Adaptive estimation for inverse problems with noisy operators

Laurent Cavalier
(Université Aix-Marseille 1 (France))

and Nicolas Hengartner
(Los Alamos National Laboratory (USA))
Many fields where inverse problems appear

- **Astronomy** (Hubble satellite)
- **Econometrics** (instrumental variables)
- **Financial mathematics** (model calibration)
Introduction

Many fields where inverse problems appear

- Astronomy (Hubble satellite)
- Econometrics (instrumental variables)
- Financial mathematics (model calibration)

Problems with indirect observations of a function that we want to reconstruct.
Inverse problems

Let $H$ and $G$ be Hilbert spaces.
Let $A$ be bounded linear operator from $H$ into $G$. 
Inverse problems

Let $H$ and $G$ be Hilbert spaces.
Let $A$ be bounded linear operator from $H$ into $G$.

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

Let $H$ and $G$ be Hilbert spaces.

Let $A$ be a bounded linear operator from $H$ into $G$.

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

“Inversion” of the operator $A$. 

Yale, 02 May 2011 – p.3/33
Inverse problems

Let $H$ and $G$ be Hilbert spaces.

Let $A$ be bounded linear operator from $H$ into $G$.

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

“Inversion” of the operator $A$.

If $A$ is not invertible, or with a non-continuous inverse, then the problem is said to be ill-posed.
Inverse problems

Let $H$ and $G$ be Hilbert spaces.
Let $A$ be bounded linear operator from $H$ into $G$.
Given $g \in G$ find $f \in H$ such that $Af = g$.
“Inversion” of the operator $A$.
If $A$ is not invertible, or with a non-continuous inverse, then the problem is said to be ill-posed.
One observes $g^\varepsilon$ a noisy version of $g$, then $f^\varepsilon = A^{-1}g^\varepsilon$ could not be close to $f$.
Importance of the notion of “noise” or “error”.

Yale, 02 May 2011 – p.3/33
Inverse problem with random noise

Let the model:

\[ Y = Af + \varepsilon \xi, \]

where \( f \in H \) (Hilbert),
\( A \) bounded lin. operator from \( H \) into \( G \) (Hil.),
\( \xi \) Gaussian white noise \( (0 < \varepsilon < 1 \) level).
Inverse problem with random noise

Let the model:

\[ Y = Af + \varepsilon \xi, \]

where \( f \in H \) (Hilbert),
\( A \) bounded lin. operator from \( H \) into \( G \) (Hil.),
\( \xi \) **Gaussian white noise** (\( 0 < \varepsilon < 1 \) level).

Reconstruct \( f \) with the observation \( Y \).
Inverse problem with random noise

Let the model:

\[ Y = Af + \varepsilon \xi, \]

where \( f \in H \) (Hilbert), 
\( A \) bounded lin. operator from H into G (Hil.),
\( \xi \) Gaussian white noise \( (0 < \varepsilon < 1 \text{ level}) \).

Reconstruct \( f \) with the observation \( Y \).

Inverse problem with random noise.
Inverse problem with random noise

Let the model:

\[ Y = Af + \varepsilon \xi, \]

where \( f \in H \) (Hilbert),
\( A \) bounded lin. operator from \( H \) into \( G \) (Hil.),
\( \xi \) Gaussian white noise \((0 < \varepsilon < 1 \text{ level})\).

Reconstruct \( f \) with the observation \( Y \).

Inverse problem with random noise.
\( A \) non invertible \( \rightarrow \) Ill-posed problem.
$A^* A$ compact operator with a known basis of eigenfunctions:
A^* A compact operator with a known basis of eigenfunctions:

\[ A^* A \phi_k = b_k^2 \phi_k. \]
Singular Value Decomposition

$A^* A$ compact operator with a known basis of eigenfunctions:

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

Singular Value Decomposition (SVD) of $A$: 

Singular Value Decomposition

$A^* A$ compact operator with a known basis of eigenfunctions:

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

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. in $H$, $\{\psi_k\}$ o.n.b in $G$ known.
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 \( \{ \xi_k \} \) i.i.d. standard Gaussian.
One obtains the 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

One obtains the 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 observation \( Y = \{Y_k\} \).

Using the \( L_2 \)–risk, equivalent to estimate \( f \).
Sequence space model

One obtains the 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 observation \( Y = \{Y_k\} \).

