldots, x_{d,i})_{i=1}^n \text{ chosen by the user and } n \text{ as small as possible.} \]

Example:

- \(m(x_1, x_2) = x_1^2 + x_2^2;\)
- \(\varepsilon \sim \mathcal{N}(0, 1).\)

Legend:

- Initial point
- Minimal point (target)
Response surface methodology for functional data

Methodology

**Goal:** minimisation of \((x_1, \ldots, x_n) \mapsto m(x_1, \ldots, x_d), \text{ unknown.}\)

**Information available:**

\[ y_i = m(x_{1,i}, \ldots, x_{d,i}) + \varepsilon_i, \quad i = 1, \ldots, n, \]

\((x_{1,i}, \ldots, x_{d,i})_{i=1}^n\) chosen by the user and \(n\) as small as possible.

**Example:**

- \(m(x_1, x_2) = x_1^2 + x_2^2;\)
- \(\varepsilon \sim \mathcal{N}(0, 1).\)

**Legend:**

- Initial point
- Minimal point (target)
- Factorial design points

Factorial 2\(^2\) design: 4 points

\(\mathbf{x}^{(0)}\)

\(\mathbf{x}_1^{(0)}\)

\(\mathbf{x}_2^{(0)}\)

\(\mathbf{x}_3^{(0)}\)

\(\mathbf{x}_4^{(0)}\)
Methodology

Goal: minimisation of \((x_1, \ldots, x_n) \mapsto m(x_1, \ldots, x_d)\), unknown.

Information available:

\[ y_i = m(x_{1,i}, \ldots, x_{d,i}) + \varepsilon_i, \quad i = 1, \ldots, n, \]

\((x_{1,i}, \ldots, x_{d,i})_{i=1}^n\) chosen by the user and \(n\) as small as possible.

Least-squares fit of a first order model:

\[ y = \beta_1 x_1 + \beta_2 x_2 + \varepsilon'. \]

Direction of steepest descent estimated:

\((-\hat{\beta}_1, -\hat{\beta}_2).\)

Example:

- \(m(x_1, x_2) = x_1^2 + x_2^2;\)
- \(\varepsilon \sim \mathcal{N}(0, 1).\)

Legend:

- Initial point
- Minimal point (target)
- Factorial design points
- Direction of descent
Methodology

**Goal:** minimisation of \((x_1, \ldots, x_n) \mapsto m(x_1, \ldots, x_d), \text{ unknown}.\)

Information available:

\[
y_i = m(x_{1,i}, \ldots, x_{d,i}) + \varepsilon_i, \ i = 1, \ldots, n, \\
(x_{1,i}, \ldots, x_{d,i})_{i=1}^n \text{ chosen by the user and } n \text{ as small as possible}.
\]

**Example:**

- \(m(x_1, x_2) = x_1^2 + x_2^2;\)
- \(\varepsilon \sim \mathcal{N}(0, 1).\)

**Legend:**

- Initial point
- Minimal point (target)
- Factorial design points
- Descent steps
Methodology

**Goal:** minimisation of $\mathbf{x} = (x_1, \ldots, x_n) \mapsto m(x_1, \ldots, x_d)$, unknown.

**Information available:**

$$y_i = m(x_{1,i}, \ldots, x_{d,i}) + \varepsilon_i, \ i = 1, \ldots, n,$$

$(x_{1,i}, \ldots, x_{d,i})_{i=1}^n$ chosen by the user and $n$ as small as possible.

**Example:**

- $m(x_1, x_2) = x_1^2 + x_2^2$;
- $\varepsilon \sim \mathcal{N}(0, 1)$.

**Legend:**

- **Initial point**
- **Minimal point (target)**
- **Factorial design points**
- **Descent steps**

Observed response:

$$y = f(x_1 - \alpha \hat{\beta}_1, x_2 - \alpha \hat{\beta}_2) + \varepsilon$$
Methodology

Goal: minimisation of \((x_1, \ldots, x_n) \mapsto m(x_1, \ldots, x_d), \text{ unknown} \).

