Control and Inverse Problems Conference

A mollifier approach to nonparametric instrumental regression

Anne VANHEMS

col-written with

Pierre MARÉCHAL and Walter Cedric SIMO TAO LEE

Monastir, May 09-11, 2022
1 Setting

2 Mollification

3 Convergence analysis

4 Simulations
Setting

Mollification

Convergence analysis

Simulations

\[ Y = h(Z) + \varepsilon \]
\[ Y = h(Z) + \varepsilon \]

- \( Y, Z, \varepsilon : \Omega \to \mathbb{R} \)
\[ Y = h(Z) + \varepsilon \]

- \( Y, Z, \varepsilon : \Omega \rightarrow \mathbb{R} \)
- \( h : \mathbb{R} \rightarrow \mathbb{R} \) is the unknown regression function
\[ Y = h(Z) + \varepsilon \]

- \( Y, Z, \varepsilon : \Omega \to \mathbb{R} \)
- \( h : \mathbb{R} \to \mathbb{R} \) is the unknown \textit{regression function}.

Problem: identify \( h \) under the assumption that \( Z \) and \( \varepsilon \) are linked: 
\[ E(\varepsilon|Z) \neq 0. \]
\[ Y = h(Z) + \varepsilon \]

- \( Y, Z, \varepsilon : \Omega \to \mathbb{R} \)
- \( h : \mathbb{R} \to \mathbb{R} \) is the unknown *regression function*

Problem: identify \( h \) under the assumption that \( Z \) and \( \varepsilon \) are linked: \( E(\varepsilon|Z) \neq 0 \). A standard strategy: introduce an *instrumental variable* \( W : \Omega \to \mathbb{R}^p \), that is linked to \( Z \) and not to \( \varepsilon \).
\[ Y = h(Z) + \varepsilon \]

- \( Y, Z, \varepsilon : \Omega \rightarrow \mathbb{R} \)
- \( h : \mathbb{R} \rightarrow \mathbb{R} \) is the unknown regression function

Problem: identify \( h \) under the assumption that \( Z \) and \( \varepsilon \) are linked: \( E(\varepsilon|Z) \neq 0 \). A standard strategy: introduce an instrumental variable \( W : \Omega \rightarrow \mathbb{R}^p \), that is linked to \( Z \) and not to \( \varepsilon \).

\[ Y = h(Z) + \varepsilon, \quad E(\varepsilon|W) = 0 \]
\[ Y = h(Z) + \varepsilon \]

- \( Y, Z, \varepsilon : \Omega \to \mathbb{R} \)
- \( h : \mathbb{R} \to \mathbb{R} \) is the unknown \textit{regression function}

Problem: identify \( h \) under the assumption that \( Z \) and \( \varepsilon \) are linked: \( E(\varepsilon|Z) \neq 0 \). A standard strategy: introduce an \textit{instrumental variable} \( W : \Omega \to \mathbb{R}^p \), that is linked to \( Z \) and not to \( \varepsilon \).

\[ Y = h(Z) + \varepsilon, \quad E(\varepsilon|W) = 0 \]

\[ E(Y|W) = E(h(Z)|W) \]
Assumption 1. The laws $P_Z, P_W, P_Y$ are absolutely continuous with respect to the Lebesgue measure $\lambda$. 

Then, the equation $E(Y | W) = E(h(Z) | W)$ reduces to the functional integral equation

$$
\int f_{YW}(y, w) y \, dy = \int f_{ZW}(z, w) h(z) \, dz,
$$

$w \in \{x | f_W(x) \neq 0\}$.

Assumption 2. The kernel $f_{ZW}$ is $\lambda \otimes \lambda$-square integrable.

Then, the above equation can be written as $g = Th$, in which $T$ is a Hilbert-Schmidt operator.

Notice that for this equation to have a solution, it is necessary that $g \in L^2(R)$.

Assumption 3. The function $g(w) = \int f_{YW}(y, w) y \, dy$ belongs to $L^2(R)$. 


**Assumption 1.** The laws $P_Z, P_W, P_Y$ are absolutely continuous with respect to the Lebesgue measure $\lambda$.

Then, the equation $E(Y|W) = E(h(Z)|W)$ reduces to the functional integral equation

$$\int f_{YW}(y, w)y \, dy = \int f_{ZW}(z, w)h(z) \, dz, \quad w \in \{x | f_W(x) \neq 0\}.$$
Assumption 1. The laws $P_Z, P_W, P_Y$ are absolutely continuous with respect to the Lebesgue measure $\lambda$.