Using the \( L_2 \) risk, equivalent to estimate \( f \).

Note that \( b_k \to 0 \) weaken the signal \( \theta_k \).
One obtains the equivalent sequence space model

\[ Y_k = b_k \theta_k + \varepsilon_k, \quad k = 1, 2, \ldots, \]

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 observation \( Y = \{Y_k\} \).

Using the \( L_2 \)-risk, equivalent to estimate \( f \).

Note that \( b_k \to 0 \) weaken the signal \( \theta_k \) → Ill-posed problem.
Examples

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

- Convolution → Fourier basis.
Examples

• Convolution $\rightarrow$ Fourier basis.
Blurred images.
Examples

- **Convolution** → Fourier basis.
  Blurred images.

- **Integration** → Estimation of the \( \beta \) derivative with direct observations.
Examples

- **Convolution** $\rightarrow$ Fourier basis. Blurred images.

- **Integration** $\rightarrow$ Estimation of the $\beta$ derivative with direct observations.

- **Direct model** $\rightarrow$ Gaussian white noise model.
Images of the Hubble satellite

HUBBLE IMAGE BEFORE WIDE FIELD AND PLANETARY CAMERA 2...

AND AFTER
An essentiel assumption is that the operator $A$ is known.
An essential assumption is that the operator $A$ is known.

This is usually not true in applications.
An essential assumption is that the operator $A$ is known.

This is usually not true in applications. Suppose that $A$ is unknown, or at least its singular values $b_k$,

$$Y_k = b_k \theta_k + \varepsilon \xi_k$$
Unknown operator

An essentiel assumption is that the operator $A$ is known.

This is usually not true in applications. Suppose that $A$ is unknown, or at least its singular values $b_k$, $Y_k = b_k \theta_k$
An essential assumption is that the operator $A$ is known.

This is usually not true in applications. Suppose that $A$ is unknown, or at least its singular values $b_k$,

$$Y_k = b_k \theta_k$$

Even without noise the parameter $\theta$ is not identifiable, if $\{b_k\}$ is unknown.
Noisy operator
Noisy operator

- **Unknown operator.** Matias (02). Model related to convolution $\rightarrow$ Logarithmic rates. Paul (05).
Noisy operator

- **Unknown operator.** Matias (02). Model related to convolution $\rightarrow$ Logarithmic rates. Paul (05).

- **Noisy operator.** Efroymovich and Kolchinskii (01), Marteau (06), C. and Raimondo (07). Noise on the operator and on the data. $\rightarrow$ Same rates of convergence.
Noisy operator

- **Unknown operator.** Matias (02). Model related to convolution $\longrightarrow$ Logarithmic rates. Paul (05).

- **Noisy operator.** Efroymovich and Kolchinskii (01), Marteau (06), C. and Raimondo (07). Noise on the operator and on the data. $\longrightarrow$ Same rates of convergence.

- **Noisy operator.** Noisy singular values. Known basis. $\longrightarrow$ More precise results.
An economic relationship is represented by

\[ Y_i = f(X_i) + U_i, \quad i = 1, \ldots, n, \]

where \( f \) has to be estimated and \( U_i \) errors.
An economic relationship is represented by

\[ Y_i = f(X_i) + U_i, \quad i = 1, \ldots, n, \]

where \( f \) has to be estimated and \( U_i \) errors.

This model does not characterize the function \( f \) if \( U \) is not constrained. The problem is solved if \( E(U|X) = 0 \).
Instrumental variables

An economic relationship is represented by

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

where $f$ has to be estimated and $U_i$ errors.

This model does not characterize the function $f$ if $U$ is not constrained. The problem is solved if $E(U|X) = 0$.

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

Another set of data, where $W_i$ are called an instrumental variables for which

$$E(U|W) = E(Y - f(X)|W) = 0.$$
Instrumental variables

Another set of data, where $W_i$ are called an instrumental variables for which

$$E(U|W) = E(Y - f(X)|W) = 0.$$ 

Estimation of the function $f$ is in fact an ill-posed inverse problems.
Another set of data, where $W_i$ are called an instrumental variables for which

$$E(U|W) = E(Y - f(X)|W) = 0.$$ 

Estimation of the function $f$ is in fact an **ill-posed inverse problems**.

