Control & Inverse Problems

Deconvolution of probability densities by mollification

Pierre Maréchal

Institut de Mathématiques de Toulouse
Université Paul Sabatier

pr.marechal@gmail.com

Monastir

May 9-11, 2022
Collaborators

THORSTEN HOHAGE
Göttingen University

LÉOPOLD SIMAR
ISBA, Université Catholique de Louvain-la-Neuve

ANNE VANHEMS
TBS and TSE, Université de Toulouse
1 Setting

2 Classical methods

3 Mollification
   - Overview
   - Approximate inverses
   - The case of deconvolution
   - Variational mollification
   - The filtering viewpoint

4 Convergence analysis
   - Consistency
   - Convergence rates
Consider the equation $Y = X + \varepsilon$ in which:

1. $Y$ is the observed random vector;
2. $X$ is the latent random vector;
3. $\varepsilon$ is a random noise vector.
Consider the equation $Y = X + \varepsilon$ in which:

1. $Y$ is the observed random vector;
2. $X$ is the latent random vector;
3. $\varepsilon$ is a random noise vector.

Standing assumptions:

(A1) $X$ and $\varepsilon$ are independent;

(A2) $Y$, $X$ and $\varepsilon$ have densities with respect to the Lebesgue measure, denoted respectively by $g$, $f$ and $\gamma$;

(A3) both $f$ and $g$ belong to $L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$. 
Consider the equation $Y = X + \varepsilon$ in which:

1. $Y$ is the observed random vector;
2. $X$ is the latent random vector;
3. $\varepsilon$ is a random noise vector.

Standing assumptions:

(A1) $X$ and $\varepsilon$ are independent;

(A2) $Y$, $X$ and $\varepsilon$ have densities with respect to the Lebesgue measure, denoted respectively by $g$, $f$ and $\gamma$;

(A3) both $f$ and $g$ belong to $L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$.

The density $f$ then satisfies the equation $T_\gamma f = g$, in which $T_\gamma$ is the convolution operator

$$T_\gamma : L^2(\mathbb{R}) \rightarrow L^2(\mathbb{R})$$

$$f \mapsto T_\gamma f := f \ast \gamma.$$
The density $g$ is in fact *unknown*, but estimated from the statistical sample $Y_1, \ldots, Y_n$. 
The density $g$ is in fact unknown, but estimated from the statistical sample $Y_1, \ldots, Y_n$. The unknown density $g$ is then replaced by a nonparametric estimator $g_n$. 
The density $g$ is in fact unknown, but estimated from the statistical sample $Y_1, \ldots, Y_n$. The unknown density $g$ is then replaced by a nonparametric estimator $g_n$.

**Notation**

- Fourier transform of an integrable function $h$ on $\mathbb{R}^d$:

\[
\hat{h}(\xi) = \int e^{-2i\pi \langle x, \xi \rangle} h(x) \, dx.
\]
The density $g$ is in fact unknown, but estimated from the statistical sample $Y_1, \ldots, Y_n$. The unknown density $g$ is then replaced by a nonparametric estimator $g_n$.

**Notation**

- Fourier transform of an integrable function $h$ on $\mathbb{R}^d$:
  \[
  \hat{h}(\xi) = \int e^{-2i\pi \langle x, \xi \rangle} h(x) \, dx.
  \]

- Corresponding Fourier-Plancherel operator on $L^2(\mathbb{R}^d)$: $U$. 
1 Setting

2 Classical methods

3 Mollification
   ■ Overview
   ■ Approximate inverses
   ■ The case of deconvolution
   ■ Variational mollification
   ■ The filtering viewpoint

4 Convergence analysis
   ■ Consistency
   ■ Convergence rates
Tikhonov regularization

\[ f_\alpha^{TK} = (T^*_\gamma T_\gamma + \alpha I)^{-1} T^*_\gamma g \quad \text{or} \quad \hat{f}_\alpha^{TK} = \frac{\tilde{\gamma}}{|\tilde{\gamma}|^2 + \alpha} \hat{g} \]
Tikhonov regularization

\[ f^{TK}_\alpha = (T^*_\gamma T_\gamma + \alpha I)^{-1} T^*_\gamma g \quad \text{or} \quad \hat{f}^{TK}_\alpha = \frac{\hat{\gamma}}{|\hat{\gamma}|^2 + \alpha \hat{g}} \]

Remarks

- Variational regularization method:
  \[ \min_f \| g - T_\gamma f \|^2 + \alpha \| f \|^2 \]
Tikhonov regularization

\[ f_{\alpha}^{TK} = (T_\gamma^* T_\gamma + \alpha I)^{-1} T_\gamma^* g \quad \text{or} \quad \hat{f}_{\alpha}^{TK} = \frac{\hat{\gamma}}{|\hat{\gamma}|^2 + \alpha} \hat{g} \]

Remarks

- Variational regularization method:

\[
\min_f \| g - T_\gamma f \|^2 + \alpha \| f \|^2
\]

- Penalizes uniformly \( |\hat{f}(\xi)|^2 \):

\[
\min_f \| \hat{g} - \hat{\gamma} \cdot \hat{f} \|^2 + \alpha \| \hat{f} \|^2
\]
Tikhonov regularization

\[ f_{\alpha}^{TK} = (T^*_\gamma T_\gamma + \alpha I)^{-1} T^*_\gamma g \quad \text{or} \quad \hat{f}_{\alpha}^{TK} = \frac{\hat{\gamma}}{\|\hat{\gamma}\|^2 + \alpha \hat{g}} \]

Remarks

- Variational regularization method:
  \[ \min_f \| g - T_\gamma f \|^2 + \alpha \| f \|^2 \]

- Penalizes uniformly \( |\hat{f}(\xi)|^2 \):
  \[ \min_f \| \hat{g} - \hat{\gamma} \cdot \hat{f} \|^2 + \alpha \| \hat{f} \|^2 \]

- Does not allow to recover a density: \( \hat{f}_{\alpha}^{TK}(0) \neq 1 \)
Deconvolution kernels

\[
\widehat{f}_h^{DK} = \frac{\hat{\phi}_h}{\hat{\gamma}} \hat{g} \quad \text{or} \quad \widehat{f}_h^{DK} = \left( T_{\gamma}^* T_{\gamma} \right)^{-1} T_{\gamma}^* C_h g
\]
Deconvolution kernels

\[ \hat{f}_h^{DK} = \frac{\hat{\phi}_h}{\hat{\gamma}} \hat{g}, \quad \text{or} \quad f_h^{DK} = (T_\gamma^* T_\gamma)^{-1} T_\gamma^* C_h \hat{g} \]

Remarks

- Restrictive assumptions on \( \phi_h \) and \( \gamma \):

\[ \sup_{\xi} \left| \frac{\phi_h(\xi)}{\hat{\gamma}(\xi)} \right| < \infty \quad \text{and} \quad \int \left| \frac{\phi_h(\xi)}{\hat{\gamma}(\xi)} \right| \, d\xi < \infty \]
Deconvolution kernels

