Inverse problems in statistics

Laurent Cavalier

(Université Aix-Marseille 1, France)

Introduction

There exist many fields where inverse problems appear

- Astronomy (Hubble satellite).
- Econometrics (instrumental variables).
- Financial mathematics (model calibration).
- Medical image processing (X-rays).

Introduction

There exist many fields where inverse problems appear

- Astronomy (Hubble satellite).
- Econometrics (instrumental variables).
- Financial mathematics (model calibration).
- Medical image processing (X-rays).

These are problems where we have indirect observations of an object (a function) that we want to reconstruct.

Inverse problems

Let H et G be Hilbert spaces.

Let A be a continuous linear operator from H into G.

Inverse problems

Let H et G be Hilbert spaces.

Let A be a continuous linear operator from H into G.

Given $g \in G$ find $f \in H$ such that $Af = g$.

Inverse problems

Let H et G be Hilbert spaces.

Let A be a continuous linear operator from H into G.

Given $g \in G$ find $f \in H$ such that $Af = g$.

Solving an inverse problem $\rightarrow$ "Inversion" of the operator A.

Inverse problems

Let H et G be Hilbert spaces.

Let A be a continuous linear operator from H into G.

Given $g \in G$ find $f \in H$ such that $Af = g$.

Solving an inverse problem $\rightarrow$ "Inversion" of the operator A.
If A^{-1} is not continuous the problem is called ill-posed.

Inverse problems

Let H et G be Hilbert spaces.

Let A be a continuous linear operator from H into G.

Given $g \in G$ find $f \in H$ such that $Af = g$.

Solving an inverse problem $\rightarrow$ "Inversion" of the operator A.
If A^{-1} is not continuous the problem is called ill-posed.

Observe g^ε a noisy version of g, then $f^\varepsilon = A^{-1}g^\varepsilon$ could be far from f.

Inverse problems

Let H et G be Hilbert spaces.

Let A be a continuous linear operator from H into G.

Given $g \in G$ find $f \in H$ such that $Af = g$.

Solving an inverse problem $\rightarrow$ "Inversion" of the operator A.
If A^{-1} is not continuous the problem is called ill-posed.

Observe g^ε a noisy version of g, then $f^\varepsilon = A^{-1} g^\varepsilon$ could be far from f.

Importance of the notion of "noise" or "error".

Linear inverse problems

Let H and G two separable Hilbert spaces. Let A be a known linear bounded operator from the space H to G.

Linear inverse problems

Let H and G two separable Hilbert spaces. Let A be a known linear bounded operator from the space H to G.

Let the model :

$$Y = Af + \varepsilon\xi,$$

where Y is the observation, $f \in H$ unknown ,
A a continuous linear operator fom H into G,
ξ is a white noise, ε corresponds to the noise level.

Linear inverse problems

Let H and G two separable Hilbert spaces. Let A be a known linear bounded operator from the space H to G.

Let the model :

$$Y = Af + \varepsilon\xi,$$

where Y is the observation, $f \in H$ unknown ,
A a continuous linear operator fom H into G,
ξ is a white noise, ε corresponds to the noise level.

Reconstruct (estimate) f with the observation Y.

Linear inverse problems

Let H and G two separable Hilbert spaces. Let A be a known linear bounded operator from the space H to G.

Let the model :

$$Y = Af + \varepsilon\xi,$$

where Y is the observation, $f \in H$ unknown ,
A a continuous linear operator fom H into G,
ξ is a white noise, ε corresponds to the noise level.

Reconstruct (estimate) f with the observation Y.

Projection of a white noise on any orthonormal basis $\{\psi_k\}$ gives a sequence of i.i.d . standard Gaussian random variables.

Discrete model of inverse problems

The standard discrete sample statistical model for linear inverse problems is

$$Y_i = Af(X_i) + \xi_i,\ i = 1, \ldots, n,$$

where $(X_1, Y_1), .., (X_n, Y_n)$ are observed (we may assume $X_i \in [0, 1]$), f is an unknown function in $L^2(0, 1)$, A is an operator from $L^2(0, 1)$ into $L^2(0, 1)$, and ξ_i are i.i.d. zero-mean Gaussian random variables of variance σ^2.

Discrete model of inverse problems

The standard discrete sample statistical model for linear inverse problems is

$$Y_i = Af(X_i) + \xi_i, \; i = 1, \dots, n,$$

where $(X_1, Y_1), .., (X_n, Y_n)$ are observed (we may assume $X_i \in [0, 1]$), f is an unknown function in $L^2(0, 1)$, A is an operator from $L^2(0, 1)$ into $L^2(0, 1)$, and ξ_i are i.i.d. zero-mean Gaussian random variables of variance σ^2.

Noise level is related to number of observations by

$$\varepsilon \approx 1/\sqrt{n}.$$

Singular value decomposition

A major property of compact operators is that they have a discrete spectrum.

Singular value decomposition

A major property of compact operators is that they have a discrete spectrum.

Suppose A^*A compact operator with a known basis of eigenfunctions in H:

$$A^*A\varphi_k = b_k^2\varphi_k.$$

Singular value decomposition

A major property of compact operators is that they have a discrete spectrum.

Suppose A^*A compact operator with a known basis of eigenfunctions in H:

$$A^*A\varphi_k = b_k^2\varphi_k.$$

Singular Value Decomposition (SVD) of A :

Singular value decomposition

A major property of compact operators is that they have a discrete spectrum.

Suppose A^*A compact operator with a known basis of eigenfunctions in H:

$$A^*A\varphi_k = b_k^2\varphi_k.$$

Singular Value Decomposition (SVD) of A :

$$A\varphi_k = b_k\psi_k, \quad A^*\psi_k = b_k\varphi_k,$$

where $b_k > 0$ are the singular values,
$\{\varphi_k\}$ o.n.b. on H, $\{\psi_k\}$ o.n.b. on G.

Singular value decomposition

A major property of compact operators is that they have a discrete spectrum.

Suppose A^*A compact operator with a known basis of eigenfunctions in H:

$$A^*A\varphi_k = b_k^2\varphi_k.$$

Singular Value Decomposition (SVD) of A :

$$A\varphi_k = b_k\psi_k, \quad A^*\psi_k = b_k\varphi_k,$$

where $b_k > 0$ are the singular values,
$\{\varphi_k\}$ o.n.b. on H, $\{\psi_k\}$ o.n.b. on G.
A linear bounded compact operator between two Hilbert spaces may really be seen as an infinite matrix.