Information available:

\[ y_i = m(x_{1,i}, \ldots, x_{d,i}) + \varepsilon_i, \ i = 1, \ldots, n, \]

\((x_{1,i}, \ldots, x_{d,i})_{i=1}^n\) chosen by the user and \(n\) as small as possible.

Example:

- \(m(x_1, x_2) = x_1^2 + x_2^2\);
- \(\varepsilon \sim N(0, 1)\).

Legend:

- Minimal point of the descent direction
- Minimal point (target)
Methodology

Goal: minimisation of \((x_1, \ldots, x_n) \mapsto m(x_1, \ldots, x_d), \text{ unknown.}\)

Information available:

\[ y_i = m(x_{1,i}, \ldots, x_{d,i}) + \varepsilon_i, \ i = 1, \ldots, n, \]

\( (x_{1,i}, \ldots, x_{d,i})_{i=1}^n \) chosen by the user and \( n \) as small as possible.

Example:

- \( m(x_1, x_2) = x_1^2 + x_2^2; \)
- \( \varepsilon \sim N(0, 1). \)

Legend:

- Minimal point of the descent direction
- Minimal point (target)
- Factorial design points
- CCD axial points
**Methodology**

**Goal:** minimisation of \((x_1, \ldots, x_n) \mapsto m(x_1, \ldots, x_d), \text{ unknown.}\)

**Information available:**

\[ y_i = m(x_{1,i}, \ldots, x_{d,i}) + \varepsilon_i, \ i = 1, \ldots, n, \]
\[ (x_{1,i}, \ldots, x_{d,i})^{n}_{i=1} \text{ chosen by the user and } n \text{ as small as possible.} \]

Least-squares fit of a second-order model:
\[ y = \beta_1 x_1 + \beta_2 x_2 + (x_1, x_2)B(x_1, x_2)^t + \varepsilon''. \]

Stationary point:
\[ (x_1^*, x_2^*) = \frac{1}{2}B^{-1}(\hat{\beta}_1, \hat{\beta}_2)^t. \]

**Example:**

- \( m(x_1, x_2) = x_1^2 + x_2^2; \)
- \( \varepsilon \sim \mathcal{N}(0, 1). \)

**Legend:**

- Minimal point of the descent direction
- Minimal point (target)
- Factorial design points
- CCD axial points
- Stationary point (estimated minimal point)
Methodology

**Goal:** minimisation of \((x_1, \ldots, x_n) \mapsto m(x_1, \ldots, x_d), \text{ unknown.} \)

**Information available:**

\[ y_i = m(x_{1,i}, \ldots, x_{d,i}) + \varepsilon_i, \; i = 1, \ldots, n, \]

\((x_{1,i}, \ldots, x_{d,i})_{i=1}^n\) chosen by the user and \(n\) as small as possible.

**Example:**

- \(m(x_1, x_2) = x_1^2 + x_2^2; \)
- \(\varepsilon \sim \mathcal{N}(0, 1).\)

**Legend:**

- Step points
- Minimal point (target)
Outline

1 Prediction in the functional linear model
   - Estimation procedure
   - Theoretical results
   - Simulation results

2 Adaptive estimation of the conditional c.d.f
   - Bias-variance decomposition of the risk
   - Bandwidth selection device
   - Optimal estimation in the minimax sense
   - Simulation study

3 Response surface methodology for functional data
   - Response surface methodology
   - Extension to the functional setting
Response surface methodology for functional data

Extension to the functional setting

Problems raised by the functional context

- First and second-order models can be defined easily but
  ... How to define functional design of experiments?
- One possible answer: combine dimension reduction with classical
  finite-dimensional design of experiments
  - \((x^{(i)} = (x_{0,1}^{(i)}, \ldots, x_{0,d}^{(i)}) \in \mathbb{R}^d, i = 1, \ldots, n_0)\) \(d\)-dimensional design of experiments;
  - \(\{\varphi_1, \ldots, \varphi_d\}\) orthonormal family of \(\mathbb{H}\)

\[
x^{(i)}_0 = x_0 + \sum_{j=1}^{d} x^{(i)}_{0,j} \varphi_j,
\]

\(\rightarrow\) functional design of experiments.