Deconvolution kernels

\[ \hat{f}_h^{\text{DK}} = \frac{\hat{\phi}_h}{\hat{\gamma}} \hat{g} \quad \text{or} \quad \hat{f}_h^{\text{DK}} = (T^{\ast} T_{\gamma})^{-1} T^{\ast} C_h g \]

Remarks

- Restrictive assumptions on \( \phi_h \) and \( \gamma \):

  \[
  \sup_{\xi} \left| \frac{\hat{\phi}_h(\xi)}{\hat{\gamma}(\xi)} \right| < \infty \quad \text{and} \quad \int \left| \frac{\hat{\phi}_h(\xi)}{\hat{\gamma}(\xi)} \right| \, d\xi < \infty
  \]

- Not \textit{a priori} a variational method, but solution of:

  \[
  \min_f \| C_h g - T_{\gamma} f \|^2
  \]
Spectral cut-off

\[ \hat{f}_{SC} = \frac{1}{|\hat{\gamma}|^2 \geq a} \hat{g} \]
Spectral cut-off

\[ \hat{f}_{SC} = \frac{1}{|\hat{\gamma}|^2 \geq a} \hat{g} \]

Remarks

- Particular case of deconvolution kernels method with

\[ \hat{\phi}_a = 1 |\hat{\gamma}|^2 \geq a \]
Spectral cut-off

\[ \hat{f}_{SC} = \frac{1}{\hat{g}} \frac{1}{|\hat{\gamma}|^2 \geq a} \hat{g} \]

Remarks

- Particular case of deconvolution kernels method with
  \[ \hat{\phi}_a = 1_{|\hat{\gamma}|^2 \geq a} \]

- Gibbs phenomena can be expected: sinc impulse response.
1 Setting

2 Classical methods

3 Mollification
   - Overview
   - Approximate inverses
   - The case of deconvolution
   - Variational mollification
   - The filtering viewpoint

4 Convergence analysis
   - Consistency
   - Convergence rates
Setting

Classical methods

Mollification

Overview

Approximate inverses

The case of deconvolution

Variational mollification

The filtering viewpoint

Convergence analysis

Consistency

Convergence rates
Consider the general ill-posed linear operator equation

\[ Tf = g \]
Consider the general ill-posed linear operator equation

\[ Tf = g \]

- \( T : F \rightarrow G \) bounded linear operator
Consider the general ill-posed linear operator equation

\[ Tf = g \]

- \( T: F \rightarrow G \) bounded linear operator
- \( F, G \) separable, infinite dimensional, Hilbert
Consider the general ill-posed linear operator equation

\[ Tf = g \]

- \( T : F \to G \) bounded linear operator
- \( F, G \) separable, infinite dimensional, Hilbert
- \( \inf \left\{ \| Tf \| \mid f \in (\ker T)^\perp, \| f \| = 1 \right\} = 0 \)
Consider the general ill-posed linear operator equation

$$ Tf = g $$

- $T : F \rightarrow G$ bounded linear operator
- $F, G$ separable, infinite dimensional, Hilbert
- $\inf \{ \| Tf \| \mid f \in (\ker T)^\perp, \| f \| = 1 \} = 0$

$T^\dagger$ is densely defined and unbounded
Consider the general ill-posed linear operator equation

\[ Tf = g \]

- \( T : F \to G \) bounded linear operator
- \( F, G \) separable, infinite dimensional, Hilbert
- \( \inf \{ \| Tf \| \mid f \in (\ker T)^\perp, \| f \| = 1 \} = 0 \)

\( T^\dagger \) is densely defined and unbounded

**Remark**

If \( T \) is injective, the last condition reduces to

\[ \inf \{ \| Tf \| \mid \| f \| = 1 \} = 0 \]
Highlights

Mollifiers were introduced in the field of Partial Differential Equations by K.O. Friedrichs:

- **K.O. Friedrichs**, *The identity of weak and strong extensions of differential operators*, Transactions of the AMS, 1944
Highlights

Mollifiers were introduced in the field of Partial Differential Equations by K.O. Friedrichs:

- **K.O. Friedrichs, The identity of weak and strong extensions of differential operators**, Transactions of the AMS, 1944

Let’s mention a well-known approximation theorem:

**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 \( f \in L^p(\mathbb{R}^n) \),

\[
\| \varphi_\beta * f - f \|_p \longrightarrow 0 \quad \text{as} \quad \beta \downarrow 0
\]
Mollifiers were used in Inverse Problems in various forms:

- Data smoothing
- Hilbert space duality
- Variational formulation

Mollifiers were used in Inverse Problems in various forms:

- **Data smoothing**
- **Hilbert space duality**
- **Variational formulation**


Mollifiers were used in Inverse Problems in various forms:

- Data smoothing
- Hilbert space duality
- Variational formulation

Variational setting: general case

\( Tf = g \)

- \( T: F \rightarrow G \) bounded linear operator, injective
- \( F, G \) separable, infinite dimensional, Hilbert
- \( \inf \left\{ \| Tf \| \mid f \in (\ker T)^\perp, \| f \| = 1 \right\} = 0 \)
Variational setting: general case

\[ Tf = g \]

- \( T : F \rightarrow G \) bounded linear operator, injective
- \( F, G \) separable, infinite dimensional, Hilbert
- \( \inf \left\{ \| Tf \| \mid f \in (\ker T)^\perp, \| f \| = 1 \right\} = 0 \)

*One must give up recovering the true object \( f \)!
$Tf = g$

- $T : F \to G$ bounded linear operator, injective
- $F, G$ separable, infinite dimensional, Hilbert
- $\inf \{ \| Tf \| \mid f \in (\ker T)^{\perp}, \| f \| = 1 \} = 0$

One must give up recovering the true object $f_\circ$!

New target: recovering $C_\beta f_\circ := \varphi_\beta \ast f_\circ$?
Variational setting: general case

\[ Tf = g \]

- \( T: F \rightarrow G \) bounded linear operator, injective
- \( F, G \) separable, infinite dimensional, Hilbert
- \( \inf \left\{ \|Tf\| \mid f \in (\ker T)^ot, \|f\| = 1 \right\} = 0 \)

One must give up recovering the true object \( f_\circ \)!

New target: recovering \( C_\beta f_\circ := \varphi_\beta * f_\circ \)?