Then, the equation $E(Y|W) = E(h(Z)|W)$ reduces to the functional integral equation

$$\int f_{YW}(y, w) y \, dy = \int f_{ZW}(z, w) h(z) \, dz, \quad w \in \{ x | f_W(x) \neq 0 \}.$$ 

Assumption 2. The kernel $f_{ZW}$ is $\lambda \otimes \lambda$-square integrable.
Assumption 1. The laws $P_Z, P_W, P_Y$ are absolutely continuous with respect to the Lebesgue measure $\lambda$.

Then, the equation $E(Y|W) = E(h(Z)|W)$ reduces to the functional integral equation

$$\int f_{YW}(y, w) y \, dy = \int f_{ZW}(z, w) h(z) \, dz, \quad w \in \{x|f_W(x) \neq 0\}.$$ 

Assumption 2. The kernel $f_{ZW}$ is $\lambda \otimes \lambda$-square integrable.

Then, the above equation can be written as $g = Th$, in which $T$ is a Hilbert-Schmidt operator.
Assumption 1. The laws $P_Z, P_W, P_Y$ are absolutely continuous with respect to the Lebesgue measure $\lambda$.

Then, the equation $E(Y|W) = E(h(Z)|W)$ reduces to the functional integral equation

$$
\int f_{YW}(y, w) y \, dy = \int f_{ZW}(z, w) h(z) \, dz, \quad w \in \{x | f_W(x) \neq 0\}.
$$

Assumption 2. The kernel $f_{ZW}$ is $\lambda \otimes \lambda$-square integrable.

Then, the above equation can be written as $g = Th$, in which $T$ is a Hilbert-Schmidt operator. Notice that for this equation to have a solution, it is necessary that $g \in L^2(\mathbb{R})$. 
Assumption 1. The laws $P_Z, P_W, P_Y$ are absolutely continuous with respect to the Lebesgue measure $\lambda$.

Then, the equation $E(Y|W) = E(h(Z)|W)$ reduces to the functional integral equation

$$\int f_{YW}(y, w)y\,dy = \int f_{ZW}(z, w)h(z)\,dz, \quad w \in \{x|f_W(x) \neq 0\}.$$ 

Assumption 2. The kernel $f_{ZW}$ is $\lambda \otimes \lambda$-square integrable.

Then, the above equation can be written as $g = Th$, in which $T$ is a Hilbert-Schmidt operator. Notice that for this equation to have a solution, it is necessary that $g \in L^2(\mathbb{R})$.

Assumption 3. The function $g(w) = \int f_{YW}(y, w)y\,dy$ belongs to $L^2(\mathbb{R})$. 
Remark
The last assumption is satisfied in particular if $E[Y^2] < \infty$ and $f_W$ is bounded.
Remark
The last assumption is satisfied in particular if $E[Y^2] < \infty$ and $f_W$ is bounded.

Remark
In practice, $g$ is estimated from observed sample, and the constraint that $g \in L^2(\mathbb{R})$ may be incorporated in the estimation process.
Remark
The last assumption is satisfied in particular if $E[Y^2] < \infty$ and $f_W$ is bounded.

Remark
In practice, $g$ is estimated from observed sample, and the constraint that $g \in L^2(\mathbb{R})$ may be incorporated in the estimation process.

Remark
Note also that the operator $T$ is unknown and it also needs to be estimated from observed sample:

$$T : h \rightarrow \int f_{ZW}(z, w)h(z)\,dz.$$
1 Setting

2 Mollification

3 Convergence analysis

4 Simulations
Mollifiers in approximation theory

Theorem

Let \( \varphi \in L^1(\mathbb{R}^n) \) be such that \( \int \varphi(x) \, dx = 1 \). For every \( \beta > 0 \), let

\[
\varphi_\beta(x) := \frac{1}{\beta^n} \varphi \left( \frac{x}{\beta} \right)
\]

Let \( p \in [1, \infty) \). Then, for every \( h \in L^p(\mathbb{R}^n) \),

\[
\| \varphi_\beta * h - h \|_p \longrightarrow 0 \quad \text{as} \quad \beta \downarrow 0
\]
Mollification