Projection on $\{\psi_k\}$

Projection of Y on $\{\psi_k\}$:

Projection on $\{\psi_k\}$

Projection of Y on $\{\psi_k\}$:

$$\langle Y, \psi_k \rangle = \langle Af, \psi_k \rangle + \varepsilon \langle \xi, \psi_k \rangle$$

Projection on $\{\psi_k\}$

Projection of Y on $\{\psi_k\}$:

$$\langle Y, \psi_k \rangle = \langle f, A^* \psi_k \rangle + \varepsilon \langle \xi, \psi_k \rangle$$

Projection on $\{\psi_k\}$

Projection of Y on $\{\psi_k\}$:

$$\langle Y, \psi_k \rangle = \langle f, A^* \psi_k \rangle + \varepsilon \langle \xi, \psi_k \rangle$$

$$= b_k \langle f, \varphi_k \rangle + \varepsilon \langle \xi, \psi_k \rangle$$

Projection on $\{\psi_k\}$

Projection of Y on $\{\psi_k\}$:

$$\langle Y, \psi_k \rangle = \langle f, A^* \psi_k \rangle + \varepsilon \langle \xi, \psi_k \rangle$$

$$= b_k \langle f, \varphi_k \rangle + \varepsilon \xi_k$$

where $\{\xi_k\}$ standard Gaussian sequence i.i.d., by projection of a white noise ξ on the o.n.b. $\{\psi_k\}$.

Sequence space model

Equivalent Sequence space model

$$y_k = b_k\theta_k + \varepsilon\xi_k, \quad k = 1, 2, \ldots,$$

where $\{\theta_k\}$ coefficients of f, $\xi_k \sim \mathcal{N}(0,1)$ i.i.d.,
$b_k \to 0$ singular values.

Sequence space model

Equivalent Sequence space model

$$y_k = b_k \theta_k + \varepsilon \xi_k, \quad k = 1, 2, \ldots,$$

where $\{\theta_k\}$ coefficients of f, $\xi_k \sim \mathcal{N}(0,1)$ i.i.d.,
$b_k \to 0$ singular values.

Estimate $\theta = \{\theta_k\}$ with the observation $Y = \{Y_k\}$.
Use L_2-risk, it is equivalent to estimate f.

Sequence space model

Equivalent Sequence space model

$$y_k = b_k\theta_k + \varepsilon\xi_k, \quad k = 1, 2, \dots,$$

where $\{\theta_k\}$ coefficients of f, $\xi_k \sim \mathcal{N}(0,1)$ i.i.d.,
$b_k \to 0$ singular values.

Estimate $\theta = \{\theta_k\}$ with the observation $Y = \{Y_k\}$.
Use L_2-risk, it is equivalent to estimate f.

Remark that $b_k \to 0$ weaken the signal θ_k.

Sequence space model

Equivalent Sequence space model

$$y_k = b_k \theta_k + \varepsilon \xi_k, \quad k = 1, 2, \dots,$$

where $\{\theta_k\}$ coefficients of f, $\xi_k \sim \mathcal{N}(0,1)$ i.i.d.,
$b_k \to 0$ singular values.

Estimate $\theta = \{\theta_k\}$ with the observation $Y = \{Y_k\}$.
Use L_2-risk, it is equivalent to estimate f.

Remark that $b_k \to 0$ weaken the signal θ_k.

$\longrightarrow$ Ill-posed problem.

Inversion

We have to “invert” in some sense the operator A.

Inversion

We have to “invert” in some sense the operator A.
Thus, we obtain the model :

$$X_k = b_k^{-1} y_k = \theta_k + \varepsilon \sigma_k \xi_k, \quad k = 1, 2, \ldots .$$

where $\sigma_k = b_k^{-1}$.

Inversion

We have to “invert” in some sense the operator A.
Thus, we obtain the model :

$$X_k = b_k^{-1} y_k = \theta_k + \varepsilon \sigma_k \xi_k, \quad k = 1, 2, \dots .$$

where $\sigma_k = b_k^{-1}$.
In the case where the problem is ill-posed the variance term grows to infinity.

Inversion

We have to “invert” in some sense the operator A.
Thus, we obtain the model :

$$X_k = b_k^{-1} y_k = \theta_k + \varepsilon \sigma_k \xi_k, \quad k = 1, 2, \dots .$$

where $\sigma_k = b_k^{-1}$.

In the case where the problem is ill-posed the variance term grows to infinity.

In this model the aim is to estimate $\{\theta_k\}$ by use of $\{X_k\}$.
When k is large the noise in X_k may then be very large, making the estimation difficult.

Inversion

We have to “invert” in some sense the operator A.
Thus, we obtain the model :

$$X_k = b_k^{-1} y_k = \theta_k + \varepsilon \sigma_k \xi_k, \quad k = 1, 2, \ldots.$$

where $\sigma_k = b_k^{-1}$.

In the case where the problem is ill-posed the variance term grows to infinity.

In this model the aim is to estimate $\{\theta_k\}$ by use of $\{X_k\}$.
When k is large the noise in X_k may then be very large, making the estimation difficult.

(see Donoho (1995), Mair and Ruymgaart (1996), Johnstone (1999) and C. and Tsybakov (2002)...).

Difficulty of inverse problems

Difficulty of inverse problems

- $\sigma_k \equiv 1$:

Direct problem.

Difficulty of inverse problems

- $\sigma_k \equiv 1$:

  Direct problem.

- $\sigma_k \sim k^\beta, \ \beta > 0$:

  Mildly ill-posed problem.

Difficulty of inverse problems

- $\sigma_k \equiv 1$:

  Direct problem.

- $\sigma_k \sim k^\beta, \ \beta > 0$:

  Mildly ill-posed problem.

- $\sigma_k \sim \exp(\beta k), \ \beta > 0$:

  Severely ill-posed problem.

Difficulty of inverse problems

- $\sigma_k \equiv 1$:

  Direct problem.

- $\sigma_k \sim k^\beta, \ \beta > 0$:

  Mildly ill-posed problem.

- $\sigma_k \sim \exp(\beta k), \ \beta > 0$:

  Severely ill-posed problem.

Parameter β is called degree of ill-posedness.

Examples

There exist many examples of operators for which the SVD is known:

Examples

There exist many examples of operators for which the SVD is known:

- Convolution.

Examples

There exist many examples of operators for which the SVD is known:

- Convolution.
- Tomography.

Examples

There exist many examples of operators for which the SVD is known:

- Convolution.
- Tomography.
- Instrumental variables.