Our purpose: do this in the framework of variational methods
Heuristics

\[ f_\circ = C_\beta f_\circ + (I - C_\beta) f_\circ \]
Heuristics

- $f_\circ = C_\beta f_\circ + (I - C_\beta) f_\circ$

  Undesired component: $(I - C_\beta) f_\circ$
Heuristics

\[ f_\circ = C_\beta f_\circ + (I - C_\beta)f_\circ \]

- Undesired component: \((I - C_\beta)f_\circ\)
- Penalty term: \(\| (I - C_\beta)f \|^2\)
Heuristics

- $f_\circ = C_\beta f_\circ + (I - C_\beta)f_\circ$
- Undesired component: $(I - C_\beta)f_\circ$
- Penalty term: $\| (I - C_\beta)f \|^2$
- Assume we can generate the data $g_\beta$ corresponding to $C_\beta f_\circ$
Heuristics

- $f_\beta = C_\beta f_\beta + (I - C_\beta)f_\beta$

- Undesired component: $(I - C_\beta)f_\beta$

- Penalty term: $\| (I - C_\beta)f \|^2$

- Assume we can generate the data $g_\beta$ corresponding to $C_\beta f_\beta$

- Then a natural choice for the fit term is $\| g_\beta - Tf \|^2$
Define the *target object* to be $C_\beta f_0$. 
Regularization scheme

- Define the *target object* to be $C_\beta f$.
- Generate, from $g \sim Tf$ an approximation $g_\beta$ of $TC_\beta f$. 
Regularization scheme

- Define the *target object* to be $C_β f_0$
- Generate, from $g \simeq Tf_0$ an approximation $g_β$ of $TC_β f_0$
- Define the *reconstructed object* $f_β$ as the solution of

$$\text{Min}_{f \in F} \frac{1}{2} \|g_β - Tf\|^2_G + \frac{1}{2} \| (I - C_β)f \|^2_F$$
Regularization scheme

- Define the target object to be $C_\beta f_\circ$
- Generate, from $g \simeq Tf_\circ$ an approximation $g_\beta$ of $TC_\beta f_\circ$
- Define the reconstructed object $f_\beta$ as the solution of

$$
\min_{f \in F} \frac{1}{2} \|g_\beta - Tf\|_G^2 + \frac{1}{2} \|(I - C_\beta)f\|_F^2
$$

$$
f_\beta := \left( T^* T + (I - C_\beta)^* (I - C_\beta) \right)^{-1} T^* g_\beta
$$
Regularization scheme

- Define the target object to be $C_\beta f_\circ$
- Generate, from $g \approx Tf_\circ$ an approximation $g_\beta$ of $TC_\beta f_\circ$
- Define the reconstructed object $f_\beta$ as the solution of

$$
\min_{f \in F} \quad \frac{1}{2}\|g_\beta - Tf\|_G^2 + \frac{1}{2}\|(I - C_\beta)f\|_F^2
$$

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

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

- Finding the regularized data $g_\beta$
Main issues

- Finding the regularized data $g_\beta$

This is achieved if we find $\Phi_\beta$ such that $\Phi_\beta T = TC_\beta$, since then

$$\Phi_\beta g = \Phi_\beta Tf = TC_\beta f$$
Main issues

- Finding the regularized data $g_\beta$

This is achieved if we find $\Phi_\beta$ such that $\Phi_\beta T = TC_\beta$, since then

$$\Phi_\beta g = \Phi_\beta Tf = TC_\beta f$$

*intertwining relationship*
Main issues

- Finding the regularized data $g_\beta$

  *This is achieved if we find $\Phi_\beta$ such that $\Phi_\beta T = TC_\beta$, since then $\Phi_\beta g = \Phi_\beta Tf = TC_\beta f$*

  *intertwining relationship*

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

- Finding the regularized data $g_\beta$

  *This is achieved if we find $\Phi_\beta$ such that $\Phi_\beta T = TC_\beta$, since then $\Phi_\beta g = \Phi_\beta Tf = TC_\beta f$*

  *intertwining relationship*

- Wellposedness for fixed $\beta > 0$

- Asymptotic behavior as $\beta \downarrow 0$
Main issues

- Finding the regularized data $g_\beta$

  This is achieved if we find $\Phi_\beta$ such that $\Phi_\beta T = TC_\beta$, since then
  
  $\Phi_\beta g = \Phi_\beta Tf = TC_\beta f$

  intertwining relationship

- Wellposedness for fixed $\beta > 0$

- Asymptotic behavior as $\beta \downarrow 0$

- Computational aspects
Intertwining relationship for deconvolution

\[ TC_\beta = \Phi_\beta T \quad \text{with} \quad C_\beta := U^{-1} \hat{\phi}_\beta U \]
Intertwining relationship for deconvolution

\[ TC_\beta = \Phi_\beta T \quad \text{with} \quad C_\beta := U^{-1} \hat{\phi}_\beta U \]

\[ T = U^{-1} \hat{k} U \]

convolution by \( k \)
Intertwining relationship for deconvolution

\[
TC_\beta = \Phi_\beta T \quad \text{with} \quad C_\beta := U^{-1} \hat{\phi}_\beta U
\]

\[
T = U^{-1} \hat{k} U
\]

\text{convolution by} \ k

Convolution operators commute

\[
TC_\beta = C_\beta T
\]
Intertwining relationship for deconvolution

\[ TC_\beta = \Phi_\beta T \quad \text{with} \quad C_\beta := U^{-1} \hat{\phi}_\beta U \]

\[ T = U^{-1} \hat{k} U \]

Convolution operators commute

\[ TC_\beta = C_\beta T \]

\[ \Phi_\beta = C_\beta \]
Another example: Fourier truncation

\[ TC_\beta = \Phi_\beta T \quad \text{with} \quad C_\beta := U^{-1} \hat{\phi}_\beta U \]
Another example: Fourier truncation

\[TC_\beta = \Phi_\beta T \quad \text{with} \quad C_\beta := U^{-1} \hat{\phi}_\beta U\]

\[T = \mathbb{1}_W U \text{ with } W \text{ bounded domain} \quad \text{[Fourier truncation]}\]
Another example: Fourier truncation

\[ TC_\beta = \Phi_\beta T \quad \text{with} \quad C_\beta := U^{-1} \hat{\phi}_\beta U \]

\[ T = 1_W U \quad \text{with} \quad W \text{ bounded domain} \quad \text{[Fourier truncation]} \]

\[ TC_\beta = 1_W U U^{-1} \hat{\phi}_\beta U = \hat{\phi}_\beta 1_W U = \hat{\phi}_\beta T \]
Another example: Fourier truncation

\[ TC_\beta = \Phi_\beta T \quad \text{with} \quad C_\beta := U^{-1} \hat{\phi}_\beta U \]

\[ T = 1_W U \quad \text{with} \quad W \text{ bounded domain} \quad \text{[Fourier truncation]} \]

\[ TC_\beta = 1_W U U^{-1} \hat{\phi}_\beta U = \hat{\phi}_\beta 1_W U = \hat{\phi}_\beta T \]

\[ \Phi_\beta = \hat{\phi}_\beta \]
One more example: the Radon transform

\[ TC_\beta = \Phi_\beta T \quad \text{with} \quad C_\beta := U^{-1} \hat{\phi}_\beta U \]
One more example: the Radon transform