Theorem
Let \( \varphi \in L^1(\mathbb{R}^n) \) be such that \( \int \varphi(x) \, dx = 1 \). For every \( \beta > 0 \), let

\[
\varphi_\beta(x) := \frac{1}{\beta^n} \varphi \left( \frac{x}{\beta} \right)
\]

Let \( p \in [1, \infty) \). Then, for every \( h \in L^p(\mathbb{R}^n) \),

\[
\| \varphi_\beta * h - h \|_p \longrightarrow 0 \quad \text{as} \quad \beta \downarrow 0
\]

Remark
The family of operators \((C_\beta)\) given by \( C_\beta h = \varphi_\beta * h \) is referred to as an approximation of unity.
Mollifiers for inverse problems

Approximate inverses and Variational mollification


Overview of approximate inverses

A function $\psi_\beta : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is a mollifier if

(i) for every $\beta > 0$ and $y \in \mathbb{R}^n$, $\psi_\beta (\cdot, y) \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$, and

$$\int_{\mathbb{R}^n} \psi_\beta(x, y) \, dy = 1$$

(ii) for every $h \in L^2(\mathbb{R}^n)$, the function $h_\beta$ defined by

$$h_\beta(y) = \langle h, \psi_\beta(\cdot, y) \rangle = \int_{\mathbb{R}^n} h(x) \psi_\beta(x, y) \, dx$$

converges to $h$ in $L^2(\mathbb{R}^n)$ as $\beta \downarrow 0$
Overview of approximate inverses

A function $\psi_\beta : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ is a \textit{mollifier} if

(i) for every $\beta > 0$ and $y \in \mathbb{R}^n$, $\psi_\beta (\cdot, y) \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$, and

$$\int_{\mathbb{R}^n} \psi_\beta (x,y) \, dy = 1$$

(ii) for every $h \in L^2(\mathbb{R}^n)$, the function $h_\beta$ defined by

$$h_\beta (y) = \langle h, \psi_\beta (\cdot, y) \rangle = \int_{\mathbb{R}^n} h(x) \psi_\beta (x,y) \, dx$$

converges to $h$ in $L^2(\mathbb{R}^n)$ as $\beta \Downarrow 0$

Assume the existence of a family of functions $(\nu_\beta (\cdot, y))$ such that

$$\forall \beta > 0, \quad \forall y \in \mathbb{R}^n, \quad T^* \nu_\beta (\cdot, y) = \psi_\beta (\cdot, y)$$
Overview of approximate inverses

A function $\psi_\beta : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ is a mollifier if

(i) for every $\beta > 0$ and $y \in \mathbb{R}^n$, $\psi_\beta(\cdot, y) \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$, and

$$\int_{\mathbb{R}^n} \psi_\beta(x, y) \, dy = 1$$

(ii) for every $h \in L^2(\mathbb{R}^n)$, the function $h_\beta$ defined by

$$h_\beta(y) = \langle h, \psi_\beta(\cdot, y) \rangle = \int_{\mathbb{R}^n} h(x) \psi_\beta(x, y) \, dx$$

converges to $h$ in $L^2(\mathbb{R}^n)$ as $\beta \downarrow 0$

Assume the existence of a family of functions $(v_\beta(\cdot, y))$ such that

$$\forall \beta > 0, \quad \forall y \in \mathbb{R}^n, \quad T^* v_\beta(\cdot, y) = \psi_\beta(\cdot, y)$$

Then $h_\beta$ is given by

$$h_\beta(y) = \langle h, T^* v_\beta(\cdot, y) \rangle = \langle Th, v_\beta(\cdot, y) \rangle = \langle g, v_\beta(\cdot, y) \rangle$$
Overview of approximate inverses

We may use the minimum norm least square solution to

\[ T^* v_\beta(\cdot, y) = \psi_\beta(\cdot, y) \]
Overview of approximate inverses

We may use the minimum norm least square solution to

$$T^* v_\beta(\cdot, y) = \psi_\beta(\cdot, y)$$

In this context, the family of mappings

$$\tilde{T}_\beta : \quad L^2(\mathbb{R}^n) \longrightarrow L^2(\mathbb{R}^n)$$

$$g \quad \mapsto \quad \langle g, v_\beta(\cdot, y) \rangle$$

is called an approximate inverse of $T$.
Overview of approximate inverses

We may use the minimum norm least square solution to

\[ T^* v_\beta (\cdot, y) = \psi_\beta (\cdot, y) \]