Not exactly our model of Gaussian white noise, but closely related. The **operator is unknown** and has to be estimated.
Instrumental variables

Another set of data, where $W_i$ are called an instrumental variables for which

$$E(U|W) = E(Y - f(X)|W) = 0.$$  

Estimation of the function $f$ is in fact an ill-posed inverse problems.

Not exactly our model of Gaussian white noise, but closely related. The operator is unknown and has to be estimated.

Florens (2003), Hall and Horowitz (2005), Chen and Reiss (2009).
One observes two sequences:
Model

One observes two sequences:

\[ Y_k = b_k \theta_k + \varepsilon \xi_k, \quad k = 1, 2, \ldots, \]

and
One observes two sequences:

\[ Y_k = b_k \theta_k + \varepsilon \xi_k, \quad k = 1, 2, \ldots, \]

and

\[ X_k = b_k + \sigma \eta_k, \quad k = 1, 2, \ldots, \]

where \( \{\xi_k\} \) and \( \{\eta_k\} \) i.i.d. Gaussian.
Model

One observes two sequences:

\[ Y_k = b_k \theta_k + \varepsilon \xi_k, \quad k = 1, 2, \ldots, \]

and

\[ X_k = b_k + \sigma \eta_k, \quad k = 1, 2, \ldots, \]

where \( \{\xi_k\} \) and \( \{\eta_k\} \) i.i.d. Gaussian.

Estimate \( \theta \) with \( \{b_k\} \) unknown.
Define a linear estimator by:

\[ \hat{\theta}_k = \hat{\theta}_k(\lambda) = \lambda_k \frac{Y_k}{b_k}, \quad k = 1, 2, \ldots, \]

where \( \{\lambda_k\} \) is a sequence.
Linear estimators

Define a linear estimator by:

\[ \hat{\theta}_k = \hat{\theta}_k(\lambda) = \lambda_k \frac{Y_k}{b_k}, \quad k = 1, 2, \ldots, \]

where \( \{\lambda_k\} \) is a sequence.

\[ \rightarrow \text{ Constructed with } \{b_k\} \text{ unknown.} \]
Define a linear estimator by:

\[ \hat{\theta}_k = \hat{\theta}_k(\lambda) = \lambda_k \frac{Y_k}{b_k}, \ k = 1, 2, \ldots, \]

where \( \{\lambda_k\} \) is a sequence.

\[ \rightarrow \text{Constructed with } \{b_k\} \text{ unknown.} \]

Its risk is:

\[ R(\lambda, \theta) = E_\theta \|\hat{\theta}(\lambda) - \theta\|^2 = \sum_{k=1}^{\infty} (1 - \lambda_k)^2 \theta_k^2 + \varepsilon^2 \frac{\lambda_k^2}{b_k^2}. \]
Classes of estimators

Standard linear classes:
Classes of estimators

Standard linear classes:

- **Projection estimators.** $\lambda_k = I(k \leq w)$. 

Classes of estimators

Standard linear classes:

- Projection estimators. $\lambda_k = I(k \leq w)$.
- Tikhonov regularization. $\lambda_k = \frac{1}{1 + (k/w)^\alpha}$, $w > 0$, $\alpha > 0$. 
Classes of estimators

Standard linear classes:

- **Projection estimators.** \( \lambda_k = I(k \leq w) \).
- **Tikhonov regularization.**
  \[
  \lambda_k = \frac{1}{1 + (k/w)^\alpha}, \quad w > 0, \quad \alpha > 0.
  \]
- **Pinsker filter.**
  \[
  \lambda_k = (1 - (k/w)^\alpha)_+, \quad w > 0, \quad \alpha > 0.
  \]
Model selection

Let $\Lambda = \{\lambda^1, \ldots, \lambda^N\}$ be a given family of linear estimators of size $N$. 
Model selection

Let $\Lambda = \{\lambda^1, \ldots, \lambda^N\}$ be a given family of linear estimators of size $N$.

The aim is to find a data-driven method for choosing the best estimator in $\Lambda$. 
Model selection

Let $\Lambda = \{\lambda^1, \ldots, \lambda^N\}$ be a given family of linear estimators of size $N$.

The aim is to find a data-driven method for choosing the best estimator in $\Lambda$.

