Adaptive estimation in functional linear models

Jan Johannes

Université catholique de Louvain

High dimensional and dependent functional data

University of Bristol, UK, September 2012

joint work with H. Cardot, F. Comte and R. Schenk

A glimpse to the essential

- Functional linear model

  Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle \cdot, \cdot \rangle)$ such that

  $$Y = \langle \beta, X \rangle + U \quad \text{with } \mathbb{E}[U \langle X, h \rangle] = 0, \quad \forall h \in \mathbb{H},$$

  with $\beta \in \mathbb{H}$ and U is an error term.

A glimpse to the essential

- Functional linear model

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

$$Y\langle X, h\rangle = \langle\beta, X\rangle\langle X, h\rangle + U\langle X, h\rangle \quad \text{with } \mathbb{E}[U\langle X, h\rangle] = 0,$$

with $\beta \in \mathbb{H}$ and U is an error term.

A glimpse to the essential

- Functional linear model

  Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

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

  with $\beta \in \mathbb{H}$ and U is an error term.

A glimpse to the essential

- Functional linear model

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

$$\langle g, h\rangle := \mathbb{E}[Y\langle X, h\rangle] = \mathbb{E}[\langle \beta, X\rangle\langle X, h\rangle] \qquad \forall h \in \mathbb{H}$$

with $\beta, g \in \mathbb{H}$.

A glimpse to the essential

- Functional linear model

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

$$\langle g, h\rangle := \mathbb{E}[Y\langle X, h\rangle] = \mathbb{E}[\langle\beta, X\rangle\langle X, h\rangle] =: \langle\Gamma\beta, h\rangle, \qquad \forall h \in \mathbb{H}$$

with $\beta, g \in \mathbb{H}$ and covariance operator $\Gamma : \mathbb{H} \to \mathbb{H}$ associated with X.

A glimpse to the essential

- Functional linear model – inverse problem

  Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle \cdot, \cdot \rangle)$ such that

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

  with $\beta, g \in \mathbb{H}$ and covariance operator $\Gamma : \mathbb{H} \to \mathbb{H}$ associated with X.

A glimpse to the essential

- Functional linear model – inverse problem

  Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

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

  with $\beta, g \in \mathbb{H}$ and covariance operator $\Gamma : \mathbb{H} \to \mathbb{H}$ associated with X.
- Observations – $(Y_i, X_i)_{1\leqslant i\leqslant n}$

A glimpse to the essential

- Functional linear model – inverse problem

  Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

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

  with $\beta, g \in \mathbb{H}$ and covariance operator $\Gamma : \mathbb{H} \to \mathbb{H}$ associated with X.

- Observations – $(Y_i, X_i)_{1\leqslant i\leqslant n}$

  Construct estimators $\widehat{g} = \frac{1}{n}\sum_{i=1}^{n} Y_i X_i$ and $\widehat{\Gamma} = \frac{1}{n}\sum_{i=1}^{n}\langle\cdot, X_i\rangle X_i$.

A glimpse to the essential

- Functional linear model – inverse problem

  Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

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

  with $\beta, g \in \mathbb{H}$ and covariance operator $\Gamma : \mathbb{H} \to \mathbb{H}$ associated with X.

- Observations – $(Y_i, X_i)_{1\leqslant i\leqslant n}$ – statistical inverse problem

  Construct estimators $\widehat{g} = \frac{1}{n}\sum_{i=1}^{n} Y_i X_i$ and $\widehat{\Gamma} = \frac{1}{n}\sum_{i=1}^{n}\langle\cdot, X_i\rangle X_i$.

A glimpse to the essential

- Functional linear model – inverse problem

  Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle \cdot, \cdot \rangle)$ such that

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

  with $\beta, g \in \mathbb{H}$ and covariance operator $\Gamma : \mathbb{H} \to \mathbb{H}$ associated with X.

- Observations – $(Y_i, X_i)_{1 \leqslant i \leqslant n}$ – statistical inverse problem

  Construct estimators $\widehat{g} = \frac{1}{n} \sum_{i=1}^{n} Y_i X_i$ and $\widehat{\Gamma} = \frac{1}{n} \sum_{i=1}^{n} \langle \cdot, X_i \rangle X_i$.

- Objective: estimate non parametrically

A glimpse to the essential

- Functional linear model – inverse problem

  Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

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

  with $\beta, g \in \mathbb{H}$ and covariance operator $\Gamma : \mathbb{H} \to \mathbb{H}$ associated with X.

- Observations – $(Y_i, X_i)_{1 \leqslant i \leqslant n}$ – statistical inverse problem

  Construct estimators $\widehat{g} = \frac{1}{n}\sum_{i=1}^{n} Y_i X_i$ and $\widehat{\Gamma} = \frac{1}{n}\sum_{i=1}^{n}\langle\cdot, X_i\rangle X_i$.

- Objective: estimate non parametrically
  - globally: the slope function β as a whole
  - locally: the value $\ell(\beta)$ of a linear functional ℓ evaluated at β

A glimpse to the essential

- Functional linear model – inverse problem

  Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

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

  with $\beta, g \in \mathbb{H}$ and covariance operator $\Gamma : \mathbb{H} \to \mathbb{H}$ associated with X.

- Observations – $(Y_i, X_i)_{1\leqslant i\leqslant n}$ – statistical inverse problem

  Construct estimators $\widehat{g} = \frac{1}{n}\sum_{i=1}^{n} Y_i X_i$ and $\widehat{\Gamma} = \frac{1}{n}\sum_{i=1}^{n}\langle\cdot, X_i\rangle X_i$.

- Objective: estimate non parametrically

  - globally: the slope function β as a whole

Bosq (2000); Ferraty & Vieu (2006); Ramsay & Silverman (2002,2005); Cardot, Ferraty & Sarda (2003); Hall & Horowitz (2007); Crambes, Kneip & Sarda (2009)

  - locally: the value $\ell(\beta)$ of a linear functional ℓ evaluated at β

Cai & Hall (2006).

Minimax-optimal estimation and adaptation

- Measure the performance of any estimator $\widetilde{\beta}$ – maximal risk

$$|\mathcal{L}(\widetilde{\beta}, \beta)|^2$$

Minimax-optimal estimation and adaptation

- Measure the performance of any estimator $\widetilde{\beta}$ – maximal risk

$$\mathbb{E}\big[\,|\mathcal{L}(\widetilde{\beta},\beta)|^2\big]$$

Minimax-optimal estimation and adaptation

- Measure the performance of any estimator $\widetilde{\beta}$ – maximal risk

$$\sup_{\beta \in \mathcal{F}} \mathbb{E}\big[\, |\mathcal{L}(\widetilde{\beta}, \beta)|^2 \big]$$

Minimax-optimal estimation and adaptation

- Measure the performance of any estimator $\widetilde{\beta}$ – maximal risk

$$\sup_{\Gamma \in \mathcal{G}} \sup_{\beta \in \mathcal{F}} \mathbb{E}\big[\,|\mathcal{L}(\widetilde{\beta}, \beta)|^2\big]$$

Minimax-optimal estimation and adaptation

- Measure the performance of any estimator $\widetilde{\beta}$ – maximal risk

$$\sup_{\Gamma \in \mathcal{G}} \sup_{\beta \in \mathcal{F}} \mathbb{E}\big[\, |\mathcal{L}(\widetilde{\beta}, \beta)|^2 \big]$$

Minimax-optimal estimation and adaptation

- Measure the performance of any estimator $\widetilde{\beta}$ – maximal risk

$$\sup_{\Gamma\in\mathcal{G}} \sup_{\beta\in\mathcal{F}} \mathbb{E}\big[\, |\mathcal{L}(\widetilde{\beta},\beta)|^2 \big]$$

- Complexity of functional linear regression – lower bound

$$\inf_{\widetilde{\beta}} \sup_{\Gamma\in\mathcal{G}} \sup_{\beta\in\mathcal{F}} \mathbb{E}\big[\, |\mathcal{L}(\widetilde{\beta},\beta)|^2 \big] \gtrsim R^*[n\,;\mathcal{F},\mathcal{G}].$$

Minimax-optimal estimation and adaptation

- Measure the performance of any estimator $\widetilde{\beta}$ – maximal risk

$$\sup_{\Gamma\in\mathcal{G}}\sup_{\beta\in\mathcal{F}}\mathbb{E}\big[\,|\mathcal{L}(\widetilde{\beta},\beta)|^2\big]$$

- Complexity of functional linear regression – lower bound

$$\inf_{\widetilde{\beta}}\sup_{\Gamma\in\mathcal{G}}\sup_{\beta\in\mathcal{F}}\mathbb{E}\big[\,|\mathcal{L}(\widetilde{\beta},\beta)|^2\big]\gtrsim R^*[n;\mathcal{F},\mathcal{G}].$$

- An estimator $\widehat{\beta}$ is called minimax-optimal if

$$\sup_{\Gamma\in\mathcal{G}}\sup_{\beta\in\mathcal{F}}\mathbb{E}\big[\,|\mathcal{L}(\widehat{\beta},\beta)|^2\big]\lesssim R^*[n;\mathcal{F},\mathcal{G}],$$

Minimax-optimal estimation and adaptation

- Measure the performance of any estimator $\widetilde{\beta}$ – maximal risk

$$\sup_{\Gamma\in\mathcal{G}}\sup_{\beta\in\mathcal{F}}\mathbb{E}\big[\,|\mathcal{L}(\widetilde{\beta},\beta)|^2\big]$$

- Complexity of functional linear regression – lower bound

$$\inf_{\widetilde{\beta}}\sup_{\Gamma\in\mathcal{G}}\sup_{\beta\in\mathcal{F}}\mathbb{E}\big[\,|\mathcal{L}(\widetilde{\beta},\beta)|^2\big]\gtrsim R^*[n\,;\mathcal{F},\mathcal{G}].$$

- An estimator $\widehat{\beta}$ is called minimax-optimal if

$$\sup_{\Gamma\in\mathcal{G}}\sup_{\beta\in\mathcal{F}}\mathbb{E}\big[\,|\mathcal{L}(\widehat{\beta},\beta)|^2\big]\lesssim R^*[n\,;\mathcal{F},\mathcal{G}],$$

and adaptive if $\widehat{\beta}$ depends neither on $\mathcal{F}$ nor $\mathcal{G}$.

Outline

- Methodology
- Background and model assumptions
- Minimax theory
  - Measure of performance
  - Lower bound: global and local risk
  - Minimax-optimal estimation
- Adaptive estimation combining model selection and Lepski's method
  - Adaptive global estimation
  - Adaptive local estimation

Outline

- Methodology
- Background and model assumptions
- Minimax theory
  - Measure of performance
  - Lower bound: global and local risk
  - Minimax-optimal estimation
- Adaptive estimation combining model selection and Lepski's method
  - Adaptive global estimation
  - Adaptive local estimation

Methodology

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

$$Y = \langle\beta, X\rangle + U \quad \text{with } \mathbb{E}[U\langle X, h\rangle] = 0, \ \forall\, h \in \mathbb{H},$$

where $\beta \in \mathbb{H}$.

Methodology

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

$$Y = \langle\beta, X\rangle + U \quad \text{with } \mathbb{E}[U\langle X, h\rangle] = 0, \ \forall\, h \in \mathbb{H},$$

where $\beta = \sum_{j=1}^{m}\langle\beta, f_j\rangle f_j$ for some ONS $\{f_j : 1 \leqslant j \leqslant m\}$ in $\mathbb{H}$;

Methodology

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle \cdot, \cdot \rangle)$ such that

$$Y = \sum_{j=1}^{m} \langle \beta, f_j \rangle \langle f_j, X \rangle + U \quad \text{with } \mathbb{E}[U \langle X, h \rangle] = 0, \forall h \in \mathbb{H}$$

where $\beta = \sum_{j=1}^{m} \langle \beta, f_j \rangle f_j$ for some ONS $\{f_j : 1 \leqslant j \leqslant m\}$ in $\mathbb{H}$;

Methodology

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

$$Y = \sum_{j=1}^{m} \langle \beta, f_j \rangle \langle f_j, X \rangle + U \quad \text{with } \mathbb{E}[U\langle X, h\rangle] = 0, \forall h \in \mathbb{H}$$

where $\beta = \sum_{j=1}^{m} \langle \beta, f_j \rangle f_j$ for some ONS $\{f_j : 1 \leqslant j \leqslant m\}$ in $\mathbb{H}$;

- param. vector $[\beta]_{\underline{m}}$ with entries $[\beta]_j := \langle \beta, f_j \rangle$, $1 \leqslant j \leqslant m$;
- rand. vector $[X]_{\underline{m}}$ with entries $[X]_j := \langle f_j, X \rangle$, $1 \leqslant j \leqslant m$;
- vector $(f)_{\underline{m}}$ with entries $(f)_j := f_j$, $1 \leqslant j \leqslant m$.

Methodology

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle \cdot, \cdot \rangle)$ such that

$$Y = [X]_{\underline{m}}^t [\beta]_{\underline{m}} + U \qquad \text{with } \mathbb{E}(U[X]_{\underline{m}}) \equiv 0,$$

where $\beta = (f)_{\underline{m}}^t [\beta]_{\underline{m}}$ for some ONS $\{f_j : 1 \leqslant j \leqslant m\}$ in $\mathbb{H}$;

- param. vector $[\beta]_{\underline{m}}$ with entries $[\beta]_j := \langle \beta, f_j \rangle, \ 1 \leqslant j \leqslant m$;
- rand. vector $[X]_{\underline{m}}$ with entries $[X]_j := \langle f_j, X \rangle, \ 1 \leqslant j \leqslant m$;
- vector $(f)_{\underline{m}}$ with entries $(f)_j := f_j, \ 1 \leqslant j \leqslant m$.

Methodology

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

$$Y = [X]^t_{\underline{m}}[\beta]_{\underline{m}} + U \quad \text{with } \mathbb{E}(U[X]_{\underline{m}}) \equiv 0,$$

where $\beta = (f)^t_{\underline{m}}[\beta]_{\underline{m}}$ for some ONS $\{f_j : 1 \leqslant j \leqslant m\}$ in $\mathbb{H}$;

- param. vector $[\beta]_{\underline{m}}$ with entries $[\beta]_j := \langle \beta, f_j \rangle$, $1 \leqslant j \leqslant m$;
- rand. vector $[X]_{\underline{m}}$ with entries $[X]_j := \langle f_j, X \rangle$, $1 \leqslant j \leqslant m$;
- vector $(f)_{\underline{m}}$ with entries $(f)_j := f_j$, $1 \leqslant j \leqslant m$.

▶ Normal equation

$$Y[X]_{\underline{m}} = [X]_{\underline{m}}[X]^t_{\underline{m}}[\beta]_{\underline{m}} + U[X]_{\underline{m}}$$

Methodology

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

$$Y = [X]_{\underline{m}}^t [\beta]_{\underline{m}} + U \qquad \text{with } \mathbb{E}(U[X]_{\underline{m}}) \equiv 0,$$

where $\beta = (f)_{\underline{m}}^t [\beta]_{\underline{m}}$ for some ONS $\{f_j : 1 \leqslant j \leqslant m\}$ in $\mathbb{H}$;

- param. vector $[\beta]_{\underline{m}}$ with entries $[\beta]_j := \langle \beta, f_j \rangle$, $1 \leqslant j \leqslant m$;
- rand. vector $[X]_{\underline{m}}$ with entries $[X]_j := \langle f_j, X \rangle$, $1 \leqslant j \leqslant m$;
- vector $(f)_{\underline{m}}$ with entries $(f)_j := f_j$, $1 \leqslant j \leqslant m$.

▸ Normal equation

$$\mathbb{E}(Y[X]_{\underline{m}}) = \mathbb{E}([X]_{\underline{m}}[X]_{\underline{m}}^t)[\beta]_{\underline{m}}$$

Methodology

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

$$Y = [X]^t_{\underline{m}}[\beta]_{\underline{m}} + U \qquad \text{with } \mathbb{E}(U[X]_{\underline{m}}) \equiv 0,$$

where $\beta = (f)^t_{\underline{m}}[\beta]_{\underline{m}}$ for some ONS $\{f_j : 1 \leqslant j \leqslant m\}$ in $\mathbb{H}$;

- param. vector $[\beta]_{\underline{m}}$ with entries $[\beta]_j := \langle \beta, f_j \rangle,\ 1 \leqslant j \leqslant m$;
- rand. vector $[X]_{\underline{m}}$ with entries $[X]_j := \langle f_j, X \rangle,\ 1 \leqslant j \leqslant m$;
- vector $(f)_{\underline{m}}$ with entries $(f)_j := f_j,\ 1 \leqslant j \leqslant m$.

▸ Normal equation

$$[g]_{\underline{m}} := \mathbb{E}(Y[X]_{\underline{m}}) = \mathbb{E}([X]_{\underline{m}}[X]^t_{\underline{m}})[\beta]_{\underline{m}} =: [\Gamma]_{\underline{m}}[\beta]_{\underline{m}}$$

Methodology

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

$$Y = [X]_{\underline{m}}^t[\beta]_{\underline{m}} + U \qquad \text{with } \mathbb{E}(U[X]_{\underline{m}}) \equiv 0,$$

where $\beta = (f)_{\underline{m}}^t[\beta]_{\underline{m}}$ for some ONS $\{f_j : 1 \leqslant j \leqslant m\}$ in $\mathbb{H}$;

- param. vector $[\beta]_{\underline{m}}$ with entries $[\beta]_j := \langle\beta, f_j\rangle,\ 1 \leqslant j \leqslant m$;
- rand. vector $[X]_{\underline{m}}$ with entries $[X]_j := \langle f_j, X\rangle,\ 1 \leqslant j \leqslant m$;
- vector $(f)_{\underline{m}}$ with entries $(f)_j := f_j,\ 1 \leqslant j \leqslant m$.