\[ TC_\beta = \Phi_\beta T \quad \text{with} \quad C_\beta := U^{-1} \hat{\phi}_\beta U \]

\[ (Tf)(\theta, s) = \int f(x) \delta(s - \langle \theta, x \rangle) \, dx \quad [\text{Radon transformation}] \]
One more example: the Radon transform

\[ TC_\beta = \Phi_\beta T \quad \text{with} \quad C_\beta := U^{-1} \hat{\phi}_\beta U \]

\[ (Tf)(\theta, s) = \int f(x) \delta(s - \langle \theta, x \rangle) \, dx \quad \text{[Radon transformation]} \]

\[ T(f_1 \ast f_2) = Tf_1 \ast Tf_2 \text{ with } \ast \text{ convolution w.r.t. } s \]
One more example: the Radon transform

\[ TC_\beta = \Phi_\beta T \quad \text{with} \quad C_\beta := U^{-1} \hat{\Phi}_\beta U \]

\[ (Tf)(\theta, s) = \int f(x) \delta(s - \langle \theta, x \rangle) \, dx \quad \text{[Radon transformation]} \]

\[ T(f_1 \ast f_2) = Tf_1 \otimes Tf_2 \quad \text{with \ convolution \ w.r.t. \ s} \]

\[ TC_\beta f = T(\varphi_\beta \ast f) = T\varphi_\beta \otimes Tf \]
One more example: the Radon transform

$$TC_\beta = \Phi_\beta T \quad \text{with} \quad C_\beta := U^{-1} \hat{\phi}_\beta U$$

$$\quad (Tf)(\theta, s) = \int f(x)\delta(s - \langle \theta, x \rangle)\,dx \quad [\text{Radon transformation}]$$

$$\quad T(f_1 * f_2) = Tf_1 \otimes Tf_2 \quad \text{with} \quad \otimes \text{convolution w.r.t.} \ s$$

$$\quad TC_\beta f = T(\varphi_\beta * f) = T\varphi_\beta \otimes Tf$$

$$\quad \Phi_\beta = (g \mapsto T\varphi_\beta \otimes g)$$
Applications under study:

- **Nonparametric instrumental regression**
  (with A. VANHEMS and W. SIMO)

- **Cauchy problem for the inhomogeneous Helmoltz equation**
  (with F. TRIKI and W. SIMO)

- **Inversion of the real Laplace Transform**
  (with F. TRIKI and W. SIMO)
1 Setting

2 Classical methods

3 Mollification
   - Overview
   - Approximate inverses
     - The case of deconvolution
   - Variational mollification
     - The filtering viewpoint

4 Convergence analysis
   - Consistency
   - Convergence rates
Consider the inverse problem $Tf = g$ in which $T: L^2(\mathbb{R}^n) \rightarrow L^2(\mathbb{R}^n)$ has unbounded pseudo-inverse as usual.
Consider the inverse problem $Tf = g$ in which $T : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ has unbounded pseudo-inverse as usual.

**Definition**

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

1. 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

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

2. for every $f \in L^2(\mathbb{R}^n)$, the function $f_\beta$ defined by

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

converges to $f$ in $L^2(\mathbb{R}^n)$ as $\beta \downarrow 0$
Now, assuming the existence of a family of functions $(v_\beta(\cdot, y))$ such that

$$\forall \beta > 0, \ \forall y \in \mathbb{R}^n, \ \ T^*v_\beta(\cdot, y) = \psi_\beta(\cdot, y) \quad (1)$$

we see that $f_\beta$ is then given by

$$f_\beta(y) = \langle f, T^*v_\beta(\cdot, y) \rangle = \langle Tf, v_\beta(\cdot, y) \rangle = \langle g, v_\beta(\cdot, y) \rangle$$
Now, assuming the existence of a family of functions \((v_\beta(\cdot,y))\) such that

\[
\forall \beta > 0, \quad \forall y \in \mathbb{R}^n, \quad T^* v_\beta(\cdot,y) = \psi_\beta(\cdot,y) \tag{1}
\]

we see that \(f_\beta\) is then given by

\[
f_\beta(y) = \langle f, T^* v_\beta(\cdot,y) \rangle = \langle Tf, v_\beta(\cdot,y) \rangle = \langle g, v_\beta(\cdot,y) \rangle
\]

More generally, if \(\psi_\beta(\cdot,y)\) belongs to \(\mathcal{D}(T^*\dagger) = \text{ran}T^* + (\text{ran}T^*)^\perp\), then the minimum norm least square solution to (1) is used instead, and denoted by \(v_\beta(\cdot,y)\) again.
Now, assuming the existence of a family of functions \((v_\beta(\cdot, y))\) such that
\[
\forall \beta > 0, \quad \forall y \in \mathbb{R}^n, \quad T^* v_\beta(\cdot, y) = \psi_\beta(\cdot, y) \quad (1)
\]
we see that \(f_\beta\) is then given by
\[
f_\beta(y) = \langle f, T^* v_\beta(\cdot, y) \rangle = \langle Tf, v_\beta(\cdot, y) \rangle = \langle g, v_\beta(\cdot, y) \rangle
\]

More generally, if \(\psi_\beta(\cdot, y)\) belongs to \(\mathcal{D}(T^*\dagger) = \text{ran} T^* + (\text{ran} T^*)^\perp\), then the minimum norm least square solution to (1) is used instead, and denoted by \(v_\beta(\cdot, y)\) again. The family of mappings
\[
\tilde{R}_\beta : \ L^2(\mathbb{R}^n) \longrightarrow L^2(\mathbb{R}^n)
\]
\[
g \longmapsto \langle g, v_\beta(\cdot, y) \rangle
\]
is then called an \textit{approximate inverse} of \(T\)
If $\psi_\beta(x, y) = \varphi_\beta(y - x)$, the function $f_\beta$ is then a convolution of $f$: 

$$f_\beta(y) = \int_{\mathbb{R}^n} f(x) \psi_\beta(x, y) \, dx = \int_{\mathbb{R}^n} f(x) \varphi_\beta(y - x) \, dx = (\varphi_\beta \ast f)(y)$$
If $\psi_\beta(x, y) = \varphi_\beta(y - x)$, the function $f_\beta$ is then a convolution of $f$:

$$f_\beta(y) = \int_{\mathbb{R}^n} f(x) \psi_\beta(x, y) \, dx = \int_{\mathbb{R}^n} f(x) \varphi_\beta(y - x) \, dx = (\varphi_\beta * f)(y)$$

The family of functions $(\varphi_\beta)$, indexed by $\beta$ in some interval of the form $(0, \beta_o]$, emulates the Dirac distribution as $\beta \downarrow 0$. 
If $\psi_\beta(x, y) = \varphi_\beta(y - x)$, the function $f_\beta$ is then a convolution of $f$:

$$f_\beta(y) = \int_{\mathbb{R}^n} f(x) \psi_\beta(x, y) \, dx = \int_{\mathbb{R}^n} f(x) \varphi_\beta(y - x) \, dx = (\varphi_\beta * f)(y)$$

The family of functions $(\varphi_\beta)$, indexed by $\beta$ in some interval of the form $(0, \beta_0]$, emulates the Dirac distribution as $\beta \downarrow 0$. It is referred to as an approximate unity.
If $\psi_\beta(x, y) = \varphi_\beta(y - x)$, the function $f_\beta$ is then a convolution of $f$:

$$f_\beta(y) = \int_{\mathbb{R}^n} f(x) \psi_\beta(x, y) \, dx = \int_{\mathbb{R}^n} f(x) \varphi_\beta(y - x) \, dx = (\varphi_\beta * f)(y)$$

The family of functions $(\varphi_\beta)$, indexed by $\beta$ in some interval of the form $(0, \beta_0]$, emulates the Dirac distribution as $\beta \downarrow 0$. It is referred to as an approximate unity.