$\rightarrow$ Model selection.
Model selection

Let $\Lambda = \{\lambda^1, \ldots, \lambda^N\}$ be a given family of linear estimators of size $N$.

The aim is to find a data-driven method for choosing the best estimator in $\Lambda$.

$\rightarrow$ Model selection.

The aim is to mimic the oracle

$$\lambda^0 = \arg \min_{\lambda \in \Lambda} R(\lambda, \theta).$$
Unbiased risk estimation

Minimize a criterion close to the risk
Unbiased risk estimation

Minimize a criterion close to the risk

\[ U[\lambda; Y] = \sum_{k=1}^{\infty} \left\{ \left( \lambda_k^2 - 2\lambda_k \right) \frac{Y_k^2 - \varepsilon^2}{b_k^2} + \varepsilon^2 \frac{\lambda_k^2}{b_k^2} \right\}, \]
Unbiased risk estimation

Minimize a criterion close to the risk

\[ U[\lambda; Y] = \sum_{k=1}^{\infty} \left\{ \left( \lambda_k^2 - 2\lambda_k \right) \frac{Y_k^2 - \varepsilon^2}{b_k^2} + \varepsilon^2 \frac{\lambda_k^2}{b_k^2} \right\}, \]

\[ E_\theta U[\lambda; Y] = R(\lambda, \theta) - \sum \theta_k^2. \]
Unbiased risk estimation

Minimize a criterion close to the risk

\[
U[\lambda; Y] = \sum_{k=1}^{\infty} \left\{ (\lambda_k^2 - 2\lambda_k) \frac{Y_k^2 - \varepsilon^2}{b_k^2} + \varepsilon^2 \frac{\lambda_k^2}{b_k^2} \right\},
\]

\[
E_{\theta}U[\lambda; Y] = R(\lambda, \theta) - \sum \theta_k^2.
\]

\(U\) is an unbiased risk estimator
(see Akaike (73), Mallows (73), Stein (81)).
Unbiased risk estimation

Minimize a criterion close to the risk

\[ U[\lambda; Y] = \sum_{k=1}^{\infty} \left\{ (\lambda_k^2 - 2\lambda_k) \frac{Y_k^2 - \varepsilon^2}{b_k^2} + \varepsilon^2 \frac{\lambda_k^2}{b_k^2} \right\}, \]

\[ E_\theta U[\lambda; Y] = R(\lambda, \theta) - \sum \theta_k^2. \]

\( U \) is an unbiased risk estimator (see Akaike (73), Mallows (73), Stein (81)).

Generalize C., Golubev, Picard et Tsybakov (02) for the case of unknown \( b_k \).
Truncated criterion

It is natural here to estimate $b_k$ by $X_k$. 
Truncated criterion

It is natural here to estimate $b_k$ by $X_k$. The main difficulty is that when $k$ is large, noise dominates.
Truncated criterion

It is natural here to estimate $b_k$ by $X_k$. The main difficulty is that when $k$ is large, noise dominates.

Truncate the sequence at

$$M = \min \{ k \leq N_0 : |X_k| \leq \sigma \log 1/\sigma \} - 1,$$

where $N_0 = \sigma^{-2}$. 
Truncated criterion

It is natural here to estimate $b_k$ by $X_k$. The main difficulty is that when $k$ is large, noise dominates.

Truncate the sequence at

$$M = \min \{ k \leq N_0 : |X_k| \leq \sigma \log 1/\sigma \} - 1,$$

where $N_0 = \sigma^{-2}$. The criterion is then

$$\bar{U}[\lambda; X, Y] = \sum_{k=1}^{M} (\lambda_k^2 - 2\lambda_k) \frac{Y_k^2 - \varepsilon^2}{X_k^2} + \varepsilon^2 \frac{\lambda_k^2}{X_k^2}.$$
Define the sequence $\lambda$ which minimizes $\bar{U}$.
Define the sequence $\lambda$ which minimizes $\bar{U}$:

$$
\lambda^* = \arg\min_{\lambda \in \Lambda} \bar{U}[\lambda; X, Y]
$$
Estimator