- Normal equation

$$[g]_{\underline{m}} := \mathbb{E}(Y[X]_{\underline{m}}) = \mathbb{E}([X]_{\underline{m}}[X]_{\underline{m}}^t)[\beta]_{\underline{m}} =: [\Gamma]_{\underline{m}}[\beta]_{\underline{m}}$$

- Solution of the normal equation

$$\beta = (f)_{\underline{m}}^t[\Gamma]_{\underline{m}}^{-1}[g]_{\underline{m}}$$

Methodology

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle \cdot, \cdot \rangle)$ such that

$$Y = [X]_{\underline{m}}^t [\beta]_{\underline{m}} + U \qquad \text{with } \mathbb{E}(U[X]_{\underline{m}}) \equiv 0,$$

where $\beta = (f)_{\underline{m}}^t [\beta]_{\underline{m}}$ for some ONS $\{f_j : 1 \leqslant j \leqslant m\}$ in $\mathbb{H}$.

- Solution of the normal equation: $\beta = (f)_{\underline{m}}^t [\Gamma]_{\underline{m}}^{-1} [g]_{\underline{m}}$ with

  $[g]_{\underline{m}} := \mathbb{E}(Y[X]_{\underline{m}}) \quad$ and $\quad [\Gamma]_{\underline{m}} := \mathbb{E}([X]_{\underline{m}} [X]_{\underline{m}}^t)$.

Methodology

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

$$Y = [X]_{\underline{m}}^t[\beta]_{\underline{m}} + U \qquad \text{with } \mathbb{E}(U[X]_{\underline{m}}) \equiv 0,$$

where $\beta = (f)_{\underline{m}}^t[\beta]_{\underline{m}}$ for some ONS $\{f_j : 1 \leqslant j \leqslant m\}$ in $\mathbb{H}$.

- Solution of the normal equation: $\beta = (f)_{\underline{m}}^t[\Gamma]_{\underline{m}}^{-1}[g]_{\underline{m}}$ with

  $[g]_{\underline{m}} := \mathbb{E}(Y[X]_{\underline{m}})$ and $[\Gamma]_{\underline{m}} := \mathbb{E}([X]_{\underline{m}}[X]_{\underline{m}}^t)$.

- Suppose an *n*-sample $(Y_i, X_i)_{1 \leqslant i \leqslant n}$

Methodology

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

$$Y = [X]^t_{\underline{m}}[\beta]_{\underline{m}} + U \quad \text{with } \mathbb{E}(U[X]_{\underline{m}}) \equiv 0,$$

where $\beta = (f)^t_{\underline{m}}[\beta]_{\underline{m}}$ for some ONS $\{f_j : 1 \leqslant j \leqslant m\}$ in $\mathbb{H}$.

- Solution of the normal equation: $\beta = (f)^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[g]_{\underline{m}}$ with

  $[g]_{\underline{m}} := \mathbb{E}(Y[X]_{\underline{m}})$ and $[\Gamma]_{\underline{m}} := \mathbb{E}([X]_{\underline{m}}[X]^t_{\underline{m}})$.

- Suppose an n-sample $(Y_i, X_i)_{1\leqslant i\leqslant n}$, then define

  $[\widehat{g}]_{\underline{m}} := \frac{1}{n}\sum_{i=1}^n (Y_i[X_i]_{\underline{m}})$ and $[\widehat{\Gamma}]_{\underline{m}} := \frac{1}{n}\sum_{i=1}^n([X_i]_{\underline{m}}[X_i]^t_{\underline{m}})$.

Methodology

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

$$Y = [X]^t_{\underline{m}}[\beta]_{\underline{m}} + U \quad \text{with } \mathbb{E}(U[X]_{\underline{m}}) \equiv 0,$$

where $\beta = (f)^t_{\underline{m}}[\beta]_{\underline{m}}$ for some ONS $\{f_j : 1 \leqslant j \leqslant m\}$ in $\mathbb{H}$.

- Solution of the normal equation: $\beta = (f)^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[g]_{\underline{m}}$ with

  $[g]_{\underline{m}} := \mathbb{E}(Y[X]_{\underline{m}})$ and $[\Gamma]_{\underline{m}} := \mathbb{E}([X]_{\underline{m}}[X]^t_{\underline{m}})$.

- Suppose an n-sample $(Y_i, X_i)_{1\leqslant i\leqslant n}$, then define

  $[\widehat{g}]_{\underline{m}} := \frac{1}{n}\sum_{i=1}^n (Y_i[X_i]_{\underline{m}})$ and $[\widehat{\Gamma}]_{\underline{m}} := \frac{1}{n}\sum_{i=1}^n ([X_i]_{\underline{m}}[X_i]^t_{\underline{m}})$.

- Estimate the solution $\widehat{\beta} := (f)^t_{\underline{m}}[\widehat{\Gamma}]^{-1}_{\underline{m}}[\widehat{g}]_{\underline{m}}$.

Methodology

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

$$Y = [X]^t_{\underline{m}}[\beta]_{\underline{m}} + U \quad \text{with } \mathbb{E}(U[X]_{\underline{m}}) \equiv 0,$$

where $\beta = (f)^t_{\underline{m}}[\beta]_{\underline{m}}$ for some ONS $\{f_j : 1 \leqslant j \leqslant m\}$ in $\mathbb{H}$.

- Solution of the normal equation: $\beta = (f)^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[g]_{\underline{m}}$ with

  $[g]_{\underline{m}} := \mathbb{E}(Y[X]_{\underline{m}})$ and $[\Gamma]_{\underline{m}} := \mathbb{E}([X]_{\underline{m}}[X]^t_{\underline{m}})$.

- Suppose an n-sample $(Y_i, X_i)_{1\leqslant i\leqslant n}$, then define

  $[\widehat{g}]_{\underline{m}} := \frac{1}{n}\sum_{i=1}^n (Y_i[X_i]_{\underline{m}})$ and $[\widehat{\Gamma}]_{\underline{m}} := \frac{1}{n}\sum_{i=1}^n ([X_i]_{\underline{m}}[X_i]^t_{\underline{m}})$.

- Estimate the solution $\widehat{\beta} := (f)^t_{\underline{m}}[\widehat{\Gamma}]^{-1}_{\underline{m}}[\widehat{g}]_{\underline{m}}$.

  Under mild assumptions $[\widehat{\beta}]_{\underline{m}}$ is consistent, asymptotic normal!

Methodology

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle \cdot, \cdot \rangle)$ such that

$$Y = [X]_{\underline{m}}^t [\beta]_{\underline{m}} + U \quad \text{with } \mathbb{E}(U[X]_{\underline{m}}) \equiv 0,$$

where $\beta = (f)_{\underline{m}}^t [\beta]_{\underline{m}}$ for some ONS $\{f_j : 1 \leqslant j \leqslant m\}$ in $\mathbb{H}$.

- Solution of the normal equation: $\beta = (f)_{\underline{m}}^t [\Gamma]_{\underline{m}}^{-1} [g]_{\underline{m}}$ with

  $[g]_{\underline{m}} := \mathbb{E}(Y[X]_{\underline{m}})$ and $[\Gamma]_{\underline{m}} := \mathbb{E}([X]_{\underline{m}}[X]_{\underline{m}}^t)$.

- Suppose an n-sample $(Y_i, X_i)_{1 \leqslant i \leqslant n}$, then define

  $[\widehat{g}]_{\underline{m}} := \frac{1}{n} \sum_{i=1}^n (Y_i [X_i]_{\underline{m}})$ and $[\widehat{\Gamma}]_{\underline{m}} := \frac{1}{n} \sum_{i=1}^n ([X_i]_{\underline{m}} [X_i]_{\underline{m}}^t)$.

- Estimate the solution $\widehat{\beta} := (f)_{\underline{m}}^t [\widehat{\Gamma}]_{\underline{m}}^{-1} [\widehat{g}]_{\underline{m}}$.

  Under mild assumptions $[\widehat{\beta}]_{\underline{m}}$ is consistent, asymptotic normal!

In general an infinite ONS $\{f_j\}_{j \in \mathbb{N}}$ is necessary to develop β!

Methodology

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

$$Y = \langle\beta, X\rangle + U \quad \text{with } \mathbb{E}(U\langle X, h\rangle) = 0, \; \forall\, h \in \mathbb{H},$$

where $\beta \in \mathbb{H}$.

Methodology

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle \cdot, \cdot \rangle)$ such that

$$Y = \langle \beta, X \rangle + U \quad \text{with } \mathbb{E}(U\langle X, h \rangle) = 0, \; \forall\, h \in \mathbb{H},$$

where $\beta \in \mathbb{H}$. Consider an infinite ONB $\{f_j\}_{j \in \mathbb{N}}$ and $m \in \mathbb{N}$.

Methodology: dimension reduction

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

$$Y = \langle \beta, X\rangle + U \quad \text{with } \mathbb{E}(U\langle X, h\rangle) = 0,\ \forall\, h \in \mathbb{H},$$

where $\beta \in \mathbb{H}$. Consider an infinite ONB $\{f_j\}_{j\in\mathbb{N}}$ and $m \in \mathbb{N}$.

- Solution of the normal equation: $\beta_m = (f)^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[g]_{\underline{m}}$ with

  $[g]_{\underline{m}} = \mathbb{E}(Y[X]_{\underline{m}})$ and $[\Gamma]_{\underline{m}} = \mathbb{E}([X]_{\underline{m}}[X]^t_{\underline{m}})$.

Methodology: dimension reduction

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

$$Y = \langle\beta, X\rangle + U \quad \text{with } \mathbb{E}(U\langle X, h\rangle) = 0,\ \forall\, h \in \mathbb{H},$$

where $\beta \in \mathbb{H}$. Consider an infinite ONB $\{f_j\}_{j\in\mathbb{N}}$ and $m \in \mathbb{N}$.

- Solution of the normal equation: $\beta_m = (f)^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[g]_{\underline{m}}$ with

  $[g]_{\underline{m}} = \mathbb{E}(Y[X]_{\underline{m}})$ and $[\Gamma]_{\underline{m}} = \mathbb{E}([X]_{\underline{m}}[X]^t_{\underline{m}})$.

- Suppose an n-sample $(Y_i, X_i)_{1\leqslant i\leqslant n}$

Methodology: dimension reduction

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle \cdot, \cdot \rangle)$ such that

$$Y = \langle \beta, X \rangle + U \quad \text{with } \mathbb{E}(U\langle X, h\rangle) = 0,\ \forall\, h \in \mathbb{H},$$

where $\beta \in \mathbb{H}$. Consider an infinite ONB $\{f_j\}_{j\in\mathbb{N}}$ and $m \in \mathbb{N}$.

- Solution of the normal equation: $\beta_m = (f)^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[g]_{\underline{m}}$ with

  $[g]_{\underline{m}} = \mathbb{E}(Y[X]_{\underline{m}})$ and $[\Gamma]_{\underline{m}} = \mathbb{E}([X]_{\underline{m}}[X]^t_{\underline{m}})$.

- Suppose an n-sample $(Y_i, X_i)_{1\leqslant i\leqslant n}$, then consider again

  $[\widehat{g}]_{\underline{m}} = \frac{1}{n}\sum_{i=1}^n (Y_i[X_i]_{\underline{m}})$ and $[\widehat{\Gamma}]_{\underline{m}} = \frac{1}{n}\sum_{i=1}^n ([X_i]_{\underline{m}}[X_i]^t_{\underline{m}})$.

Methodology: dimension reduction

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

$$Y = \langle \beta, X\rangle + U \quad \text{with } \mathbb{E}(U\langle X, h\rangle) = 0,\ \forall\, h \in \mathbb{H},$$

where $\beta \in \mathbb{H}$. Consider an infinite ONB $\{f_j\}_{j\in\mathbb{N}}$ and $m \in \mathbb{N}$.

- Solution of the normal equation: $\beta_m = (f)^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[g]_{\underline{m}}$ with

  $$[g]_{\underline{m}} = \mathbb{E}(Y[X]_{\underline{m}}) \quad \text{and} \quad [\Gamma]_{\underline{m}} = \mathbb{E}([X]_{\underline{m}}[X]^t_{\underline{m}}).$$

- Suppose an n-sample $(Y_i, X_i)_{1\leqslant i\leqslant n}$, then consider again

  $$[\widehat{g}]_{\underline{m}} = \tfrac{1}{n}\textstyle\sum_{i=1}^n (Y_i[X_i]_{\underline{m}}) \quad \text{and} \quad [\widehat{\Gamma}]_{\underline{m}} = \tfrac{1}{n}\textstyle\sum_{i=1}^n ([X_i]_{\underline{m}}[X_i]^t_{\underline{m}}).$$

- Estimate globally β with $\widehat{\beta}_m := (f)^t_{\underline{m}}[\widehat{\Gamma}]^{-1}_{\underline{m}}[\widehat{g}]_{\underline{m}}$.

Methodology: dimension reduction & thresholding

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle \cdot, \cdot \rangle)$ such that

$$Y = \langle \beta, X \rangle + U \quad \text{with } \mathbb{E}(U\langle X, h \rangle) = 0, \; \forall\, h \in \mathbb{H},$$

where $\beta \in \mathbb{H}$. Consider an infinite ONB $\{f_j\}_{j\in\mathbb{N}}$ and $m \in \mathbb{N}$.

- Solution of the normal equation: $\beta_m = (f)^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[g]_{\underline{m}}$ with

  $[g]_{\underline{m}} = \mathbb{E}(Y[X]_{\underline{m}})$ and $[\Gamma]_{\underline{m}} = \mathbb{E}([X]_{\underline{m}}[X]^t_{\underline{m}})$.

- Suppose an n-sample $(Y_i, X_i)_{1 \leqslant i \leqslant n}$, then consider again

  $[\widehat{g}]_{\underline{m}} = \frac{1}{n}\sum_{i=1}^n (Y_i[X_i]_{\underline{m}})$ and $[\widehat{\Gamma}]_{\underline{m}} = \frac{1}{n}\sum_{i=1}^n ([X_i]_{\underline{m}}[X_i]^t_{\underline{m}})$.

- Estimate globally β with $\widehat{\beta}_m := (f)^t_{\underline{m}}[\widehat{\Gamma}]^{-1}_{\underline{m}}[\widehat{g}]_{\underline{m}}\, \mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{m}}\|_s \leqslant n\}}$.

Methodology: dimension reduction & thresholding

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle \cdot, \cdot \rangle)$ such that

$$Y = \langle \beta, X \rangle + U \quad \text{with } \mathbb{E}(U\langle X, h\rangle) = 0,\ \forall\, h \in \mathbb{H},$$

where $\beta \in \mathbb{H}$. Consider an infinite ONB $\{f_j\}_{j\in\mathbb{N}}$ and $m \in \mathbb{N}$.

- Solution of the normal equation: $\beta_m = (f)^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[g]_{\underline{m}}$ with

  $[g]_{\underline{m}} = \mathbb{E}(Y[X]_{\underline{m}}) \quad \text{and} \quad [\Gamma]_{\underline{m}} = \mathbb{E}([X]_{\underline{m}}[X]^t_{\underline{m}}).$

- Suppose an n-sample $(Y_i, X_i)_{1\leqslant i\leqslant n}$, then consider again

  $[\widehat{g}]_{\underline{m}} = \frac{1}{n}\sum_{i=1}^n(Y_i[X_i]_{\underline{m}}) \quad \text{and} \quad [\widehat{\Gamma}]_{\underline{m}} = \frac{1}{n}\sum_{i=1}^n([X_i]_{\underline{m}}[X_i]^t_{\underline{m}}).$

- Estimate globally β with $\widehat{\beta}_m := (f)^t_{\underline{m}}[\widehat{\Gamma}]^{-1}_{\underline{m}}[\widehat{g}]_{\underline{m}}\, \mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{m}}\|_s \leqslant n\}}$.
- Estimate locally $\ell(\beta)$ with $\widehat{\ell}_m := \ell(\widehat{\beta}_m)$.

Methodology: dimension reduction & thresholding

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

$$Y = \langle \beta, X\rangle + U \quad \text{with } \mathbb{E}(U\langle X, h\rangle) = 0,\ \forall\, h \in \mathbb{H},$$

where $\beta \in \mathbb{H}$. Consider an infinite ONB $\{f_j\}_{j\in\mathbb{N}}$ and $m \in \mathbb{N}$.

- Solution of the normal equation: $\beta_m = (f)^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[g]_{\underline{m}}$ with

  $$[g]_{\underline{m}} = \mathbb{E}(Y[X]_{\underline{m}}) \quad \text{and} \quad [\Gamma]_{\underline{m}} = \mathbb{E}([X]_{\underline{m}}[X]^t_{\underline{m}}).$$

- Suppose an n-sample $(Y_i, X_i)_{1\leqslant i\leqslant n}$, then consider again

  $$[\widehat{g}]_{\underline{m}} = \tfrac{1}{n}\textstyle\sum_{i=1}^n (Y_i[X_i]_{\underline{m}}) \quad \text{and} \quad [\widehat{\Gamma}]_{\underline{m}} = \tfrac{1}{n}\textstyle\sum_{i=1}^n ([X_i]_{\underline{m}}[X_i]^t_{\underline{m}}).$$

- Estimate globally β with $\widehat{\beta}_m := (f)^t_{\underline{m}}[\widehat{\Gamma}]^{-1}_{\underline{m}}[\widehat{g}]_{\underline{m}}\, \mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{m}}\|_s \leqslant n\}}$.
- Estimate locally $\ell(\beta)$ with $\widehat{\ell}_m := \ell(\widehat{\beta}_m) = [\ell]^t_{\underline{m}}[\widehat{\Gamma}]^{-1}_{\underline{m}}[\widehat{g}]_{\underline{m}}\, \mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{m}}\|_s \leqslant n\}}$.