A standard way to produce such an approximation of unity is to choose an integrable function $\varphi$ and to define $\varphi_\beta$ by

$$\varphi_\beta(x) := \frac{1}{\beta^n} \varphi \left( \frac{x}{\beta} \right), \quad x \in \mathbb{R}^n$$
1 Setting

2 Classical methods

3 Mollification
   • Overview
   • Approximate inverses
   • The case of deconvolution
   • Variational mollification
   • The filtering viewpoint

4 Convergence analysis
   • Consistency
   • Convergence rates
Recall the DK solution:

\[
\hat{f}_h^{DK} = \frac{\hat{\phi}_h}{\gamma} \hat{g}.
\]
Recall the DK solution:
\[ \hat{f}_{h}^{DK} = \frac{\hat{\phi}_{h}}{\hat{\gamma}} \hat{g}. \]

We readily see that it correspond to Murio’s *data smooing* approach.
Recall the DK solution:

\[ f_h^{DK} = \frac{\phi_h}{\gamma} \hat{g}. \]

We readily see that it correspond to Murio’s *data smoothing* approach. We will see below that it is also Louis & Maass’ *approximate inverse* solution, but that the variational approach differs, and is in fact preferable.
Approximate inverses

\[ T = T_\gamma = U^* \hat{\gamma} U \]
Approximate inverses

\[ T = T_\gamma = U^*\hat{\gamma}U \]

\[ T^* = U^*\hat{\gamma}U \]
Approximate inverses

- $T = T_\gamma = U^* \hat{\gamma} U$
- $T^* = U^* \hat{\gamma} U$
- $(T^*)^{-1} = U^* \frac{1}{\hat{\gamma}} U$
Approximate inverses

\[ T = T_\gamma = U^* \hat{\gamma} U \]

\[ T^* = U^* \tilde{\gamma} U \]

\[ (T^*)^{-1} = U^* \frac{1}{\tilde{\gamma}} U \]

Recall: for every \( \beta > 0 \) and every \( y \in \mathbb{R}^n \),

\[ T^* \nu_\beta (\cdot, y) = \psi_\beta (\cdot, y) \]
Approximate inverses

\[ T = T_\gamma = U^* \hat{\gamma} U \]

\[ T^* = U^* \tilde{\gamma} U \]

\[ (T^*)^{-1} = U^* \frac{1}{\tilde{\gamma}} U \]

Recall: for every \( \beta > 0 \) and every \( y \in \mathbb{R}^n \),

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

\[ U^* \tilde{\gamma} U v_\beta (\cdot, y) = \psi_\beta (\cdot, y) \]
Approximate inverses

- \( T = T_\gamma = U^* \hat{\gamma} U \)
- \( T^* = U^* \tilde{\gamma} U \)
- \( (T^*)^{-1} = U^* \frac{1}{\tilde{\gamma}} U \)

Recall: for every \( \beta > 0 \) and every \( y \in \mathbb{R}^n \),

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

\[
U^* \tilde{\gamma} U v_\beta(\cdot, y) = \psi_\beta(\cdot, y)
\]
\[ Uv_{\beta}(\cdot, y)(\xi) = \left[ \frac{1}{\hat{\gamma}} U \psi_{\beta}(\cdot, y) \right](\xi) \]
\[ = \frac{1}{\hat{\gamma}(\xi)} \int e^{-2i\pi \langle x, \xi \rangle} \varphi_{\beta}(y - x) \, dx \]
\[ = \frac{1}{\hat{\gamma}(\xi)} \int e^{-2i\pi \langle (y - x'), \xi \rangle} \varphi_{\beta}(x') \, dx' \]
\[ = \frac{1}{\hat{\gamma}(\xi)} e^{-2i\pi \langle y, \xi \rangle} \hat{\varphi}_{\beta}(-\xi) \]
\[ = e^{-2i\pi \langle y, \xi \rangle} \frac{\hat{\varphi}_{\beta}(\xi)}{\hat{\gamma}(\xi)} \]
The approximate inverse solution is then given by:

\[ f_\beta(y) = \langle g, v_\beta(\cdot, y) \rangle \]
The approximate inverse solution is then given by:

\[
f_\beta(y) = \langle g, v_\beta(\cdot, y) \rangle \\
= \langle Ug, Uv_\beta(\cdot, y) \rangle
\]
The approximate inverse solution is then given by:

\[ f_\beta(y) = \langle g, v_\beta(\cdot, y) \rangle = \langle Ug, Uv_\beta(\cdot, y) \rangle = \int \hat{g}(\xi)e^{2i\pi \langle y, \xi \rangle} \frac{\hat{\phi}_\beta(\xi)}{\hat{\gamma}(\xi)} d\xi \]
The approximate inverse solution is then given by:

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

\[ = \langle Ug, Uv_\beta(\cdot, y) \rangle \]

\[ = \int \hat{g}(\xi) e^{2i\pi \langle y, \xi \rangle} \frac{\hat{\phi}_\beta(\xi)}{\hat{\gamma}(\xi)} d\xi \]

\[ = \left( U^* \frac{\hat{\phi}_\beta}{\hat{\gamma}} Ug \right)(y) \]
The approximate inverse solution is then given by:

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

\[
= \langle Ug, Uv_\beta(\cdot, y) \rangle
\]

\[
= \int \hat{g}(\xi)e^{2i\pi\langle y, \xi \rangle} \frac{\hat{\phi}_\beta(\xi)}{\hat{\gamma}(\xi)} d\xi
\]

\[
= \left( U^* \frac{\hat{\phi}_\beta}{\hat{\gamma}} Ug \right)(y)
\]

The approximate inverse operator is to be compared with the inverse of \( T : L^2(\mathbb{R}^n) \rightarrow \text{ran } T \), which is given by

\[
T^{-1} = U^* \frac{1}{\hat{\gamma}} U
\]
1 Setting

2 Classical methods

3 Mollification
   - Overview
   - Approximate inverses
   - The case of deconvolution
   - Variational mollification
   - The filtering viewpoint