In this context, the family of mappings

\[ \tilde{T}_\beta : \quad L^2 (\mathbb{R}^n) \rightarrow L^2 (\mathbb{R}^n) \]

\[ g \mapsto \langle g, v_\beta (\cdot, y) \rangle \]

is called an approximate inverse of \( T \)

If \( \psi_\beta (x, y) = \phi_\beta (y - x) \), the function \( h_\beta \) is then a convolution of \( h \):

\[ h_\beta (y) = \int_{\mathbb{R}^n} h(x) \psi_\beta (x, y) \, dx = \int_{\mathbb{R}^n} h(x) \phi_\beta (y - x) \, dx = (\phi_\beta * h)(y) \]
Mollification in variational form

\[ Th = g \]

- \( T: L^2(V) \rightarrow L^2(\mathbb{R}^p) \) is bounded linear and injective. where

\[
L^2(V) = \{ h \in L^2(\mathbb{R}) \mid \text{supp} h \subset V \}, \quad V \text{ bounded.}
\]
Mollification in variational form

\[ Th = g \]

- \( T : L^2(V) \to L^2(\mathbb{R}^p) \) is bounded linear and injective. where
  \[ L^2(V) = \{ h \in L^2(\mathbb{R}) \mid \text{supp} \ h \subset V \} , \quad V \text{ bounded.} \]

- \( \inf \{ \| Th \| \mid h \in (\ker T)^\perp , \| h \| = 1 \} = 0 \)
Mollification in variational form

\[ Th = g \]

- \( T : L^2(V) \rightarrow L^2(\mathbb{R}^p) \) is bounded linear and injective. where

\[ L^2(V) = \left\{ h \in L^2(\mathbb{R}) \mid \text{supp} \, h \subset V \right\}, \quad V \text{ bounded}. \]

- \( \inf \left\{ \| Th \| \mid h \in (\ker T)^\perp, \| h \| = 1 \right\} = 0 \)

**Principle**: Due to the ill-posedness of the problem, give up recovering the true object \( h^\dagger \) but instead try to recover a smooth version of \( h^\dagger \), namely

\[ C_\beta h^\dagger = \varphi_\beta \ast h^\dagger \]
Heuristics

- \( h_\circ = C_\beta h^\dagger + (I - C_\beta)h^\dagger \)
Heuristics

- $h_\circ = C_\beta h^\dagger + (I - C_\beta)h^\dagger$
- Undesired component: $(I - C_\beta)h^\dagger$
Heuristics

- $h_\circ = C_\beta h^\dagger + (I - C_\beta)h^\dagger$
- Undesired component: $(I - C_\beta)h^\dagger$
- Penalty term: $\| (I - C_\beta)h \|^2$
• $h_\circ = C_\beta h^\dagger + (I - C_\beta)h^\dagger$
• Undesired component: $(I - C_\beta)h^\dagger$
• Penalty term: $\| (I - C_\beta)h \|^2$
• A natural choice for the fit term is $\| g - Th \|^2$
Regularization scheme

- Define the *target object* to be $C_\beta h^\dagger$
Regularization scheme

- Define the *target object* to be \( C_\beta h^\dagger \)
- Define the *reconstructed object* \( h_\beta \) as the solution of

\[
\min_{h \in L^2(V)} \frac{1}{2} \| g - Th \|_{L^2(\mathbb{R}^p)} + \frac{1}{2} \| (I - C_\beta)h \|_{L^2(\mathbb{R})}
\]

• Regard \((C_\beta)\beta \in (0, 1]\) as an approximation of unity, and consider the asymptotic behavior as \( \beta \downarrow 0 \)
Regularization scheme

- Define the target object to be $C_\beta h^*$
- Define the reconstructed object $h_\beta$ as the solution of

$$
\text{Min}_{h \in L^2(V)} \frac{1}{2} \left\| g - Th \right\|_{L^2(\mathbb{R}^p)} + \frac{1}{2} \left\| (I - C_\beta)h \right\|_{L^2(\mathbb{R})}
$$

$$
h_\beta := (T^*T + (I - C_\beta)^*(I - C_\beta))^{-1} T^* g
$$
Regularization scheme