Methodology: adaptive estimation

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle \cdot, \cdot \rangle)$ such that

$$Y = \langle \beta, X \rangle + U \quad \text{ with } \mathbb{E}(U\langle X, h \rangle) = 0, \ \forall\, h \in \mathbb{H},$$

where $\beta \in \mathbb{H}$. Consider an infinite ONB $\{f_j\}_{j \in \mathbb{N}}$ and $m \in \mathbb{N}$.

- Thresholded projection estimator: $\widehat{\beta}_m := (f)^t_{\underline{m}}[\widehat{\Gamma}]^{-1}_{\underline{m}}[\widehat{g}]_{\underline{m}} \, \mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{m}}\|_s \leqslant n\}}$

Methodology: adaptive estimation

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

$$Y = \langle \beta, X\rangle + U \quad \text{with } \mathbb{E}(U\langle X, h\rangle) = 0,\ \forall\, h \in \mathbb{H},$$

where $\beta \in \mathbb{H}$. Consider an infinite ONB $\{f_j\}_{j\in\mathbb{N}}$ and $m \in \mathbb{N}$.

- Thresholded projection estimator: $\widehat{\beta}_m := (f)^t_{\underline{m}}[\widehat{\Gamma}]^{-1}_{\underline{m}}[\widehat{g}]_{\underline{m}}\, \mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{m}}\|_s \leqslant n\}}$
- Select m by using a penalized minimum contrast criterion

Methodology: adaptive estimation

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

$$Y = \langle \beta, X\rangle + U \quad \text{with } \mathbb{E}(U\langle X, h\rangle) = 0,\ \forall\, h \in \mathbb{H},$$

where $\beta \in \mathbb{H}$. Consider an infinite ONB $\{f_j\}_{j\in\mathbb{N}}$.

- Thresholded projection estimator: $\widehat{\beta}_{\widehat{m}} := (f)^t_{\underline{\widehat{m}}}[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}[\widehat{g}]_{\underline{\widehat{m}}}\, \mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}\|_s \leqslant n\}}$
- Select $\widehat{m}$ by using a penalized minimum contrast criterion

$$\widehat{m} := \mathop{\arg\min}_{1\leqslant m\leqslant \widehat{M}} \left\{ \text{Contrast}_m + \widehat{\text{pen}}_m \right\}.$$

Methodology: adaptive estimation

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

$$Y = \langle \beta, X\rangle + U \qquad \text{with } \mathbb{E}(U\langle X, h\rangle) = 0,\ \forall\, h \in \mathbb{H},$$

where $\beta \in \mathbb{H}$. Consider an infinite ONB $\{f_j\}_{j\in\mathbb{N}}$.

- Thresholded projection estimator: $\widehat{\beta}_{\widehat{m}} := (f)^t_{\underline{\widehat{m}}}[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}[\widehat{g}]_{\underline{\widehat{m}}}\,\mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}\|_s \leqslant n\}}$
- Select $\widehat{m}$ by using a penalized minimum contrast criterion

$$\widehat{m} := \mathop{\arg\min}_{1\leqslant m\leqslant \widehat{M}} \left\{ \text{Contrast}_m + \widehat{\mathrm{pen}}_m \right\}.$$

- Measure its performance: $\sup_{\Gamma\in\mathcal{G}} \sup_{\beta\in\mathcal{F}} \mathbb{E}\big[\,|\mathcal{L}(\widehat{\beta}_{\widehat{m}}, \beta)|^2\big]$

Methodology: adaptive estimation

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle \cdot, \cdot \rangle)$ such that

$$Y = \langle \beta, X \rangle + U \quad \text{with } \mathbb{E}(U\langle X, h\rangle) = 0, \ \forall\, h \in \mathbb{H},$$

where $\beta \in \mathbb{H}$. Consider an infinite ONB $\{f_j\}_{j \in \mathbb{N}}$.

- Thresholded projection estimator: $\widehat{\beta}_{\widehat{m}} := (f)^t_{\underline{\widehat{m}}} [\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}} [\widehat{g}]_{\underline{\widehat{m}}} \, \mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}\|_s \leqslant n\}}$
- Select $\widehat{m}$ by using a penalized minimum contrast criterion

$$\widehat{m} := \arg\min_{1 \leqslant m \leqslant \widehat{M}} \left\{ \text{Contrast}_m + \widehat{\mathrm{pen}}_m \right\}.$$

- Measure its performance: $\sup_{\Gamma \in \mathcal{G}} \sup_{\beta \in \mathcal{F}} \mathbb{E}\big[\, |\mathcal{L}(\widehat{\beta}_{\widehat{m}}, \beta)|^2 \big]$ then

$$\text{Contrast}_m := \sup_{m \leqslant k \leqslant \widehat{M}} \left\{ |\mathcal{L}(\widehat{\beta}_k, \widehat{\beta}_m)|^2 - \widehat{\mathrm{pen}}_k \right\}$$

which is inspired by the recent work of Goldenshluger and Lepski [2011].

Methodology: adaptive estimation

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

$$Y = \langle \beta, X\rangle + U \quad \text{with } \mathbb{E}(U\langle X, h\rangle) = 0,\ \forall\, h \in \mathbb{H},$$

where $\beta \in \mathbb{H}$. Consider an infinite ONB $\{f_j\}_{j\in\mathbb{N}}$.

- Thresholded projection estimator: $\widehat{\beta}_{\widehat{m}} := (f)^t_{\underline{\widehat{m}}}[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}[\widehat{g}]_{\underline{\widehat{m}}}\, \mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}\|_s \leqslant n\}}$
- Select $\widehat{m}$ by using a penalized minimum contrast criterion

$$\widehat{m} := \mathop{\arg\min}_{1\leqslant m\leqslant \widehat{M}} \left\{\text{Contrast}_m + \widehat{\text{pen}}_m\right\}.$$

- Measure its performance: $\sup_{\Gamma\in\mathcal{G}} \sup_{\beta\in\mathcal{F}} \mathbb{E}\big[\,|\mathcal{L}(\widehat{\beta}_{\widehat{m}}, \beta)|^2\big]$ then

$$\text{Contrast}_m := \sup_{m\leqslant k\leqslant \widehat{M}} \left\{|\mathcal{L}(\widehat{\beta}_k, \widehat{\beta}_m)|^2 - \widehat{\text{pen}}_k\right\}$$

which is inspired by the recent work of Goldenshluger and Lepski [2011].

Outline

- Methodology
- Background and model assumptions
- Minimax theory
  - Measure of performance
  - Lower bound: global and local risk
  - Minimax-optimal estimation
- Adaptive estimation combining model selection and Lepski's method
  - Adaptive global estimation
  - Adaptive local estimation

Basic model assumptions

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

$$Y = \langle \beta, X \rangle + U \quad \text{with } \mathbb{E}[U\langle X, h\rangle] = 0, \ \forall\, h \in \mathbb{H},$$

where $\beta \in \mathbb{H}$ is an unknown slope function,

Basic model assumptions

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle \cdot, \cdot \rangle)$ such that

$$Y = \langle \beta, X \rangle + U \quad \text{with } \mathbb{E}[U \langle X, h \rangle] = 0, \ \forall\, h \in \mathbb{H},$$

where $\beta \in \mathbb{H}$ is an unknown slope function,

Assumption A1.

- X is a centered Gaussian regressor with $\mathbb{E}\|X\|^2 < \infty$,
- U is a centered Gaussian error term with $\mathbb{E}U^2 < \infty$,
- X and U are independent ($X \perp\!\!\!\perp U$).

Basic model assumptions

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle \cdot, \cdot \rangle)$ such that

$$Y = \langle \beta, X \rangle + U \qquad \text{with } \mathbb{E}[U\langle X, h\rangle] = 0, \ \forall\, h \in \mathbb{H},$$

where $\beta \in \mathbb{H}$ is an unknown slope function,

Assumption A1.

- X is a centered Gaussian regressor with $\mathbb{E}\|X\|^2 < \infty$,
- U is a centered Gaussian error term with $\mathbb{E}U^2 < \infty$,
- X and U are independent ($X \perp\!\!\!\perp U$).

Define the covariance operator Γ associated with X

- $\Gamma : \mathbb{H} \to \mathbb{H}$ with $f \mapsto \langle \Gamma f, h\rangle := \mathbb{E}[\langle f, X\rangle\langle X, h\rangle], \quad \forall\, h \in \mathbb{H}$,

which is positive semi-definite and nuclear.

Basic model assumptions

Let $Y \in \mathbb{R}$ be a r.v. and X be a random function in $(\mathbb{H}, \langle \cdot, \cdot \rangle)$ such that

$$Y = \langle \beta, X \rangle + U \quad \text{with } \mathbb{E}[U\langle X, h\rangle] = 0, \ \forall\, h \in \mathbb{H},$$

where $\beta \in \mathbb{H}$ is an unknown slope function,

Assumption A1.

- X is a centered Gaussian regressor with $\mathbb{E}\|X\|^2 < \infty$,
- U is a centered Gaussian error term with $\mathbb{E}U^2 < \infty$,
- X and U are independent ($X \perp\!\!\!\perp U$).

Define the covariance operator Γ associated with X

- $\Gamma : \mathbb{H} \to \mathbb{H}$ with $f \mapsto \langle \Gamma f, h\rangle := \mathbb{E}[\langle f, X\rangle \langle X, h\rangle], \quad \forall\, h \in \mathbb{H}$,

which is positive semi-definite and nuclear. Rewrite the model equation

- $\langle g, h\rangle := \mathbb{E}[Y\langle X, h\rangle] = \mathbb{E}[\langle \beta, X\rangle\langle X, h\rangle] =: \langle \Gamma\beta, h\rangle, \quad \forall\, h \in \mathbb{H}$.

Ill-posed inverse problem

Let Y be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

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

where $\beta, g \in \mathbb{H}$ and $\Gamma : \mathbb{H} \to \mathbb{H}$ is nuclear.

Ill-posed inverse problem

Let Y be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

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

where $\beta, g \in \mathbb{H}$ and $\Gamma : \mathbb{H} \to \mathbb{H}$ is nuclear.

Hadamard (1932): An inverse problem $g = \Gamma\beta$ is well posed, if:

- a solution β exists,
- the solution β is unique and
- the solution β depends continuously on g.

Ill-posed inverse problem

Let Y be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

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

where $\beta, g \in \mathbb{H}$ and $\Gamma : \mathbb{H} \to \mathbb{H}$ is nuclear.

Hadamard (1932): An inverse problem $g = \Gamma\beta$ is well posed, if:

- a solution β exists,
- the solution β is unique and
- the solution β depends continuously on g.

Engel, Hanke & Neubauer (2000); Cardot, Ferraty & Sarda (2003)

Ill-posed inverse problem

Let Y be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

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

where $\beta, g \in \mathbb{H}$ and $\Gamma : \mathbb{H} \to \mathbb{H}$ is nuclear.

Hadamard (1932): An inverse problem $g = \Gamma\beta$ is well posed, if:

- a solution β exists,
- the solution β is unique and
- the solution β depends continuously on g.

Assumption A2. There exists an unique solution of the equation $g = \Gamma\beta$.

Ill-posed inverse problem

Let Y be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

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

where $\beta, g \in \mathbb{H}$ and $\Gamma : \mathbb{H} \to \mathbb{H}$ is nuclear.

Hadamard (1932): An inverse problem $g = \Gamma\beta$ is well posed, if:

- a solution β exists,
- the solution β is unique and
- the solution β depends continuously on g.

Assumption A2. There exists an unique solution of the equation $g = \Gamma\beta$.

- Since the covariance operator Γ is nuclear, its (generalized) inverse Γ^{-1} is not continuous as long as the range of Γ is infinite dimensional.

Ill-posed inverse problem

Let Y be a r.v. and X be a random function in $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ such that

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

where $\beta, g \in \mathbb{H}$ and $\Gamma : \mathbb{H} \to \mathbb{H}$ is nuclear.

Hadamard (1932): An inverse problem $g = \Gamma\beta$ is well posed, if:

- a solution β exists,
- the solution β is unique and
- the solution β depends continuously on g.

Assumption A2. There exists an unique solution of the equation $g = \Gamma\beta$.

- Since the covariance operator Γ is nuclear, its (generalized) inverse Γ^{-1} is not continuous as long as the range of Γ is infinite dimensional.

Reconstruction of the slope function β is an ill-posed inverse problem.

Minimal regularity conditions

Let $\{f_j\}_{j\geqslant 1}$ be a pre-specified ONB in $\mathbb{H}$ and $[h]_j := \langle h, f_j \rangle$ for all $h \in \mathbb{H}$.

Minimal regularity conditions

Let $\{f_j\}_{j\geqslant 1}$ be a pre-specified ONB in $\mathbb{H}$ and $[h]_j := \langle h, f_j\rangle$ for all $h \in \mathbb{H}$.

Given a strictly positive sequence of weights $w := (w_j)_{j\geqslant 1}$ define

- the weigthed norm $\|h\|_w^2 := \sum_{j\geqslant 1} w_j\, [h]_j^2$ for $h \in \mathbb{H}$;

Minimal regularity conditions

Let $\{f_j\}_{j\geqslant 1}$ be a pre-specified ONB in $\mathbb{H}$ and $[h]_j := \langle h, f_j \rangle$ for all $h \in \mathbb{H}$.

Given a strictly positive sequence of weights $w := (w_j)_{j\geqslant 1}$ define

- the weigthed norm $\|h\|_w^2 := \sum_{j\geqslant 1} w_j\,[h]_j^2$ for $h \in \mathbb{H}$;
- the completion $\mathcal{F}_w$ of $\mathbb{H}$ with regard to $\|\cdot\|_w$;

Minimal regularity conditions

Let $\{f_j\}_{j\geqslant 1}$ be a pre-specified ONB in $\mathbb{H}$ and $[h]_j := \langle h, f_j\rangle$ for all $h \in \mathbb{H}$.

Given a strictly positive sequence of weights $w := (w_j)_{j\geqslant 1}$ define

- the weigthed norm $\|h\|_w^2 := \sum_{j\geqslant 1} w_j\,[h]_j^2$ for $h \in \mathbb{H}$;
- the completion $\mathcal{F}_w$ of $\mathbb{H}$ with regard to $\|\cdot\|_w$;
- the ellipsoid $\mathcal{F}_w^r := \left\{h \in \mathcal{F}_w : \|h\|_w^2 \leqslant r\right\}$ with radius $r > 0$.

Minimal regularity conditions

Let $\{f_j\}_{j\geqslant 1}$ be a pre-specified ONB in $\mathbb{H}$ and $[h]_j := \langle h, f_j\rangle$ for all $h \in \mathbb{H}$.

Given a strictly positive sequence of weights $w := (w_j)_{j\geqslant 1}$ define

- the weigthed norm $\|h\|_w^2 := \sum_{j\geqslant 1} w_j\,[h]_j^2$ for $h \in \mathbb{H}$;
- the completion $\mathcal{F}_w$ of $\mathbb{H}$ with regard to $\|\cdot\|_w$;
- the ellipsoid $\mathcal{F}_w^r := \left\{h \in \mathcal{F}_w : \|h\|_w^2 \leqslant r\right\}$ with radius $r > 0$.

Assumption A3. (Minimal regularity assumptions)

(i) The slope function β belongs to $\mathcal{F}_b^\rho$ for some non-decreasing, unbounded sequence of weights $b := (b_j)_{j\geqslant 1}$ with $b_1 = 1$, and $\rho > 0$.

Minimal regularity conditions

Let $\{f_j\}_{j\geqslant 1}$ be a pre-specified ONB in $\mathbb{H}$ and $[h]_j := \langle h, f_j \rangle$ for all $h \in \mathbb{H}$.

Minimal regularity conditions

Let $\{f_j\}_{j\geqslant 1}$ be a pre-specified ONB in $\mathbb{H}$ and $[h]_j := \langle h, f_j \rangle$ for all $h \in \mathbb{H}$.

Given a sequence of weights $w := (w_j)_{j\geqslant 1} > 0$ and a constant $r \geqslant 1$ define

- the set $\mathcal{G}$ of all strictly positive nuclear operators defined on $\mathbb{H}$ and

Minimal regularity conditions

Let $\{f_j\}_{j\geqslant 1}$ be a pre-specified ONB in $\mathbb{H}$ and $[h]_j := \langle h, f_j \rangle$ for all $h \in \mathbb{H}$.

Given a sequence of weights $w := (w_j)_{j\geqslant 1} > 0$ and a constant $r \geqslant 1$ define

- the set $\mathcal{G}$ of all strictly positive nuclear operators defined on $\mathbb{H}$ and
- its subset $\mathcal{G}_w^r := \big\{\Gamma \in \mathcal{G} : r^{-2}\|h\|_{w^2}^2 \leqslant \|\Gamma h\|^2 \leqslant r^2\|h\|_{w^2}^2,\ \forall\, h \in \mathbb{H}\big\}$.

Minimal regularity conditions

Let $\{f_j\}_{j\geqslant 1}$ be a pre-specified ONB in $\mathbb{H}$ and $[h]_j := \langle h, f_j\rangle$ for all $h \in \mathbb{H}$.

Given a sequence of weights $w := (w_j)_{j\geqslant 1} > 0$ and a constant $r \geqslant 1$ define

- the set $\mathcal{G}$ of all strictly positive nuclear operators defined on $\mathbb{H}$ and
- its subset $\mathcal{G}_w^r := \big\{\Gamma \in \mathcal{G} : r^{-2}\|h\|_{w^2}^2 \leqslant \|\Gamma h\|^2 \leqslant r^2\|h\|_{w^2}^2,\ \forall\, h \in \mathbb{H}\big\}$.

Preliminary observations.

