

DISCUSSION PAPER

2018/28

A mollifier approach to the deconvolution of probability densities

Maréchal, P., Simar, L. and A. Vanhems

A mollifier approach to the deconvolution of probability densities

P. Maréchal¹, L. Simar², and A. Vanhems³

¹*Université de Toulouse, IMT and TSE. pr.marechal@gmail.com*

²*Institut de Statistique, Biostatistique et Sciences Actuarielles, Université Catholique de Louvain-la-Neuve, Belgium. leopold.simar@uclouvain.be*

³*Toulouse Business School, France. a.vanhems@tbs-education.fr*

November 14, 2018

Abstract

In this paper, we use a mollifier approach to regularize the deconvolution, which has been used in research fields like medical imaging, tomography, astrophysics but, to the best of our knowledge, never in statistics or econometrics. We show that the analysis of this new regularization method offers a unifying and generalizing frame in order to compare the benefits of various different filter-type techniques like deconvolution kernels, Tikhonov or spectral cut-off method. In particular, the mollifier approach allows to relax some restrictive assumptions required for the deconvolution problem, and has better stabilizing properties compared to spectral cutoff and Tikhonov. We prove the asymptotic convergence of our estimator and provide simulations analysis to compare the finite sample properties of our estimator with respect to the well-known methods.

1 Introduction

Deconvolution is a classical issue in statistics and econometrics and is well-known to be an ill-posed problem. Various regularization methods have been proposed, among which we find the deconvolution kernels (see Stefanski and Carroll [18], Fan [12], Carroll and Hall [6], Devroye [10] among others), the Tikhonov regularization (as in Carrasco and Florens [5]), and the spectral cutoff (see Mair and Ruymgaart [14])

and Johannes [13]). There exists also streams of the literature in deconvolution using projection-based methods (as in Comte et al. [7–9], or in Van Rooij and Ruymgaart [20], Efromovich [11], Pensky and Vidakovic [17], Butucea and Matias [3], Meister [16] among others).

In this paper, we propose a different method to regularize the deconvolution problem, which uses a regularization principle introduced in the deterministic setting, and has been applied in several fields signal and image processing (deconvolution of images in astronomy, computerized tomography as in Maréchal et al. [15]). To the best of our knowledge, this regularization principle has never been applied in the stochastic setting. We refer to it as the *regularization by mollification*, or merely as the *mollification*. As we shall see, this regularization approach shares features with each of the aforementioned methodologies, but it offers definite advantages, notably in terms of flexibility, and also in terms of the usual tradeoff between stability and fidelity to the model.

In order to compare the benefits of our new regularization method with various approaches, we consider the class of *filter-type techniques*, which provides us with a unifying framework and encompasses the deconvolution kernels, the spectral cutoff, Tikhonov regularization and mollification. In this paper, we shall:

1. introduce the mollification method for the deconvolution of random variables;
2. show that mollification allows to relax some restrictive assumptions required by the deconvolution kernels, while having better stabilizing properties than the Tikhonov regularization, and better morphological properties than the spectral cutoff (in terms of impulse response).
3. prove the asymptotic convergence of the corresponding estimator;
4. provide simulations to compare the finite sample properties of our estimator with respect to the well-known methods.

The paper is organized as follows. In Section 2, we review the analysis the of deconvolution problem and we give an overview of some well-known methodologies. In Section 3, we present the mollification approach to the deconvolution problem. In Section 4, we propose a unifying framework, whose main interest will be to provide us with criteria for the comparison of their respective performances. In Section 5, we consider numerical aspects of both the reconstruction and the estimation of its stability, and we illustrate our approach by means of simulations. We finally conclude in Section 6 and provide as an appendix basic results on inverse problems and Fourier analysis.

2 Setting

In this section, we present the deconvolution problem in $L^2(\mathbb{R})$ and recall the reasons for its illposedness. We also recall some standard regularization methods for deconvolution: the deconvolution kernels approach, the Tikhonov regularization and the spectral cutoff. All these methods share with mollification that they can be interpreted in terms of filtering.

2.1 Model and ill-posedness

Consider the equation

$$Y = X + \varepsilon, \quad (1)$$

in which Y is the observed random variable, X is the latent random variable, and ε is a random noise. Throughout, we make the following assumptions:

- (A1) the random variables X and ε are independent;
- (A2) all random variables Y , X and ε have densities with respect to the Lebesgue measure, denoted respectively by $g_\circ$, $f_\circ$ and γ ;
- (A3) both $f_\circ$ and $g_\circ$ belong to $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, and $\gamma \in L^1(\mathbb{R})$.

We are mostly interested here in the case where γ is known, either from modeling or from empirical observations. In this case, the density of X satisfies the equation

$$T_\gamma f = g_\circ, \quad (2)$$

in which T_γ is the convolution operator

$$\begin{aligned} T_\gamma: \quad L^2(\mathbb{R}) &\longrightarrow L^2(\mathbb{R}) \\ f &\longmapsto T_\gamma f := f * \gamma. \end{aligned}$$

The density $g_\circ$ is usually not known, but estimated from the statistical sample $Y_1, \dots, Y_n$. The unknown density $g_\circ$ is then replaced by a nonparametric estimator g_n using, for example, a kernel approach.

We denote by U the Fourier-Plancherel operator on $L^2(\mathbb{R})$, and the Fourier transform of a function $f(x)$ is denoted by either $Uf(\xi)$ or $\hat{f}(\xi)$ (see Section 7.3 in the appendix). Under the mild assumption that the set $\{\xi \in \mathbb{R} \mid \hat{\gamma}(\xi) = 0\}$ has Lebesgue measure zero, the operator T_γ is injective. As a matter of fact,

$$T_\gamma f = 0 \Leftrightarrow f * \gamma = 0 \Leftrightarrow \hat{f} \cdot \hat{\gamma} = 0 \Leftrightarrow \hat{f} = 0 \Leftrightarrow f = 0,$$

where equality in $L^2(\mathbb{R})$ (that is, almost everywhere) is meant.

Remark 1. The assumption that $\{\xi \in \mathbb{R} \mid \hat{\gamma}(\xi) = 0\}$ has Lebesgue measure zero is much less stringent than imposing that $|\hat{\gamma}(\xi)| > 0$ for all ξ in $\mathbb{R}$, as it is assumed in Stefanski and Carroll [18] or Johannes [13]. For example, a uniform density function γ is now permitted: its Fourier transform is equal to the sinc function, which vanishes at infinitely many points, but still satisfies our condition. This assumption also generalizes the condition required by Carrasco and Florens [4] that $\hat{\gamma}$ may have isolated zeros.

The dark side of things is that the pseudoinverse $T_\gamma^\dagger$ of T_γ is unbounded, as a consequence of Proposition 11 in the appendix (Section 7.1). As a matter of fact, we have:

$$\inf_{\|f\|^2=1} \|T_\gamma f\|^2 = \inf_{\|Uf\|^2=1} \|U\gamma \cdot Uf\|^2 = 0.$$

Here, the first equality follows by the Plancherel theorem. As for the second equality, it is easily obtained from the Riemann-Lebesgue lemma, which says that $U\gamma$ is continuous and vanishes at infinity, by observing that, for every $\alpha \in \mathbb{R}$,

$$U(e^{2i\pi\alpha x} f(x))(\xi) = Uf(\xi - \alpha).$$

Remark 2. Further insight is provided by spectral analysis. It can be shown that the spectrum of T_γ is the closure of the set of values taken by $\hat{\gamma}$, and that its point spectrum $\sigma_p(T_\gamma)$ is empty. For example, if γ is Gaussian, then $\sigma(T_\gamma) = [0, 1]$. At all events, since $\hat{\gamma}$ is continuous and vanishes at infinity by the Riemann-Lebesgue lemma, $\sigma(T_\gamma)$ must contain zero. ■

Remark 3. In the case where it is *a priori* known that the unknown density $f_\circ$ is supported in some compact interval I , we may also consider the convolution operator on $L^2(I)$:

$$\begin{aligned} T_\gamma: L^2(I) &\longrightarrow L^2(\mathbb{R}) \\ f &\longmapsto T_\gamma f := f * \gamma. \end{aligned}$$

This operator is, of course, the restriction of the previous one to the closed subspace $L^2(I)$ of $L^2(\mathbb{R})$, and there is no need to denote this restriction differently. Again, T_γ is injective if and only if the set $\{\xi \in \mathbb{R} \mid \hat{\gamma}(\xi) = 0\}$ has Lebesgue measure zero. Moreover, it is worth observing that, under the extra assumption that γ is square integrable, the above operator is Hilbert-Schmidt. This places us in the familiar framework of operator equations with compact operators. The following then holds:

- (1) $T_\gamma^* T_\gamma$ is diagonalizable in a Hilbert basis $(f_k)_{k \in \mathbb{N}^*}$.
- (2) $\sigma(T_\gamma^* T_\gamma) = \sigma_p(T_\gamma^* T_\gamma)$ is the set of values taken by a sequence $\lambda_1 \geq \lambda_2 \geq \dots > 0$ which converges to zero.
- (3) $\text{ran } T_\gamma$ is not closed and $T_\gamma^\dagger: (\text{ran } T_\gamma)^\perp \rightarrow L^2(I)$ is unbounded.

Here, T^* denotes, as usual, the adjoint of T . ■

We will show that, in the general setting described above, the mollification solution is consistent, and therefore it extends the frameworks considered in Stefanski and Carroll [18], Johannes [13] or Carrasco and Florens [4].