4 Convergence analysis
   - Consistency
   - Convergence rates
Min\[ f \in L^2(\mathbb{R}) \]
\[
\frac{1}{2} \left\| C_\beta g - T_\gamma f \right\|^2 + \frac{1}{2} \left\| (I - C_\beta)f \right\|^2
\]
Deconvolution

Pierre Maréchal

Setting
Classical methods

Mollification
Overview
Approximate inverses
The case of deconvolution

Variational mollification
The filtering viewpoint

Convergence analysis
Consistency
Convergence rates

\[
\min_{f \in L^2(\mathbb{R})} \frac{1}{2} \| C_\beta g - T_\gamma f \|^2 + \frac{1}{2} \| (I - C_\beta) f \|^2
\]

\[
f_\beta := (T_\gamma^* T_\gamma + (I - C_\beta)^* (I - C_\beta))^{-1} T_\gamma^* C_\beta g
\]
Deconvolution

Pierre Maréchal

Setting

Classical methods

Mollification

Overview

Approximate inverses

The case of deconvolution

Variational mollification

The filtering viewpoint

Convergence analysis

Consistency

Convergence rates

Minimize $f \in L^2(\mathbb{R})$

$$\frac{1}{2} \| C_\beta g - T_\gamma f \|^2 + \frac{1}{2} \| (I - C_\beta) f \|^2$$

$$f_\beta := (T_\gamma^* T_\gamma + (I - C_\beta)^* (I - C_\beta))^{-1} T_\gamma^* C_\beta g$$

$$\hat{f}_\beta = \frac{\bar{\gamma} \hat{\phi}_\beta}{|\hat{\gamma}|^2 + |1 - \hat{\phi}_\beta|^2} \cdot \hat{g}$$
Stability in $L^2(\mathbb{R})$

$$f_\beta = (T_\gamma^* T_\gamma + (I - C_\beta)^*(I - C_\beta))^{-1} T_\gamma^* C_\beta g$$
Stability in $L^2(\mathbb{R})$

$$f_\beta = (T_\gamma^* T_\gamma + (I - C_\beta)^*(I - C_\beta))^{-1} T_\gamma^* C_\beta g$$

Thus $f_\beta$ depends continuously on $g$ if and only if the operator

$$T_\gamma^* T_\gamma + (I - C_\beta)^*(I - C_\beta)$$

has a bounded inverse
Stability in $L^2(\mathbb{R})$

$$f_\beta = \left(T_\gamma^* T_\gamma + (I - C_\beta)^*(I - C_\beta)\right)^{-1} T_\gamma^* C_\beta g$$

Thus $f_\beta$ depends continuously on $g$ if and only if the operator

$$T_\gamma^* T_\gamma + (I - C_\beta)^*(I - C_\beta)$$

has a bounded inverse

$$T_\gamma^* T_\gamma + \alpha(I - C_\beta)^*(I - C_\beta) = U^* \left(|\hat{\gamma}|^2 + |1 - \hat{\phi}_\beta|^2\right) U$$
Stability in $L^2(\mathbb{R})$

\[ f_\beta = \left( T_\gamma^* T_\gamma + (I - C_\beta)^*(I - C_\beta) \right)^{-1} T_\gamma^* C_\beta g \]

Thus $f_\beta$ depends continuously on $g$ if and only if the operator

\[ T_\gamma^* T_\gamma + (I - C_\beta)^*(I - C_\beta) \]

has a bounded inverse

\[ T_\gamma^* T_\gamma + \alpha (I - C_\beta)^*(I - C_\beta) = U^* \left( |\hat{\gamma}|^2 + |1 - \hat{\phi}_\beta|^2 \right) U \]

Make sure that $|\hat{\gamma}|^2 + |1 - \hat{\phi}_\beta|^2$ remains bounded away from zero!
Example: Lévy kernels

\[ \hat{\phi}(\xi) = \exp\left(-|\xi|^s\right) \text{ with } s \in ]0, 2] \]
Example: Lévy kernels

\[ \hat{\phi}(\xi) = \exp\left(-|\xi|^s\right) \text{ with } s \in ]0, 2] \]

\[ \varphi = U^{-1} \hat{\phi} \]
Example: Lévy kernels

\[ \hat{\varphi}(\xi) = \exp\left(-|\xi|^s\right) \text{ with } s \in ]0, 2]\]

\[ \varphi = U^{-1}\hat{\varphi} \]

**Proposition**

*The above defined function \( \varphi \) is positive, even, decreasing on \( \mathbb{R}_+ \) and of class \( C^\infty \)*
1 Setting

2 Classical methods

3 Mollification
   ■ Overview
   ■ Approximate inverses
   ■ The case of deconvolution
   ■ Variational mollification
   ■ The filtering viewpoint

4 Convergence analysis
   ■ Consistency
   ■ Convergence rates
Overview

\[ \hat{f}_{\text{REG}} = \Phi \hat{g} \]
Overview

\( \hat{f}_{\text{REG}} = \Phi \hat{g} \)

<table>
<thead>
<tr>
<th>Setting</th>
<th>Classical methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overview</td>
<td>Approximate inverses</td>
</tr>
<tr>
<td>The case of deconvolution</td>
<td>Variational mollification</td>
</tr>
<tr>
<td>The filtering viewpoint</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mollification</th>
<th>Overview</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approximate inverses</td>
<td>Variational mollification</td>
</tr>
<tr>
<td>The filtering viewpoint</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Convergence analysis</th>
<th>Consistency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Convergence rates</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Functional to be minimized</th>
<th>Filter ( \Phi )</th>
</tr>
</thead>
<tbody>
<tr>
<td>TK</td>
<td>( \frac{1}{2} | g - \gamma * f |^2 + \frac{\alpha}{2} | f |^2 )</td>
</tr>
<tr>
<td>DK</td>
<td>( \frac{1}{2} | \phi_h * g - \gamma * f |^2 )</td>
</tr>
<tr>
<td>SC</td>
<td>( \frac{1}{2} | \mathbb{1}_{{\gamma \geq a}} * g - \gamma * f |^2 )</td>
</tr>
<tr>
<td>MO</td>
<td>( \frac{1}{2} | \phi_\beta * g - \gamma * f |^2 + \frac{1}{2} | f - \phi_\beta * f |^2 )</td>
</tr>
</tbody>
</table>
1 Setting

2 Classical methods

3 Mollification
   - Overview
   - Approximate inverses
   - The case of deconvolution
   - Variational mollification
   - The filtering viewpoint

4 Convergence analysis
   - Consistency
   - Convergence rates
The density $g$ is unknown and is estimated using a sample of observations $Y_1, \ldots, Y_n$. 
The density $g$ is unknown and is estimated using a sample of observations $Y_1, \ldots, Y_n$.