- Define the target object to be $C_\beta h^+$
- Define the reconstructed object $h_\beta$ as the solution of

$$\min_{h \in L^2(V)} \frac{1}{2} \| g - Th \|_{L^2(\mathbb{R}^p)} + \frac{1}{2} \| (I - C_\beta)h \|_{L^2(\mathbb{R})}$$

$$h_\beta := (T^* T + (I - C_\beta)^* (I - C_\beta))^{-1} T^* g$$

- Regard $(C_\beta)_{\beta \in (0,1]}$ as an approximation of unity, and consider the asymptotic behavior as $\beta \downarrow 0$
Main issues

- Wellposedness for fixed $\beta > 0$
Main issues

- Wellposedness for fixed $\beta > 0$
- Asymptotic behavior as $\beta \downarrow 0$
Main issues

• Wellposedness for fixed $\beta > 0$
• Asymptotic behavior as $\beta \downarrow 0$
• Computational aspects
1 Setting

2 Mollification

3 Convergence analysis

4 Simulations
Theorem (Consistency)

Assume that $h^\dagger \in L^2(V) \cap H^s(\mathbb{R})$, that $g = Th^\dagger$ and let

$$h_\beta := (T^*T + (I - C_\beta)^*(I - C_\beta))^{-1}T^*g.$$ 

Then $(h_\beta)_{\beta \in (0,1]}$ is bounded and weakly compact in $L^2(V)$. 


**Theorem (Consistency)**

Assume that $h^* \in L^2(V) \cap H^s(\mathbb{R})$, that $g = Th^*$ and let

$$h_\beta := (T^*T + (I - C_\beta)^*(I - C_\beta))^{-1}T^*g.$$  

Then $(h_\beta)_{\beta \in (0,1]}$ is bounded and weakly compact in $L^2(V)$.

Moreover, for every sequence $(\beta_n)_n$ converging to 0,

- $h_{\beta_n} \rightharpoonup h^*$;
- $\lim_{R \to \infty} \sup_{n \in \mathbb{N}} \int_{\|x\| > R} |h_{\beta_n}(x)|^2 \, dx = 0$;
- $\sup_{n \in \mathbb{N}} \|T_\delta h_{\beta_n} - h_{\beta_n}\|_{L^2(\mathbb{R})} \to 0$ as $\delta \to 0$. 
**Theorem (Consistency)**

Assume that $h^\dagger \in L^2(V) \cap H^s(\mathbb{R})$, that $g = Th^\dagger$ and let

$$h_\beta := (T^*T + (I - C_\beta)^*(I - C_\beta))^{-1}T^*g.$$  

Then $(h_\beta)_{\beta \in (0,1]}$ is bounded and weakly compact in $L^2(V)$.

Moreover, for every sequence $(\beta_n)_n$ converging to 0,

- $h_\beta_n \rightharpoonup h^\dagger$;
- $\lim_{R \to \infty} \sup_{n \in \mathbb{N}} \int_{\|x\| > R} |h_\beta_n(x)|^2 \, dx = 0$;
- $\sup_{n \in \mathbb{N}} \|\mathcal{T}_\delta h_\beta_n - h_\beta_n\|_{L^2(\mathbb{R})} \to 0$ as $\delta \to 0$.

Consequently, by the Fréchet-Kolmogorov theorem,

$$h_\beta \to h^\dagger \text{ as } \beta \to 0.$$
Theorem (Consistency under approximate setting)

With the notation and assumptions of the previous theorem, let \( T_n \) and \( g_n \) be approximations of \( T \) and \( g \):

\[
T_n \rightarrow T \text{ and } g_n \rightarrow g \text{ as } n \rightarrow \infty.
\]

Let

\[
h_{\beta,n} := (T_n^*T_n + (I - C_\beta)^* (I - C_\beta))^{-1} T_n^* g_n
\]

\[
h_{\beta,n}^\dagger := (T_n^*T_n + (I - C_\beta)^* (I - C_\beta))^{-1} T_n^* T_n h_{\beta,n}^\dagger.
\]
Theorem (Consistency under approximate setting)
With the notation and assumptions of the previous theorem, let $T_n$ and $g_n$ be approximations of $T$ and $g$:

$$T_n \to T \text{ and } g_n \to g \text{ as } n \to \infty.$$ 

Let

$$h_{\beta,n} := (T_n^* T_n + (I - C_\beta)^* (I - C_\beta))^{-1} T_n^* g_n$$

$$h_{\beta,n}^\dagger := (T_n^* T_n + (I - C_\beta)^* (I - C_\beta))^{-1} T_n^* T_n h^\dagger.$$ 

Then

(i) $\| h^\dagger - h_{\beta,n}^\dagger \|_{L^2} \to 0$ as $\beta \downarrow 0$;

(ii) $\| h_{\beta,n}^\dagger - h_{\beta,n} \|_{L^2} \leq \bar{C} \beta^{-2s} \left( \| (T_n - T) h^\dagger \|_{L^2} + \| g - g_n \|_{L^2} \right)$. 