... How can we define the directions \(\{\varphi_1, \ldots, \varphi_d\}\) ?

- Possible basis of approximation
  - Fixed basis: Fourier, \(B\)-splines, wavelets,...
  - If a training sample exists: data driven basis
    - PCA basis;
    - PLS basis Wold (1975), Preda and Saporta (2005), Delaigle and Hall (2012): allows to take into account the interaction between \(x\) and \(y\).
Problems raised by the functional context

- First and second-order models can be defined easily but
  ... How to define functional design of experiments?

- One possible answer: combine dimension reduction with classical finite-dimensional design of experiments
  - \((x^{(i)}_0 = (x^{(i)}_{0,1}, \ldots, x^{(i)}_{0,d}) \in \mathbb{R}^d, i = 1, \ldots, n_0)\) \(d\)-dimensional design of experiments;
  - \(\{\varphi_1, \ldots, \varphi_d\}\) orthonormal family of \(\mathbb{H}\)

\[
x^{(i)}_0 = x_0 + \sum_{j=1}^{d} x^{(i)}_{0,j} \varphi_j,
\]

\(\longrightarrow\) functional design of experiments.
  ... How can we define the directions \(\{\varphi_1, \ldots, \varphi_d\}\) ?

- Possible basis of approximation
  - Fixed basis: Fourier, B-splines, wavelets,...
  - If a training sample exists: data driven basis
    - PCA basis;
    - PLS basis Wold (1975), Preda and Saporta (2005), Delaigle and Hall (2012): allows to take into account the interaction between \(x\) and \(y\).
Problems raised by the functional context

- First and second-order models can be defined easily but ...
  ... How to define functional design of experiments?
- One possible answer: combine dimension reduction with classical
  finite-dimensional design of experiments
  - \((x_0^{(i)} = (x_{0,1}^{(i)}, \ldots, x_{0,d}^{(i)}) \in \mathbb{R}^d, i = 1, \ldots, n_0)\) \(d\)-dimensional design of experiments;
  - \(\{\varphi_1, \ldots, \varphi_d\}\) orthonormal family of \(H\)

\[
  x_o^{(i)} = x_0 + \sum_{j=1}^{d} x_{0,j}^{(i)} \varphi_j,
\]

\(\rightarrow\) functional design of experiments.

... How can we define the directions \(\{\varphi_1, \ldots, \varphi_d\}\) ?

- Possible basis of approximation
  - Fixed basis: Fourier, B-splines, wavelets, ...
  - If a training sample exists: data driven basis
    - PCA basis;
    - PLS basis Wold (1975), Preda and Saporta (2005), Delaigle and Hall (2012): allows to take into account the interaction between \(x\) and \(y\).
Example of functional design of experiments
Factorial $2^d$ design in $\mathbb{H} = \mathbb{L}^2([0, 1])$

$d = 2, 16$ curves

$d = 4, 32$ curves

$d = 8, 280$ curves

$X$ brownian motion, $Y = \|X - f\|^2 + \varepsilon, f(t) = \cos(4\pi t) + 3 \sin(\pi t) + 10, \varepsilon \sim \mathcal{N}(0, 0.01)$

$^1$calculated from $(X_i)_{i=1}^{500}$

$^2$calculated from $(X_i, Y_i)_{i=1}^{500}$
Example of functional design of experiments

Central Composite Designs in $\mathbb{H} = \mathbb{L}^2([0, 1])$

- $d = 2$, 4 curves
- $d = 4$, 16 curves
- $d = 8$, 256 curves

$X$ brownian motion, $Y = ||X - f||^2 + \varepsilon, f(t) = \cos(4\pi t) + 3\sin(\pi t) + 10, \varepsilon \sim \mathcal{N}(0, 0.01)$