Circular convolution

The framework of deconvolution is perhaps one of the most well-known inverse problem. It is used in many applications as econometrics, physics, astronomy, medical image processing. For example, it corresponds to the problem of a blurred signal that one wants to recover from indirect data.

Circular convolution

The framework of deconvolution is perhaps one of the most well-known inverse problem. It is used in many applications as econometrics, physics, astronomy, medical image processing. For example, it corresponds to the problem of a blurred signal that one wants to recover from indirect data.

Consider the following convolution operator

$$Af(t) = r * f(t) = \int_0^1 r(t-x)f(x)dx, \quad x \in [0,1],$$

where r is a known 1-periodic symetric real convolution kernel in $L^2[0,1]$. In this model, A is a linear bounded self-adjoint operator from $L^2[0,1]$ to $L^2[0,1]$.

Blurred cameraman

(a)

(b)

Convolution model

Define then the following model

$$Y(t) = r * f(t) + \varepsilon \, \xi(t), \quad x \in [0,1],$$

where Y is observed, f is an unknown periodic function in $L^2[0,1]$ and $\xi(t)$ is a white noise on $L^2[0,1]$.

Convolution model

Define then the following model

$$Y(t) = r * f(t) + \varepsilon\, \xi(t), \quad x \in [0, 1],$$

where Y is observed, f is an unknown periodic function in $L^2[0, 1]$ and $\xi(t)$ is a white noise on $L^2[0, 1]$.

The SVD basis is then clearly here the Fourier basis $\{\varphi_k(t)\}$.

Convolution model

Define then the following model

$$Y(t) = r * f(t) + \varepsilon\, \xi(t), \quad x \in [0,1],$$

where Y is observed, f is an unknown periodic function in $L^2[0,1]$ and $\xi(t)$ is a white noise on $L^2[0,1]$.

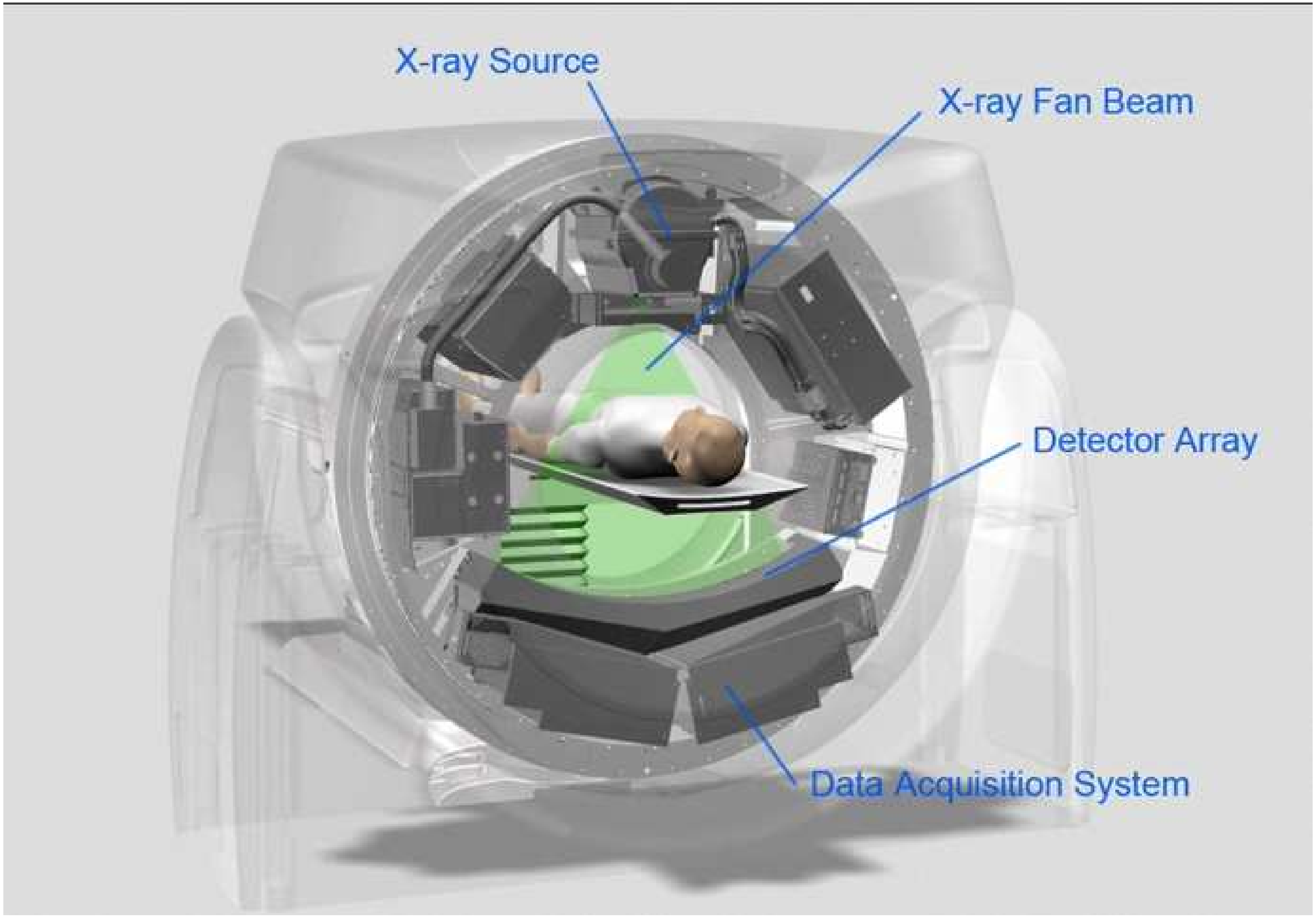
The SVD basis is then clearly here the Fourier basis $\{\varphi_k(t)\}$.

We make the projection on $\{\varphi_k(t)\}$, in the Fourier domain, and obtain

$$y_k = b_k \theta_k + \varepsilon \xi_k,$$

where $b_k = \sqrt{2}\int_0^1 r(x)\cos(2\pi k x)dx$ for even k, θ_k are the Fourier coefficients of f, and ξ_k are i.i.d. $\mathcal{N}(0,1)$.

Tomography scan



Instrumental variables

An economic relationship between a response variable Y and a vector X of explanatory variables is represented by

$$Y_i = f(X_i) + U_i, \ i = 1, \ldots, n,$$

where f has to be estimated and U_i are the errors.

Instrumental variables

An economic relationship between a response variable Y and a vector X of explanatory variables is represented by

$$Y_i = f(X_i) + U_i, \; i = 1, \ldots, n,$$

where f has to be estimated and U_i are the errors. This model does not characterize the function f if U is not constrained. The problem is solved if $\mathbf{E}(U|X) = 0$.