Corollary

In the above setting, there exists a parameter choice rule $\beta(n) \to 0$ as $n \to \infty$ such that

$$\|h^\dagger - h_{\beta,n}\|_{L^2} \to 0 \text{ as } n \to \infty.$$
Corollary

In the above setting, there exists a parameter choice rule $\beta(n) \to 0$ as $n \to \infty$ such that

$$\| h^\dagger - h_{\beta,n} \|_{L^2} \to 0 \text{ as } n \to \infty.$$  

For example, if $\| (T - T_n)h^\dagger \| = \| g - g_n \| = O(1/n)$, then $\beta(n) = n^{-\tau/(2s)}$ with $\tau < 1$ is a converging a priori selection rule.
1 Setting

2 Mollification

3 Convergence analysis

4 Simulations
In the simulations, we compare 5 methods:

- **Mollification**: \( h_{\beta,n} = (T_n^* T_n + (I - C_\beta)^*(I - C_\beta))^{-1} T_n^* g_n; \)
In the simulations, we compare 5 methods:

- **Mollification**: $h_{\beta,n} = (T^*T_n + (I - C_\beta)^*(I - C_\beta))^{-1}T_n^*g_n$;
- **Tikhonov**: $h_{\alpha,n} = (T_n^*T_n + \alpha I)^{-1}T_n^*g_n$;

Where $C_\beta$ is the mollification operator and $g_n$ is the noisy data.
In the simulations, we compare 5 methods:

- **Mollification**: $h_{\beta,n} = \left(T_n^*T_n + (I - C_\beta)^*(I - C_\beta)\right)^{-1}T_n^*g_n$;
- **Tikhonov**: $h_{\alpha,n} = \left(T_n^*T_n + \alpha I\right)^{-1}T_n^*g_n$;
- **Generalized Tikhonov**: $h_{\alpha,n} = \left(T_n^*T_n + \alpha D_2^*D_2\right)^{-1}T_n^*g_n$;
In the simulations, we compare 5 methods:

- **Mollification**: \( h_{\beta,n} = (T_n^*T_n + (I - C_\beta)(I - C_\beta))^{-1}T_n^*g_n; \)
- **Tikhonov**: \( h_{\alpha,n} = (T_n^*T_n + \alpha I)^{-1}T_n^*g_n; \)
- **Generalized Tikhonov**: \( h_{\alpha,n} = (T_n^*T_n + \alpha D_2^*D_2)^{-1}T_n^*g_n; \)
- **Landweber**: \( h_{k+1,n} = T_n^*(g_n - \mu T_nh_{k,n}) + h_{k,n}, \mu < 1, k = 1, 2, \ldots; \)
In the simulations, we compare 5 methods:

- **Mollification**: \( h_{\beta,n} = \left( T_n^* T_n + (I - C_{\beta} )^* (I - C_{\beta} ) \right)^{-1} T_n^* g_n \);
- **Tikhonov**: \( h_{\alpha,n} = \left( T_n^* T_n + \alpha I \right)^{-1} T_n^* g_n \);
- **Generalized Tikhonov**: \( h_{\alpha,n} = \left( T_n^* T_n + \alpha D_2^* D_2 \right)^{-1} T_n^* g_n \);
- **Landweber**: \( h_{k+1,n} = T_n^* ( g_n - \mu T_n h_{k,n} ) + h_{k,n}, \mu < 1, \)
  \( k = 1, 2, ... \);
- **Spectral cut-off**: \( h_{k,n} = \sum_{j=1}^{k} \frac{1}{\sigma_j} \langle v_j, g_n \rangle u_j, \)
  \( k = 1, 2, ..., N. \)
For each regularization method, we computed the reconstruction error

$$\| h^\dagger - h_{\text{reg.par},n} \|$$
For each regularization method, we computed the reconstruction error