2.2 Overview of filter-type regularization methods

Recall that U is unitary, that is, that $U^{-1} = U^*$ where U^* is the adjoint of U . The operator T_γ can then be written as $T_\gamma = U^* [\hat{\gamma}] U$, in which $[\hat{\gamma}]$ denotes the operator of multiplication by $\hat{\gamma}$. More precisely, consider the function $\varphi \in L^\infty(\mathbb{R})$ and let $[\varphi] : L^2(\mathbb{R}) \rightarrow L^2(\mathbb{R})$ be the *multiplication operator* defined by

$$([\varphi] f)(\xi) = \varphi(\xi) \cdot f(\xi), \quad \xi \in \mathbb{R}.$$

Then, the inverse of $T_\gamma : L^2(\mathbb{R}) \rightarrow \text{ran } T_\gamma$ is given by

$$T_\gamma^{-1} = U^{-1} \left[\frac{1}{\hat{\gamma}} \right] U. \quad (3)$$

As mentioned previously, the unboundedness of $1/\hat{\gamma}$ yields that of T_γ^{-1} (and $T_\gamma^\dagger$). We call *filter-type method* any regularization methods which acts explicitly in the Fourier domain in order to bound the multiplication operation. The corresponding regularized solution f_{REG} is defined by

$$f_{\text{REG}} = U^{-1} [\Phi] U g,$$

or, equivalently, by

$$\hat{f}_{\text{REG}} = \Phi \cdot \hat{g},$$

in which the filter Φ depends on regularization parameters. To this class of methods pertain the well-known deconvolution kernels, the spectral cutoff, but also the Tikhonov regularization. We now briefly review these approaches and connect these well known methods to the filter Φ . We will later on introduce the mollification approach, which also belongs to the class of filter-type methods and which shares features with both the deconvolution kernels and the Tikhonov regularization.

Deconvolution kernels

The deconvolution kernels were introduced by Stefanski and Carroll [18] in the late eighties. In essence, the deconvolution kernel estimator stabilizes the reconstruction by bounding the function $1/\hat{\gamma}$. More precisely, using our notation system, and ignoring (temporarily) the discrete aspects carried by the estimation process, the reconstructed density can be written as

$$f_{\text{DK}} = U^{-1} \left[\frac{\hat{\varphi}_h}{\hat{\gamma}} \right] U g \quad (4)$$

or, equivalently, by

$$\hat{f}_{\text{DK}}(\xi) = \frac{\hat{\varphi}(h\xi)}{\hat{\gamma}(\xi)} \hat{g}(\xi) = \frac{\hat{\varphi}_h(\xi)}{\hat{\gamma}(\xi)} \hat{g}(\xi), \quad (5)$$

where φ_h is defined as

$$\varphi_h(x) := \frac{1}{h} \varphi\left(\frac{x}{h}\right) \quad \text{with } h > 0.$$

Here, φ is the *deconvolution kernel* (denoted by K in the original paper Stefanski and Carroll [18] and in the subsequent literature) and h is the regularization parameter. Therefore, the filter Φ is here defined by: $\Phi(\xi) = \hat{\varphi}_h(\xi)/\hat{\gamma}(\xi)$.

We see right away that, for the solution to be well-defined and stable, it is necessary that

- (a) the function $\hat{\gamma}$ does not vanish, and
- (b) for every $h > 0$, the function $\xi \mapsto \hat{\varphi}(h\xi)/\hat{\gamma}(\xi)$ is bounded.

Notice that, whenever γ is symmetric, $\hat{\gamma}$ is real, hence Condition (a) entails strict positivity of $\hat{\gamma}$. As a matter of fact, the Riemann-Lebesgue lemma tells us that $\hat{\gamma}$ is continuous and vanishes at infinity (and of course $\hat{\gamma}(0) = 1$).

At all events, the above assumptions are somewhat restrictive, obviously, and constitute a serious limitation of the methodology. We will see later on that an interesting property of mollification is to allow for relaxation of these assumptions and allow for any combinations of densities γ and φ .

Remark 4. In the original paper by Stefanski and Carroll [18], it was also requested that, for every $h > 0$,

$$\int \left| \frac{\hat{\varphi}_h(\xi)}{\hat{\gamma}(\xi)} \right| d\xi < \infty.$$

This assumption ensures that the function $(\hat{\varphi}_h(\xi)/\hat{\gamma}(\xi))\hat{g}(\xi)$ in Equation (5) is integrable, allowing for application of the inverse Fourier *integral* transform. But the latter function is square integrable anyways, allowing for the application of the inverse Fourier-Plancherel operator. This is why we do not need the above assumption.

Remark 5. In Stefanski and Carroll [18], the estimation formula reads

$$f_{\text{DK},n} = U^{-1} \frac{\hat{\varphi}_h}{\hat{\gamma}} \hat{g}_n \quad \text{with} \quad \hat{g}_n(\xi) = \frac{1}{n} \sum_{j=1}^n e^{-2\pi i \xi Y_j}.$$

The function $\hat{g}_n$ is the Fourier transform, in the sense of distributions, of the empirical distribution $n^{-1} \sum_j \delta_{Y_j}$. We stress that this goes beyond the functional framework under consideration.

Spectral cutoff

In the spectral cutoff method (see e.g. Johannes [13]), the Fourier transform of the reconstructed density is merely truncated whenever $\hat{\gamma}(\xi)$ falls below a threshold $a > 0$, which plays the role of the regularization parameter. The spectral cutoff may be regarded as a special case of the deconvolution kernels: the solution is defined as

$$f_{\text{SC}} = U^{-1} \left[\frac{\mathbb{1}_{\{|\hat{\gamma}|^2 \geq a\}}}{\hat{\gamma}} \right] Ug$$

or, equivalently, as

$$\hat{f}_{\text{SC}}(\xi) = \frac{\mathbb{1}_{\{|\hat{\gamma}|^2 \geq a\}}(\xi)}{\hat{\gamma}(\xi)} \hat{g}(\xi),$$

which is understood as zero whenever $\hat{\gamma}(\xi) < a$. Here, $\mathbb{1}_S$ denotes the indicator function of the set S , that is, the function which takes the value 1 if the argument belong to S and zero otherwise. Therefore, the spectral cutoff is also a filter-type regularization method, with

$$\Phi(\xi) = \begin{cases} \frac{\mathbb{1}_{\{|\hat{\gamma}|^2 \geq a\}}(\xi)}{\hat{\gamma}(\xi)} & \text{if } \hat{\gamma}(\xi) \geq a, \\ 0 & \text{otherwise.} \end{cases}$$

We can also remark that the inverse Fourier transform of $\mathbb{1}_{\{|\hat{\gamma}|^2 \geq a\}}$ is not a density function and that its behavior is similar to the well-known sinc function. We will see in Section 5 that this specific choice introduces additional perturbations in the reconstruction, known as Gibbs phenomena.

Tikhonov regularization

The Tikhonov regularization (see Tikhonov and Arsenin [19]) has been applied and studied in the context of econometrics in Carrasco et al. [5] and in Carrasco and Florens [4]. The Tikhonov solution is defined by $f_{\text{TK}} = (T_\gamma^* T_\gamma + \alpha I)^{-1} T_\gamma^* g$ or, equivalently, by

$$f_{\text{TK}} = U^{-1} \left[\frac{\bar{\hat{\gamma}}}{|\hat{\gamma}|^2 + \alpha} \right] Ug.$$

Expressed in the Fourier domain, we get:

$$\hat{f}_{\text{TK}}(\xi) = \frac{\overline{\hat{\gamma}(\xi)}}{|\hat{\gamma}(\xi)|^2 + \alpha} \hat{g}(\xi).$$

The solution is well-defined for every positive value of the regularization parameter α . Therefore, the Tikhonov method can be regarded as a filter-type technique (for deconvolution), with

$$\Phi(\xi) = \frac{\overline{\hat{\gamma}(\xi)}}{|\hat{\gamma}(\xi)|^2 + \alpha}.$$

Note that Carrasco and Florens [4] define their solution in weighted L^2 -spaces in order to recover the compactness of the operator T_γ , which we do not need in our context.

Remark 6. The Tikhonov approach is usually introduced in variational form: f_{TK} is solution to the optimization problem

$$\text{Min} \quad \mathcal{F}_{\text{TK}} := \frac{1}{2} \|g - T_\gamma f\|^2 + \frac{\alpha}{2} \|f\|^2.$$

3 Mollification

In this section, we present the mollification approach and show its consistency. Its main interest will appear in Section 4, via comparison to the other methods.

3.1 Definition

Given an integrable function φ , we consider the family $(\varphi_\beta)_{\beta>0}$ defined by

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

Such a family is named an *approximation of unity* (see Section 7.2 in the appendix). We then define the mollification solution to our deconvolution problem by

$$f_{\text{MO}} = U^{-1} \left[\frac{\overline{\hat{\gamma}} \hat{\varphi}_\beta}{|\hat{\gamma}|^2 + |1 - \hat{\varphi}_\beta|^2} \right] U g$$

or, equivalently, by

$$\hat{f}_{\text{MO}}(\xi) = \frac{\overline{\hat{\gamma}(\xi)} \hat{\varphi}_\beta(\xi)}{|\hat{\gamma}(\xi)|^2 + |1 - \hat{\varphi}_\beta(\xi)|^2} \hat{g}(\xi). \quad (6)$$

We observe right away that, as in the Tikhonov solution, the solution is well-defined for every positive value of the regularization parameter β . As a matter of fact, the Riemann-Lebesgue lemma implies that both functions $\hat{\gamma}$ and $\hat{\varphi}_\beta$ are continuous and vanish at infinity, so that the denominator in (6) can't vanish (recall that $\hat{\gamma}(0) = 1 = \hat{\varphi}_\beta(0)$).

Another feature that is shared with the Tikhonov solution is that its form actually derives from a variational formulation. It is readily seen in Alibaud et al. [1] and Bonnefond and Maréchal [2] that f_{MO} is the solution to the optimization problem

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

in which C_β denotes the operator of convolution with φ_β . Clearly, f_{MO} depends continuously on g . As a matter of fact, the Fourier-Plancherel operator is an isometry and the multiplication operator by the bounded filter function

$$\Phi(\xi) := \frac{\overline{\hat{\gamma}(\xi)} \hat{\varphi}_\beta(\xi)}{|\hat{\gamma}(\xi)|^2 + |1 - \hat{\varphi}_\beta(\xi)|^2}$$

has norm equal to the L^∞ -norm of Φ (see the appendix, theorems 19 and 21). In the next paragraph, we shall prove the consistency of the corresponding estimator.

3.2 Consistency

For the sake of clarity, we shall denote the solution f_{MO} by f_β instead, emphasizing the dependence on the regularization parameter β . The following result establishes that the estimator which we have obtained is consistent.