Let $(\lambda_j)_{j\geqslant 1}$ denote the sequence of eigenvalues of $\Gamma \in \mathcal{G}_w^r$, then

Minimal regularity conditions

Let $\{f_j\}_{j\geqslant 1}$ be a pre-specified ONB in $\mathbb{H}$ and $[h]_j := \langle h, f_j\rangle$ for all $h \in \mathbb{H}$.

Given a sequence of weights $w := (w_j)_{j\geqslant 1} > 0$ and a constant $r \geqslant 1$ define

- the set $\mathcal{G}$ of all strictly positive nuclear operators defined on $\mathbb{H}$ and
- its subset $\mathcal{G}_w^r := \big\{\Gamma \in \mathcal{G} : r^{-2}\|h\|^2_{w^2} \leqslant \|\Gamma h\|^2 \leqslant r^2\|h\|^2_{w^2},\ \forall\, h \in \mathbb{H}\big\}$.

Preliminary observations.

Let $(\lambda_j)_{j\geqslant 1}$ denote the sequence of eigenvalues of $\Gamma \in \mathcal{G}_w^r$, then

- $r^{-1}w_j \leqslant \langle \Gamma f_j, f_j\rangle \leqslant r\, w_j$ $\qquad j \geqslant 1.$

Minimal regularity conditions

Let $\{f_j\}_{j\geqslant 1}$ be a pre-specified ONB in $\mathbb{H}$ and $[h]_j := \langle h, f_j\rangle$ for all $h \in \mathbb{H}$.

Given a sequence of weights $w := (w_j)_{j\geqslant 1} > 0$ and a constant $r \geqslant 1$ define

- the set $\mathcal{G}$ of all strictly positive nuclear operators defined on $\mathbb{H}$ and
- its subset $\mathcal{G}_w^r := \big\{\Gamma \in \mathcal{G} : r^{-2}\|h\|_{w^2}^2 \leqslant \|\Gamma h\|^2 \leqslant r^2\|h\|_{w^2}^2,\ \forall\, h \in \mathbb{H}\big\}$.

Preliminary observations.

Let $(\lambda_j)_{j\geqslant 1}$ denote the sequence of eigenvalues of $\Gamma \in \mathcal{G}_w^r$, then

- $r^{-1}w_j \leqslant \langle \Gamma f_j, f_j\rangle \leqslant r\, w_j$ and $r^{-1}w_j \leqslant \lambda_j \leqslant r\, w_j$, $j \geqslant 1$.

Minimal regularity conditions

Let $\{f_j\}_{j\geqslant 1}$ be a pre-specified ONB in $\mathbb{H}$ and $[h]_j := \langle h, f_j\rangle$ for all $h \in \mathbb{H}$.

Given a sequence of weights $w := (w_j)_{j\geqslant 1} > 0$ and a constant $r \geqslant 1$ define

- the set $\mathcal{G}$ of all strictly positive nuclear operators defined on $\mathbb{H}$ and
- its subset $\mathcal{G}_w^r := \big\{\Gamma \in \mathcal{G} : r^{-2}\|h\|_{w^2}^2 \leqslant \|\Gamma h\|^2 \leqslant r^2\|h\|_{w^2}^2,\ \forall\, h \in \mathbb{H}\big\}$.

Assumption A3. (Minimal regularity assumptions)

(ii) The covariance operator Γ belongs to $\mathcal{G}_\gamma^d$ for some non-increasing, summable sequence of weights $\gamma := (\gamma_j)_{j\geqslant 1}$ with $\gamma_1 = 1$, and $d > 0$.

Outline

- Methodology
- Background and model assumptions
- Minimax theory
  - Measure of performance
  - Lower bound: global and local risk
  - Minimax-optimal estimation
- Adaptive estimation combining model selection and Lepski's method
  - Adaptive global estimation
  - Adaptive local estimation

Minimax-optimal estimation

- Measure the performance of any estimator $\widetilde{\beta}$ – maximal risk

$$\sup_{\Gamma\in\mathcal{G}_\gamma^{\boldsymbol{d}}}\ \sup_{\beta\in\mathcal{F}_{\boldsymbol{b}}^{\rho}} \mathbb{E}\big[\,|\mathcal{L}(\widetilde{\beta},\beta)|^2\big]$$

Minimax-optimal estimation

- Measure the performance of any estimator $\widetilde{\beta}$ – maximal risk

$$\sup_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{F}_b^\rho} \mathbb{E}\big[\,|\mathcal{L}(\widetilde{\beta}, \beta)|^2\big]$$

- Complexity of functional linear regression – lower bound

$$\inf_{\widetilde{\beta}} \sup_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{F}_b^\rho} \mathbb{E}\big[\,|\mathcal{L}(\widetilde{\beta}, \beta)|^2\big] \gtrsim R^*[n\,; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d].$$

Minimax-optimal estimation

- Measure the performance of any estimator $\widetilde{\beta}$ – maximal risk

$$\sup_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{F}_b^\rho} \mathbb{E}\big[\,|\mathcal{L}(\widetilde{\beta},\beta)|^2\big]$$

- Complexity of functional linear regression – lower bound

$$\inf_{\widetilde{\beta}} \sup_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{F}_b^\rho} \mathbb{E}\big[\,|\mathcal{L}(\widetilde{\beta},\beta)|^2\big] \gtrsim R^*[n\,; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d].$$

- $\widehat{\beta}_{m_n^*} = (f)^t_{\underline{m}_n^*}[\widehat{\Gamma}]^{-1}_{\underline{m}_n^*}[\widehat{g}]_{\underline{m}_n^*}\,\mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{m}_n^*}\| \leqslant n\}}$ is called minimax-optimal if

$$\sup_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{F}_b^\rho} \mathbb{E}\big[\,|\mathcal{L}(\widehat{\beta}_{m_n^*},\beta)|^2\big] \lesssim R^*[n\,; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d].$$

Outline

- Methodology
- Background and model assumptions
- Minimax theory
  - Measure of performance
  - Lower bound: global and local risk
  - Minimax-optimal estimation
- Adaptive estimation combining model selection and Lepski's method
  - Adaptive global estimation
  - Adaptive local estimation

Global measure of performance

Let $\widetilde{\beta}$ be an estimator of the slope function β. Consider the loss

- $\mathcal{L}(\widetilde{\beta}, \beta) = \|\widetilde{\beta} - \beta\|_\omega$ for some $\omega := (\omega_j) > 0$.

Global measure of performance

Let $\widetilde{\beta}$ be an estimator of the slope function β. Consider the loss

- $\mathcal{L}(\widetilde{\beta}, \beta) = \|\widetilde{\beta} - \beta\|_\omega$ for some $\omega := (\omega_j) > 0$.

Illustrations. Consider $\mathbb{H} = L^2[0,1]$.

Mean integrated squared error (c.f. Hall & Horowitz (2007))

Global measure of performance

Let $\widetilde{\beta}$ be an estimator of the slope function β. Consider the loss

- $\mathcal{L}(\widetilde{\beta}, \beta) = \|\widetilde{\beta} - \beta\|_\omega$ for some $\omega := (\omega_j) > 0$.

Illustrations. Consider $\mathbb{H} = L^2[0,1]$.

$(\omega_j = 1)$ *Mean integrated squared error* (c.f. Hall & Horowitz (2007))

Global measure of performance

Let $\widetilde{\beta}$ be an estimator of the slope function β. Consider the loss

- $\mathcal{L}(\widetilde{\beta}, \beta) = \|\widetilde{\beta} - \beta\|_\omega$ for some $\omega := (\omega_j) > 0$.

Illustrations. Consider $\mathbb{H} = L^2[0,1]$.

$(\omega_j = 1)$ *Mean integrated squared error* (c.f. Hall & Horowitz (2007))

MISE of the s-th derivative $\widetilde{\beta}^{(s)}$ of $\widetilde{\beta}$, i.e.,

$\mathbb{E}\|\widetilde{\beta}^{(s)} - \beta^{(s)}\|^2 \sim \mathbb{E}\|\widetilde{\beta} - \beta\|_\omega^2$ if $\{f_j\}$ is the trigonometric basis.

Global measure of performance

Let $\widetilde{\beta}$ be an estimator of the slope function β. Consider the loss

- $\mathcal{L}(\widetilde{\beta}, \beta) = \|\widetilde{\beta} - \beta\|_\omega$ for some $\omega := (\omega_j) > 0$.

Illustrations. Consider $\mathbb{H} = L^2[0,1]$.

$(\omega_j = 1)$ *Mean integrated squared error* (c.f. Hall & Horowitz (2007))

$(\omega_j = j^{2s})$ *MISE of the s-th derivative* $\widetilde{\beta}^{(s)}$ of $\widetilde{\beta}$, i.e.,

$\mathbb{E}\|\widetilde{\beta}^{(s)} - \beta^{(s)}\|^2 \sim \mathbb{E}\|\widetilde{\beta} - \beta\|_\omega^2$ if $\{f_j\}$ is the trigonometric basis.

Global measure of performance

Let $\widetilde{\beta}$ be an estimator of the slope function β. Consider the loss

- $\mathcal{L}(\widetilde{\beta}, \beta) = \|\widetilde{\beta} - \beta\|_\omega$ for some $\omega := (\omega_j) > 0$.

Illustrations. Consider $\mathbb{H} = L^2[0,1]$.

$(\omega_j = 1)$ *Mean integrated squared error* (c.f. Hall & Horowitz (2007))

$(\omega_j = j^{2s})$ *MISE of the s-th derivative* $\widetilde{\beta}^{(s)}$ of $\widetilde{\beta}$, i.e.,

$\mathbb{E}\|\widetilde{\beta}^{(s)} - \beta^{(s)}\|^2 \sim \mathbb{E}\|\widetilde{\beta} - \beta\|_\omega^2$ if $\{f_j\}$ is the trigonometric basis.

Mean prediction error (c.f. Crambes et al. (2009)), i.e.,

$\mathbb{E}\left[|\langle\widetilde{\beta}, X_{n+1}\rangle - \langle\beta, X_{n+1}\rangle|^2 \,\middle|\, X_1, \dots, X_n\right] \sim \mathbb{E}\|\widetilde{\beta} - \beta\|_\omega^2$ if $\Gamma \in \mathcal{G}_\gamma^d$.

Global measure of performance

Let $\widetilde{\beta}$ be an estimator of the slope function β. Consider the loss

- $\mathcal{L}(\widetilde{\beta}, \beta) = \|\widetilde{\beta} - \beta\|_\omega$ for some $\omega := (\omega_j) > 0$.

Illustrations. Consider $\mathbb{H} = L^2[0,1]$.

$(\omega_j = 1)$ *Mean integrated squared error* (c.f. Hall & Horowitz (2007))

$(\omega_j = j^{2s})$ *MISE of the s-th derivative* $\widetilde{\beta}^{(s)}$ of $\widetilde{\beta}$, i.e.,

$\mathbb{E}\|\widetilde{\beta}^{(s)} - \beta^{(s)}\|^2 \sim \mathbb{E}\|\widetilde{\beta} - \beta\|_\omega^2$ if $\{f_j\}$ is the trigonometric basis.

$(\omega_j = \gamma_j)$ *Mean prediction error* (c.f. Crambes et al. (2009)), i.e.,

$\mathbb{E}\Big[|\langle \widetilde{\beta}, X_{n+1}\rangle - \langle \beta, X_{n+1}\rangle|^2 \Big| X_1, \dots, X_n\Big] \sim \mathbb{E}\|\widetilde{\beta} - \beta\|_\omega^2$ if $\Gamma \in \mathcal{G}_\gamma^d$.

Local measure of performance

Let $\widetilde{\beta}$ be an estimator of the slope function β. Consider the loss

- $\mathcal{L}(\widetilde{\beta}, \beta) = |\ell(\widetilde{\beta}) - \ell(\beta)|$ for some linear functional ℓ.

Local measure of performance

Let $\widetilde{\beta}$ be an estimator of the slope function β. Consider the loss

- $\mathcal{L}(\widetilde{\beta}, \beta) = |\ell(\widetilde{\beta}) - \ell(\beta)|$ for some linear functional ℓ.

Illustrations. Consider $\mathbb{H} = L^2[0, 1]$.

- *Point evaluation* at $t_0 \in [0, 1]$, i.e.,

  $\beta(t_0) = \ell(\beta) = \sum_{j \geqslant 1} [\beta]_j [\ell]_j$ with $[\ell]_j = f_j(t_0)$;

Local measure of performance

Let $\widetilde{\beta}$ be an estimator of the slope function β. Consider the loss

- $\mathcal{L}(\widetilde{\beta}, \beta) = |\ell(\widetilde{\beta}) - \ell(\beta)|$ for some linear functional ℓ.

Illustrations. Consider $\mathbb{H} = L^2[0, 1]$.

- *Point evaluation* at $t_0 \in [0, 1]$, i.e.,

  $\beta(t_0) = \ell(\beta) = \sum_{j\geqslant 1}[\beta]_j[\ell]_j$ with $[\ell]_j = f_j(t_0)$;

- *Average value* of β over an interval $[0, a]$, i.e.,

  $a^{-1}\int_0^a \beta(t)dt = \ell(\beta) = \sum_{j\geqslant 1}[\beta]_j[\ell]_j$ with $[\ell]_j = a^{-1}\int_0^a f_j(t)dt$;

Local measure of performance

Let $\widetilde{\beta}$ be an estimator of the slope function β. Consider the loss

- $\mathcal{L}(\widetilde{\beta}, \beta) = |\ell(\widetilde{\beta}) - \ell(\beta)|$ for some linear functional ℓ.

Illustrations. Consider $\mathbb{H} = L^2[0,1]$.

- *Point evaluation* at $t_0 \in [0,1]$, i.e.,
  $\beta(t_0) = \ell(\beta) = \sum_{j\geqslant 1}[\beta]_j[\ell]_j$ with $[\ell]_j = f_j(t_0)$;
- *Average value* of β over an interval $[0,a]$, i.e.,
  $a^{-1}\int_0^a \beta(t)dt = \ell(\beta) = \sum_{j\geqslant 1}[\beta]_j[\ell]_j$ with $[\ell]_j = a^{-1}\int_0^a f_j(t)dt$;
- *Weighted average derivative* of β, i.e.,
  $\int_0^1 \beta'(t)w(t)dt = \ell(\beta) = \sum_{j\geqslant 1}[\beta]_j[\ell]_j$ with $[\ell]_j = \int_0^1 f_j'(t)w(t)dt$.

Measure of performance

Let $\widetilde{\beta}$ be an estimator of the slope function β. Consider the maximal risk

- global: $\sup_{\Gamma \in \mathcal{G}_\gamma^{\boldsymbol{d}}} \sup_{\beta \in \mathcal{F}_{\boldsymbol{b}}^\rho} \mathbb{E}\|\widetilde{\beta} - \beta\|_\omega^2$ for some $\omega := (\omega_j) > 0$,
- local: $\sup_{\Gamma \in \mathcal{G}_\gamma^{\boldsymbol{d}}} \sup_{\beta \in \mathcal{F}_{\boldsymbol{b}}^\rho} \mathbb{E}|\ell(\widetilde{\beta}) - \ell(\beta)|^2$ for some linear functional ℓ.

Measure of performance

Let $\widetilde{\beta}$ be an estimator of the slope function β. Consider the maximal risk

- global: $\sup_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{F}_b^\rho} \mathbb{E}\|\widetilde{\beta} - \beta\|_\omega^2$ for some $\omega := (\omega_j) > 0$,
- local: $\sup_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{F}_b^\rho} \mathbb{E}|\ell(\widetilde{\beta}) - \ell(\beta)|^2$ for some linear functional ℓ.

Assumption A4. (Admissible measure of performance)

global: $\omega := (\omega_j)_{j \geqslant 1}$ is a strictly positive sequence with $\omega_1 = 1$ such that $b/\omega = (b_j/\omega_j)_{j \geqslant 1}$ is non-decreasing with limit zero.

Measure of performance

Let $\widetilde{\beta}$ be an estimator of the slope function β. Consider the maximal risk

- global: $\sup_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{F}_b^\rho} \mathbb{E}\|\widetilde{\beta} - \beta\|_\omega^2$ for some $\omega := (\omega_j) > 0$,
- local: $\sup_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{F}_b^\rho} \mathbb{E}|\ell(\widetilde{\beta}) - \ell(\beta)|^2$ for some linear functional ℓ.

Assumption A4. (Admissible measure of performance)

global: $\omega := (\omega_j)_{j \geqslant 1}$ is a strictly positive sequence with $\omega_1 = 1$ such that $b/\omega = (b_j/\omega_j)_{j \geqslant 1}$ is non-decreasing with limit zero.

local: $\{f_j\}_{j \geqslant 1}$ belongs to the domain of ℓ, i.e., $[\ell]_j := \ell(f_j)$ is well-defined, and the sequence $([\ell]_j)_{j \geqslant 1}$ satisfies $\sum_{j \geqslant 1} [\ell]_j^2 b_j^{-1} < \infty$.

Illustration

Let $\{f_j\}$ be the trigonometric basis in $L^2[0,1]$ and $\mathcal{W}_p^\rho$ an ellipsoid in the *Sobolev space of p-times differentiable periodic functions*.

Illustration

Let $\{f_j\}$ be the trigonometric basis in $L^2[0,1]$ and $\mathcal{W}_p^\rho$ an ellipsoid in the *Sobolev space of p-times differentiable periodic functions*.

- Assume that $\beta \in \mathcal{W}_p^\rho$ for some $p > 0$, i.e., $b_j = j^{-2p}$.

Illustration

Let $\{f_j\}$ be the trigonometric basis in $L^2[0,1]$ and $\mathcal{W}_p^\rho$ an ellipsoid in the *Sobolev space of p-times differentiable periodic functions*.