- Calculated from $(X_i)_{i=1}^{500}$
- Calculated from $(X_i, Y_i)_{i=1}^{500}$
Methodology
Adaptation to a functional context

**Goal:** minimisation of $x \mapsto m(x)$, unknown.

**Example:**
- $m(x) = \| x - f \|^2$ with
  
  $f(t) = \cos(4\pi t) + 3 \sin(\pi t) + 10$;

  $\varepsilon \sim \mathcal{N}(0, 10)$. 
Methodology

Adaptation to a functional context

Goal: minimisation of $x \mapsto m(x)$, unknown.

Example:
- $m(x) = \|x - f\|^2$ with $f(t) = \cos(4\pi t) + 3 \sin(\pi t) + 10$;
- $\varepsilon \sim \mathcal{N}(0, 10)$.

Legend:
- Initial point
Methodology

Adaptation to a functional context

**Goal:** minimisation of $x \mapsto m(x)$, unknown.

**Example:**
- $m(x) = \|x - f\|^2$ with
  - $f(t) = \cos(4\pi t) + 3 \sin(\pi t) + 10$;  
- $\varepsilon \sim \mathcal{N}(0, 10)$.

**Legend:**
- Initial point
- Minimal point $f(t)$ (target)
Methodology
Adaptation to a functional context

Goal: minimisation of $x \mapsto m(x)$, unknown.

Example: 
- $m(x) = \|x - f\|^2$ with $f(t) = \cos(4\pi t) + 3 \sin(\pi t) + 10$;
- $\varepsilon \sim \mathcal{N}(0, 10)$.

Legend:
- Initial point
- Minimal point $f(t)$ (target)
- $2^8$ factorial design

$^5$directions: PLS basis calculated from $(X_i, m(X_i) + \varepsilon_i)_{i=1}^{500}$ ($X_i$ brownian motion)
**Methodology**

Adaptation to a functional context

**Goal:** minimisation of $x \mapsto m(x)$, unknown.

\[ m(x) = \|x - f\|^2 \]

**Example:**
- $m(x) = \|x - f\|^2$ with 
  \[ f(t) = \cos(4\pi t) + 3 \sin(\pi t) + 10; \]
- $\varepsilon \sim \mathcal{N}(0, 10)$. 

**Legend:**
- Initial point
- Minimal point $f(t)$ (target)

Least-squares fit of a first order model $\rightarrow$ estimation of direction of steepest descent
**Methodology**

**Adaptation to a functional context**

**Goal:** minimisation of $x \mapsto m(x)$, **unknown**.

**Example:**
- $m(x) = \|x - f\|^2$ with
  - $f(t) = \cos(4\pi t) + 3 \sin(\pi t) + 10$;
- $\varepsilon \sim \mathcal{N}(0, 10)$.

**Legend:**
- Initial point
- Minimal point $f(t)$ (target)
- Points of the descent direction

---

**Observed response on descent path:**
Methodology

Adaptation to a functional context

**Goal:** minimisation of $x \mapsto m(x)$, unknown.

**Example:**
- $m(x) = \|x - f\|^2$ with
  - $f(t) = \cos(4\pi t) + 3 \sin(\pi t) + 10$;
- $\varepsilon \sim \mathcal{N}(0, 10)$.

**Legend:**
- Minimal point of the descent direction
- Minimal point $f(t)$ (target)
- Points of the descent direction
**Methodology**

Adaptation to a functional context

**Goal:** minimisation of $x \mapsto m(x)$, unknown.

**Example:**
- $m(x) = \|x - f\|^2$ with $f(t) = \cos(4\pi t) + 3 \sin(\pi t) + 10$;
- $\varepsilon \sim \mathcal{N}(0, 10)$.