Theorem 7. 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 $f_{n,\beta}$ denote the mollified solution corresponding to data g_n . There then exist a sequence $\beta_n \downarrow 0$ such that

$$E \|f_{n,\beta_n} - f_\circ\| \longrightarrow 0 \quad \text{as } n \rightarrow \infty.$$

The proof of the theorem relies on the following simple lemma:

Lemma 8. Let $c: (0, 1] \rightarrow \mathbb{R}_+$ be any function and let (α_n) be any sequence of positive numbers which converges to zero. Then, there exists a sequence (β_n) such that

- (1) $\beta_n \downarrow 0$ as $n \rightarrow \infty$;
- (2) $c(\beta_n)\alpha_n \rightarrow 0$ as $n \rightarrow \infty$.

PROOF. Let $(\beta^{(k)}) \in (0, 1]^{\mathbb{N}^*}$ be strictly decreasing, converging to zero. Since (α_n) converges to zero, for every $k \in \mathbb{N}^*$, there exists $n_k \in \mathbb{N}^*$ such that $c(\beta^{(k)})\alpha_n \leq \beta^{(k)}$ for all $n \geq n_k$. Clearly, we can choose (n_k) to be strictly increasing. Define $(\beta_n) \in (0, 1]^{\mathbb{N}^*}$ by $\beta_n = 1$ if $n < n_1$ and, for $k \geq 1$, $\beta_n = \beta^{(k)}$ if $n_k \leq n < n_{k+1}$. Then (β_n) has the desired properties. ■

PROOF OF THE THEOREM. By the triangle inequality, we have:

$$E \|f_{n,\beta} - f_\circ\| \leq E \|f_{n,\beta} - f_\beta\| + \|f_\beta - f_\circ\|.$$

Let us first control the size of the deterministic part. By Parseval's theorem,

$$\begin{aligned} \|f_\beta - f_\circ\|^2 &= \left\| (T_\gamma^* T_\gamma + (I - C_\beta)^*(I - C_\beta))^{-1} T_\gamma^* C_\beta T_\gamma f_\circ - f_\circ \right\|^2 \\ &= \left\| \left(\frac{|\hat{\gamma}|^2 \hat{\varphi}_\beta}{|\hat{\gamma}|^2 + |1 - \hat{\varphi}_\beta|^2} - 1 \right) \hat{f}_\circ \right\|^2 \\ &= \int \left| \frac{|\hat{\gamma}|^2 \hat{\varphi}_\beta}{|\hat{\gamma}|^2 + |1 - \hat{\varphi}_\beta|^2} - 1 \right|^2 |\hat{f}_\circ|^2 d\xi \end{aligned}$$

Since the integrand is dominated by the integrable function $\xi \mapsto |\hat{f}_\circ|^2$, and since it converges pointwise to zero as $\beta \downarrow 0$ (recall that $\hat{\varphi}_\beta(\xi) = \hat{\varphi}(\beta\xi)$), Lebesgue's dominated convergence theorem shows that

$$\|f_\beta - f_\circ\|^2 \longrightarrow 0 \quad \text{as } \beta \rightarrow 0.$$

Let us now deal with the stochastic term. Using again Parseval's theorem, we have:

$$\begin{aligned}
E \|f_{n,\beta} - f_\beta\|^2 &= E \left\| (T_\gamma^* T_\gamma + (I - C_\beta)^*(I - C_\beta))^{-1} T_\gamma^* C_\beta (g_n - g) \right\|^2 \\
&= E \left\| \left(\frac{\bar{\gamma} \hat{\varphi}_\beta}{|\hat{\gamma}|^2 + |1 - \hat{\varphi}_\beta|^2} \right) (\hat{g}_n - \hat{g}) \right\|^2 \\
&\leq c(\beta) E \|g_n - g\|^2,
\end{aligned}$$

in which

$$c(\beta) := \sup_{\xi} \left(\frac{\bar{\gamma}(\xi) \hat{\varphi}_\beta(\xi)}{|\hat{\gamma}(\xi)|^2 + |1 - \hat{\varphi}_\beta(\xi)|^2} \right)^2 = \left(\frac{\bar{\gamma}(\xi) \hat{\varphi}(\beta\xi)}{|\hat{\gamma}(\xi)|^2 + |1 - \hat{\varphi}(\beta\xi)|^2} \right)^2.$$

Using the Riemann-Lebesgue lemma, it is easy to see that, for every fixed $\beta > 0$, the function inside the supremum takes the value 1 at $\xi = 0$, vanishes at infinity, and is continuous. Consequently, the supremum is always finite (and greater than or equal to 1). Now, applying Lemma 8 with $\alpha_n = E \|g_n - g\|^2$ yields the desired result. ■

4 A framework for comparison

The purpose of this section is to put the various approaches in the same framework and, by doing so, to derive theoretical and numerical tools for their comparison.

4.1 Unifying framework

The mollification method shares with the Tikhonov regularization that it is primarily variational, and with kernel type methods that filtering is in force. In fact, kernel-type methods may also be regarded as variational methods.

For simplicity of the notation, T_γ is now denoted by T . We already observed that all deconvolution methods under consideration are of filtering type. Let

$$\mathcal{F}(f) := \frac{1}{2} \|Pg - Tf\|^2 + \frac{\alpha}{2} \|Hf\|^2, \quad (7)$$

and let us retrieve the aforementioned methods as the minimization of particular instances of it (by specifying the operators P and H and the parameter α). From the basic theory of least squares, the unique minimizer of $\mathcal{F}$ is

$$\bar{f} = (T^*T + \alpha H^*H)^{-1} T^*Pg. \quad (8)$$

Since $T = U^{-1} [\hat{\gamma}] U$, it is readily seen that $T^* = U^{-1} [\bar{\gamma}] U$ and $T^*T = U^{-1} [|\hat{\gamma}|^2] U$. Therefore,

- (1) letting $H = P = I$ obviously yields the Tikhonov functional;

- (2) the choice $H = I - C_\beta$, $P = C_\beta = U^{-1}[\hat{\varphi}_\beta]U$ and $\alpha = 1$ yields the Mollification functional;
- (3) letting $\alpha = 0$ and $P = C_h$ (where C_h denotes the operator of convolution with φ_h) yields the functional $\mathcal{F}(f) = \|C_h g - T f\|^2 / 2$, whose unique minimizer is

$$T^\dagger C_h g = (T^* T)^{-1} T^* C_h g = U^{-1} \left[\frac{\hat{\varphi}_h}{\hat{\gamma}} \right] U g,$$

which turns out to be the deconvolution kernel solution f_{DK} .

- (4) letting $\alpha = 0$ and P be the convolution by $U^{-1} \mathbb{1}_{|\hat{\gamma}|^2 \geq a}$ yields the spectral cutoff solution f_{SC} .

Notice first that the convolution kernels corresponding to the above three filters need not be real (if γ is not symmetric) or even positive, if they are real. Notice also that, in the Tikhonov case, $\Phi(0) = (1 + \alpha)^{-1} \neq 1$, which implies the additional drawback that the integral of the reconstructed function will not be equal to one, as one should expect from a density. We may then consider the modified Tikhonov filter

$$\Phi(\xi) = \frac{(1 + \alpha) \bar{\hat{\gamma}}(\xi)}{|\hat{\gamma}(\xi)|^2 + \alpha}.$$

In this modified version, we merely let $P = [1 + \alpha]$ instead of $P = I = [1]$. As for the mollification approach, it would also make sense to consider a version with $P = I$, letting the regularization be operated by $H = I - C_\beta$ only. The corresponding filter is easily shown to be

$$\Phi(\xi) = \frac{\bar{\hat{\gamma}}(\xi)}{|\hat{\gamma}(\xi)|^2 + |1 - \hat{\varphi}_\beta(\xi)|^2},$$

and we shall refer to the corresponding method as the modified mollification. Finally, notice that the particular kernel corresponding to the spectral cutoff method is the inverse Fourier transform of the function $\mathbb{1}_{\{|\hat{\gamma}|^2 \geq a\}} / \hat{\gamma}$, and that the regularization parameter is now $a > 0$. Table 1 gives an overview of the functionals and filters associated with each regularization method.

4.2 Comparisons

Deconvolution kernels versus mollification

We emphasize that, obviously, the regularization parameters h and β have the same interpretation.

The first obvious limitation of the deconvolution kernels is the restriction imposed by the decrease of $\hat{\gamma}$ and its strict positivity. For example, if γ is Gaussian, the decrease of $\hat{\varphi}$ at infinity should be faster, which discards many deconvolution kernels. An even

	Functional $\mathcal{F}$	Filter Φ
TK	$\frac{1}{2} \ g - \gamma * f\ ^2 + \frac{\alpha}{2} \ f\ ^2$	$\frac{\bar{\gamma}}{ \hat{\gamma} ^2 + \alpha}$
MT	$\frac{1}{2} \ (1 + \alpha)g - \gamma * f\ ^2 + \frac{\alpha}{2} \ f\ ^2$	$\frac{(1 + \alpha)\bar{\gamma}}{ \hat{\gamma} ^2 + \alpha}$
MO	$\frac{1}{2} \ \varphi_\beta * g - \gamma * f\ ^2 + \frac{1}{2} \ f - \varphi_\beta * f\ ^2$	$\frac{\bar{\gamma}\hat{\varphi}_\beta}{ \hat{\gamma} ^2 + 1 - \hat{\varphi}_\beta ^2}$
MM	$\frac{1}{2} \ g - \gamma * f\ ^2 + \frac{1}{2} \ f - \varphi_\beta * f\ ^2$	$\frac{\bar{\gamma}}{ \hat{\gamma} ^2 + 1 - \hat{\varphi}_\beta ^2}$
DK	$\frac{1}{2} \ \varphi_h * g - \gamma * f\ ^2$	$\frac{\hat{\varphi}_h}{\hat{\gamma}}$
SC	$\frac{1}{2} \left\ \mathbf{1}_{\{ \hat{\gamma} ^2 \geq a\}} * g - \gamma * f \right\ ^2$	$\frac{\mathbf{1}_{\{ \hat{\gamma} ^2 \geq a\}}}{\hat{\gamma}}$

Table 1: Overview of regularization methods for the deconvolution problem: TK stands for Tikhonov, MT for modified Tikhonov, MO for mollification, MM for modified mollification, DK for the deconvolution kernels and SC for the spectral cutoff.

more extreme example is provided by the convolution kernel $\gamma(x) = \text{sinc}^2(\pi x)$. Its Fourier transform is the triangle function

$$\hat{\gamma}(\xi) = \begin{cases} \xi + 1 & \text{if } x \in [-1, 0), \\ -\xi + 1 & \text{if } x \in [0, 1), \\ 0 & \text{elsewhere.} \end{cases}$$

Here, the convolution operator T_γ fails to be injective, as a consequence of the fact that the support of $\hat{\gamma}$ is the interval $[-1, 1]$, and the deconvolution kernel solution cannot be defined. By contrast, the mollification solution is defined for any mollifier $\varphi \in L^1(\mathbb{R})$ and any positive value of β .