- Assume that $\beta \in \mathcal{W}_p^\rho$ for some $p > 0$, i.e., $b_j = j^{-2p}$.
- Suppose $\Gamma \in \mathcal{G}_\gamma^d$ and two cases for γ:

(p) polynomial decay, i.e., $\gamma_j = j^{-2a}$, $a > 1/2$;

Illustration

Let $\{f_j\}$ be the trigonometric basis in $L^2[0,1]$ and $\mathcal{W}_p^\rho$ an ellipsoid in the *Sobolev space of p-times differentiable periodic functions*.

- Assume that $\beta \in \mathcal{W}_p^\rho$ for some $p > 0$, i.e., $b_j = j^{-2p}$.
- Suppose $\Gamma \in \mathcal{G}_\gamma^d$ and two cases for γ:

(p) polynomial decay, i.e., $\gamma_j = j^{-2a}$, $a > 1/2$;

(e) exponential decay, i.e., $\gamma_j = \exp(-j^{2a} + 1)$, $a > 0$.

Illustration

Let $\{f_j\}$ be the trigonometric basis in $L^2[0,1]$ and $\mathcal{W}_p^\rho$ an ellipsoid in the *Sobolev space of p-times differentiable periodic functions*.

- Assume that $\beta \in \mathcal{W}_p^\rho$ for some $p > 0$, i.e., $b_j = j^{-2p}$.
- Suppose $\Gamma \in \mathcal{G}_\gamma^d$ and two cases for γ:

(p) polynomial decay, i.e., $\gamma_j = j^{-2a}$, $a > 1/2$;

(e) exponential decay, i.e., $\gamma_j = \exp(-j^{2a} + 1)$, $a > 0$.

- global risk: $\sup_{\Gamma\in\mathcal{G}_\gamma^d} \sup_{\beta\in\mathcal{W}_p^\rho} \mathbb{E}\|\widetilde{\beta}^{(s)} - \beta^{(s)}\|^2$ for some $p > s \geqslant 0$,

$$: \sup_{\Gamma\in\mathcal{G}_\gamma^d} \sup_{\beta\in\mathcal{W}_p^\rho} \mathbb{E}\Big[|\langle\widetilde{\beta} - \beta, X_{n+1}\rangle|^2 \,\Big|\, X_1, \dots, X_n\Big],$$

Illustration

Let $\{f_j\}$ be the trigonometric basis in $L^2[0,1]$ and $\mathcal{W}_p^\rho$ an ellipsoid in the *Sobolev space of p-times differentiable periodic functions*.

- Assume that $\beta \in \mathcal{W}_p^\rho$ for some $p > 0$, i.e., $b_j = j^{-2p}$.
- Suppose $\Gamma \in \mathcal{G}_\gamma^d$ and two cases for γ:

(p) polynomial decay, i.e., $\gamma_j = j^{-2a}$, $a > 1/2$;

(e) exponential decay, i.e., $\gamma_j = \exp(-j^{2a} + 1)$, $a > 0$.

- global risk: $\sup_{\Gamma\in\mathcal{G}_\gamma^d} \sup_{\beta\in\mathcal{W}_p^\rho} \mathbb{E}\|\widetilde{\beta}^{(s)} - \beta^{(s)}\|^2$ for some $p > s \geqslant 0$,

  $$: \sup_{\Gamma\in\mathcal{G}_\gamma^d} \sup_{\beta\in\mathcal{W}_p^\rho} \mathbb{E}\Big[|\langle\widetilde{\beta} - \beta, X_{n+1}\rangle|^2 \,\Big|\, X_1, \dots, X_n\Big],$$

- local risk: $\sup_{\Gamma\in\mathcal{G}_\gamma^d} \sup_{\beta\in\mathcal{W}_p^\rho} \mathbb{E}|\widetilde{\beta}(t_0) - \beta(t_0)|^2$ for $t_0 \in [0,1]$ and $p > 1/2$.

Illustration

Let $\{f_j\}$ be the trigonometric basis in $L^2[0,1]$ and $\mathcal{W}_p^\rho$ an ellipsoid in the *Sobolev space of p-times differentiable periodic functions*.

- Assume that $\beta \in \mathcal{W}_p^\rho$ for some $p > 0$, i.e., $b_j = j^{-2p}$.
- Suppose $\Gamma \in \mathcal{G}_\gamma^d$ and two cases for γ:

(p) polynomial decay, i.e., $\gamma_j = j^{-2a}$, $a > 1/2$;

(e) exponential decay, i.e., $\gamma_j = \exp(-j^{2a} + 1)$, $a > 0$.

- global risk: $\sup_{\Gamma\in\mathcal{G}_\gamma^d} \sup_{\beta\in\mathcal{W}_p^\rho} \mathbb{E}\|\widetilde{\beta}^{(s)} - \beta^{(s)}\|^2$ for some $p > s \geqslant 0$,

$$: \sup_{\Gamma\in\mathcal{G}_\gamma^d} \sup_{\beta\in\mathcal{W}_p^\rho} \mathbb{E}\Big[|\langle\widetilde{\beta} - \beta, X_{n+1}\rangle|^2 \,\Big|\, X_1, \dots, X_n\Big],$$

- local risk: $\sup_{\Gamma\in\mathcal{G}_\gamma^d} \sup_{\beta\in\mathcal{W}_p^\rho} \mathbb{E}|\widetilde{\beta}(t_0) - \beta(t_0)|^2$ for $t_0 \in [0,1]$ and $p > 1/2$.

The Assumptions A3 and A4 are satisfied.

Outline

- Methodology
- Background and model assumptions
- Minimax theory
  - Measure of performance
  - Lower bound: global and local risk
  - Minimax-optimal estimation
- Adaptive estimation combining model selection and Lepski's method
  - Adaptive global estimation
  - Adaptive local estimation

Lower bound: global and local risk

Theorem (Cardot & JJ (2010), JJ & Schenk (2011))
Assume an iid n-sample of (Y, X).

Lower bound: global and local risk

Theorem (Cardot & JJ (2010), JJ & Schenk (2011))
Assume an iid n-sample of (Y, X). Under Assumption A1-A4 we have:

$$\inf_{\widetilde{\beta}} \inf_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{F}_b^\rho} \mathbb{E}\|\widetilde{\beta} - \beta\|_\omega^2 \gtrsim R_\omega^*[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] := \min_{m \geqslant 1} \Big\{ R_\omega^m[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] \Big\}$$

Lower bound: global and local risk

Theorem (Cardot & JJ (2010), JJ & Schenk (2011))
Assume an iid n-sample of (Y, X). Under Assumption A1-A4 we have:

$$\inf_{\widetilde{\beta}} \inf_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{F}_b^\rho} \mathbb{E}\|\widetilde{\beta} - \beta\|_\omega^2 \gtrsim R_\omega^*[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] := \min_{m \geqslant 1} \Big\{ R_\omega^m[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] \Big\}$$

with $R_\omega^m[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] := \max\Big(\frac{\omega_m}{b_m}, \frac{1}{n} \sum_{j=1}^m \frac{\omega_j}{\gamma_j}\Big)$,

Lower bound: global and local risk

Theorem (Cardot & JJ (2010), JJ & Schenk (2011))
Assume an iid n-sample of (Y, X). Under Assumption A1-A4 we have:

$$\inf_{\widetilde{\beta}} \inf_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{F}_b^\rho} \mathbb{E}\|\widetilde{\beta} - \beta\|_\omega^2 \gtrsim R_\omega^*[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] := \min_{m \geqslant 1} \Big\{ R_\omega^m[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] \Big\}$$

$$\text{with } R_\omega^m[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] := \max\Big(\frac{\omega_m}{b_m}, \frac{1}{n} \sum_{j=1}^m \frac{\omega_j}{\gamma_j} \Big),$$

Lower bound: global and local risk

Theorem (Cardot & JJ (2010), JJ & Schenk (2011))
Assume an iid n-sample of (Y, X). Under Assumption A1-A4 we have:

$$\inf_{\widetilde{\beta}} \inf_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{F}_b^\rho} \mathbb{E}\|\widetilde{\beta} - \beta\|_\omega^2 \gtrsim R_\omega^*[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] := \min_{m \geqslant 1} \Big\{ R_\omega^m[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] \Big\}$$

$$\text{with } R_\omega^m[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] := \max\Big(\frac{\omega_m}{b_m}, \frac{1}{n} \sum_{j=1}^m \frac{\omega_j}{\gamma_j} \Big),$$

$$\inf_{\widetilde{\ell}} \inf_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{F}_b^\rho} \mathbb{E}|\widetilde{\ell} - \ell(\beta)|^2 \gtrsim R_\ell^\diamond[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] := \min_{m \geqslant 1} \Big\{ R_\ell^m[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] \Big\}$$

Lower bound: global and local risk

Theorem (Cardot & JJ (2010), JJ & Schenk (2011))
Assume an iid n-sample of (Y, X). Under Assumption A1-A4 we have:

$$\inf_{\widetilde{\beta}} \inf_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{F}_b^\rho} \mathbb{E}\|\widetilde{\beta} - \beta\|_\omega^2 \gtrsim R_\omega^*[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] := \min_{m \geqslant 1}\Big\{ R_\omega^m[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] \Big\}$$

$$\text{with } R_\omega^m[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] := \max\Big(\frac{\omega_m}{b_m}, \frac{1}{n} \sum_{j=1}^m \frac{\omega_j}{\gamma_j} \Big),$$

$$\inf_{\widetilde{\ell}} \inf_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{F}_b^\rho} \mathbb{E}|\widetilde{\ell} - \ell(\beta)|^2 \gtrsim R_\ell^\diamond[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] := \min_{m \geqslant 1}\Big\{ R_\ell^m[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] \Big\}$$

$$\text{with } R_\ell^m[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] := \max\Big\{ \sum_{j>m} \frac{[\ell]_j^2}{b_j}, \max\left(\frac{\gamma_m}{b_m}, \frac{1}{n}\right) \sum_{j=1}^m \frac{[\ell]_j^2}{\gamma_j} \Big\}.$$

Lower bound: global and local risk

Theorem (Cardot & JJ (2010), JJ & Schenk (2011))
Assume an iid n-sample of (Y, X). Under Assumption A1-A4 we have:

$$\inf_{\widetilde{\beta}} \inf_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{F}_b^\rho} \mathbb{E}\|\widetilde{\beta} - \beta\|_\omega^2 \gtrsim R_\omega^*[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] := \min_{m \geqslant 1} \Big\{ R_\omega^m[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] \Big\}$$

$$\text{with } R_\omega^m[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] := \max\Big(\frac{\omega_m}{b_m}, \frac{1}{n} \sum_{j=1}^m \frac{\omega_j}{\gamma_j} \Big),$$

$$\inf_{\widetilde{\ell}} \inf_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{F}_b^\rho} \mathbb{E}|\widetilde{\ell} - \ell(\beta)|^2 \gtrsim R_\ell^\diamond[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] := \min_{m \geqslant 1} \Big\{ R_\ell^m[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] \Big\}$$

$$\text{with } R_\ell^m[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] := \max\Big\{ \sum_{j > m} \frac{[\ell]_j^2}{b_j}, \max\Big(\frac{\gamma_m}{b_m}, \frac{1}{n}\Big) \sum_{j=1}^m \frac{[\ell]_j^2}{\gamma_j} \Big\}.$$

Outline

- Methodology
- Background and model assumptions
- Minimax theory
  - Measure of performance
  - Lower bound: global and local risk
  - Minimax-optimal estimation
- Adaptive estimation combining model selection and Lepski's method
  - Adaptive global estimation
  - Adaptive local estimation

Dimension reduction

Consider the equation $g = \Gamma\beta$. Given the ONB $\{f_j\}_{j\geqslant 1}$ and $m \in \mathbb{N}$ define $\beta_m := (f)_{\underline{m}}^t[\Gamma]_{\underline{m}}^{-1}[g]_{\underline{m}}$ with $[g]_{\underline{m}} = \mathbb{E}(Y[X]_{\underline{m}})$ and $[\Gamma]_{\underline{m}} = \mathbb{E}([X]_{\underline{m}}[X]_{\underline{m}}^t)$.

Dimension reduction

Consider the equation $g = \Gamma\beta$. Given the ONB $\{f_j\}_{j\geqslant 1}$ and $m \in \mathbb{N}$ define $\beta_m := (f)^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[g]_{\underline{m}}$ with $[g]_{\underline{m}} = \mathbb{E}(Y[X]_{\underline{m}})$ and $[\Gamma]_{\underline{m}} = \mathbb{E}([X]_{\underline{m}}[X]^t_{\underline{m}})$.

Let $\mathbb{H}_m$ denote the linear subspace spanned by $\{f_1, \dots, f_m\}$.

Dimension reduction

Consider the equation $g = \Gamma\beta$. Given the ONB $\{f_j\}_{j\geqslant 1}$ and $m \in \mathbb{N}$ define $\beta_m := (f)^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[g]_{\underline{m}}$ with $[g]_{\underline{m}} = \mathbb{E}(Y[X]_{\underline{m}})$ and $[\Gamma]_{\underline{m}} = \mathbb{E}([X]_{\underline{m}}[X]^t_{\underline{m}})$.

Let $\mathbb{H}_m$ denote the linear subspace spanned by $\{f_1, \dots, f_m\}$. Since $\Gamma > 0$,

$$\|g - \Gamma\beta_m\| \leqslant \|g - \Gamma h\|, \quad \forall h \in \mathbb{H}_m,$$

Dimension reduction: linear Galerkin approach

Consider the equation $g = \Gamma\beta$. Given the ONB $\{f_j\}_{j\geqslant 1}$ and $m \in \mathbb{N}$ define $\beta_m := (f)_{\underline{m}}^t[\Gamma]_{\underline{m}}^{-1}[g]_{\underline{m}}$ with $[g]_{\underline{m}} = \mathbb{E}(Y[X]_{\underline{m}})$ and $[\Gamma]_{\underline{m}} = \mathbb{E}([X]_{\underline{m}}[X]_{\underline{m}}^t)$. Let $\mathbb{H}_m$ denote the linear subspace spanned by $\{f_1, \dots, f_m\}$. Since $\Gamma > 0$,

$$\|g - \Gamma\beta_m\| \leqslant \|g - \Gamma h\|, \quad \forall h \in \mathbb{H}_m,$$

and hence β_m is called Galerkin solution of $g = \Gamma\beta$ (c.f. Natterer (1977)).

Dimension reduction: linear Galerkin approach

Consider the equation $g = \Gamma\beta$. Given the ONB $\{f_j\}_{j\geqslant 1}$ and $m \in \mathbb{N}$ define $\beta_m := (f)^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[g]_{\underline{m}}$ with $[g]_{\underline{m}} = \mathbb{E}(Y[X]_{\underline{m}})$ and $[\Gamma]_{\underline{m}} = \mathbb{E}([X]_{\underline{m}}[X]^t_{\underline{m}})$.

Lemma (Bias due to dimension reduction)

Suppose Assumption A2-A4 and let $\Gamma \in \mathcal{G}^d_\gamma$.

Dimension reduction: linear Galerkin approach

Consider the equation $g = \Gamma\beta$. Given the ONB $\{f_j\}_{j\geqslant 1}$ and $m \in \mathbb{N}$ define $\beta_m := (f)^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[g]_{\underline{m}}$ with $[g]_{\underline{m}} = \mathbb{E}(Y[X]_{\underline{m}})$ and $[\Gamma]_{\underline{m}} = \mathbb{E}([X]_{\underline{m}}[X]^t_{\underline{m}})$.

Lemma (Bias due to dimension reduction)

Suppose Assumption A2-A4 and let $\Gamma \in \mathcal{G}^d_\gamma$.

- $\forall\, \beta \in \mathcal{F}^\rho_b : \|\beta_m - \beta\|^2_\omega \lesssim \frac{\omega_m}{b_m} \max\Big\{1, \frac{\gamma_m^2}{\omega_m}\Big\}$.

Dimension reduction: linear Galerkin approach

Consider the equation $g = \Gamma\beta$. Given the ONB $\{f_j\}_{j\geqslant 1}$ and $m \in \mathbb{N}$ define

$\beta_m := (f)^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[g]_{\underline{m}}$ with $[g]_{\underline{m}} = \mathbb{E}(Y[X]_{\underline{m}})$ and $[\Gamma]_{\underline{m}} = \mathbb{E}([X]_{\underline{m}}[X]^t_{\underline{m}})$.

Lemma (Bias due to dimension reduction)

Suppose Assumption A2-A4 and let $\Gamma \in \mathcal{G}^d_\gamma$.

- $\forall\,\beta \in \mathcal{F}^\rho_b : \|\beta_m - \beta\|^2_\omega \lesssim \frac{\omega_m}{b_m}\max\Big\{1, \frac{\gamma^2_m}{\omega_m}\Big\}$.
- $\forall\,\beta \in \mathcal{F}^\rho_b : |\ell(\beta_m) - \ell(\beta)|^2 \lesssim \max\Big\{\sum_{j>m}\frac{[\ell]^2_j}{b_j}, \frac{\gamma_m}{b_m}\sum_{j=1}^{m}\frac{[\ell]^2_j}{\gamma_j}\Big\}$.

Dimension reduction: linear Galerkin approach

Consider the equation $g = \Gamma\beta$. Given the ONB $\{f_j\}_{j\geqslant 1}$ and $m \in \mathbb{N}$ define $\beta_m := (f)_{\underline{m}}^t [\Gamma]_{\underline{m}}^{-1} [g]_{\underline{m}}$ with $[g]_{\underline{m}} = \mathbb{E}(Y[X]_{\underline{m}})$ and $[\Gamma]_{\underline{m}} = \mathbb{E}([X]_{\underline{m}}[X]_{\underline{m}}^t)$.