We denote by $g_n$ (resp. $\hat{g}_n$) the estimator of $g$ (resp. $\hat{g}$).
The density $g$ is unknown and is estimated using a sample of observations $Y_1, ..., Y_n$.

We denote by $g_n$ (resp. $\hat{g}_n$) the estimator of $g$ (resp. $\hat{g}$).

Our objective is now to study the convergence of $f_{\beta,n}$ to $f$. 
The density $g$ is unknown and is estimated using a sample of observations $Y_1, \ldots, Y_n$.

We denote by $g_n$ (resp. $\hat{g}_n$) the estimator of $g$ (resp. $\hat{g}$).

Our objective is now to study the convergence of $f_{\beta,n}$ to $f$

$$f_{\beta,n} = (T_\gamma^* T_\gamma + (I - C_\beta)^* (I - C_\beta))^{-1} T_\gamma^* C_\beta g_n$$
Theorem (Consistency)

Assume \( g_n \) is a consistent nonparametric estimator of \( g \), that is, that \( E\| g_n - g \| \) goes to zero as \( n \) goes to infinity. Let \( f_{\beta_n} \) denote the mollified solution corresponding to data \( g_n \). There then exist a sequence \( \beta_n \downarrow 0 \) such that

\[
E\| f_{\beta_n,n} - f_\circ \| \longrightarrow 0 \quad \text{as} \quad n \to \infty.
\]
1 Setting

2 Classical methods

3 Mollification
   - Overview
   - Approximate inverses
   - The case of deconvolution
   - Variational mollification
   - The filtering viewpoint

4 Convergence analysis
   - Consistency
   - Convergence rates
Assumption (filter shape assumption)

There exists a strictly decreasing differentiable function $\Phi : [0, \infty) \to \mathbb{R}$ such that

$$\forall \xi \in \mathbb{R}^d, \quad \hat{\phi}(\xi) = \Phi(|\xi|)$$

and constants $s, C_\Phi \in \mathbb{R}_+^*$ with the following properties:

$$\forall t \in [0, 1], \quad \frac{1}{2} \leq \Phi(t) \leq 1$$

$$\forall t \in [0, 1], \quad C_\Phi^{-1} t^s \leq 1 - \Phi(t) \leq C_\Phi t^s$$

$$\forall t \in [0, 1], \quad |\Phi'(t)| \leq C_\Phi t^{s-1}$$

$$\int_0^{\infty} \Phi(t)^2 t^{d-1} dt < \infty.$$
Definition

For a function $f \in L^2(\mathbb{R}^d)$ let

$$e_f(t) := \int_{|\xi| > t} |\hat{f}(\xi)|^2 \, d\xi, \quad t > 0.$$  

The **Besov-Nikolskiĭ space** $B^u_{2,\infty}(\mathbb{R}^d)$ of smoothness index $u > 0$ is the set of all $f \in L^2(\mathbb{R}^d)$ for which

$$\|f\|_{B^u_{2,\infty}} := \left( \sup_{t>0} (1 + t)^{2u} e_f(t) \right)^{1/2}$$

is finite.
Definition

For a function $f \in L^2(\mathbb{R}^d)$ let

$$e_f(t) := \int_{|\xi| > t} |\hat{f}(\xi)|^2 \, d\xi, \quad t > 0.$$ 

The **Besov-Nikolskiĭ space** $B^u_{2,\infty}(\mathbb{R}^d)$ of smoothness index $u > 0$ is the set of all $f \in L^2(\mathbb{R}^d)$ for which

$$\|f\|_{B^u_{2,\infty}} := \left( \sup_{t > 0} (1 + t)^{2u} e_f(t) \right)^{1/2}$$

is finite.

With the above norm, $B^u_{2,\infty}(\mathbb{R}^d)$ is a Banach space.
Definition

For a function $f \in L^2(\mathbb{R}^d)$ let

$$e_f(t) := \int_{|\xi|>t} |\hat{f}(\xi)|^2 \, d\xi, \quad t > 0.$$ 

The Besov-Nikolskiĭ space $B^u_{2,\infty}(\mathbb{R}^d)$ of smoothness index $u > 0$ is the set of all $f \in L^2(\mathbb{R}^d)$ for which

$$\|f\|_{B^u_{2,\infty}} := \left( \sup_{t>0} (1 + t)^{2u} e_f(t) \right)^{1/2}$$

is finite.

With the above norm, $B^u_{2,\infty}(\mathbb{R}^d)$ is a Banach space. The Sobolev space $H^u(\mathbb{R}^d)$ is a subspace of $B^u_{2,\infty}(\mathbb{R}^d)$ since

$$(1 + t)^{2u} e_f(t) \leq \int_{|\xi|>t} (1 + |\xi|)^{2u} |\hat{f}(\xi)|^2 \, d\xi \leq \|f\|_{H^u}^2.$$
Throughout, we use the standard decomposition

\[ f_{\beta,n} - f = f_{\beta,n} - f_{\beta} + f_{\beta} - f. \]
Throughout, we use the standard decomposition

\[ f_{\beta,n} - f = f_{\beta,n} - f_{\beta} + f_{\beta} - f. \]

For convenience, we refer to

- the deterministic error \( \|f_{\beta} - f\| \) as the bias (or regularization bias);
Throughout, we use the standard decomposition

\[ f_{\beta,n} - f = f_{\beta,n} - f_{\beta} + f_{\beta} - f. \]

For convenience, we refer to

- the deterministic error \( \|f_{\beta} - f\| \) as the bias (or regularization bias);
- the statistical quadratic error \( E(\|f_{\beta,n} - f_{\beta}\|^2) \) as the variance.
Ordinary smoothness

We assume here that the density $\gamma$ of $\varepsilon$ satisfies the following *ordinary smoothness condition*:

$$C^{-1} (1 + |\xi|)^{-a} \leq |\hat{\gamma}(\xi)| \leq C (1 + |\xi|)^{-a}, \quad \xi \in \mathbb{R}^d,$$

for some $a > 0$ and $C \geq 1$. 
Ordinary smoothness

We assume here that the density $\gamma$ of $\varepsilon$ satisfies the following *ordinary smoothness condition*:

$$C^{-1} (1 + |\xi|)^{-a} \leq |\hat{\gamma}(\xi)| \leq C (1 + |\xi|)^{-a}, \quad \xi \in \mathbb{R}^d,$$

for some $a > 0$ and $C \geq 1$. In this case, the problem is mildly ill-posed.
Deconvolution
Pierre Maréchal

Setting
Classical methods

Mollification
Overview
Approximate inverses
The case of deconvolution
Variational mollification
The filtering viewpoint

Convergence analysis
Consistency
Convergence rates

Ordinary smoothness

We assume here that the density \( \gamma \) of \( \varepsilon \) satisfies the following *ordinary smoothness condition*:

\[
C^{-1} (1 + |\xi|)^{-a} \leq |\hat{\gamma}(\xi)| \leq C (1 + |\xi|)^{-a}, \quad \xi \in \mathbb{R}^d,
\]

for some \( a > 0 \) and \( C \geq 1 \). In this case, the problem is mildly ill-posed.