Moreover, in the variational formulation of the deconvolution kernels, the regularization appears only in the fit term, so that the optimization problem remains ill-posed. This may be an obstacle to the introduction of additional constraints, such as the positivity of the reconstruction, since such a constraint can be introduced only in the variational form. On the contrary, the variational form of the mollification approach is stable, thanks to the regularization term, and such constraint may therefore be safely introduced. The only price to be paid would then be a different numerical strategy

(based on optimization).

We emphasize here that deconvolution kernels and mollification both aim at reconstructing an explicit object, namely $\varphi_\beta * f_\circ$. The latter may be referred to as the *target object*. In addition, as it will be illustrated in Section 5, when the deconvolution kernels regularization is well-defined, the performance of the mollification approach is similar.

Spectral cutoff versus mollification

Unlike the case of deconvolution kernels, the spectral cutoff solution remains defined when $\hat{\gamma}$ vanishes. The target object, in the sense defined above, is $\psi_a * f_\circ$, with

$$\psi_a = U^{-1} \mathbb{1}_{\{|\hat{\gamma}|^2 \geq a\}}.$$

The function ψ_a can be regarded as a *target* impulse response of the reconstruction. Its definition relies not only on the regularization parameter a , but also on the shape of $\hat{\gamma}$. We stress here that, as the inverse Fourier transform of some indicator function, this impulse response may have poor morphological properties, and may result in oscillations (like in the Gibbs phenomenon) of the reconstructed density. These oscillations may incidentally produce significant negative parts, which is a serious drawback for probability densities. By contrast, the mollification approach enables to choose an *apodized* target impulse response, by avoiding sharp edges in the Fourier domain. This will be illustrated in Section 5.

Notice at last that, as in the case of the deconvolution kernels, the variational form of the spectral cutoff has no regularization term. Again, this may be an obstacle to the introduction of additional constraints, such as the positivity of the reconstruction (see the discussion in the previous paragraph).

Tikhonov versus mollification

Unlike deconvolution kernels, spectral cutoff and mollification, Tikhonov regularization does not appeal to any target object, which is a conceptual drawback. The regularization is *uniformly* exercised in the Fourier domain, as can be seen from the variational formulation. As a matter of fact, using the well-known fact that the Fourier-Plancherel operator is an isometry, the Tikhonov solution is readily seen to be the minimizer of

$$\mathcal{F}_{TK}(f) = \frac{1}{2} \left\| \hat{g} - \hat{\gamma} \cdot \hat{f} \right\|^2 + \frac{\alpha}{2} \left\| \hat{f} \right\|^2.$$

The penalty term attracts $\hat{f}$ towards zero everywhere. This contradicts the action of the fit term in the low frequency domain, where both $\hat{g}$ and $\hat{\gamma}$ are not expected to be close to zero. This opposition between the fit and regularization terms may induce an unfavorable tradeoff between stability and fidelity to the model. The mollification

approach avoids this pitfall by introducing a *smooth disjunction* of the realms of action of the fit and regularization terms, as can be seen from the transposition of the mollification functional in the Fourier domain:

$$\mathcal{F}_{MO}(f) = \frac{1}{2} \left\| \hat{\varphi}_\beta \cdot \hat{g} - \hat{\gamma} \cdot \hat{f} \right\|^2 + \frac{1}{2} \left\| (1 - \hat{\varphi}_\beta) \hat{f} \right\|^2.$$

The resulting improvement in the tradeoff between stability and fidelity to the initial model equation will be illustrated by means of simulations in Section 5.

A nice aspect of Tikhonov, that is shared with mollification, is that, unlike the deconvolution kernels and the spectral cutoff, the variational formulation is stabilized, which opens the way to the introduction of the positivity constraint.

5 Numerical aspects and simulations

Having derived, in the previous sections, a common framework for all reconstruction methods under consideration, we now proceed to develop tools for their assessment and comparison, in terms of the tradeoff between stability and fidelity. In all cases, $f_{\text{REG}} = U^{-1}[\Phi] U g_o$, in which the filter Φ depends on regularization parameters. If g_o is replaced by its estimated value g_n , the corresponding reconstruction is denoted by $f_{\text{REG},n}$. Otherwise expressed, $f_{\text{REG},n} := U^{-1}[\Phi] U g_n$.

5.1 Assessment of the various regularization methods

We now define the quantities to be use for the assessment of the various regularization methods.

Concerning the fidelity, a significant quantity is the *reconstruction error*

$$\begin{aligned} f_{\text{REG},n} - f_o &= (f_{\text{REG},n} - f_{\text{REG}}) + (f_{\text{REG}} - f_o) \\ &= U^{-1}[\Phi] U (g_n - g_o) + (f_{\text{REG}} - f_o). \end{aligned} \quad (9)$$

In the right hand side, the first and second terms will be respectively called the *statistical error* and *regularization error*. The L^2 norm of the reconstruction error, referred to as *reconstruction-RISE* (Root Integrated Square Error), will be one important indicator to compare the performances of the main four regularization methods: Deconvolution Kernels, Spectral cutoff, Tikhonov and Mollification.

Obviously the reconstruction-RISE depends on the value of the regularization parameter of the considered regularization method (either a , α , h , or β) and their range of variations as well as their impact on the solution (through the chosen regularization method) may not be easily comparable. That is why we have also introduced a common object to evaluate the stability of the reconstruction, that we define below.

In our setting, the reconstructed density depends linearly on the data g , and that the error on the data $g_n - g_o$ is potentially amplified by the action of the reconstruction

operator $U^{-1}[\Phi]U$ by a factor equal to its operator norm (see Equation (9) as an illustration). Since U is unitary, the operator norm of the reconstruction operator $U^{-1}[\Phi]U$ is equal to the L^∞ -norm of Φ (see Section 7.4 in the appendix). The stability of the reconstruction can then be estimated via the computation of $\|\Phi\|_\infty$ as a function of the regularization parameters. We may refer to $\|\Phi\|_\infty$ as an *instability index*.

Therefore, in the next graphs, we compare the performance of the various regularization methods by comparing the variation of the reconstruction-RISE with respect to the instability index, whose level can be arbitrarily fixed for all regularization methods. To any fixed level of the instability index corresponds a fixed value of the regularization parameter.

Notice that the reconstruction error depend on the true value $f_\circ$, unobserved in practice. The *residual*

$$g_n - T_\gamma f_{\text{REG},n}$$

may then serve the purpose of evaluating the fidelity to the original model. Some additional plots to evaluate the performance of the regularization methods are provided using the *residual RISE* with respect to the instability index.

5.2 Two examples with a simulated sample

We first illustrate the comparison between the different approaches in two examples denoted Case I and Case II. In the two examples, the signal f is a Beta(3, 2) density function, rescaled for having the support $[-2, 2]$, note that its standard deviation is $\sigma_f = 0.8$. In both cases, $n = 500$. In Case I, the noise γ is a Cauchy(0, σ) where σ is the scale, i.e. $\text{Cauchy}(0, \sigma) = \sigma t_{(1)}$ where $t_{(1)}$ is a student- t with one degree of freedom. In Case II, the noise γ is a $N(0, \sigma^2)$ where $\sigma = 0.5 * \sigma_f = 0.40$. For Case I, the standard deviation does not exist, we choose the scale $\sigma = 0.20$ such that this Cauchy has an $IQR = 0.40$.

5.2.1 Case I

For Case I, we choose for the Mollification (MOL) and the Deconvolution Kernel (DK) techniques a normal density with regularization parameter β and h respectively. So, in this case, the four approaches described above can be used, we expect a quiet similar behavior between the MOL et the DK approaches. Figure 1 shows how the instability index vary with the regularization parameter in the 4 methods. In all the cases, as it should, the methods are more stable when increasing the regularization parameter. Figure 2 shows in our particular sample of size $n = 500$ the components of the error in the reconstruction of the signal f as defined in equation (9). As expected in the 4 cases, the regularization error increases with the regularization parameter but the opposite is true for the statistical error. By combining the two we have the dashed-dot

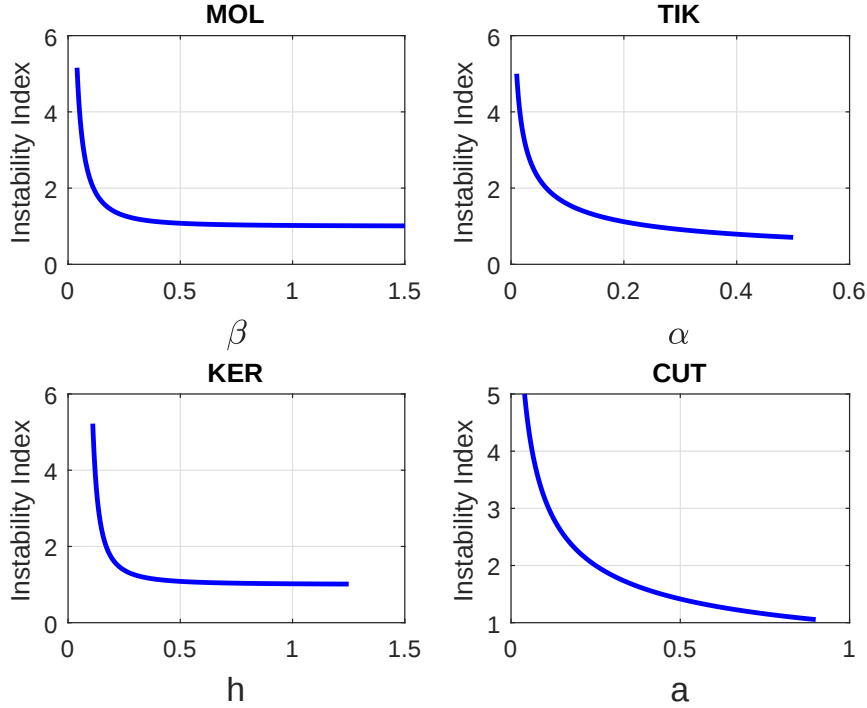


Figure 1: Case I: The instability index as a function of the regularization parameters.

line of the Root-Mean Squared Errors (RISE) of the total error which has a minimum for an “optimal” value of the parameters. Direct and precise calculations gives the values shown in Table 2. We observe that for this particular sample of size $n = 500$ the best performance is achieved by the Mollification approach. We will see in the Monte-Carlo experiments below if this is confirmed when many samples are considered and also for sample sizes $n = 100$ and 1000 .