**Legend:**
- Minimal point of the descent direction
- Minimal point $f(t)$ (target)
**Methodology**

Adaptation to a functional context

**Goal:** minimisation of $x \mapsto m(x)$, unknown.

**Example:**

- $m(x) = \|x - f\|^2$ with 
  
  $f(t) = \cos(4\pi t) + 3 \sin(\pi t) + 10$;

- $\varepsilon \sim \mathcal{N}(0, 10)$.

**Legend:**

- Minimal point of the descent direction
- Minimal point $f(t)$ (target)
- } Central Composite Design${}^5$

Directions: PLS basis calculated from $(X_i, m(X_i) + \varepsilon_i)_{i=1}^{500}$ ($X_i$ brownian motion, $d = 8$)

${}^5$
Methodology

Adaptation to a functional context

**Goal:** minimisation of $x \mapsto m(x)$, unknown.

**Example:**
- $m(x) = \|x - f\|^2$ with
  - $f(t) = \cos(4\pi t) + 3 \sin(\pi t) + 10$;
  - $\varepsilon \sim \mathcal{N}(0, 10)$.

**Legend:**
- Minimal point of the descent direction
- Minimal point $f(t)$ (target)
- } Central Composite Design$^5$

$^5$directions: PLS basis calculated from $(X_i, m(X_i) + \varepsilon_i)_{i=1}^{500}$ ($X_i$ brownian motion, $d = 8$)

Least-squares fit of a second order model $\rightarrow$ estimation of stationary point
**Methodology**

Adaptation to a functional context

**Goal:** minimisation of \( x \mapsto m(x) \), unknown.

**Example:**

\[
m(x) = \| x - f \|^2 \quad \text{with} \quad f(t) = \cos(4\pi t) + 3 \sin(\pi t) + 10;
\]

\[\varepsilon \sim \mathcal{N}(0, 10).\]

**Legend:**

- Minimal point of the descent direction
- Minimal point \( f(t) \) (target)
- Stationary point (estimation of the minimal point)
**Methodology**

Adaptation to a functional context

**Goal:** minimisation of $x \mapsto m(x)$, unknown.

**Example:**
- $m(x) = \|x - f\|^2$ with $f(t) = \cos(4\pi t) + 3 \sin(\pi t) + 10$;
- $\varepsilon \sim \mathcal{N}(0, 10)$.

**Legend:**
- Step points
- Minimal point $f(t)$ (target)
Conclusion

- **Model selection for functional principal component regression**
  -> faster and more stable than usual cross-validation
  ... with non-asymptotic control of the prediction error.

- **Bandwidth selection for kernel estimation**
  -> first adaptive estimation procedure in *nonparametric* estimation for functional data
  -> precise lower bounds and convergence rates.

  → both estimation procedures leads to **minimax optimal** estimators.

- **First attempt to adapt Response Surface Methodology to functional data.**
  -> definition of functional design of experiments.
Conclusion

- **Model selection for functional principal component regression**
  → faster and more stable than usual cross-validation
  ...
  ... with non-asymptotic control of the prediction error.

- **Bandwidth selection for kernel estimation**
  → first adaptive estimation procedure in *nonparametric* estimation for functional data
  → precise lower bounds and convergence rates.

  → both estimation procedures leads to **minimax optimal** estimators.

- **First attempt to adapt Response Surface Methodology to functional data.**
  → definition of functional design of experiments.
Conclusion

- **Model selection for functional principal component regression**
  → faster and more stable than usual cross-validation
  ... with non-asymptotic control of the prediction error.

- **Bandwidth selection for kernel estimation**
  → first adaptive estimation procedure in *nonparametric* estimation for functional data
  → precise lower bounds and convergence rates.
  → both estimation procedures leads to minimax optimal estimators.

- **First attempt to adapt Response Surface Methodology to functional data.**
  → definition of functional design of experiments.
Conclusion

- **Model selection for functional principal component regression**
  → faster and more stable than usual cross-validation
  ... with non-asymptotic control of the prediction error.