Lemma (Bias due to dimension reduction)

Suppose Assumption A2-A4 and let $\Gamma \in \mathcal{G}_\gamma^d$.

- $\forall\, \beta \in \mathcal{F}_b^\rho : \|\beta_m - \beta\|_\omega^2 \lesssim \frac{\omega_m}{b_m} \max\Big\{1, \frac{\gamma_m^2}{\omega_m}\Big\}$.
- $\forall\, \beta \in \mathcal{F}_b^\rho : |\ell(\beta_m) - \ell(\beta)|^2 \lesssim \max\Big\{\sum_{j>m} \frac{[\ell]_j^2}{b_j}, \frac{\gamma_m}{b_m} \sum_{j=1}^m \frac{[\ell]_j^2}{\gamma_j}\Big\}$.

Assumption A5. The sequences $(\frac{\gamma_j^2}{\omega_j})_{j\geqslant 1}$ and $(\frac{j^4\gamma_j}{b_j})_{j\geqslant 1}$ are bounded.

Minimax-optimal global and local estimation

Theorem (Cardot & JJ (2010), JJ & Schenk (2010))
Assume an iid n-sample of (Y, X).

Minimax-optimal global and local estimation

Theorem (Cardot & JJ (2010), JJ & Schenk (2010))
Assume an iid n-sample of (Y, X). Under Assumption A1-A5 we have:

$$\sup_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{F}_b^\rho} \mathbb{E}\|\widehat{\beta}_{m_n^*} - \beta\|_\omega^2 \lesssim R_\omega^*[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d]$$

Minimax-optimal global and local estimation

Theorem (Cardot & JJ (2010), JJ & Schenk (2010))
Assume an iid n-sample of (Y, X). Under Assumption A1-A5 we have:

$$\sup_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{F}_b^\rho} \mathbb{E}\|\widehat{\beta}_{m_n^*} - \beta\|_\omega^2 \lesssim R_\omega^*[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] = R_\omega^{m_n^*}[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d],$$

$$\text{with } m_n^* = \arg\min_{m \geqslant 1} \left\{ R_\omega^m[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] \right\},$$

Minimax-optimal global and local estimation

Theorem (Cardot & JJ (2010), JJ & Schenk (2010))
Assume an iid n-sample of (Y, X). Under Assumption A1-A5 we have:

$$\sup_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{F}_b^\rho} \mathbb{E}\|\widehat{\beta}_{m_n^*} - \beta\|_\omega^2 \lesssim R_\omega^*[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] = R_\omega^{m_n^*}[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d],$$

$$\text{with } m_n^* = \arg\min_{m \geqslant 1} \left\{ R_\omega^m[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] \right\},$$

$$\sup_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{F}_b^\rho} \mathbb{E}|\widehat{\ell}_{m_n^\diamond} - \ell(\beta)|^2 \lesssim R_\ell^\diamond[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d]$$

Minimax-optimal global and local estimation

Theorem (Cardot & JJ (2010), JJ & Schenk (2010))
Assume an iid n-sample of (Y, X). Under Assumption A1-A5 we have:

$$\sup_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{F}_b^\rho} \mathbb{E}\|\widehat{\beta}_{m_n^*} - \beta\|_\omega^2 \lesssim R_\omega^*[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] = R_\omega^{m_n^*}[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d],$$

$$\text{with } m_n^* = \arg\min_{m \geqslant 1} \{ R_\omega^m[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] \},$$

$$\sup_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{F}_b^\rho} \mathbb{E}|\widehat{\ell}_{m_n^\diamond} - \ell(\beta)|^2 \lesssim R_\ell^\diamond[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] = R_\ell^{m_n^\diamond}[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d].$$

$$\text{with } m_n^\diamond = \arg\min_{m \geqslant 1} \{ R_\ell^m[n; \mathcal{F}_b^\rho, \mathcal{G}_\gamma^d] \}.$$

Illustration: minimax-optimal global and local estimation

Minimax-optimality. Under Assumption A1-A2 we have in case of

Illustration: minimax-optimal global and local estimation

Minimax-optimality. Under Assumption A1-A2 we have in case of

(p) for $p + a \geqslant 2$ with $m_n^* \sim n^{1/(2p+2a+1)}$ and $m_n^\diamond \sim n^{1/(2p+2a)}$

- $\sup_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{W}_p^\rho} \mathbb{E}\|\widehat{\beta}_{m_n^*}^{(s)} - \beta^{(s)}\|^2 \lesssim n^{-2(p-s)/(2p+2a+1)}$.
- $\sup_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{W}_p^\rho} \mathbb{E}[\langle \widehat{\beta}_{m_n^*} - \beta, X_{n+1}\rangle^2 \,|\, X_1, \dots] \lesssim n^{-2(p+a)/(2p+2a+1)}$

(e) with $m_n^* \sim (\log n)^{1/(2a)}$ and $m_n^\diamond \sim (\log n)^{1/(2a)}$

- $\sup_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{W}_p^\rho} \mathbb{E}\|\widehat{\beta}_{m_n^*}^{(s)} - \beta^{(s)}\|^2 \lesssim (\log n)^{-(p-s)/a}$.
- $\sup_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{W}_p^\rho} \mathbb{E}[\langle \widehat{\beta}_{m_n^*} - \beta, X_{n+1}\rangle^2 \,|\, X_1, \dots] \lesssim n^{-1}(\log n)^{1/(2a)}$

Illustration: minimax-optimal global and local estimation

Minimax-optimality. Under Assumption A1-A2 we have in case of

(p) for $p + a \geqslant 2$ with $m_n^* \sim n^{1/(2p+2a+1)}$ and $m_n^\diamond \sim n^{1/(2p+2a)}$

- $\sup_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{W}_p^\rho} \mathbb{E}\|\widehat{\beta}_{m_n^*}^{(s)} - \beta^{(s)}\|^2 \lesssim n^{-2(p-s)/(2p+2a+1)}$.
- $\sup_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{W}_p^\rho} \mathbb{E}[\langle \widehat{\beta}_{m_n^*} - \beta, X_{n+1}\rangle^2 \,|\, X_1, \dots] \lesssim n^{-2(p+a)/(2p+2a+1)}$
- $\sup_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{W}_p^\rho} \mathbb{E}|\widehat{\beta}_{m_n^\diamond}(t_0) - \beta(t_0)|^2 \lesssim n^{-(2p-1)/(2p+2a)}$.

(e) with $m_n^* \sim (\log n)^{1/(2a)}$ and $m_n^\diamond \sim (\log n)^{1/(2a)}$

- $\sup_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{W}_p^\rho} \mathbb{E}\|\widehat{\beta}_{m_n^*}^{(s)} - \beta^{(s)}\|^2 \lesssim (\log n)^{-(p-s)/a}$.
- $\sup_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{W}_p^\rho} \mathbb{E}[\langle \widehat{\beta}_{m_n^*} - \beta, X_{n+1}\rangle^2 \,|\, X_1, \dots] \lesssim n^{-1}(\log n)^{1/(2a)}$
- $\sup_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{W}_p^\rho} \mathbb{E}|\widehat{\beta}_{m_n^\diamond}(t_0) - \beta(t_0)|^2 \lesssim (\log n)^{-(2p-1)/(2a)}$.

Illustration: minimax-optimal global and local estimation

Minimax-optimality. Under Assumption A1-A2 we have in case of

(p) for $p + a \geqslant 2$ with $m_n^* \sim n^{1/(2p+2a+1)}$ and $m_n^\diamond \sim n^{1/(2p+2a)}$

- $\sup_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{W}_p^\rho} \mathbb{E}\|\widehat{\beta}^{(s)}_{m_n^*} - \beta^{(s)}\|^2 \lesssim n^{-2(p-s)/(2p+2a+1)}$.
- $\sup_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{W}_p^\rho} \mathbb{E}[\langle \widehat{\beta}_{m_n^*} - \beta, X_{n+1}\rangle^2 \,|\, X_1, \dots] \lesssim n^{-2(p+a)/(2p+2a+1)}$
- $\sup_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{W}_p^\rho} \mathbb{E}|\widehat{\beta}_{m_n^\diamond}(t_0) - \beta(t_0)|^2 \lesssim n^{-(2p-1)/(2p+2a)}$.

(e) with $m_n^* \sim (\log n)^{1/(2a)}$ and $m_n^\diamond \sim (\log n)^{1/(2a)}$

- $\sup_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{W}_p^\rho} \mathbb{E}\|\widehat{\beta}^{(s)}_{m_n^*} - \beta^{(s)}\|^2 \lesssim (\log n)^{-(p-s)/a}$.
- $\sup_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{W}_p^\rho} \mathbb{E}[\langle \widehat{\beta}_{m_n^*} - \beta, X_{n+1}\rangle^2 \,|\, X_1, \dots] \lesssim n^{-1}(\log n)^{1/(2a)}$
- $\sup_{\Gamma \in \mathcal{G}_\gamma^d} \sup_{\beta \in \mathcal{W}_p^\rho} \mathbb{E}|\widehat{\beta}_{m_n^\diamond}(t_0) - \beta(t_0)|^2 \lesssim (\log n)^{-(2p-1)/(2a)}$.

Outline

- Methodology
- Background and model assumptions
- Minimax theory
  - Measure of performance
  - Lower bound: global and local risk
  - Minimax-optimal estimation
- Adaptive estimation combining model selection and Lepski's method
  - Adaptive global estimation
  - Adaptive local estimation

Adaptive estimation

$$\widehat{\beta}_{\widehat{m}} := (f)^t_{\underline{\widehat{m}}}[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}[\widehat{g}]_{\underline{\widehat{m}}}\, \mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}\|_s \leqslant n\}} \text{ with } \widehat{m} := \underset{1\leqslant m\leqslant \widehat{M}}{\arg\min}\left\{\text{Contrast}_m + \widehat{\mathrm{pen}}_m\right\},$$

$$\text{Contrast}_m := \max_{m\leqslant k\leqslant \widehat{M}}\left\{|\mathcal{L}(\widehat{\beta}_k, \widehat{\beta}_m)|^2 - \widehat{\mathrm{pen}}_k\right\}$$

Adaptive estimation

$$\widehat{\beta}_{\widehat{m}} := (f)^t_{\underline{\widehat{m}}}[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}[\widehat{g}]_{\underline{\widehat{m}}}\, \mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}\|_s \leqslant n\}} \text{ with } \widehat{m} := \arg\min_{1\leqslant m\leqslant \widehat{M}} \left\{\text{Contrast}_m + \widehat{\text{pen}}_m\right\},$$

$$\text{Contrast}_m := \max_{m\leqslant k\leqslant \widehat{M}} \left\{|\mathcal{L}(\widehat{\beta}_k, \widehat{\beta}_m)|^2 - \widehat{\text{pen}}_k\right\}$$

Lemma

If $(\widehat{\text{pen}}_1, \dots, \widehat{\text{pen}}_{\widehat{M}})$ is non-decreasing, then for all $1 \leqslant m \leqslant \widehat{M}$ we have

$$|\mathcal{L}(\widehat{\beta}_{\widehat{m}}, \beta)|^2 \leqslant 7\,\widehat{\text{pen}}_m + 78\,\text{bias}^2_m + 42 \max_{m\leqslant k\leqslant \widehat{M}} \left(|\mathcal{L}(\widehat{\beta}_k, \beta_k)|^2 - \frac{1}{6}\,\widehat{\text{pen}}_k\right)_+$$

with $\text{bias}_m = \sup_{k\geqslant m} |\mathcal{L}(\beta_k, \beta)|$, $m \geqslant 1$.

Outline

- Methodology
- Background and model assumptions
- Minimax theory
  - Measure of performance
  - Lower bound: global and local risk
  - Minimax-optimal estimation
- Adaptive estimation combining model selection and Lepski's method
  - Adaptive global estimation
  - Adaptive local estimation

Adaptive global estimation

$$\widehat{\beta}_{\widetilde{m}} := (f)^t_{\underline{\widetilde{m}}}[\widehat{\Gamma}]^{-1}_{\underline{\widetilde{m}}}[\widehat{g}]_{\underline{\widetilde{m}}}\,\mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{\widetilde{m}}}\|_s \leqslant n\}} \text{ with } \widetilde{m} := \arg\min_{1\leqslant m\leqslant M}\left\{\text{Contrast}_m + \mathrm{pen}_m\right\},$$

Adaptive global estimation

$$\widehat{\beta}_{\widetilde{m}} := (f)^t_{\underline{\widetilde{m}}}[\widehat{\Gamma}]^{-1}_{\underline{\widetilde{m}}}[\widehat{g}]_{\underline{\widetilde{m}}}\, \mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{\widetilde{m}}}\|_s \leqslant n\}} \text{ with } \widetilde{m} := \arg\min_{1\leqslant m\leqslant M} \left\{\text{Contrast}_m + \mathrm{pen}_m\right\},$$

$$\text{Contrast}_m = \sup_{m\leqslant k\leqslant M}\left\{\|\widehat{\beta}_k - \widehat{\beta}_m\|^2_\omega - \mathrm{pen}_k\right\} \text{ and } \mathrm{pen}_m := c\,\sigma^2_m\,\delta_m\,n^{-1}$$

Adaptive global estimation

$$\widehat{\beta}_{\widetilde{m}} := (f)^t_{\underline{\widetilde{m}}}[\widehat{\Gamma}]^{-1}_{\underline{\widetilde{m}}}[\widehat{g}]_{\underline{\widetilde{m}}}\, \mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{\widetilde{m}}}\|_s \leqslant n\}} \text{ with } \widetilde{m} := \arg\min_{1\leqslant m\leqslant M} \left\{\text{Contrast}_m + \mathrm{pen}_m\right\},$$

$$\text{Contrast}_m = \sup_{m\leqslant k\leqslant M} \left\{\|\widehat{\beta}_k - \widehat{\beta}_m\|^2_\omega - \mathrm{pen}_k\right\} \text{ and } \mathrm{pen}_m := c\, \sigma^2_m\, \delta_m\, n^{-1}$$

- $\mathbb{V}\mathrm{ar}(U + \langle \beta - \beta_m, X\rangle) \leqslant \sigma^2_m$

Adaptive global estimation

$$\widehat{\beta}_{\widetilde{m}} := (f)^t_{\underline{\widetilde{m}}}[\widehat{\Gamma}]^{-1}_{\underline{\widetilde{m}}}[\widehat{g}]_{\underline{\widetilde{m}}}\, \mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{\widetilde{m}}}\|_s \leqslant n\}} \text{ with } \widetilde{m} := \mathop{\arg\min}_{1\leqslant m\leqslant M} \left\{\text{Contrast}_m + \text{pen}_m\right\},$$

$$\text{Contrast}_m = \sup_{m\leqslant k\leqslant M} \left\{\|\widehat{\beta}_k - \widehat{\beta}_m\|^2_\omega - \text{pen}_k\right\} \text{ and } \text{pen}_m := c\, \sigma^2_m\, \delta_m\, n^{-1}$$

- $\mathbb{V}\mathrm{ar}(U + \langle \beta - \beta_m, X\rangle) \leqslant \sigma^2_m := 2\left\{\mathbb{E}Y^2 + [g]^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[g]_{\underline{m}}\right\}$

Adaptive global estimation

$$\widehat{\beta}_{\widetilde{m}} := (f)^t_{\underline{\widetilde{m}}}[\widehat{\Gamma}]^{-1}_{\underline{\widetilde{m}}}[\widehat{g}]_{\underline{\widetilde{m}}}\,\mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{\widetilde{m}}}\|_s \leqslant n\}} \text{ with } \widetilde{m} := \arg\min_{1\leqslant m\leqslant M}\left\{\text{Contrast}_m + \mathrm{pen}_m\right\},$$

$$\text{Contrast}_m = \sup_{m\leqslant k\leqslant M}\left\{\|\widehat{\beta}_k - \widehat{\beta}_m\|^2_\omega - \mathrm{pen}_k\right\} \text{ and } \mathrm{pen}_m := c\,\sigma^2_m\,\delta_m\,n^{-1}$$

- $\mathbb{V}\mathrm{ar}(U + \langle\beta - \beta_m, X\rangle) \leqslant \sigma^2_m := 2\left\{\mathbb{E}Y^2 + [g]^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[g]_{\underline{m}}\right\}$
- $\delta_m := m\Delta_m \frac{\log(\Delta_m \vee m+2)}{\log(m+2)}, \quad \Delta_m := \max_{1\leqslant k\leqslant m}\|[\nabla_\omega]^{1/2}_{\underline{k}}[\Gamma]^{-1}_{\underline{k}}[\nabla_\omega]^{1/2}_{\underline{k}}\|_s,$

Adaptive global estimation

$$\widehat{\beta}_{\widetilde{m}} := (f)^t_{\underline{\widetilde{m}}}[\widehat{\Gamma}]^{-1}_{\underline{\widetilde{m}}}[\widehat{g}]_{\underline{\widetilde{m}}}\, \mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{\widetilde{m}}}\|_s \leqslant n\}} \text{ with } \widetilde{m} := \arg\min_{1\leqslant m\leqslant M} \{\text{Contrast}_m + \text{pen}_m\},$$

$$\text{Contrast}_m = \sup_{m\leqslant k\leqslant M} \left\{\|\widehat{\beta}_k - \widehat{\beta}_m\|^2_\omega - \text{pen}_k\right\} \text{ and } \text{pen}_m := c\,\sigma^2_m\,\delta_m\,n^{-1}$$