Method	Parameter	Rec-RISE
MOL	$\beta = 0.30346$	0.042010
TIK	$\alpha = 0.12912$	0.109681
KER	$h = 0.37571$	0.043268
CUT	$a = 0.36546$	0.077186

Table 2: Case I: Reconstruction-RISE for the 4 methods and values of optimal regularization parameter.

One way to compare the performances of the four approaches is to plot the various components of the RISE and the residual RISE versus the reached instability index. The results are shown in Figure 3. Note that the top-left panel allows to find for each

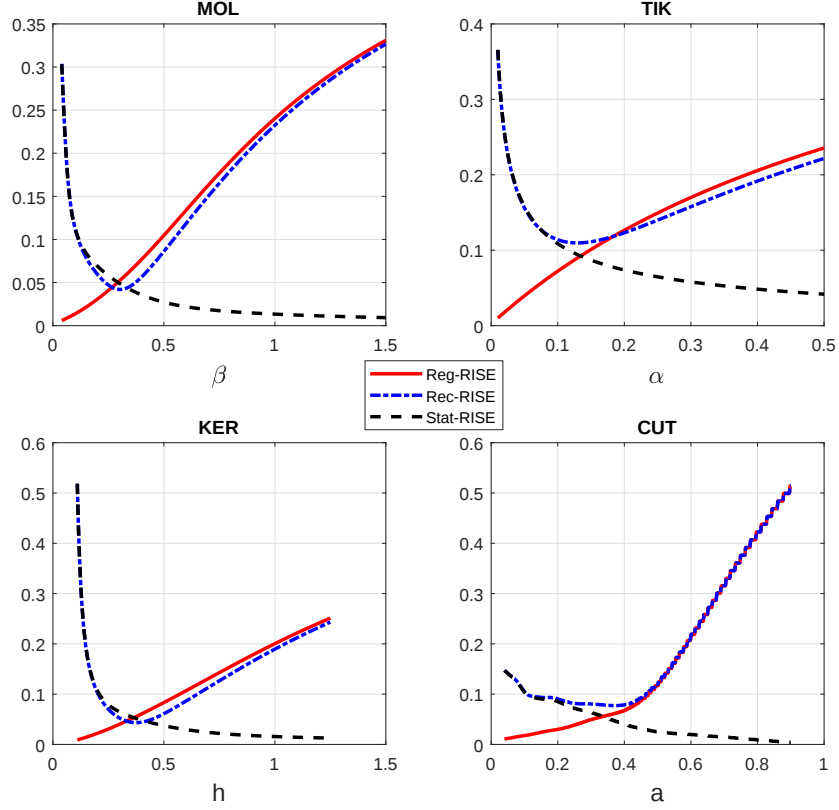


Figure 2: Case I: The Root-Mean Squared Errors (RISE) and its components in the 4 approaches. “Rec” is related to the reconstruction error (the total error), “Reg” is for its regularization part and “Stat” is for the statistical part of the error (see equation (9)).

method the optimal instability index minimizing the reconstruction RISE and is only available in a simulated data framework (the true f is known) whereas, the bottom-right panel with the residuals is available in a real data framework and could be used to find the optimal method for a given level of the instability index.

Finally, for this particular sample of size $n = 500$ it is interesting to show how the signal f can be reconstructed in the various approaches (here the optimal values of the regularization parameters have been used). This is shown in Figure 4 where the various estimates are displayed along with the true value of the density f . We see how irregular is the estimate in the Tikhonov approach and the Gibbs effect of the cutoff approach for values of $|x|$ greater than 2. Again, we see that the Mollification and the Kernel Deconvolution approaches give very similar and good results.

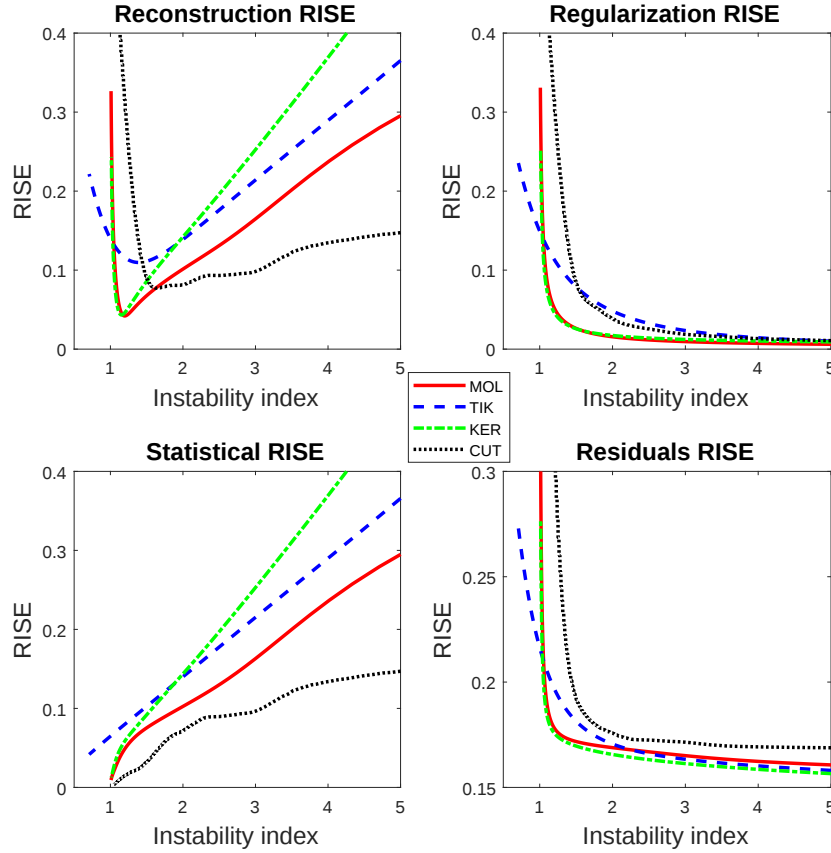


Figure 3: Case I: Various components of the RISE and the corresponding values of the residual RISE as a function of the instability index.

5.2.2 Case II

For Case II, where we have a Gaussian noise, we illustrate the flexibility of the Mollification approach by choosing for the mollifying density $\varphi_\beta(x)$ a uniform on the interval $[-\beta/2, +\beta/2]$. Note that here this density cannot be used as regularization density for the Deconvolution Kernel method (the filter is unbounded). So we only have here the 3 methods MOL, TIK and CUT to be compared. Figure 5 displays how the instability index vary as a function of the regularization parameter in this setup. Figure 6 shows in our particular sample of size $n = 500$ the components of the error in the reconstruction of the signal f . We have a similar behavior for the 3 approaches as we observed in Case I above: the regularization error increases with the regularization parameter but the opposite is true for the statistical error. By combining the two we have the U -shape curve of reconstruction RISE. The optimal value of the regularization parameters are provided in Table 3, indicating again for this particular sample a better behavior of the Mollification, in term of the total RISE.

As for Case I above the 3 approaches can be compared on the same picture by

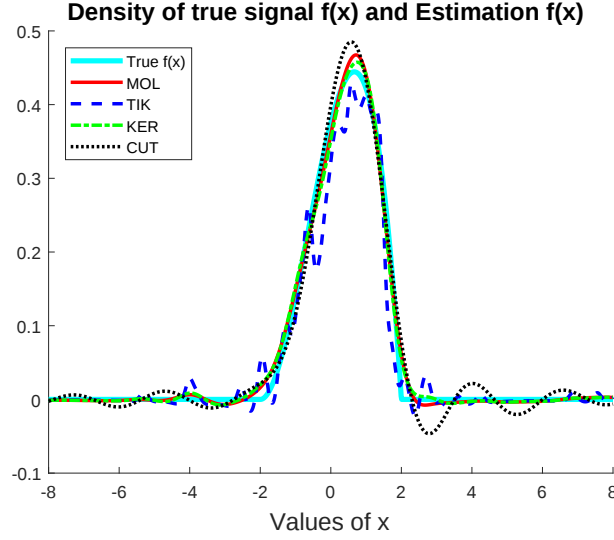


Figure 4: Case I: Different estimates of the signal f and its true value.

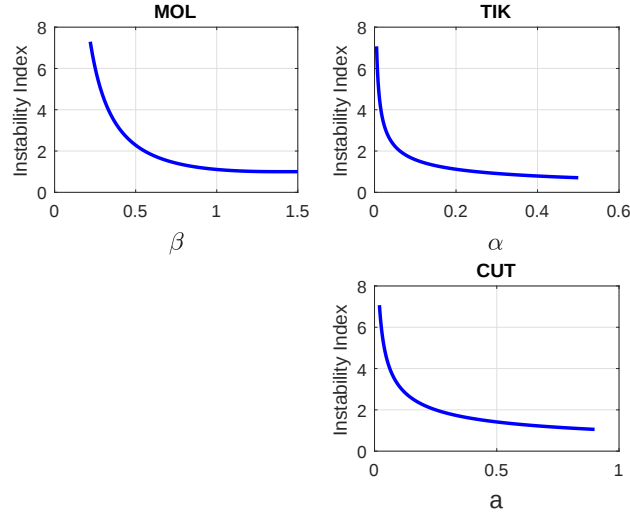


Figure 5: Case II: The instability index as a function of the regularization parameters.

representing the components of the RISE and the residual Rise as a function of the instability index. The results are displayed in Figure 7. In this case we see for the total-RISE and the residual RISE a better behavior of the Mollification approach. The final estimates of the signal f obtained by the reconstruction, in this particular sample of size $n = 500$ are displayed in Figure 8. We used the optimal values of the regularization parameters given in Table 3. Again we see the nice behavior of the Mollification estimate, the irregularities of the Tikhonov one and the Gibbs effect for

Method	Parameter	Rec-RISE
MOL	$\beta = 0.75895$	0.042248
TIK	$\alpha = 0.05462$	0.066442
CUT	$a = 0.17218$	0.053963

Table 3: Case II: Reconstruction-RISE for the 3 methods and values of optimal regularization parameter.