Instrumental variables

An economic relationship between a response variable Y and a vector X of explanatory variables is represented by

$$Y_i = f(X_i) + U_i, \; i = 1, \ldots, n,$$

where f has to be estimated and U_i are the errors.
This model does not characterize the function f if U is not constrained. The problem is solved if $\mathbf{E}(U|X) = 0$.

In many structural econometrics models some components of X are endogeneous.

Instrumental variables

An economic relationship between a response variable Y and a vector X of explanatory variables is represented by

$$Y_i = f(X_i) + U_i, \ i = 1, \ldots, n,$$

where f has to be estimated and U_i are the errors.
This model does not characterize the function f if U is not constrained. The problem is solved if $\mathbf{E}(U|X) = 0$.

In many structural econometrics models some components of X are endogeneous.

If Y denotes wages and X, level of education, among other variables. The error U includes, ability, not observed, but influences wages.

Instrumental variables

An economic relationship between a response variable Y and a vector X of explanatory variables is represented by

$$Y_i = f(X_i) + U_i,\ i = 1, \ldots, n,$$

where f has to be estimated and U_i are the errors.
This model does not characterize the function f if U is not constrained. The problem is solved if $\mathbf{E}(U|X) = 0$.

In many structural econometrics models some components of X are endogeneous.

If Y denotes wages and X, level of education, among other variables. The error U includes, ability, not observed, but influences wages.
High ability tends to have high level of education, then education and ability are correlated, and thus X and U also.

Instrumental variables

Nevertheless, suppose that we observe another set of data, W_i where W is called an instrumental variable for which

$$\mathbf{E}(U|W) = \mathbf{E}(Y - f(X)|W) = 0.$$

Instrumental variables

Nevertheless, suppose that we observe another set of data, W_i where W is called an instrumental variable for which

$$\mathbf{E}(U|W) = \mathbf{E}(Y - f(X)|W) = 0.$$

This equation characterizes f by a Fredholm equation of the first kind. Estimation of the function f is in fact an ill-posed inverse problems.

Instrumental variables

Nevertheless, suppose that we observe another set of data, W_i where W is called an instrumental variable for which

$$\mathbf{E}(U|W) = \mathbf{E}(Y - f(X)|W) = 0.$$

This equation characterizes f by a Fredholm equation of the first kind. Estimation of the function f is in fact an ill-posed inverse problems.

Not exactly our model of Gaussian white noise, but closely related.

Instrumental variables

Nevertheless, suppose that we observe another set of data, W_i where W is called an instrumental variable for which

$$\mathbf{E}(U|W) = \mathbf{E}(Y - f(X)|W) = 0.$$

This equation characterizes f by a Fredholm equation of the first kind. Estimation of the function f is in fact an ill-posed inverse problems.

Not exactly our model of Gaussian white noise, but closely related.

Inverse problems have been the topic of many articles in the econometrics literature, see Florens (2003), Hall and Horowitz (2005), Chen and Reiss (2009).

Inverse problem and sequence space

Let the model :

$$Y = Af + \varepsilon\xi,$$

where Y is the observation, $f \in H$ unknown ,
A a continuous linear compact operator fom H into G,
ξ is a white noise, ε is the noise level.

Inverse problem and sequence space

Let the model :

$$Y = Af + \varepsilon\xi,$$

where Y is the observation, $f \in H$ unknown ,
A a continuous linear compact operator fom H into G,
ξ is a white noise, ε is the noise level.

By using the SVD, we obtain the equivalent sequence space model :

$$X_k = \theta_k + \varepsilon\sigma_k\xi_k, \quad k = 1, 2, \ldots.$$

where $\sigma_k \to \infty$.

Inverse problem and sequence space

Let the model :

$$Y = Af + \varepsilon\xi,$$

where Y is the observation, $f \in H$ unknown ,
A a continuous linear compact operator fom H into G,
ξ is a white noise, ε is the noise level.

By using the SVD, we obtain the equivalent sequence space model :

$$X_k = \theta_k + \varepsilon\sigma_k\xi_k, \quad k = 1, 2, \ldots .$$

where $\sigma_k \to \infty$.

The aim is to estimate (reconstruct) the function f (or the sequence $\{\theta_k\}$) by use of observations.

Linear estimators

Consider here a specific family of estimators.

Linear estimators

Consider here a specific family of estimators.
Let $\lambda = (\lambda_1, \lambda_2, \ldots)$ be a sequence of nonrandom weights.
Every sequence λ defines a linear estimator

$$\hat{\theta}(\lambda) = (\hat{\theta}_1, \hat{\theta}_2, \ldots), \quad \hat{\theta}_k = \lambda_k X_k$$

Linear estimators

Consider here a specific family of estimators.
Let $\lambda = (\lambda_1, \lambda_2, \ldots)$ be a sequence of nonrandom weights.
Every sequence λ defines a linear estimator

$$\hat{\theta}(\lambda) = (\hat{\theta}_1, \hat{\theta}_2, \ldots), \quad \hat{\theta}_k = \lambda_k X_k$$

and

$$\hat{f}(\lambda) = \sum_{k=1}^{\infty} \hat{\theta}_k \, \varphi_k.$$

Linear estimators

Consider here a specific family of estimators.
Let $\lambda = (\lambda_1, \lambda_2, \ldots)$ be a sequence of nonrandom weights.
Every sequence λ defines a linear estimator

$$\hat{\theta}(\lambda) = (\hat{\theta}_1, \hat{\theta}_2, \ldots), \quad \hat{\theta}_k = \lambda_k X_k$$

and

$$\hat{f}(\lambda) = \sum_{k=1}^{\infty} \hat{\theta}_k \, \varphi_k.$$

The L^2-risk of a linear estimator is

$$E\|\hat{f}(\lambda) - f\|^2 = R(\theta, \lambda) = \sum_{k=1}^{\infty} (1 - \lambda_k)^2 \theta_k^2 + \varepsilon^2 \sum_{k=1}^{\infty} \sigma_k^2 \lambda_k^2.$$

Classes of linear estimators

- Projection estimators (spectral cut-off),

$$\lambda_k = I(k \le N), \; N > 0.$$

Classes of linear estimators

- Projection estimators (spectral cut-off),

$$\lambda_k = I(k \le N), \ N > 0.$$

- Tikhonov regularization (penalized),

$$\lambda_k = \frac{1}{1 + \gamma \sigma_k^{2\alpha}}, \ \alpha \ge 1, \quad \gamma > 0.$$