**Theorem (bound on bias term)**

Suppose the above filter shape assumption and the ordinary smoothness condition are satisfied. Then for \( 0 < u < s + a \) the following statements are equivalent:

\[
\begin{align*}
\triangleright & \quad f \in B^u_{2,\infty} (\mathbb{R}^d); \\
\text{sup} & \quad \beta^{-\frac{su}{s+a}} \|f - f_\beta\| < \infty. \\
\text{sup} \quad 0<\beta\leq1
\end{align*}
\]

Moreover, \( \|f - f_\beta\| = O(\beta^s) \) as \( \beta \downarrow 0 \) if \( f \in H^{s+a}(\mathbb{R}^d) \).
Proposition (bound on the variance term)

We have

$$E\left(\|f_{\beta,n} - f_{\beta}\|^2\right) \leq \frac{2}{n} \|\Phi_{\beta}\|^2_{L^2(\mathbb{R}^d)}.$$ 

In particular, if the filter shape assumption and the ordinary smoothness condition are satisfied and $4s \geq d - 2a$, then

$$E\left(\|f_{\beta,n} - f_{\beta}\|^2\right) = O\left(\frac{1}{n}\beta^{-\frac{s(d+2a)}{s+a}}\right).$$
Ordinary smoothness

Now we can state an order-optimal bound on the mean integrated square error in terms of the sample size.
Ordinary smoothness

Now we can state an order-optimal bound on the mean integrated square error in terms of the sample size. We write \( \psi_1(x) \sim \psi_2(x) \) as \( x \to x_0 \) for two positive functions \( \psi_1 \) and \( \psi_2 \) if \( \liminf_{x \to x_0} \frac{\psi_1(x)}{\psi_2(x)} > 0 \) and \( \limsup_{x \to x_0} \frac{\psi_1(x)}{\psi_2(x)} < \infty \).
Ordinary smoothness

Now we can state an order-optimal bound on the mean integrated square error in terms of the sample size. We write $\psi_1(x) \sim \psi_2(x)$ as $x \to x_0$ for two positive functions $\psi_1$ and $\psi_2$ if $\liminf_{x \to x_0} \frac{\psi_1(x)}{\psi_2(x)} > 0$ and $\limsup_{x \to x_0} \frac{\psi_1(x)}{\psi_2(x)} < \infty$.

Theorem (convergence rate)

Suppose the above filter shape assumption and the ordinary smoothness condition are satisfied, that $4s \geq d - 2a$ and that $f \in B_{2,\infty}^u(\mathbb{R}^d)$ for some $0 < u < s + a$ or $f \in H^{s+a}(\mathbb{R}^d)$ for $u = s + a$. Then, for

$$\beta \sim n^{-\frac{s+a}{2su+s(d+2a)}},$$

we obtain the optimal rate

$$\mathbb{E} \left( \|f_{\beta,n} - f\|^2 \right) = O \left( n^{-\frac{u}{u+a+d/2}} \right) \quad \text{as} \quad n \to \infty.$$
Super-smoothness

We assume here that the density $\gamma$ of $\varepsilon$ satisfies the following super-smoothness condition:

$$C^{-1} \exp (-\kappa |\xi|^a) \leq |\hat{\gamma}(\xi)|^2 \leq C \exp (-\kappa |\xi|^a), \quad \xi \in \mathbb{R}^d.$$  

for some constants $a, \kappa > 0$ and $C \geq 1$.  

We assume here that the density $\gamma$ of $\varepsilon$ satisfies the following super-smoothness condition:
Super-smoothness

We assume here that the density $\gamma$ of $\varepsilon$ satisfies the following super-smoothness condition:

$$C^{-1} \exp\left(-\kappa |\xi|^a\right) \leq |\hat{\gamma}(\xi)|^2 \leq C \exp\left(-\kappa |\xi|^a\right), \quad \xi \in \mathbb{R}^d.$$

for some constants $a, \kappa > 0$ and $C \geq 1$. In this case, the problem is severely ill-posed.
We assume here that the density $\gamma$ of $\varepsilon$ satisfies the following super-smoothness condition:

$$C^{-1} \exp\left(-\kappa |\xi|^a\right) \leq |\hat{\gamma}(\xi)|^2 \leq C \exp\left(-\kappa |\xi|^a\right), \quad \xi \in \mathbb{R}^d.$$ 

for some constants $a, \kappa > 0$ and $C \geq 1$. In this case, the problem is severely ill-posed. Note that $a = 2$ corresponds to Gaussian errors $\varepsilon$ and $a = 1$ to Cauchy errors.
Super-smoothness

We assume here that the density $\gamma$ of $\epsilon$ satisfies the following **super-smoothness condition**:

$$C^{-1} \exp\left(-\kappa |\xi|^a\right) \leq |\hat{\gamma}(\xi)|^2 \leq C \exp\left(-\kappa |\xi|^a\right), \quad \xi \in \mathbb{R}^d.$$ 

for some constants $a, \kappa > 0$ and $C \geq 1$. In this case, the problem is severely ill-posed. Note that $a = 2$ corresponds to Gaussian errors $\epsilon$ and $a = 1$ to Cauchy errors.

**Theorem (bound on bias term)**

Suppose that the filter shape assumption with $s > \frac{1}{2}$ and the super-smoothness condition are satisfied. Then the following statements are equivalent for $u > 0$:

$$f \in B_{2,\infty}^u(\mathbb{R}^d),$$

$$\sup_{0<\beta<1} (-\ln \beta)^{u/a} \|f - f\beta\| < \infty.$$
Proposition (bound on variance term)

If the super-smoothness condition holds true, then for any $b > 4s$ the statistical error satisfies

$$E\left( \| f_{\beta,n} - E(f_{\beta,n}) \|^2 \right) = O\left( \frac{1}{n \beta^{2b}} \right).$$
Super-smoothness

Proposition (bound on variance term)

If the super-smoothness condition holds true, then for any $b > 4s$ the statistical error satisfies

$$E \left( \left\| f_{\beta,n} - E(f_{\beta,n}) \right\|^2 \right) = O \left( \frac{1}{n} \beta^{-2b} \right).$$

Combining the previous results yields the following logarithmic convergence rates with respect to the sample size:

Theorem (convergence rate)

Suppose that the filter shape assumption with $s > \frac{1}{2}$ and the super-smoothness condition are satisfied. Let $f \in B^u_{2,\infty}(\mathbb{R}^d)$ for some $u > 0$ and let $\beta = \frac{1}{n}$. Then

$$E \left( \left\| f_{\beta,n} - f \right\|^2 \right) = O \left( (\ln n)^{-2u/a} \right) \quad \text{as } n \to \infty.$$
Thankyou !