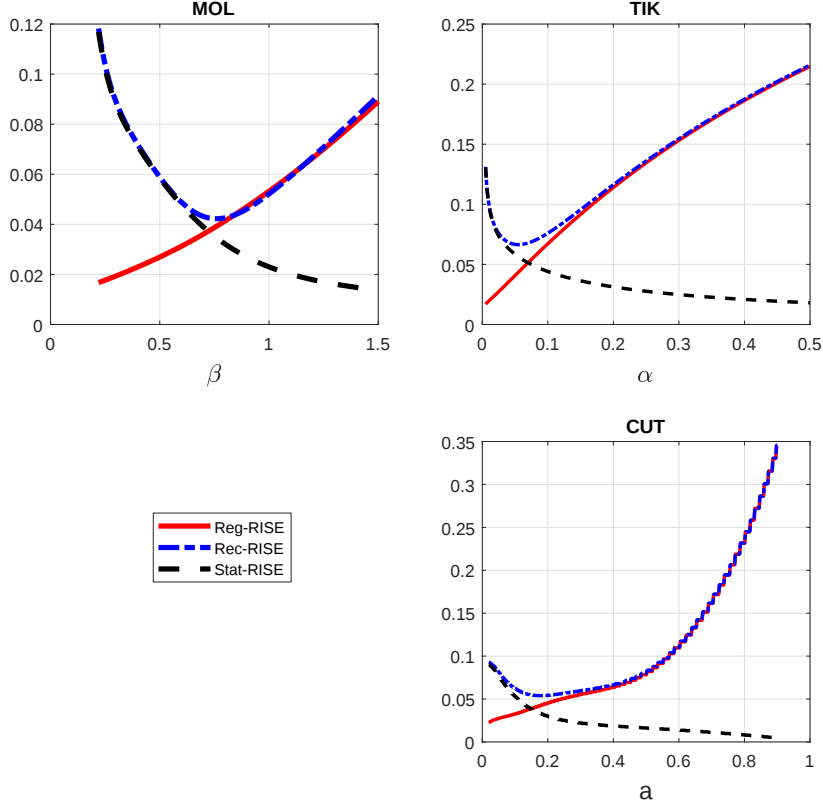


Figure 6: Case II: The Root-Mean Squared Errors (RISE) and its components in the 3 approaches. “Rec” is related to the reconstruction error (the total error), “Reg” is for its regularization part and “Stat” is for the statistical part of the error.

the cutoff approach.

5.3 Monte-Carlo Experiments

Now, for the two cases illustrated in the preceding section, we perform a Monte-Carlo experiment by simulating a large number of times, say M , a sample of size n and

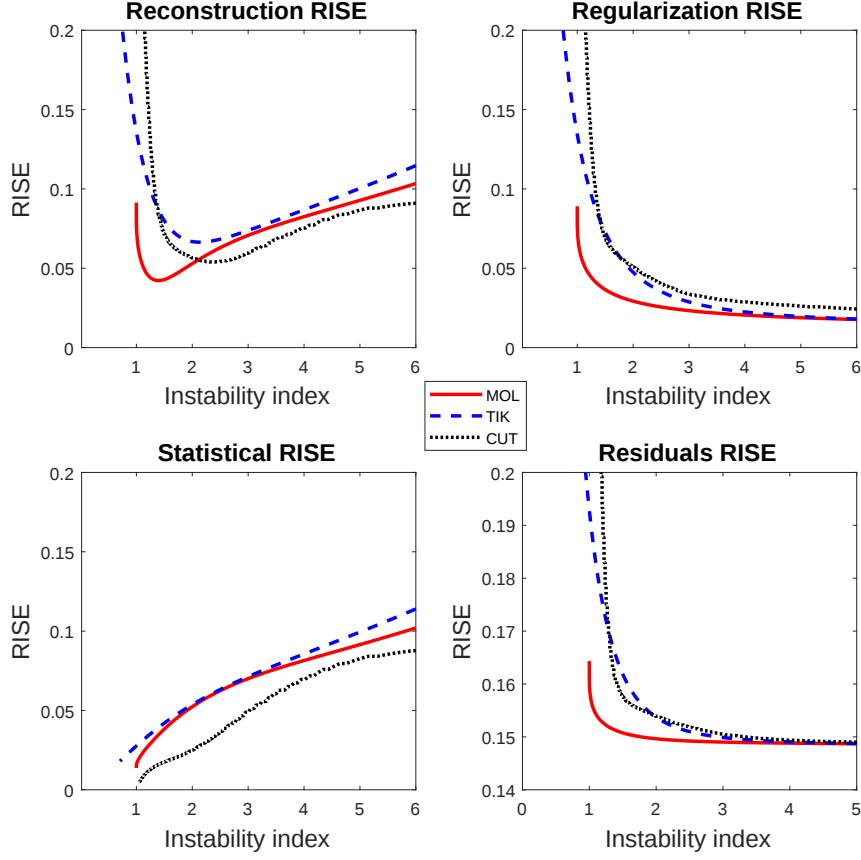


Figure 7: Case II: Various components of the RISE and the corresponding values of the residual RISE as a function of the instability index.

compare the performance of the different approaches by recording the achieved minimal reconstruction RISE for each approach, and then averaging over the M simulated samples. The results are displayed in Table 5. For each case the Table provide the Monte-Carlo estimator of the Average optimal RISE for the reconstruction error of the signal f , defined in (9), computed over $M = 1000$ simulated samples, i.e.

$$ARISE = M^{-1} \sum_{m=1}^M \|f_{\text{REG},n,m} - f_{\circ}\|,$$

where $f_{\text{REG},n,m}$ is the reconstruction obtained with the sample m of size n computed with the optimal regularization parameter obtained by minimizing the reconstruction RISE. To appreciate if the differences are significant we also provide the Monte-Carlo

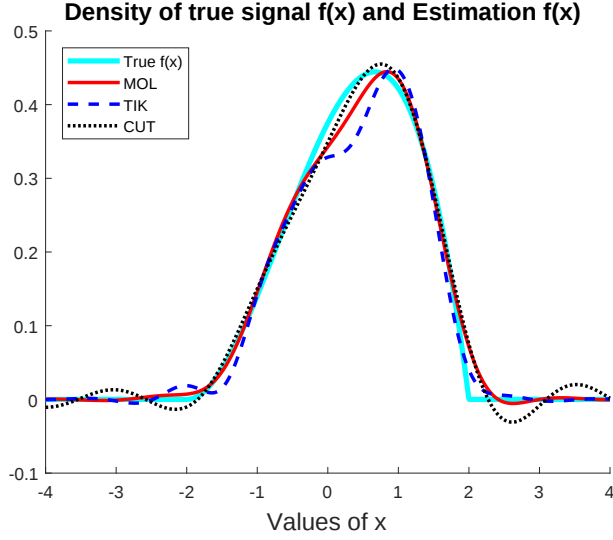


Figure 8: Case I: Different estimates of the signal f and its true value.

standard deviation of this estimator, i.e.

$$Std_{ARISE} = \sqrt{[M(M-1)]^{-1} \sum_{m=1}^M [\|f_{\text{REG},n,m} - f_{\circ}\| - ARISE]^2},$$

The Table also gives for each case the average of the M optimal values of the regularization parameters.

Looking to the table we see that in both Cases I and II, for each method, as expected, the $ARISE$ decreases when the sample size increases. In all the cases in the Table, the MOL approach provides better results than the other regularization techniques, and the difference is significant (compared with the respective standard errors). The cutoff method seems to be better than the Deconvolution kernel (when available) and than the Tikhonov which appears to be the less reliable method. Of course these general comments apply only for the two chosen scenarios, but added with the theoretical comparisons made above, this Monte-Carlo exercise seems to advocate for using Mollification techniques in deconvolution problems.

6 Conclusion and perspectives

We have introduced the mollification approach to the deconvolution of probability densities. We have established the consistency of the corresponding estimator. By placing mollification in the framework of *filter-type* methods, we have compared it, both theoretically and numerically, with various other methods, namely the deconvolution kernels, the spectral cutoff and the Tikhonov regularization. This comparison

CASE I:

Method:	MOL	TIK	KER	CUT
$n = 100$				
$ARISE$	0.100038	0.209369	0.106937	0.105383
Std_{ARISE}	0.000941	0.000898	0.000952	0.000813
$\overline{reg.par.}$	0.358822	0.238026	0.460649	0.373794
$n = 500$				
$ARISE$	0.064398	0.134029	0.070904	0.068135
Std_{ARISE}	0.000470	0.000543	0.000481	0.000450
$\overline{reg.par.}$	0.257730	0.124067	0.344160	0.272226
$n = 1000$				
$ARISE$	0.052165	0.108078	0.058255	0.055289
Std_{ARISE}	0.000355	0.000426	0.000366	0.000359
$\overline{reg.par.}$	0.221208	0.094325	0.303763	0.225243

Table 4: Case I: Monte-Carlo (MC) performances of the 4 methods over $M = 1000$ replications of samples of size $n = 100, 500$ and 1000 , respectively. $ARISE$ is the Monte-Carlo average of the reached optimal reconstruction RISE, and Std_{ARISE} is its Monte-Carlo standard deviation. $\overline{reg.par.}$ is the mean of the respective optimal values of the regularization parameters.

reveals notably that the mollification enables to substantially extend the domain of applicability of the deconvolution kernels, while providing better performances than all methods under consideration in terms of the tradeoff between fidelity and stability of the reconstruction. Mollification inherits from advantages of both the deconvolution kernels (in particular a *target object* is clearly defined) and the Tikhonov regularization (in particular the flexibility brought by the variational formulation).

We have focused the present work on important practical issues regarding the reconstruction. Although of central interest in the theory of inverse problems, we have not addressed here the question of convergence rates, whose study is deferred to a forthcoming publication. From this practical viewpoint, the assessment of the relative performances of mollification and other methods, it appears that mollification brings significant improvements to all other methods. Although our numerical simulation are necessarily attached to particular cases, their interpretation corroborates some theoretical evidence.