Classes of linear estimators

- Projection estimators (spectral cut-off),

$$\lambda_k = I(k \le N),\ N > 0.$$

- Tikhonov regularization (penalized),

$$\lambda_k = \frac{1}{1+\gamma\sigma_k^{2\alpha}},\ \alpha \ge 1,\ \ \gamma > 0.$$

- Landweber iteration,

$$\lambda_k = 1 - (1-\sigma_k^{-2})^n, n \in N.$$

Classes of linear estimators

- Projection estimators (spectral cut-off),

$$\lambda_k = I(k \leq N), \;\; N > 0.$$

- Tikhonov regularization (penalized),

$$\lambda_k = \frac{1}{1 + \gamma \sigma_k^{2\alpha}}, \;\; \alpha \geq 1, \quad \gamma > 0.$$

- Landweber iteration,

$$\lambda_k = 1 - (1 - \sigma_k^{-2})^n, n \in N.$$

$\longrightarrow$ Choice of N, γ or n?

Ellipsoid of coefficients

Assume that f belongs to functional class corresponding to ellipsoids Θ in space of coefficients $\{\theta_k\}$:

$$\Theta = \Theta(a, L) = \left\{ \theta : \sum_{k=1}^{\infty} a_k^2 \theta_k^2 \leq L \right\},$$

where $a = \{a_k\}$, where $a_k > 0, a_k \to \infty$ and $L > 0$.

Ellipsoid of coefficients

Assume that f belongs to functional class corresponding to ellipsoids Θ in space of coefficients $\{\theta_k\}$:

$$\Theta = \Theta(a, L) = \left\{ \theta : \sum_{k=1}^{\infty} a_k^2 \theta_k^2 \leq L \right\},$$

where $a = \{a_k\}$, where $a_k > 0, a_k \to \infty$ and $L > 0$.

For large values of k coefficients θ_k will be decreasing with k and then be small.

Ellipsoid of coefficients

Assume that f belongs to functional class corresponding to ellipsoids Θ in space of coefficients $\{\theta_k\}$:

$$\Theta = \Theta(a, L) = \left\{ \theta : \sum_{k=1}^{\infty} a_k^2 \theta_k^2 \leq L \right\},$$

where $a = \{a_k\}$, where $a_k > 0, a_k \to \infty$ and $L > 0$.

For large values of k coefficients θ_k will be decreasing with k and then be small.

Assumptions on coefficients θ_k usually related to properties (smoothness) on f.

Sobolev classes

Introduce the Sobolev classes

$$\mathcal{W}(\alpha, L) = \left\{ f = \sum_{k=1}^{\infty} \theta_k \varphi_k : \theta \in \Theta(\alpha, L) \right\}$$

Sobolev classes

Introduce the Sobolev classes

$$\mathcal{W}(\alpha, L) = \left\{ f = \sum_{k=1}^{\infty} \theta_k \varphi_k : \theta \in \Theta(\alpha, L) \right\}$$

where $\Theta(\alpha, L) = \Theta(a, L)$ with $a = \{a_k\}$ polynomial such that $a_1 = 0$ and

$$a_k = \begin{cases} (k-1)^\alpha & \text{for } k \text{ odd}, \\ k^\alpha & \text{for } k \text{ even}, \end{cases} \qquad k = 2, 3, \dots,$$

where $\alpha > 0$, $L > 0$. We have also

Sobolev classes

Introduce the Sobolev classes

$$\mathcal{W}(\alpha, L) = \left\{ f = \sum_{k=1}^{\infty} \theta_k \varphi_k : \theta \in \Theta(\alpha, L) \right\}$$

where $\Theta(\alpha, L) = \Theta(a, L)$ with $a = \{a_k\}$ polynomial such that $a_1 = 0$ and

$$a_k = \begin{cases} (k-1)^\alpha & \text{for } k \text{ odd}, \\ k^\alpha & \text{for } k \text{ even}, \end{cases} \qquad k = 2, 3, \ldots,$$

where $\alpha > 0$, $L > 0$. We have also

$$\mathcal{W}(\alpha, L) = \left\{ f \text{ periodic } : \int_0^1 (f^{(\alpha)}(t))^2 dt \leq \pi^{2\alpha} L \right\}.$$

Rates of convergence

Function f has Fourier coefficients in some ellipsoid, and the problem is mildly, severely ill-posed or even direct. Rates appear in the following table :

Rates of convergence

Function f has Fourier coefficients in some ellipsoid, and the problem is mildly, severely ill-posed or even direct. Rates appear in the following table :

Problem/Functions	Sobolev
Direct problem	$\varepsilon^{\frac{4\alpha}{2\alpha+1}}$
Mildly ill-posed	$\varepsilon^{\frac{4\alpha}{2\alpha+2\beta+1}}$
Severely ill-posed	$(\log \frac{1}{\varepsilon})^{-2\alpha}$

Comments

- Rates usually depend on smoothness α of function f and on degree of ill-posedness β.

Comments

- Rates usually depend on smoothness α of function f and on degree of ill-posedness β.
- When β increases rates are slower.

Comments

- Rates usually depend on smoothness α of function f and on degree of ill-posedness β.
- When β increases rates are slower.
- In direct model, standard rates for nonparametric estimation.
  $\longrightarrow$ For example, $2\alpha/(2\alpha+1)$ with Sobolev classes.

Comments

- To attain optimal rate with projection estimator, choose N corresponding to optimal trade-off between bias and variance.

Comments

- To attain optimal rate with projection estimator, choose N corresponding to optimal trade-off between bias and variance.
- In minimax sense, optimal choice for N. However, choice depends on smoothness α and on degree of ill-posedness β.

Comments

- To attain optimal rate with projection estimator, choose N corresponding to optimal trade-off between bias and variance.
- In minimax sense, optimal choice for N. However, choice depends on smoothness α and on degree of ill-posedness β.
- Even if operator A (and its degree β) is known, no real meaning to consider smoothness of f as known.

Comments

- To attain optimal rate with projection estimator, choose N corresponding to optimal trade-off between bias and variance.
- In minimax sense, optimal choice for N. However, choice depends on smoothness α and on degree of ill-posedness β.
- Even if operator A (and its degree β) is known, no real meaning to consider smoothness of f as known.
- Notion of adaptation and oracle inequalities, i.e. how to choose bandwidth N without prior assumptions on f.

Oracle

Consider now a linked, but different point of view.
Assume that a class of estimators is fixed, i.e. that the class of possible weights $\lambda \in \Lambda$ is given (projection, Tikhonov,...).