$$\| h^\dagger - h_{\text{reg} \cdot \text{par} \cdot n} \|$$

- Mollification: $$\| h^\dagger - h_{\beta \cdot n} \|$$
- Tikhonov and Generalized Tikhonov: $$\| h^\dagger - h_{\alpha \cdot n} \|$$
- Landweber: $$\| h^\dagger - h_{k \cdot n} \|$$
- Spectral Cutoff: $$\| h^\dagger - h_{k \cdot n} \|$$
We consider a sample size $n = 3000$ and we generate the data as follows:
We consider a sample size $n = 3000$ and we generate the data as follows:

- we generate independently $(W, \varepsilon)$ from a uniform distribution $\mathcal{U}(-0.5, 0.5)$
We consider a sample size $n = 3000$ and we generate the data as follows:

- we generate independently $(W, \varepsilon)$ from a uniform distribution $\mathcal{U}(-0.5, 0.5)$
- We generate $Z$ as:
  \[ Z = m(W) + 0.5\varepsilon, \quad m(W) = 0.25 + 0.5W \]
We consider a sample size $n = 3000$ and we generate the data as follows:

- we generate independently $(W, \varepsilon)$ from a uniform distribution $\mathcal{U}(-0.5, 0.5)$
- We generate $Z$ as: $Z = m(W) + 0.5\varepsilon$, $m(W) = 0.25 + 0.5W$
- $Y$ is defined as $Y = h_l(Z) + \varepsilon$, $l = 1, 2$
We consider a sample size $n = 3000$ and we generate the data as follows:

- we generate independently $(W, \varepsilon)$ from a uniform distribution $\mathcal{U}(-0.5, 0.5)$
- We generate $Z$ as:
  \[ Z = m(W) + 0.5\varepsilon, \quad m(W) = 0.25 + 0.5W \]
- $Y$ is defined as $Y = h_l(Z) + \varepsilon$, \quad $l = 1, 2$
  - **Smooth case**: $h_1 = f(0.4, 0.1) + f(0.65, 0.075)$ truncated to the interval $[0, 1]$, where the function $f(\mu, \sigma)$ is the p.d.f of the gaussian of mean $\mu$ and standard deviation $\sigma$. 
We consider a sample size $n = 3000$ and we generate the data as follows:

- we generate independently $(W, \varepsilon)$ from a uniform distribution $\mathcal{U}(-0.5, 0.5)$

- We generate $Z$ as:
  $$Z = m(W) + 0.5\varepsilon, \quad m(W) = 0.25 + 0.5W$$

- $Y$ is defined as $Y = h_l(Z) + \varepsilon, \quad l = 1, 2$
  - **Smooth case**: $h_1 = f(0.4, 0.1) + f(0.65, 0.075)$ truncated to the interval $[0, 1]$, where the function $f(\mu, \sigma)$ is the p.d.f of the gaussian of mean $\mu$ and standard deviation $\sigma$.
  - **Nonsmooth case**: $h_2(t) = \exp(-|x - 0.5|)$
• The mollifier $\varphi_\beta$ is the centered gaussian kernel with standard deviation $\beta$;
• The mollifier $\varphi_\beta$ is the centered gaussian kernel with standard deviation $\beta$;

• The discretization of the problem is done by projection onto 100 finite dimensional basis of gate functions on $[0, 1]$. The finite dimensional equation is given by

$$M_{ji} = \mathbb{E}[\phi_i(Z) \psi_j(W)], \text{ and}$$

$$\bar{r}_j = \mathbb{E}[Y \psi_j(W)], \ i,j = 1, \ldots, 100.$$
The data

- $Y_i$ vs $Z_i$ for $\varphi_1$
- $Y_i$ vs $Z_i$ for $\varphi_2$
**Figure**: Error versus regularization parameter
Best approximation (smooth case)

Figure: Comparison best approximation of each regularization method for the function $\varphi_1$ in case of projection onto gate functions.
Best approximation (nonsmooth case)

Figure: Comparison best approximation of each regularization method for the function $\varphi_2$ in case of projection onto gate functions.
A Monte-Carlo experiment

Monte-Carlo performances of the 5 methods with $M = 1000$

Figure: Results of Monte Carlo simulation for the functions $\varphi_1$ and $\varphi_2$. 
Thank you for your attention!