CASE II:

Method:	MOL	TIK	CUT
$n = 100$			
$ARISE$	0.093505	0.152658	0.098631
Std_{ARISE}	0.000920	0.001044	0.000793
$\overline{reg.par.}$	1.179047	0.161683	0.389871
$n = 500$			
$ARISE$	0.061859	0.092822	0.065483
Std_{ARISE}	0.000467	0.000638	0.000418
$\overline{reg.par.}$	0.856686	0.081881	0.222219
$n = 1000$			
$ARISE$	0.051734	0.074589	0.055102
Std_{ARISE}	0.000376	0.000492	0.000350
$\overline{reg.par.}$	0.739565	0.062007	0.154203

Table 5: Case II: Monte-Carlo (MC) performances of the different methods over $M = 1000$ replications of samples of size $n = 100, 500$ and 1000 , respectively. $ARISE$ is the Monte-Carlo average of the reached optimal reconstruction RISE, and Std_{ARISE} is its Monte-Carlo standard deviation. $\overline{reg.par.}$ is the mean of the respective optimal values of the regularization parameters.

7 Appendix: reminders

7.1 Least squares and ill-posedness

Let F and G be Hilbert spaces, and let $T: F \rightarrow G$ be a linear mapping. Recall that $F = \ker T \oplus \overline{\text{ran } T^*}$ and $G = \ker T^* \oplus \overline{\text{ran } T}$. Otherwise expressed, the mapping T induces an orthogonal decomposition of F and G .

Theorem 9. Let F and G be Hilbert spaces and let $T: F \rightarrow G$ be a linear mapping. Let P denote the orthogonal projection onto $\overline{\text{ran } A}$. Let $g \in G$ and let $\tilde{g} := Pg$. Then the following are equivalent: équivalentes:

- (1) $\tilde{g} = Tf_{\circ}$;
- (2) $x_{\circ}$ minimizes the function $f \mapsto \|g - Tf\|$;
- (3) $T^*g = T^*Tf_{\circ}$.

The equation $T^*g = T^*Tx_{\circ}$ is the so-called *normal equation*. Notice that if $g \in \text{ran } T + (\text{ran } T)^{\perp}$, then $Pg \in \text{ran } T$. In this case, $T^{-1}(Pg)$ is nonempty and, from the previous theorem,

$$T^{-1}(Pg) = \{f \in F \mid T^*g = T^*Tg\}.$$

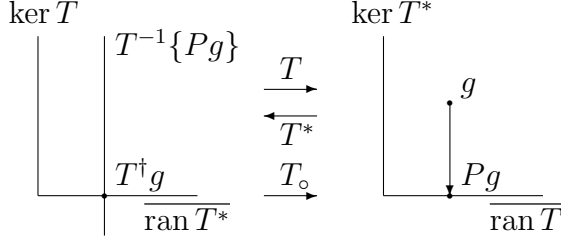


Figure 9: Orthogonal decomposition of F and G induced by T , and construction of the pseudo-inverse.

Therefore $T^{-1}(Pg)$ is an affine subspace parallel to $\ker T^*T = \ker T$. Now, let $T_\circ: (\ker T)^\perp \rightarrow \text{ran } T$ be the linear mapping obtained by restricting T to $(\ker T)^\perp$. The mapping $T_\circ$ is bijective by construction. We then define the *pseudo-inverse* of T as the linear mapping

$$\begin{aligned} T^\dagger: \mathcal{D}(T^\dagger) &\longrightarrow F \\ g &\longmapsto T^\dagger g := T_\circ^{-1}Pg, \end{aligned}$$

where $\mathcal{D}(T^\dagger) := \text{ran } T + (\text{ran } T)^\perp = \text{ran } T + \ker T^*$. It is then easy to prove that:

Proposition 10. For every $g \in \mathcal{D}(T^\dagger)$, $T^\dagger g$ belongs to $T^{-1}(Pg)$, and $T^\dagger g$ minimizes the function $f \mapsto \|f\|$ over $T^{-1}(Pg)$.

We refer to $T^\dagger g$ as the *minimum norm least square solution* of the linear equation $Tf = g$. The orthogonal decomposition of F and G induced by T and the definition of the pseudo-inverse $T^\dagger$ are illustrated in Figure 7.1.

In this setting, the linear equation $Tf = g$ is said to be *ill-posed* whenever the pseudo-inverse $T^\dagger$ is unbounded, yielding unstable minimum norm least square solution. The following holds:

Proposition 11. If $\inf \{ \|Tf\| \mid f \in (\ker T)^\perp, \|f\| = 1 \} = 0$, then $T^\dagger$ is unbounded.

PROOF. Pick (f_n) in $(\ker T)^\perp$ with $\|f_n\| = 1$ and $\|Tf_n\| \rightarrow 0$ as $n \rightarrow \infty$. Define $g_n = Tf_n / \|Tf_n\|$. Since $T^\dagger T$ is the orthogonal projection onto $(\ker T)^\perp$,

$$\|T^\dagger g_n\| = \frac{\|T^\dagger Tf_n\|}{\|Tf_n\|} = \frac{1}{\|Tf_n\|} \rightarrow \infty \quad \text{as } n \rightarrow \infty. \blacksquare$$

It is worth noticing that the range of T cannot be closed if the problem is ill-posed and that, consequently, the domain $\mathcal{D}(T^\dagger)$ is then a proper subspace of G . As a matter of fact, from the Open Mapping Theorem applied to $T_\circ$, the closedness of $\text{ran } T$ would imply the boundedness of $T_\circ$, which would imply in turn the boundedness of $T^\dagger$.

7.2 Convolution

If f and g are two functions from $\mathbb{R}$ into $\mathbb{R}$, we write

$$(f * g)(x) := \int f(x - y)g(y) \, dy$$

whenever the integral is well defined. This requires that the function

$$y \mapsto |f(x - y)g(y)|$$

be integrable. A simple change of variable shows that, whenever this is the case, the function $y \mapsto |g(x - y)f(y)|$ is also integrable, and that $(g * f)(x) = (f * g)(x)$.

We denote by $\mathcal{C}_o(\mathbb{R})$ the vector space of all real-valued continuous functions with compact support. If $f, g \in \mathcal{C}_o(\mathbb{R})$, then $f * g$ is everywhere defined and actually $f * g \in \mathcal{C}_o(\mathbb{R})$. More generally, if $p, q \in [1, \infty]$ are conjugate exponents (i.e. $p^{-1} + q^{-1} = 1$) and if $f \in L^p(\mathbb{R})$ and $g \in L^q(\mathbb{R})$, then $f * g$ is everywhere defined, bounded by $\|f\|_p \|g\|_q$, and uniformly continuous. In the case where $p \in (1, \infty)$, $f * g$ goes to zero at infinity. If now $f \in L^1(\mathbb{R})$ and $g \in L^p(\mathbb{R})$, where $p \in [1, \infty]$, then the function $\varphi_x: y \mapsto f(x - y)g(y)$ is integrable for almost every $x \in \mathbb{R}$ and the function $f * g$ (which is almost everywhere defined) belongs to $L^p(\mathbb{R})$ and satisfies $\|f * g\|_p \leq \|f\|_1 \|g\|_p$. In this setting, the following approximation theorem holds:

Theorem 12. Let $\varphi \in L^1(\mathbb{R})$ be such that $\int \varphi(x) \, dx = 1$. For every $\varepsilon > 0$, let

$$\varphi_\varepsilon(x) := \frac{1}{\varepsilon} \varphi\left(\frac{x}{\varepsilon}\right).$$

At last, let $p \in [1, \infty)$. Then, for every $f \in L^p(\mathbb{R})$,

$$\|f * \varphi_\varepsilon - f\|_p \longrightarrow 0 \quad \text{as } \varepsilon \longrightarrow 0.$$

The family of functions $(\varphi_\varepsilon)_{\varepsilon > 0}$ is referred to as an *approximation of unity*. Recall that the Schwartz space $\mathcal{S}(\mathbb{R})$ is the vector space of all functions φ in $\mathcal{C}^\infty(\mathbb{R})$ such that, for every $\alpha, \beta \in \mathbb{N}$,

$$\sup \left\{ \left| x^\alpha \varphi^{(\beta)}(x) \right| \mid x \in \mathbb{R} \right\} < \infty,$$

in which $\varphi^{(\beta)}$ denotes the β -th derivative of φ . Clearly, $\mathcal{C}_o^\infty(\mathbb{R}) \subset \mathcal{S}(\mathbb{R}) \subset L^p(\mathbb{R})$ for every $p \in [1, \infty]$. The Gaussian function $x \mapsto e^{-x^2}$ is well-known example of a function in $\mathcal{S}(\mathbb{R})$ which does not belong to $\mathcal{C}_o^\infty(\mathbb{R})$. The space $\mathcal{C}_o^\infty(\mathbb{R})$ is dense in $L^p(\mathbb{R})$ for every $p \in [1, \infty)$, which obviously implies that $\mathcal{S}(\mathbb{R})$ is dense in $L^p(\mathbb{R})$ for every $p \in [1, \infty)$.

7.3 Fourier transforms

Let $f \in L^1(\mathbb{R})$. The function $x \mapsto e^{-2i\pi x\xi} f(x)$ is obviously integrable for every $\xi \in \mathbb{R}$ and we denote by $\hat{f}$ and $\check{f}$ the functions respectively given by

$$\hat{f}(\xi) := \int e^{-2i\pi x\xi} f(x) dx \quad \text{and} \quad \check{f}(\xi) := \int e^{2i\pi x\xi} f(x) dx.$$

The linear mappings $f \mapsto \hat{f}$ and $f \mapsto \check{f}$ are denoted by U and $\bar{U}$, respectively.

Theorem 13. Let $f, g \in L^1(\mathbb{R})$. Then $f * g \in L^1(\mathbb{R})$ from the above and

$$U(f * g) = Uf \cdot Ug.$$

The Schwartz space $\mathcal{S}(\mathbb{R})$ turns out to be stable under Fourier transformation.

Theorem 14. If $\varphi \in \mathcal{S}(\mathbb{R})$, then $U\varphi \in \mathcal{S}(\mathbb{R})$. Moreover, $U\varphi^{(n)}(\xi) = (2i\pi\xi)^n \cdot U\varphi(\xi)$ and $(U\varphi)^{(n)}(\xi) = U[(-2i\pi x)^n \varphi](\xi)$ for every $n \in \mathbb{N}$.