- $\mathbb{V}\text{ar}(U + \langle\beta - \beta_m, X\rangle) \leqslant \sigma^2_m := 2\{\mathbb{E}Y^2 + [g]^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[g]_{\underline{m}}\}$
- $\delta_m := m\Delta_m \frac{\log(\Delta_m \vee m+2)}{\log(m+2)}, \quad \Delta_m := \max_{1\leqslant k\leqslant m} \|[\nabla_\omega]^{1/2}_{\underline{k}}[\Gamma]^{-1}_{\underline{k}}[\nabla_\omega]^{1/2}_{\underline{k}}\|_s,$
- $M := \arg\min_{2\leqslant m\leqslant M^\omega} \left\{m\|[\Gamma]^{-1}_{\underline{m}}\|_s\|[\nabla_\omega]_{\underline{m}}\|_s > \frac{n}{1+\log n}\right\} - 1$

Adaptive global estimation

$$\widehat{\beta}_{\widetilde{m}} := (f)^t_{\underline{\widetilde{m}}}[\widehat{\Gamma}]^{-1}_{\underline{\widetilde{m}}}[\widehat{g}]_{\underline{\widetilde{m}}}\,\mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{\widetilde{m}}}\|_s \leqslant n\}} \text{ with } \widetilde{m} := \arg\min_{1\leqslant m\leqslant M}\left\{\text{Contrast}_m + \text{pen}_m\right\},$$

$$\text{Contrast}_m = \sup_{m\leqslant k\leqslant M}\left\{\|\widehat{\beta}_k - \widehat{\beta}_m\|^2_\omega - \text{pen}_k\right\} \text{ and } \text{pen}_m := c\,\sigma^2_m\,\delta_m\,n^{-1}$$

- $\mathbb{V}\mathrm{ar}(U + \langle\beta - \beta_m, X\rangle) \leqslant \sigma^2_m := 2\{\mathbb{E}Y^2 + [g]^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[g]_{\underline{m}}\}$
- $\delta_m := m\Delta_m\frac{\log(\Delta_m \vee m+2)}{\log(m+2)}, \quad \Delta_m := \max_{1\leqslant k\leqslant m}\|[\nabla_\omega]^{1/2}_{\underline{k}}[\Gamma]^{-1}_{\underline{k}}[\nabla_\omega]^{1/2}_{\underline{k}}\|_s,$
- $M := \arg\min_{2\leqslant m\leqslant M^\omega}\left\{m\|[\Gamma]^{-1}_{\underline{m}}\|_s\|[\nabla_\omega]_{\underline{m}}\|_s > \frac{n}{1+\log n}\right\} - 1$

Theorem (Comte & JJ (2011)) Under Assumption A1-A5 we have:

- $\sup_{\Gamma\in\mathcal{G}^d_\gamma}\sup_{\beta\in\mathcal{F}^\rho_b}\mathbb{E}\|\widehat{\beta}_{\widetilde{m}} - \beta\|^2_\omega \lesssim \min_{1\leqslant m\leqslant M^-}\left\{\max\left(\frac{\omega_m}{b_m}, \frac{\delta_m}{n}\right)\right\} + \frac{C(\mathcal{G}^d_\gamma, \mathcal{F}^\rho_b)}{n}$

Adaptive global estimation

$$\widehat{\beta}_{\widetilde{m}} := (f)^t_{\underline{\widetilde{m}}}[\widehat{\Gamma}]^{-1}_{\underline{\widetilde{m}}}[\widehat{g}]_{\underline{\widetilde{m}}}\,\mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{\widetilde{m}}}\|_s \leqslant n\}} \text{ with } \widetilde{m} := \arg\min_{1\leqslant m\leqslant M}\left\{\text{Contrast}_m + \mathrm{pen}_m\right\},$$

$$\text{Contrast}_m = \sup_{m\leqslant k\leqslant M}\left\{\|\widehat{\beta}_k - \widehat{\beta}_m\|^2_\omega - \mathrm{pen}_k\right\} \text{ and } \mathrm{pen}_m := c\,\sigma^2_m\,\delta_m\,n^{-1}$$

- $\mathbb{V}\mathrm{ar}(U + \langle \beta - \beta_m, X\rangle) \leqslant \sigma^2_m := 2\{\mathbb{E}Y^2 + [g]^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[g]_{\underline{m}}\}$
- $\delta_m := m\Delta_m \frac{\log(\Delta_m \vee m+2)}{\log(m+2)}, \quad \Delta_m := \max_{1\leqslant k\leqslant m} \|[\nabla_\omega]^{1/2}_{\underline{k}}[\Gamma]^{-1}_{\underline{k}}[\nabla_\omega]^{1/2}_{\underline{k}}\|_s,$
- $M := \arg\min_{2\leqslant m\leqslant M^\omega}\left\{m\|[\Gamma]^{-1}_{\underline{m}}\|_s\|[\nabla_\omega]_{\underline{m}}\|_s > \frac{n}{1+\log n}\right\} - 1$

Theorem (Comte & JJ (2011)) Under Assumption A1-A5 we have:

- $\sup_{\Gamma\in\mathcal{G}^d_\gamma}\sup_{\beta\in\mathcal{F}^\rho_b}\mathbb{E}\|\widehat{\beta}_{\widetilde{m}} - \beta\|^2_\omega \lesssim \min_{1\leqslant m\leqslant M^-}\left\{\max\left(\frac{\omega_m}{b_m}, \frac{\delta_m}{n}\right)\right\} + \frac{C(\mathcal{G}^d_\gamma, \mathcal{F}^\rho_b)}{n}$

Remember: $R^*_\omega[n;\mathcal{F}^\rho_b,\mathcal{G}^d_\gamma] = \min_{1\leqslant m<\infty}\left\{\max\left(\frac{\omega_m}{b_m}, \sum_{j=1}^m \frac{\omega_j}{n\gamma_j}\right)\right\}$

Adaptive global estimation

$$\widehat{\beta}_{\widehat{m}} := (f)^t_{\underline{\widehat{m}}}[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}[\widehat{g}]_{\underline{\widehat{m}}}\,\mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}\|_s \leqslant n\}} \text{ with } \widehat{m} := \underset{1\leqslant m\leqslant \widehat{M}}{\arg\min}\left\{\text{Contrast}_m + \widehat{\text{pen}}_m\right\},$$

$$\text{Contrast}_m := \sup_{m\leqslant k\leqslant \widehat{M}}\left\{\|\widehat{\beta}_k - \widehat{\beta}_m\|^2_\omega - \widehat{\text{pen}}_k\right\} \text{ and } \widehat{\text{pen}}_m := 14c\,\widehat{\sigma}^2_m\,\widehat{\delta}_m\,n^{-1}$$

Adaptive global estimation

$$\widehat{\beta}_{\widehat{m}} := (f)^t_{\underline{\widehat{m}}}[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}[\widehat{g}]_{\underline{\widehat{m}}}\,\mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}\|_s \leqslant n\}} \text{ with } \widehat{m} := \arg\min_{1\leqslant m\leqslant \widehat{M}}\left\{\text{Contrast}_m + \widehat{\mathrm{pen}}_m\right\},$$

$$\text{Contrast}_m := \sup_{m\leqslant k\leqslant \widehat{M}}\left\{\|\widehat{\beta}_k - \widehat{\beta}_m\|^2_\omega - \widehat{\mathrm{pen}}_k\right\} \text{ and } \widehat{\mathrm{pen}}_m := 14c\,\widehat{\sigma}^2_m\,\widehat{\delta}_m\,n^{-1}$$

- $\sigma^2_m := 2\left\{\mathbb{E}Y^2 + [g]^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[g]_{\underline{m}}\right\}$

Adaptive global estimation

$$\widehat{\beta}_{\widehat{m}} := (f)^t_{\underline{\widehat{m}}}[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}[\widehat{g}]_{\underline{\widehat{m}}}\, \mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}\|_s \leqslant n\}} \text{ with } \widehat{m} := \arg\min_{1\leqslant m\leqslant \widehat{M}} \left\{\text{Contrast}_m + \widehat{\text{pen}}_m\right\},$$

$$\text{Contrast}_m := \sup_{m\leqslant k\leqslant \widehat{M}} \left\{\|\widehat{\beta}_k - \widehat{\beta}_m\|^2_\omega - \widehat{\text{pen}}_k\right\} \text{ and } \widehat{\text{pen}}_m := 14c\, \widehat{\sigma}^2_m\, \widehat{\delta}_m\, n^{-1}$$

- $\widehat{\sigma}^2_m := 2\left\{n^{-1}\sum_{i=1}^n Y_i^2 + [\widehat{g}]^t_{\underline{m}}[\widehat{\Gamma}]^{-1}_{\underline{m}}[\widehat{g}]_{\underline{m}}\right\}$

Adaptive global estimation

$$\widehat{\beta}_{\widehat{m}} := (f)^t_{\underline{\widehat{m}}}[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}[\widehat{g}]_{\underline{\widehat{m}}}\, \mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}\|_s \leqslant n\}} \text{ with } \widehat{m} := \arg\min_{1\leqslant m\leqslant \widehat{M}} \left\{\text{Contrast}_m + \widehat{\mathrm{pen}}_m\right\},$$

$$\text{Contrast}_m := \sup_{m\leqslant k\leqslant \widehat{M}} \left\{\|\widehat{\beta}_k - \widehat{\beta}_m\|^2_\omega - \widehat{\mathrm{pen}}_k\right\} \text{ and } \widehat{\mathrm{pen}}_m := 14c\,\widehat{\sigma}^2_m\,\widehat{\delta}_m\, n^{-1}$$

- $\widehat{\sigma}^2_m := 2\{n^{-1}\sum_{i=1}^n Y_i^2 + [\widehat{g}]^t_{\underline{m}}[\widehat{\Gamma}]^{-1}_{\underline{m}}[\widehat{g}]_{\underline{m}}\}$
- $\widehat{\delta}_m := m\widehat{\Delta}_m \frac{\log(\widehat{\Delta}_m \vee m+2)}{\log(m+2)}, \quad \widehat{\Delta}_m := \max_{1\leqslant k\leqslant m} \|[\nabla_\omega]^{1/2}_{\underline{k}}[\widehat{\Gamma}]^{-1}_{\underline{k}}[\nabla_\omega]^{1/2}_{\underline{k}}\|_s,$

Adaptive global estimation

$$\widehat{\beta}_{\widehat{m}} := (f)^t_{\underline{\widehat{m}}}[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}[\widehat{g}]_{\underline{\widehat{m}}}\, \mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}\|_s \leqslant n\}} \text{ with } \widehat{m} := \arg\min_{1\leqslant m\leqslant \widehat{M}} \left\{\mathrm{Contrast}_m + \widehat{\mathrm{pen}}_m\right\},$$

$$\mathrm{Contrast}_m := \sup_{m\leqslant k\leqslant \widehat{M}} \left\{\|\widehat{\beta}_k - \widehat{\beta}_m\|^2_\omega - \widehat{\mathrm{pen}}_k\right\} \text{ and } \widehat{\mathrm{pen}}_m := 14c\,\widehat{\sigma}^2_m\,\widehat{\delta}_m\, n^{-1}$$

- $\widehat{\sigma}^2_m := 2\{n^{-1}\sum_{i=1}^n Y_i^2 + [\widehat{g}]^t_{\underline{m}}[\widehat{\Gamma}]^{-1}_{\underline{m}}[\widehat{g}]_{\underline{m}}\}$
- $\widehat{\delta}_m := m\widehat{\Delta}_m \frac{\log(\widehat{\Delta}_m \vee m+2)}{\log(m+2)}, \quad \widehat{\Delta}_m := \max_{1\leqslant k\leqslant m} \|[\nabla_\omega]^{1/2}_{\underline{k}}[\widehat{\Gamma}]^{-1}_{\underline{k}}[\nabla_\omega]^{1/2}_{\underline{k}}\|_s,$
- $\widehat{M} := \arg\min_{2\leqslant m\leqslant M^\omega} \left\{m\|[\widehat{\Gamma}]^{-1}_{\underline{m}}\|_s\|[\nabla_\omega]_{\underline{m}}\|_s > \frac{n}{\log n}\right\} - 1$

Adaptive global estimation

$$\widehat{\beta}_{\widehat{m}} := (f)^t_{\underline{\widehat{m}}}[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}[\widehat{g}]_{\underline{\widehat{m}}}\, \mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}\|_s \leqslant n\}} \text{ with } \widehat{m} := \arg\min_{1\leqslant m\leqslant \widehat{M}} \left\{\text{Contrast}_m + \widehat{\text{pen}}_m\right\},$$

$$\text{Contrast}_m := \sup_{m\leqslant k\leqslant \widehat{M}} \left\{\|\widehat{\beta}_k - \widehat{\beta}_m\|^2_\omega - \widehat{\text{pen}}_k\right\} \text{ and } \widehat{\text{pen}}_m := 14c\,\widehat{\sigma}^2_m\,\widehat{\delta}_m\, n^{-1}$$

- $\widehat{\sigma}^2_m := 2\{n^{-1}\sum_{i=1}^n Y_i^2 + [\widehat{g}]^t_{\underline{m}}[\widehat{\Gamma}]^{-1}_{\underline{m}}[\widehat{g}]_{\underline{m}}\}$
- $\widehat{\delta}_m := m\widehat{\Delta}_m \frac{\log(\widehat{\Delta}_m \vee m+2)}{\log(m+2)}, \quad \widehat{\Delta}_m := \max_{1\leqslant k\leqslant m} \|[\nabla_\omega]^{1/2}_{\underline{k}}[\widehat{\Gamma}]^{-1}_{\underline{k}}[\nabla_\omega]^{1/2}_{\underline{k}}\|_s,$
- $\widehat{M} := \arg\min_{2\leqslant m\leqslant M^\omega} \left\{m\|[\widehat{\Gamma}]^{-1}_{\underline{m}}\|_s\|[\nabla_\omega]_{\underline{m}}\|_s > \frac{n}{\log n}\right\} - 1$

Theorem (Comte & JJ (2011)) Under Assumption A1-A5 we have:

- $$\sup_{\Gamma\in\mathcal{G}^d_\gamma}\sup_{\beta\in\mathcal{F}^\rho_b} \mathbb{E}\|\widehat{\beta}_{\widehat{m}} - \beta\|^2_\omega \lesssim \min_{1\leqslant m\leqslant M^-}\left\{\max\left(\frac{\omega_m}{b_m}, \frac{\delta_m}{n}\right)\right\} + \frac{C(\mathcal{G}^d_\gamma, \mathcal{F}^\rho_b)}{n}$$

Adaptive global estimation

$$\widehat{\beta}_{\widehat{m}} := (f)^t_{\underline{\widehat{m}}}[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}[\widehat{g}]_{\underline{\widehat{m}}}\,\mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}\|_s\leqslant n\}} \text{ with } \widehat{m} := \arg\min_{1\leqslant m\leqslant \widehat{M}}\left\{\text{Contrast}_m + \widehat{\text{pen}}_m\right\},$$

$$\text{Contrast}_m := \sup_{m\leqslant k\leqslant \widehat{M}}\left\{\|\widehat{\beta}_k - \widehat{\beta}_m\|^2_\omega - \widehat{\text{pen}}_k\right\} \text{ and } \widehat{\text{pen}}_m := 14c\,\widehat{\sigma}^2_m\,\widehat{\delta}_m\,n^{-1}$$

- $\widehat{\sigma}^2_m := 2\{n^{-1}\sum_{i=1}^n Y_i^2 + [\widehat{g}]^t_{\underline{m}}[\widehat{\Gamma}]^{-1}_{\underline{m}}[\widehat{g}]_{\underline{m}}\}$
- $\widehat{\delta}_m := m\widehat{\Delta}_m\frac{\log(\widehat{\Delta}_m\vee m+2)}{\log(m+2)}, \quad \widehat{\Delta}_m := \max_{1\leqslant k\leqslant m}\|[\nabla_\omega]^{1/2}_{\underline{k}}[\widehat{\Gamma}]^{-1}_{\underline{k}}[\nabla_\omega]^{1/2}_{\underline{k}}\|_s,$
- $\widehat{M} := \arg\min_{2\leqslant m\leqslant M^\omega}\left\{m\|[\widehat{\Gamma}]^{-1}_{\underline{m}}\|_s\|[\nabla_\omega]_{\underline{m}}\|_s > \frac{n}{\log n}\right\} - 1$

Theorem (Comte & JJ (2011)) Under Assumption A1-A5 we have:

- $$\sup_{\Gamma\in\mathcal{G}^d_\gamma}\sup_{\beta\in\mathcal{F}^\rho_b}\mathbb{E}\|\widehat{\beta}_{\widehat{m}} - \beta\|^2_\omega \lesssim \min_{1\leqslant m\leqslant M^-}\left\{\max\left(\frac{\omega_m}{b_m}, \frac{\delta_m}{n}\right)\right\} + \frac{C(\mathcal{G}^d_\gamma,\mathcal{F}^\rho_b)}{n}$$

Remember: $R^*_\omega[n;\mathcal{F}^\rho_b,\mathcal{G}^d_\gamma] = \min_{1\leqslant m<\infty}\left\{\max\left(\frac{\omega_m}{b_m}, \sum_{j=1}^m \frac{\omega_j}{n\gamma_j}\right)\right\}$

Illustration

Let $\{f_j\}$ be the trigonometric basis in $L^2[0,1]$ and $\mathcal{W}_p^\rho$ an ellipsoid in the *Sobolev space of p-times differentiable periodic functions*.