Oracle

Consider now a linked, but different point of view.
Assume that a class of estimators is fixed, i.e. that the class of possible weights $\lambda \in \Lambda$ is given (projection, Tikhonov,...).
Define the oracle λ^0 as

$$R(\theta, \lambda^0) = \inf_{\lambda \in \Lambda} R(\theta, \lambda).$$

The oracle corresponds to the best possible choice in Λ, i.e. the one which minimizes the risk.

Oracle

Consider now a linked, but different point of view.
Assume that a class of estimators is fixed, i.e. that the class of possible weights $\lambda \in \Lambda$ is given (projection, Tikhonov,...).
Define the oracle λ^0 as

$$R(\theta, \lambda^0) = \inf_{\lambda \in \Lambda} R(\theta, \lambda).$$

The oracle corresponds to the best possible choice in Λ, i.e. the one which minimizes the risk.

However, this is not an estimator since the risk depends on the unknown θ, the oracle will depend also.

Oracle

Consider now a linked, but different point of view.
Assume that a class of estimators is fixed, i.e. that the class of possible weights $\lambda \in \Lambda$ is given (projection, Tikhonov,...).
Define the oracle λ^0 as

$$R(\theta, \lambda^0) = \inf_{\lambda \in \Lambda} R(\theta, \lambda).$$

The oracle corresponds to the best possible choice in Λ, i.e. the one which minimizes the risk.

However, this is not an estimator since the risk depends on the unknown θ, the oracle will depend also.

An oracle is the best in the family, but it knows the true θ.

Unbiased risk estimation

A very natural idea in statistics is to estimate this unknown risk using the available data, and then to minimize this estimator of the risk.

Unbiased risk estimation

A very natural idea in statistics is to estimate this unknown risk using the available data, and then to minimize this estimator of the risk.

A classical approach to this minimization problem is based on the principle of unbiased risk estimation (URE) (Stein (1981)).

Unbiased risk estimation

A very natural idea in statistics is to estimate this unknown risk using the available data, and then to minimize this estimator of the risk.

A classical approach to this minimization problem is based on the principle of unbiased risk estimation (URE) (Stein (1981)).

This method goes back to Akaike Information Criteria (AIC) in Akaike (1973) and Mallows C_p (1973).

Unbiased risk estimation

A very natural idea in statistics is to estimate this unknown risk using the available data, and then to minimize this estimator of the risk.

A classical approach to this minimization problem is based on the principle of unbiased risk estimation (URE) (Stein (1981)).

This method goes back to Akaike Information Criteria (AIC) in Akaike (1973) and Mallows C_p (1973).

Originally, the URE was in the context of regression estimation. Nowadays, it is a basic adaptation tool for many statistical models.

Unbiased risk estimation

A very natural idea in statistics is to estimate this unknown risk using the available data, and then to minimize this estimator of the risk.

A classical approach to this minimization problem is based on the principle of unbiased risk estimation (URE) (Stein (1981)).

This method goes back to Akaike Information Criteria (AIC) in Akaike (1973) and Mallows C_p (1973).

Originally, the URE was in the context of regression estimation. Nowadays, it is a basic adaptation tool for many statistical models.

This idea appears also in all the cross-validation techniques.

URE in inverse problems

For inverse problems, this method was studied in C., Golubev, Picard and Tsybakov (2002), where exact oracle inequalities were obtained.

URE in inverse problems

For inverse problems, this method was studied in C., Golubev, Picard and Tsybakov (2002), where exact oracle inequalities were obtained.

In this setting, the functional

$$U(X,\lambda) = \sum_{k=1}^{\infty} (1-\lambda_k)^2 (X_k^2 - \varepsilon^2 \sigma_k^2) + \varepsilon^2 \sum_{k=1}^{\infty} \sigma_k^2 \lambda_k^2$$

URE in inverse problems

For inverse problems, this method was studied in C., Golubev, Picard and Tsybakov (2002), where exact oracle inequalities were obtained.

In this setting, the functional

$$U(X,\lambda) = \sum_{k=1}^{\infty} (1-\lambda_k)^2 (X_k^2 - \varepsilon^2 \sigma_k^2) + \varepsilon^2 \sum_{k=1}^{\infty} \sigma_k^2 \lambda_k^2$$

is an unbiased estimator of $R(\theta,\lambda)$:

$$R(\theta,\lambda) = \mathbf{E}_\theta U(X,\lambda), \ \forall \lambda.$$

Data-driven choice

Unbiased risk estimation suggests to minimize over $\lambda \in \Lambda$ the functional $U(X, \lambda)$ in place of $R(\theta, \lambda)$.

Data-driven choice

Unbiased risk estimation suggests to minimize over $\lambda \in \Lambda$ the functional $U(X, \lambda)$ in place of $R(\theta, \lambda)$.

This leads to the following data-driven choice of λ:

$$\lambda^\star = \arg\min_{\lambda \in \Lambda} U(X, \lambda).$$

Data-driven choice

Unbiased risk estimation suggests to minimize over $\lambda \in \Lambda$ the functional $U(X, \lambda)$ in place of $R(\theta, \lambda)$.

This leads to the following data-driven choice of λ:

$$\lambda^\star = \arg\min_{\lambda \in \Lambda} U(X, \lambda).$$

Define then the estimator $\theta^\star$ by

$$\theta_k^\star = \lambda_k^\star X_k.$$

Assumptions

Denote

$$S = \left(\frac{\max_{\lambda \in \Lambda} \sum_{k=1}^{\infty} \sigma_k^4 \lambda_k^2}{\min_{\lambda \in \Lambda} \sum_{k=1}^{\infty} \sigma_k^4 \lambda_k^2} \right)^{1/2}.$$

Let the following assumptions hold.