- **Bandwidth selection for kernel estimation**
  → first adaptive estimation procedure in nonparametric estimation for functional data
  → precise lower bounds and convergence rates.
  → both estimation procedures leads to minimax optimal estimators.

- **First attempt to adapt Response Surface Methodology to functional data.**
  → definition of functional design of experiments.
Conclusion

- **Model selection for functional principal component regression**
  → faster and more stable than usual cross-validation
  ... with non-asymptotic control of the prediction error.

- **Bandwidth selection for kernel estimation**
  → first adaptive estimation procedure in *nonparametric* estimation for functional data
  → precise lower bounds and convergence rates.

→ both estimation procedures leads to minimax optimal estimators.

- **First attempt to adapt Response Surface Methodology to functional data.**
  → definition of functional design of experiments.
Conclusion

- **Model selection for functional principal component regression**
  → faster and more stable than usual cross-validation
  ... with non-asymptotic control of the prediction error.

- **Bandwidth selection for kernel estimation**
  → first adaptive estimation procedure in *nonparametric* estimation for functional data
  → precise lower bounds and convergence rates.

→ both estimation procedures leads to *minimax optimal* estimators.

- **First attempt to adapt Response Surface Methodology to functional data.**
  → definition of functional design of experiments.
Conclusion

- **Model selection for functional principal component regression**
  → faster and more stable than usual cross-validation
  ... with non-asymptotic control of the prediction error.
- **Bandwidth selection for kernel estimation**
  → first adaptive estimation procedure in *nonparametric* estimation for functional data
  → precise lower bounds and convergence rates.
  → both estimation procedures leads to minimax optimal estimators.
- **First attempt to adapt Response Surface Methodology to functional data.**
  → definition of functional design of experiments.
Conclusion

- **Model selection for functional principal component regression**
  - faster and more stable than usual cross-validation
  - ... with non-asymptotic control of the prediction error.
- **Bandwidth selection for kernel estimation**
  - first adaptive estimation procedure in *nonparametric* estimation for functional data
  - precise lower bounds and convergence rates.
  - both estimation procedures leads to **minimax optimal** estimators.

- **First attempt to adapt Response Surface Methodology to functional data.**
  - definition of functional design of experiments.
Conclusion

- **Model selection for functional principal component regression**
  - faster and more stable than usual cross-validation
  - ... with non-asymptotic control of the prediction error.

- **Bandwidth selection for kernel estimation**
  - first adaptive estimation procedure in *nonparametric* estimation for functional data
  - precise lower bounds and convergence rates.
  - both estimation procedures leads to **minimax optimal** estimators.

- **First attempt to adapt Response Surface Methodology to functional data.**
  - definition of functional design of experiments.
Perspectives

- **Response surface methodology**: minimisation of the probability of failure of a nuclear reactor vessel (CEA Cadarache);

- **Functional single-index model**: \( Y = g(\langle \beta, X \rangle) + \varepsilon \). Is it possible to define a projection based estimator which is adaptive?

- **Kernel estimators in high/infinite dimension (with Gaëlle Chagny)**:
  - How to choose relevant metrics for kernels?
  - Theoretical study of resulting estimators.

- **Functional linear model**: Adaptive parameter selection for the roughness regularization method.

\[
\hat{\beta}_\rho \in \arg \min_{f \in S} \left\{ \frac{1}{n} \sum_{i=1}^{n} (Y_i - \langle f, X_i \rangle)^2 + \rho \| f \|_S^2 \right\},
\]

with \( \rho \) a smoothing parameter, \( S \subset \mathcal{H} \) and \( \| \cdot \|_S \) a seminorm on \( S \).
Thank you for your attention!

- Penalized contrast estimation in functional linear models with circular data. É. Brunel and A. Roche, accepted for publication in *Statistics*.


- Response surface methodology for functional data : application to nuclear safety. Work in progress.