Define the sequence $\lambda$ which minimizes $\bar{U}$:

$$\lambda^* = \arg \min_{\lambda \in \Lambda} \bar{U}[\lambda; X, Y]$$

and then the estimator

$$\theta^* = \begin{cases} 
\lambda^* \frac{Y_k}{X_k} & k \leq M, \\
0 & k > M.
\end{cases}$$
Oracle inequality

**Theorem 1.** Suppose that $b_k \sim k^{-\beta}$, $\beta \geq 0$. For any $B$ large enough, $c > 0$,

$$E_\theta \|\theta^* - \theta\|^2 \leq (1 + cB^{-1})R(\lambda^0, \theta) + c\varepsilon^2(B(\log N))^{2\beta+1} + \Gamma(\theta) + \Omega,$$
Theorem 1. Suppose that $b_k \sim k^{-\beta}$, $\beta \geq 0$. For any $B$ large enough, $c > 0$,

$$E_\theta \| \theta^* - \theta \|^2 \leq (1 + cB^{-1})R(\lambda^0, \theta) + c\varepsilon^2 (B(\log N))^{2\beta+1} + \Gamma(\theta) + \Omega,$$

If the remainder terms are small, 

$\longrightarrow$ estimator almost as good as the oracle.
Two sequences of observations

Solution:
Two sequences of observations

Solution:
We need a second sequence of observations

\[ X_k = b_k + \sigma \eta_k, \quad k = 1, 2, \ldots, \]

where \( \{\eta_k\} \) i.i.d. standard Gaussian indep. of \( \{\xi_k\} \),

\[ 0 < \sigma < 1 \] noise level.
Two sequences of observations

Solution:
We need a second sequence of observations

\[ X_k = b_k + \sigma \eta_k, \quad k = 1, 2, \ldots, \]

where \( \{\eta_k \} \) i.i.d. standard Gaussian indep. of \( \{\xi_k \} \),

\( 0 < \sigma < 1 \) noise level.

\[ \rightarrow \text{Observe noisy singular values} \]
Example of model:

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

where \( \{Y(t), t \in [0, 1]\} \) is observed, 
\( g \) unknown convolution kernel,
\( f \) periodic signal in \( L^2[0, 1] \),
\( \xi(t) \) Gaussian white noise, \( 0 < \varepsilon < 1 \) noise level.
Fourier basis

The SVD is known.
Let \( \{ \varphi_k(t) \} \) be the real trigonometric basis:
Fourier basis

The SVD is known.
Let \( \{ \varphi_k(t) \} \) be the real trigonometric basis:
The convolution model is equivalent to the sequence space model:

\[
Y_k = b_k \theta_k + \varepsilon \xi_k, \quad k = 1, 2, \ldots,
\]

where \( \xi_k = \langle \xi, \varphi_k \rangle \) Gaussian i.i.d.
\( \theta_k \) Fourier coefficients of \( f \)
\( b_k \) Fourier coefficients of \( g \) → unknown
Test the Fourier basis

Suppose that we can put each element of the Fourier basis through the filter.
Test the Fourier basis

Suppose that we can put each element of the Fourier basis through the filter.

Each $\varphi_k$ is a specific function $f$.
The coefficient $\theta_k$ is then 1.
Suppose that we can put each element of the Fourier basis through the filter.

Each $\varphi_k$ is a specific function $f$. The coefficient $\theta_k$ is then 1.

One obtains the model:

$$X_k = b_k + \sigma \eta_k, \quad k = 1, 2, \ldots,$$

where $\sigma = \varepsilon$. 

Test the Fourier basis
Minimax estimation

The oracle inequality in Theorem 1 has a real meaning when the remainder terms are small.
Minimax estimation

The oracle inequality in Theorem 1 has a real meaning when the remainder terms are small.

\[ \Omega \text{ converges to } 0 \text{ very fastly with } \sigma. \]
Minimax estimation

The oracle inequality in Theorem 1 has a real meaning when the remainder terms are small.

Ω converges to 0 very fastly with σ.

Control \( \Gamma(\theta) \). This term corresponds to the difference between the risk and the risk truncated at \( M \).
Minimax estimation

The oracle inequality in Theorem 1 has a real meaning when the remainder terms are small.

\( \Omega \) converges to 0 very fastly with \( \sigma \).

Control \( \Gamma(\theta) \). This term corresponds to the difference between the risk and the risk truncated at \( M \).