Assumptions

Denote

$$S = \left(\frac{\max_{\lambda\in\Lambda} \sum_{k=1}^{\infty} \sigma_k^4 \lambda_k^2}{\min_{\lambda\in\Lambda} \sum_{k=1}^{\infty} \sigma_k^4 \lambda_k^2} \right)^{1/2}.$$

Let the following assumptions hold. For any $\lambda \in \Lambda$

$$0 < \sum_{k=1}^{\infty} \sigma_k^2 \lambda_k^2 < \infty, \qquad \max_{\lambda\in\Lambda} \sup_k |\lambda_k| \le 1.$$

Assumptions

Denote

$$S = \left(\frac{\max_{\lambda \in \Lambda} \sum_{k=1}^{\infty} \sigma_k^4 \lambda_k^2}{\min_{\lambda \in \Lambda} \sum_{k=1}^{\infty} \sigma_k^4 \lambda_k^2} \right)^{1/2}.$$

Let the following assumptions hold. For any $\lambda \in \Lambda$

$$0 < \sum_{k=1}^{\infty} \sigma_k^2 \lambda_k^2 < \infty, \qquad \max_{\lambda \in \Lambda} \sup_{k} |\lambda_k| \leq 1.$$

There exists a constant $C_1 > 0$ such that, uniformly in $\lambda \in \Lambda$,

$$\sum_{k=1}^{\infty} \sigma_k^4 \lambda_k^2 \leq C_1 \sum_{k=1}^{\infty} \sigma_k^4 \lambda_k^4.$$

Oracle inequality for URE

Theorem. Suppose $\sigma_k \sim k^\beta,\ \beta \geq 0$.
Assume that Λ is finite with cardinality D and belongs to the family of Projection, Tikhonov or Pinsker estimators.

There exist constants $\gamma, C^* > 0$ such that $\forall \theta \in \ell_2$, we have for B large enough,

$$\mathbf{E}_\theta \|\theta^\star - \theta\|^2 \leq (1 + \gamma B^{-1}) \min_{\lambda \in \Lambda} R(\theta, \lambda) + BC^* \varepsilon^2 (\log(DS))^{2\beta+1}.$$

Oracle inequality for URE

Theorem. Suppose $\sigma_k \sim k^\beta,\ \beta \geq 0$.
Assume that Λ is finite with cardinality D and belongs to the family of Projection, Tikhonov or Pinsker estimators.

There exist constants $\gamma, C^* > 0$ such that $\forall \theta \in \ell_2$, we have for B large enough,

$$\mathbf{E}_\theta \|\theta^\star - \theta\|^2 \leq (1 + \gamma B^{-1}) \min_{\lambda \in \Lambda} R(\theta, \lambda) + B C^* \varepsilon^2 (\log(DS))^{2\beta+1}.$$

The data-driven choice by URE mimics the oracle.

Simulations

Discrete model : inverse problem.

$$Y(i) = g * f\left(\frac{i}{m}\right) + \varepsilon\,\sqrt{m}\xi(i), \quad i = 1, \dots, m,$$

Simulations

Discrete model : inverse problem.

$$Y(i) = g * f\left(\frac{i}{m}\right) + \varepsilon\sqrt{m}\xi(i), \quad i = 1, \ldots, m,$$

where

$$f(t) = 0.5n(t, 0.4, 0.12) + 0.5n(t, 0.7, 0.08),$$

and

$$g(t) = \exp(-10|t - 0.5|), \quad \beta \sim 2.$$

Simulations

Discrete model : inverse problem.

$$Y(i) = g * f\left(\frac{i}{m}\right) + \varepsilon\,\sqrt{m}\xi(i), \quad i = 1, \ldots, m,$$

where

$$f(t) = 0.5n(t, 0.4, 0.12) + 0.5n(t, 0.7, 0.08),$$

and

$$g(t) = \exp(-10|t - 0.5|), \quad \beta \sim 2.$$

Here $m = 1000$ et $\varepsilon^{-2} = 10^5$. Signal/Noise = 100. Estimator by truncated Fourier series.

Simulations

Discrete model : inverse problem.

$$Y(i) = g * f\left(\frac{i}{m}\right) + \varepsilon\,\sqrt{m}\xi(i), \quad i = 1, \dots, m,$$

where

$$f(t) = 0.5n(t, 0.4, 0.12) + 0.5n(t, 0.7, 0.08),$$

and

$$g(t) = \exp(-10|t - 0.5|), \quad \beta \sim 2.$$

Here $m = 1000$ et $\varepsilon^{-2} = 10^5$. Signal/Noise = 100. Estimator by truncated Fourier series.
With

$$\varepsilon^2 \sum_{|k| \leq N} \sigma_k^2 \leq 1/\log(1/\varepsilon).$$

Simulations

Discrete model : inverse problem.

$$Y(i) = g * f\left(\frac{i}{m}\right) + \varepsilon\sqrt{m}\xi(i), \quad i = 1, \ldots, m,$$