Illustration

Let $\{f_j\}$ be the trigonometric basis in $L^2[0,1]$ and $\mathcal{W}_p^\rho$ an ellipsoid in the *Sobolev space of p-times differentiable periodic functions*.

$$\widehat{\beta}_{\widehat{m}} := (f)^t_{\underline{\widehat{m}}}[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}[\widehat{g}]_{\underline{\widehat{m}}}\,\mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}\|_s \leqslant n\}} \text{ with } \widehat{m} := \arg\min_{1\leqslant m\leqslant \widehat{M}} \left\{\text{Contrast}_m + \widehat{\mathrm{pen}}_m\right\},$$

$$\text{Contrast}_m := \sup_{m\leqslant k\leqslant \widehat{M}} \left\{\|\widehat{\beta}^{(s)}_k - \widehat{\beta}^{(s)}_m\|^2 - \widehat{\mathrm{pen}}_k\right\}.$$

Illustration

Let $\{f_j\}$ be the trigonometric basis in $L^2[0,1]$ and $\mathcal{W}_p^\rho$ an ellipsoid in the *Sobolev space of p-times differentiable periodic functions*.

$$\widehat{\beta}_{\widehat{m}} := (f)^t_{\underline{\widehat{m}}}[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}[\widehat{g}]_{\underline{\widehat{m}}}\, \mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}\|_s \leqslant n\}} \text{ with } \widehat{m} := \mathop{\arg\min}_{1\leqslant m\leqslant \widehat{M}} \left\{\text{Contrast}_m + \widehat{\text{pen}}_m\right\},$$

$$\text{Contrast}_m := \sup_{m\leqslant k\leqslant \widehat{M}} \left\{\|\widehat{\beta}^{(s)}_k - \widehat{\beta}^{(s)}_m\|^2 - \widehat{\text{pen}}_k\right\}.$$

Minimax-optimality. Under Assumption A1-A2 we have in case of

(p) for $p + a \geqslant 2$

- global: $\sup_{\Gamma\in\mathcal{G}_\gamma^d} \sup_{\beta\in\mathcal{W}_p^\rho} \mathbb{E}\|\widehat{\beta}^{(s)}_{\widehat{m}} - \beta^{(s)}\|^2 \lesssim n^{-2(p-s)/(2p+2a+1)}$.

(e)

- global: $\sup_{\Gamma\in\mathcal{G}_\gamma^d} \sup_{\beta\in\mathcal{W}_p^\rho} \mathbb{E}\|\widehat{\beta}^{(s)}_{\widehat{m}} - \beta^{(s)}\|^2 \lesssim (\log n)^{-(p-s)/a}$.

Outline

- Methodology
- Background and model assumptions
- Minimax theory
  - Measure of performance
  - Lower bound: global and local risk
  - Minimax-optimal estimation
- Adaptive estimation combining model selection and Lepski's method
  - Adaptive global estimation
  - Adaptive local estimation

Adaptive local estimation

$$\widehat{\ell}_{\widehat{m}} := [\ell]^t_{\underline{\widehat{m}}}[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}[\widehat{g}]_{\underline{\widehat{m}}}\,\mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}\|_s \leqslant n\}} \text{ with } \widehat{m} := \arg\min_{1\leqslant m\leqslant \widehat{M}} \left\{\text{Contrast}_m + \widehat{\text{pen}}_m\right\},$$

Adaptive local estimation

$$\widehat{\ell}_{\widehat{m}} := [\ell]^t_{\underline{\widehat{m}}}[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}[\widehat{g}]_{\underline{\widehat{m}}}\,\mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}\|_s \leqslant n\}} \text{ with } \widehat{m} := \arg\min_{1\leqslant m\leqslant \widehat{M}} \left\{\text{Contrast}_m + \widehat{\mathrm{pen}}_m\right\},$$

$$\text{Contrast}_m := \sup_{m\leqslant k\leqslant \widehat{M}} \left\{|\widehat{\ell}_k - \widehat{\ell}_m|^2 - \widehat{\mathrm{pen}}_k\right\} \text{ and } \widehat{\mathrm{pen}}_m := 7c\,\widehat{\sigma}^2_m\,\widehat{\delta}_m\,\frac{1+\log n}{n}$$

Adaptive local estimation

$$\widehat{\ell}_{\widehat{m}} := [\ell]^t_{\underline{\widehat{m}}}[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}[\widehat{g}]_{\underline{\widehat{m}}}\,\mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}\|_s \leqslant n\}} \text{ with } \widehat{m} := \arg\min_{1\leqslant m\leqslant \widehat{M}} \left\{\text{Contrast}_m + \widehat{\mathrm{pen}}_m\right\},$$

$$\text{Contrast}_m := \sup_{m\leqslant k\leqslant \widehat{M}} \left\{|\widehat{\ell}_k - \widehat{\ell}_m|^2 - \widehat{\mathrm{pen}}_k\right\} \text{ and } \widehat{\mathrm{pen}}_m := 7c\,\widehat{\sigma}^2_m\,\widehat{\delta}_m\,\frac{1+\log n}{n}$$

- $\widehat{\sigma}^2_m := 2\{n^{-1}\sum_{i=1}^n Y_i^2 + [\widehat{g}]^t_{\underline{m}}[\widehat{\Gamma}]^{-1}_{\underline{m}}[\widehat{g}]_{\underline{m}}\}$

Adaptive local estimation

$$\widehat{\ell}_{\widehat{m}} := [\ell]^t_{\underline{\widehat{m}}} [\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}} [\widehat{g}]_{\underline{\widehat{m}}} \mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}\|_s \leqslant n\}} \text{ with } \widehat{m} := \arg\min_{1\leqslant m \leqslant \widehat{M}} \left\{ \text{Contrast}_m + \widehat{\text{pen}}_m \right\},$$

$$\text{Contrast}_m := \sup_{m\leqslant k\leqslant \widehat{M}} \left\{ |\widehat{\ell}_k - \widehat{\ell}_m|^2 - \widehat{\text{pen}}_k \right\} \text{ and } \widehat{\text{pen}}_m := 7c\, \widehat{\sigma}^2_m\, \widehat{\delta}_m\, \frac{1+\log n}{n}$$

- $\widehat{\sigma}^2_m := 2\{n^{-1}\sum_{i=1}^n Y_i^2 + [\widehat{g}]^t_{\underline{m}} [\widehat{\Gamma}]^{-1}_{\underline{m}} [\widehat{g}]_{\underline{m}}\}$
- $\widehat{\delta}_m := \max_{1\leqslant k\leqslant m} [\ell]^t_{\underline{k}} [\widehat{\Gamma}]^{-1}_{\underline{k}} [\ell]_{\underline{k}},$

Adaptive local estimation

$$\widehat{\ell}_{\widehat{m}} := [\ell]^t_{\underline{\widehat{m}}}[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}[\widehat{g}]_{\underline{\widehat{m}}}\, \mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}\|_s \leqslant n\}} \text{ with } \widehat{m} := \underset{1\leqslant m\leqslant \widehat{M}}{\arg\min}\left\{\text{Contrast}_m + \widehat{\text{pen}}_m\right\},$$

$$\text{Contrast}_m := \sup_{m\leqslant k\leqslant \widehat{M}} \left\{|\widehat{\ell}_k - \widehat{\ell}_m|^2 - \widehat{\text{pen}}_k\right\} \text{ and } \widehat{\text{pen}}_m := 7c\,\widehat{\sigma}^2_m\,\widehat{\delta}_m\,\frac{1+\log n}{n}$$

- $\widehat{\sigma}^2_m := 2\{n^{-1}\sum_{i=1}^n Y_i^2 + [\widehat{g}]^t_{\underline{m}}[\widehat{\Gamma}]^{-1}_{\underline{m}}[\widehat{g}]_{\underline{m}}\}$
- $\widehat{\delta}_m := \max_{1\leqslant k\leqslant m} [\ell]^t_{\underline{k}}[\widehat{\Gamma}]^{-1}_{\underline{k}}[\ell]_{\underline{k}}$,
- $\widehat{M} := \underset{2\leqslant m\leqslant M^\ell}{\arg\min}\left\{\|[\widehat{\Gamma}]^{-1}_{\underline{m}}\|_s\,[\ell]^t_{\underline{m}}[\ell]_{\underline{m}} > \frac{n}{1+\log n}\right\} - 1$

Adaptive local estimation

$$\widehat{\ell}_{\widehat{m}} := [\ell]^t_{\underline{\widehat{m}}}[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}[\widehat{g}]_{\underline{\widehat{m}}}\,\mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}\|_s \leqslant n\}} \text{ with } \widehat{m} := \arg\min_{1\leqslant m\leqslant \widehat{M}} \left\{\text{Contrast}_m + \widehat{\mathrm{pen}}_m\right\},$$

$$\text{Contrast}_m := \sup_{m\leqslant k\leqslant \widehat{M}} \left\{|\widehat{\ell}_k - \widehat{\ell}_m|^2 - \widehat{\mathrm{pen}}_k\right\} \text{ and } \widehat{\mathrm{pen}}_m := 7c\,\widehat{\sigma}^2_m\,\widehat{\delta}_m\,\frac{1+\log n}{n}$$

- $\widehat{\sigma}^2_m := 2\{n^{-1}\sum_{i=1}^n Y_i^2 + [\widehat{g}]^t_{\underline{m}}[\widehat{\Gamma}]^{-1}_{\underline{m}}[\widehat{g}]_{\underline{m}}\}$
- $\widehat{\delta}_m := \max_{1\leqslant k\leqslant m} [\ell]^t_{\underline{k}}[\widehat{\Gamma}]^{-1}_{\underline{k}}[\ell]_{\underline{k}},$
- $\widehat{M} := \arg\min_{2\leqslant m\leqslant M^\ell} \left\{\|[\widehat{\Gamma}]^{-1}_{\underline{m}}\|_s\,[\ell]^t_{\underline{m}}[\ell]_{\underline{m}} > \frac{n}{1+\log n}\right\} - 1$

Theorem (JJ & Schenk (2011)) Under Assumption A1-A5 we have:

- $\sup_{\Gamma\in\mathcal{G}^d_\gamma}\sup_{\beta\in\mathcal{F}^\rho_b} \mathbb{E}|\widehat{\ell}_{\widehat{m}} - \ell(\beta)|^2_\omega \lesssim R^\diamond_\ell[\frac{n}{1+\log n};\mathcal{F}^\rho_b,\mathcal{G}^d_\gamma] + \frac{C(\mathcal{G}^d_\gamma,\mathcal{F}^\rho_b)}{n}$

Adaptive local estimation

$$\widehat{\ell}_{\widehat{m}} := [\ell]^t_{\underline{\widehat{m}}}[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}[\widehat{g}]_{\underline{\widehat{m}}}\,\mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}\|_s \leqslant n\}} \text{ with } \widehat{m} := \mathop{\arg\min}_{1\leqslant m\leqslant \widehat{M}} \Big\{\text{Contrast}_m + \widehat{\text{pen}}_m\Big\},$$

$$\text{Contrast}_m := \sup_{m\leqslant k\leqslant \widehat{M}} \Big\{|\widehat{\ell}_k - \widehat{\ell}_m|^2 - \widehat{\text{pen}}_k\Big\} \text{ and } \widehat{\text{pen}}_m := 7c\,\widehat{\sigma}^2_m\,\widehat{\delta}_m\,\frac{1+\log n}{n}$$

- $\widehat{\sigma}^2_m := 2\{n^{-1}\sum_{i=1}^n Y_i^2 + [\widehat{g}]^t_{\underline{m}}[\widehat{\Gamma}]^{-1}_{\underline{m}}[\widehat{g}]_{\underline{m}}\}$
- $\widehat{\delta}_m := \max_{1\leqslant k\leqslant m} [\ell]^t_{\underline{k}}[\widehat{\Gamma}]^{-1}_{\underline{k}}[\ell]_{\underline{k}}$,
- $\widehat{M} := \mathop{\arg\min}_{2\leqslant m\leqslant M^\ell} \Big\{\|[\widehat{\Gamma}]^{-1}_{\underline{m}}\|_s\,[\ell]^t_{\underline{m}}[\ell]_{\underline{m}} > \frac{n}{1+\log n}\Big\} - 1$

Theorem (JJ & Schenk (2011)) Under Assumption A1-A5 we have:

- $\sup_{\Gamma\in\mathcal{G}^d_\gamma}\sup_{\beta\in\mathcal{F}^\rho_b} \mathbb{E}|\widehat{\ell}_{\widehat{m}} - \ell(\beta)|^2_\omega \lesssim R^\diamond_\ell[\frac{n}{1+\log n};\mathcal{F}^\rho_b,\mathcal{G}^d_\gamma] + \frac{C(\mathcal{G}^d_\gamma,\mathcal{F}^\rho_b)}{n}$

Remember: $R^\diamond_\ell[n;\mathcal{F}^\rho_b,\mathcal{G}^d_\gamma]$ is the minimax-rate.

Illustration

Let $\{f_j\}$ be the trigonometric basis in $L^2[0,1]$ and $\mathcal{W}_p^\rho$ an ellipsoid in the *Sobolev space of p-times differentiable periodic functions*.

Illustration

Let $\{f_j\}$ be the trigonometric basis in $L^2[0,1]$ and $\mathcal{W}_p^\rho$ an ellipsoid in the *Sobolev space of p-times differentiable periodic functions*.

$$\widehat{\ell}_{\widehat{m}} := [\ell]^t_{\underline{\widehat{m}}}[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}[\widehat{g}]_{\underline{\widehat{m}}}\, \mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}\|_s \leqslant n\}} \text{ with } \widehat{m} := \arg\min_{1\leqslant m\leqslant \widehat{M}} \big\{\text{Contrast}_m + \widehat{\mathrm{pen}}_m\big\},$$

$$\text{Contrast}_m := \sup_{m\leqslant k\leqslant \widehat{M}} \big\{|\widehat{\ell}_k - \widehat{\ell}_m|^2 - \widehat{\mathrm{pen}}_k\big\}$$

Illustration

Let $\{f_j\}$ be the trigonometric basis in $L^2[0,1]$ and $\mathcal{W}_p^\rho$ an ellipsoid in the *Sobolev space of p-times differentiable periodic functions*.

$$\widehat{\ell}_{\widehat{m}} := [\ell]^t_{\underline{\widehat{m}}}[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}[\widehat{g}]_{\underline{\widehat{m}}}\,\mathbb{1}_{\{\|[\widehat{\Gamma}]^{-1}_{\underline{\widehat{m}}}\|_s \leqslant n\}} \text{ with } \widehat{m} := \underset{1\leqslant m \leqslant \widehat{M}}{\arg\min}\left\{\text{Contrast}_m + \widehat{\mathrm{pen}}_m\right\},$$

$$\text{Contrast}_m := \sup_{m\leqslant k\leqslant \widehat{M}} \left\{|\widehat{\ell}_k - \widehat{\ell}_m|^2 - \widehat{\mathrm{pen}}_k\right\}$$

Minimax-optimality. Under Assumption A1-A2 we have in case of

(p) for $p + a \geqslant 2$

- local: $\sup_{\Gamma\in\mathcal{G}_\gamma^d}\sup_{\beta\in\mathcal{W}_p^\rho}\mathbb{E}|\widehat{\beta}_{\widehat{m}}(t_0) - \beta(t_0)|^2 \lesssim \left(\frac{n}{\log n}\right)^{-(2p-1)/(2p+2a)}$.

(e)

- local: $\sup_{\Gamma\in\mathcal{G}_\gamma^d}\sup_{\beta\in\mathcal{W}_p^\rho}\mathbb{E}|\widehat{\beta}_{\widehat{m}}(t_0) - \beta(t_0)|^2 \lesssim (\log n)^{-(2p-1)/(2a)}$.

Summary

In functional linear regression: $Y = \langle \beta, X \rangle + U$ and $\mathbb{E}(UX) = 0$

- a lower bound is derived globally and locally considering (abstract) smoothness conditions and general link conditions;

Summary

In functional linear regression: $Y = \langle \beta, X \rangle + U$ and $\mathbb{E}(UX) = 0$

- a lower bound is derived globally and locally considering (abstract) smoothness conditions and general link conditions;
- an estimator of β is proposed and its minimax-optimality is shown
  - requiring an optimal choice of a dimension parameter

Summary

In functional linear regression: $Y = \langle \beta, X \rangle + U$ and $\mathbb{E}(UX) = 0$

- a lower bound is derived globally and locally considering (abstract) smoothness conditions and general link conditions;
- an estimator of β is proposed and its minimax-optimality is shown
  - requiring an optimal choice of a dimension parameter
- a minimax-optimal data-driven estimator of β is constructed
  - combining model selection and Lepski's method;

Summary

In functional linear regression: $Y = \langle \beta, X \rangle + U$ and $\mathbb{E}(UX) = 0$

- a lower bound is derived globally and locally considering (abstract) smoothness conditions and general link conditions;
- an estimator of β is proposed and its minimax-optimality is shown
  - requiring an optimal choice of a dimension parameter
- a minimax-optimal data-driven estimator of β is constructed
  - combining model selection and Lepski's method;
- the results are illustrated using classical smoothness assumptions.

Summary

In functional linear regression: $Y = \langle \beta, X \rangle + U$ and $\mathbb{E}(UX) = 0$

- a lower bound is derived globally and locally considering (abstract) smoothness conditions and general link conditions;
- an estimator of β is proposed and its minimax-optimality is shown
  - requiring an optimal choice of a dimension parameter
- a minimax-optimal data-driven estimator of β is constructed
  - combining model selection and Lepski's method;
- the results are illustrated using classical smoothness assumptions.

Extensions:

- ♢ Sparse irregular repeated noisy measurements of $X(\cdot)$;
- ♢ Structured or unstructured sparse representation of $\beta(\cdot)$;
- ♢ Observational dependence.

References

Cardot & Johannes (2010)
Thresholding projection estimators in functional linear models.
Journal of Multivariate Analysis, 101:395–408

Johannes & Schenk (2012)
On rate optimal local estimation in functional linear regression.
Discussion paper, Université catholique de Louvain (under revision).

Comte & Johannes (2012)
Adaptive functional linear regression.
To appear in The Annals of Statistics.

Johannes & Schenk (2012)
Adaptive estimation of linear functionals in functional linear models.
To appear in Mathematical Methods of Statistics.