\( \rightarrow \) One can obtain minimax results.
Classes of coefficients

An important point is the smoothness of $f$ related to the properties of $\theta$. 
Classes of coefficients

An important point is the smoothness of $f$ related to the properties of $\theta$.

A classical hypothesis:

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

where $a_k \to \infty$ et $L > 0$. 
Classes of coefficients

An important point is the smoothness of $f$ related to the properties of $\theta$.

A classical hypothesis:

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

where $a_k \to \infty$ et $L > 0$.

- Sobolev classes: $a_k = k^\alpha$, $\alpha > 0$. 
Theorem 2. Let $\Lambda$ be a “not too large” family of Projection, Tikhonov or Pinsker which attains the optimal rate, $\sigma = O(\varepsilon)$. 
Adaptive estimation

Theorem 2. Let $\Lambda$ be a “not too large” family of Projection, Tikhonov or Pinsker which attains the optimal rate, $\sigma = O(\varepsilon)$.

Then for any $\alpha, L > 0$ we have

$$\sup_{\theta \in \Theta} E_\theta \|\theta^* - \theta\|^2 \leq (1 + o(1)) \min_{\lambda \in \Lambda} \sup_{\theta \in \Theta} R(\lambda, \theta),$$

when $\varepsilon \to 0$. 
Theorem 2. Let $\Lambda$ be a “not too large” family of Projection, Tikhonov or Pinsker which attains the optimal rate, $\sigma = O(\varepsilon)$.

Then for any $\alpha, L > 0$ we have

$$\sup_{\theta \in \Theta} E_{\theta} \|\theta^* - \theta\|^2 \leq (1 + o(1)) \min_{\lambda \in \Lambda} \sup_{\theta \in \Theta} R(\lambda, \theta),$$

when $\varepsilon \to 0$.

Adaptive estimator which attains the optimal rate of convergence.
If \( f \) is smooth, then the truncation at \( M \) has no influence, even in the constant, since it happens after the optimal choice.
Comments

- If \( f \) is smooth \( \rightsquigarrow \) The truncation at \( M \) has no influence, even in the constant, since it happens after the optimal choice.

- We directly use the noisy singular values as the true one \( \rightsquigarrow \) No price to pay for not knowing \( b_k \).
Discrete model:

\[ Y(i) = g * f \left( \frac{i}{N} \right) + \varepsilon \sqrt{N} \xi(i), \quad i = 1, \ldots, N, \]
Simulations

Discrete model:

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

where

\[ g(t) = n(t, 0.5, 0.02), \quad \beta \sim 0.5. \]
Simulations

Discrete model:

\[ Y(i) = g \ast f \left( \frac{i}{N} \right) + \varepsilon \sqrt{N} \xi(i), \quad i = 1, \ldots, N, \]

where

\[ g(t) = n(t, 0.5, 0.02), \quad \beta \sim 0.5. \]

Estimation by truncated Fourier series

\[ f^* = \sum_{|k| \leq W^*} \frac{Y_k}{X_k} \varphi_k. \]
True fonction $f$. Estimator $f^*$. 
Oracle by projection. Estimator $f^*$. Estimator with known singular values.
Conclusion
Conclusion

- Close results between noisy and non-noisy cases.
Conclusion

- Close results between noisy and non-noisy cases.
- Limitation on the family size.
  Penalty. Loss of generality.
Conclusion

- Close results between noisy and non-noisy cases.
Conclusion

- Close results between noisy and non-noisy cases.
- Limitation on the family size.
  Penalty. Loss of generality.
  → Rather unstable method.
- Risk hull method (RHM).
  (C. and Golubev (06))
Conclusion

- Close results between noisy and non-noisy cases.

- Limitation on the family size.
  Penalty. Loss of generality.
  \[\rightarrow\text{Rather unstable method.}\]

- Risk hull method (RHM).
  (C. and Golubev (06))
  \[\rightarrow\text{Generalisation for the noisy case.}\]
  (Marteau (09))
Conclusion

- Close results between noisy and non-noisy cases.
- Limitation on the family size.
  Penalty. Loss of generality.
  → Rather unstable method.
- Risk hull method (RHM).
  (C. and Golubev (06))
  → Generalisation for the noisy case.
  (Marteau (09))
- Extension to instrumental variables?