where

$$f(t) = 0.5n(t, 0.4, 0.12) + 0.5n(t, 0.7, 0.08),$$

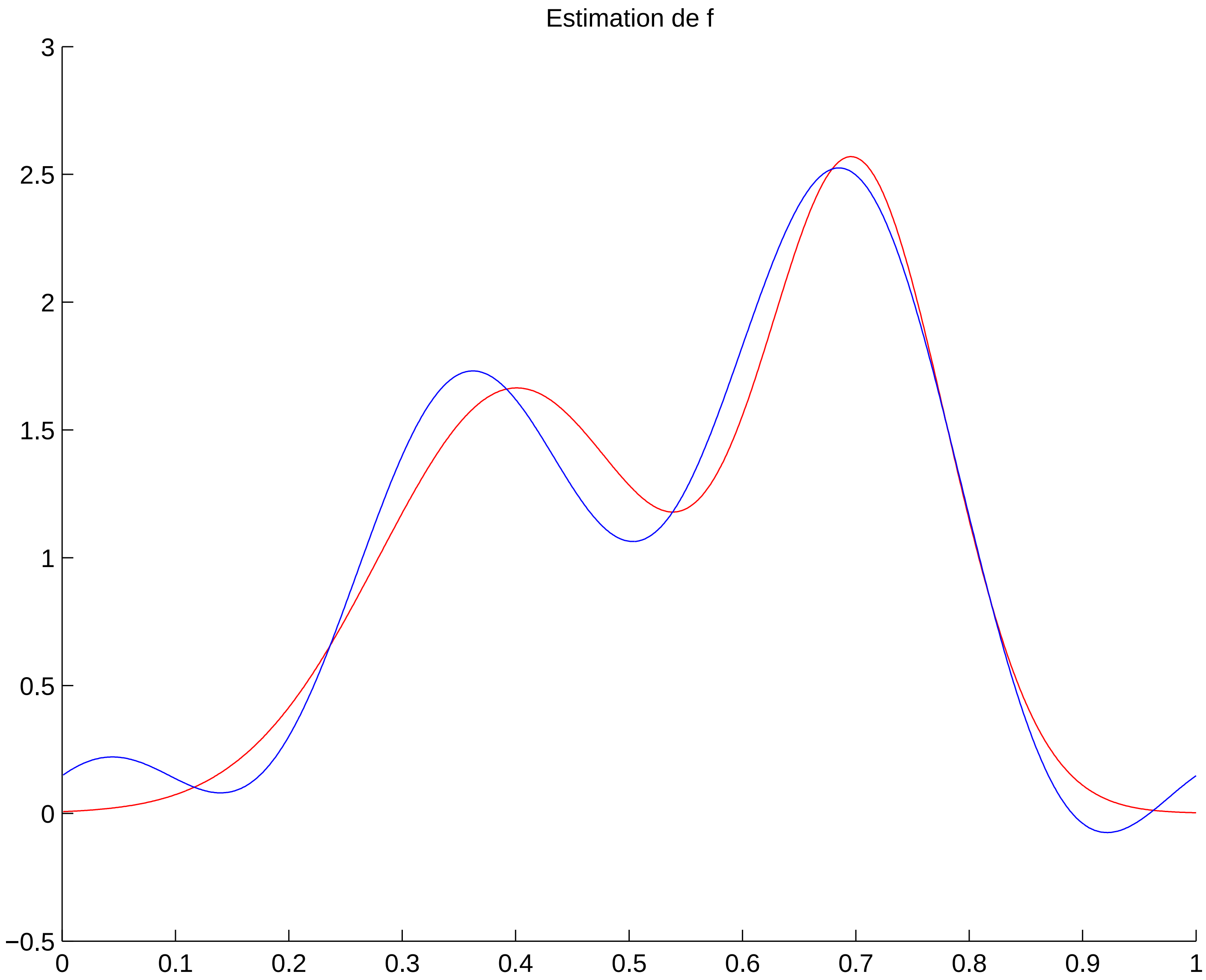
and

$$g(t) = \exp(-10|t - 0.5|), \quad \beta \sim 2.$$

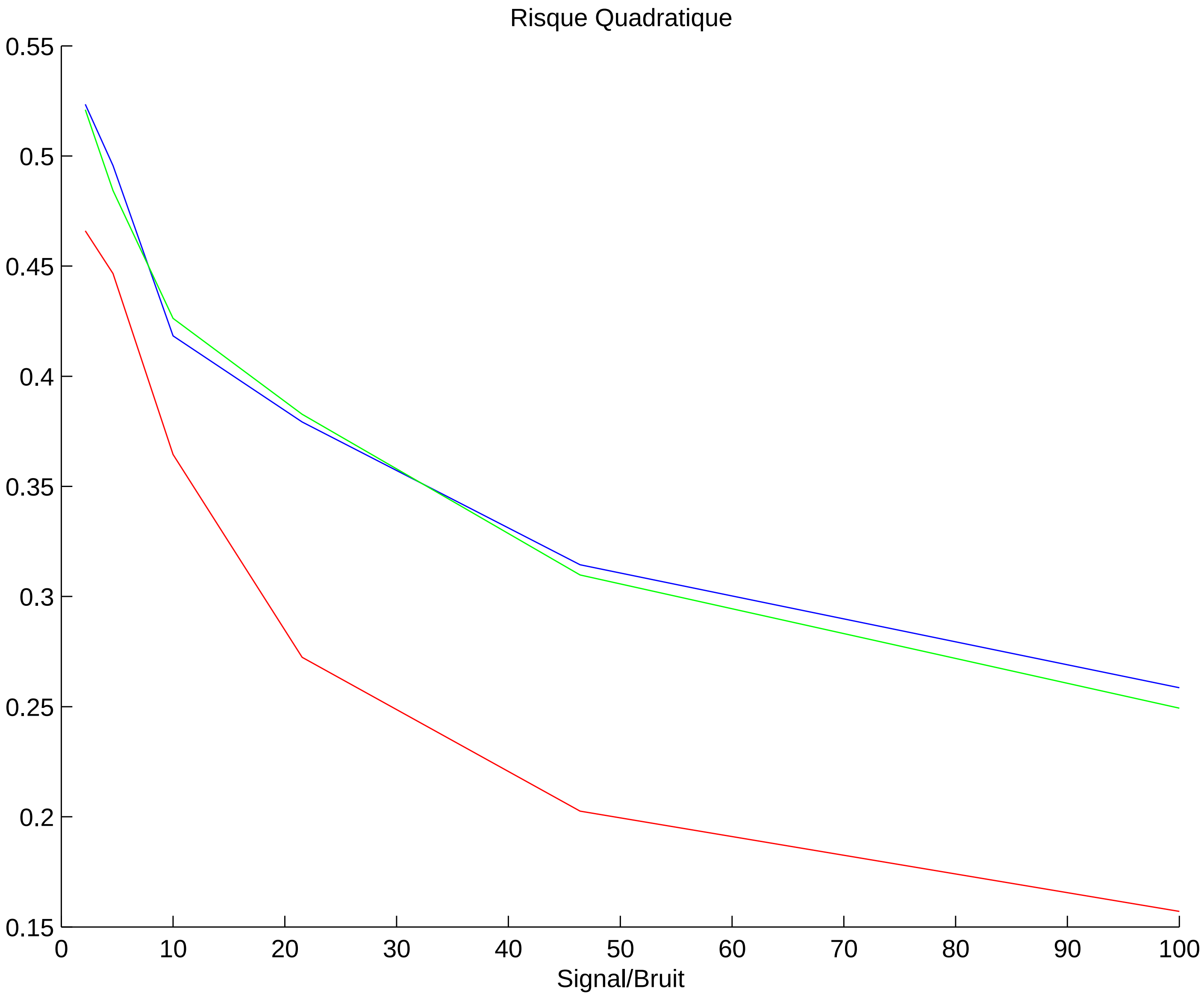
Here $m = 1000$ et $\varepsilon^{-2} = 10^5$. Signal/Noise = 100. Estimator by truncated Fourier series.
With

$$\varepsilon^2 \sum_{|k| \leq N} \sigma_k^2 \leq 1/\log(1/\varepsilon).$$

True function f. Estimator $f^{\star}$.



Oracle by projection. Estimator $f^\star$.



Comments

- Simulations correspond more or less to theory.

Comments

- Simulations correspond more or less to theory.
- Limitation on the size of the family.

Comments

- Simulations correspond more or less to theory.
- Limitation on the size of the family.
- Method not always stable enough.

Comments

- Simulations correspond more or less to theory.
- Limitation on the size of the family.
- Method not always stable enough.
- Need for stronger penalties than the URE penalty (or AIC).

Comments

- Simulations correspond more or less to theory.
- Limitation on the size of the family.
- Method not always stable enough.
- Need for stronger penalties than the URE penalty (or AIC).
- Different method called Risk Hull Method, defined in C. and Golubev (2006).