Here, $[(-2i\pi x)^n]$ denotes the operator of multiplication by the function $x \mapsto (-2i\pi x)^n$.

Theorem 15 (Riemann-Lebesgue). Let $f \in L^1(\mathbb{R})$. Then Uf is bounded and $\|Uf\|_\infty \leq \|f\|_1$. Moreover, $Uf(\xi)$ is continuous and goes to zero as $|\xi| \rightarrow \infty$.

Theorem 16. Let $f \in \mathcal{S}(\mathbb{R})$. Then, for every $x \in \mathbb{R}$,

$$f(x) = \int e^{2i\pi x\xi} \hat{f}(\xi) d\xi.$$

Otherwise expressed, $U: \mathcal{S}(\mathbb{R}) \rightarrow \mathcal{S}(\mathbb{R})$ is bijective, and $U^{-1} = \bar{U}$.

Corollary 17. Let f and g be two functions of $\mathcal{S}(\mathbb{R})$. Then fg and $f * g$ belong to $\mathcal{S}(\mathbb{R})$.

The mapping

$$\begin{aligned} \langle \cdot, \cdot \rangle : \quad \mathcal{S}(\mathbb{R}) \times \mathcal{S}(\mathbb{R}) &\longrightarrow \mathbb{C} \\ (f, g) &\longmapsto \int f \bar{g} \end{aligned}$$

is a Hermitian product, which turns $\mathcal{S}(\mathbb{R})$ into an inner product space. The linear mappings U and U^{-1} are adjoint to each other, since

$$\begin{aligned} \langle Uf, g \rangle &= \int \left(\int e^{-2i\pi x\xi} f(x) dx \right) \bar{g}(\xi) d\xi \\ &= \int f(x) \left(\int e^{-2i\pi x\xi} \bar{g}(\xi) d\xi \right) dx \\ &= \int f(x) \overline{\int e^{2i\pi x\xi} g(\xi) d\xi} dx \\ &= \langle f, U^{-1}g \rangle, \end{aligned}$$

in which the second equality stems from the Fubini Theorem. Using the density of $\mathcal{S}(\mathbb{R})$ in $L^2(\mathbb{R})$, we can extend by closure both U and $\overline{U}$. This process can be described by the following proposition.

Proposition 18. Let F, G be Banach spaces whose norms are denoted by $\|\cdot\|$, and let E be a dense subspace of F . Let $A: E \rightarrow G$ be a linear mapping such that:

$$\exists \kappa, K > 0: \forall x \in E, \kappa \|x\| \leq \|Ax\| \leq K \|x\|.$$

Then, there exists a unique continuous linear mapping $\mathbb{A}: F \rightarrow G$ whose restriction to E coincides with A . This mapping is referred to as the *extension by closure* of A . Moreover, we have:

- (1) $\forall x \in F, \kappa \|x\| \leq \|\mathbb{A}x\| \leq K \|x\|$;
- (2) the range of $\mathbb{A}$ is the closure of the range of A ;
- (3) $\mathbb{A}: F \rightarrow \text{ran } \mathbb{A}$ is bijective and bicontinuous, and $\mathbb{A}^{-1}$ is the extension by closure of $A^{-1}: \text{ran } A \rightarrow E$.

Theorem 19. For every $\varphi \in \mathcal{S}(\mathbb{R})$, $\|\hat{\varphi}\|^2 = \|\varphi\|^2$. Consequently, the Fourier transformation $U: \mathcal{S}(\mathbb{R}) \rightarrow \mathcal{S}(\mathbb{R})$ can be extended to a unique continuous linear mapping $\mathbb{U}: L^2(\mathbb{R}) \rightarrow L^2(\mathbb{R})$, which is an isometry of $L^2(\mathbb{R})$:

$$\forall f \in L^2(\mathbb{R}), \quad \|\mathbb{U}f\| = \|f\|.$$

The operator $\mathbb{U}$ is referred to as the Fourier-Plancherel operator. It is a Hilbert space isomorphism since, from the polarization identity,

$$\forall f, g \in L^2(\mathbb{R}), \quad \langle f, g \rangle = \langle \mathbb{U}f, \mathbb{U}g \rangle.$$

The Fourier-Plancherel operator $\mathbb{U}$ is merely denoted by U , throughout. We stress that for functions in $L^2(\mathbb{R})$ that are not integrable, the Fourier integral may not be defined. However, the following holds:

Theorem 20. (1) Suppose that $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. Then $\mathbb{U}f = \hat{f}$, where $\hat{f}$ is identified to its equivalence class in $L^2(\mathbb{R})$.

- (2) Suppose now that f is any member of $L^2(\mathbb{R})$. For every $R > 0$, let

$$\begin{aligned} \Phi_R(\xi) &:= \int_{-R}^R e^{-2i\pi x\xi} f(x) dx = \mathbb{U}\mathbb{1}_{[-R,R]}f(\xi), \\ F_R(x) &:= \int_{-R}^R e^{2i\pi x\xi} (\mathbb{U}f)(\xi) d\xi = \mathbb{U}^{-1}\mathbb{1}_{[-R,R]}\mathbb{U}f(x). \end{aligned}$$

Then $\|\Phi_R - \mathbb{U}f\|_2 \rightarrow 0$ and $\|F_R - f\|_2 \rightarrow 0$ as $R \rightarrow \infty$.

7.4 The multiplication operator

Theorem 21. Let φ be in $L^\infty(\mathbb{R})$ and let $M: L^2(\mathbb{R}) \rightarrow L^2(\mathbb{R})$ be the multiplication operator defined by

$$Mf = \varphi \cdot f.$$

Then, $\|M\| = \|\varphi\|_\infty$, in which $\|M\|$ denotes the operator norm of M .

PROOF. We can assume without loss of generality that $\|\varphi\|_\infty > 0$. Let λ denote the Lebesgue measure on $\mathbb{R}$. Clearly,

$$\int |\varphi f|^2 d\lambda \leq \|\varphi\|_\infty^2 \int |f|^2 d\lambda = \|\varphi\|_\infty^2 \|f\|_2^2,$$

so that $\|M\| \leq \|\varphi\|_\infty$. In order to obtain the opposite inequality, let $\varepsilon > 0$ be fixed. The Lebesgue measure being σ -finite, one can find $A \subset \mathbb{R}$ measurable such that

- (i) $0 < \lambda(A) < \lambda(\mathbb{R}) = \infty$;
- (ii) for every $x \in A$, $\varphi(x) \geq \|\varphi\|_\infty - \varepsilon$.

As a matter of fact, let $A_\circ := \{x \in \mathbb{R} \mid |\varphi(x)| \geq \|\varphi\|_\infty - \varepsilon\}$. Then $\lambda(A_\circ) > 0$ (for otherwise one would have $\|\varphi\|_\infty \leq \|\varphi\|_\infty - \varepsilon$). If $\lambda(A_\circ) < \infty$, then just take $A = A_\circ$. Otherwise, consider an increasing sequence (B_n) such that $\lambda(B_n) > 0$ for all n and $\cup_n B_n = \mathbb{R}$. The sequence $(A_\circ \cap B_n)$ is increasing, with limit $A_\circ$. For $n_\circ$ sufficiently large, $\lambda(A_\circ \cap B_{n_\circ}) > 0$, since $\lambda(A_\circ \cap B_n) \rightarrow \lambda(A_\circ) > 0$. Thus take $A = A_\circ \cap B_{n_\circ}$ in this case.

Now, let $f = \mathbb{1}_A / \sqrt{\lambda(A)}$. Then $f \in L^2(\mathbb{R})$ and $\|f\|_2 = 1$, so that

$$\|M\|^2 \geq \|Mf\|_2^2 = \int |\varphi f|^2 d\lambda = \frac{1}{\lambda(A)} \int_A |\varphi|^2 d\lambda \geq (\|\varphi\|_\infty - \varepsilon)^2.$$

Since ε can be chosen arbitrarily small, we have shown that $\|M\| \geq \|\varphi\|_\infty$. ■

References

- [1] Nathaël Alibaud, Pierre Maréchal, and Yaowaluk Saesor. A variational approach to the inversion of truncated fourier operators. *Inverse Problems*, 25(4):045002, 2009.
- [2] Xavier Bonnefond and Pierre Maréchal. A variational approach to the inversion of some compact operators. *Pacific journal of optimization*, 5(1):97–110, 2009.
- [3] Cristina Butucea and Catherine Matias. Minimax estimation of the noise level and of the deconvolution density in a semiparametric convolution model. *Bernoulli*, 11(2):309–340, 2005.

- [4] Marine Carrasco and Jean-Pierre Florens. A spectral method for deconvolving a density. *Econometric Theory*, 27(3):546–581, 2011.
- [5] Marine Carrasco, Jean-Pierre Florens, and Eric Renault. Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization. *Handbook of econometrics*, 6:5633–5751, 2007.
- [6] Raymond J Carroll and Peter Hall. Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association*, 83(404):1184–1186, 1988.
- [7] Fabienne Comte, Jérôme Dedecker, and Marie-Luce Taupin. Adaptive density deconvolution with dependent inputs. *Mathematical methods of Statistics*, 17(2):87, 2008.
- [8] Fabienne Comte, Yves Rozenholc, and M-L Taupin. Finite sample penalization in adaptive density deconvolution. *Journal of Statistical Computation and Simulation*, 77(11):977–1000, 2007.
- [9] Fabienne Comte, Yves Rozenholc, and Marie-Luce Taupin. Penalized contrast estimator for adaptive density deconvolution. *Canadian Journal of Statistics*, 34(3):431–452, 2006.
- [10] Luc Devroye. Consistent deconvolution in density estimation. *Canadian Journal of Statistics*, 17(2):235–239, 1989.
- [11] Sam Efromovich. Density estimation for the case of supersmooth measurement error. *Journal of the American Statistical Association*, 92(438):526–535, 1997.
- [12] Jianqing Fan. Global behavior of deconvolution kernel estimates. *Statistica Sinica*, pages 541–551, 1991.
- [13] Jan Johannes. Deconvolution with unknown error distribution. *The Annals of Statistics*, 37(5A):2301–2323, 2009.
- [14] Bernard A Mair and Frits H Ruymgaart. Statistical inverse estimation in hilbert scales. *SIAM Journal on Applied Mathematics*, 56(5):1424–1444, 1996.
- [15] Pierre Maréchal, Dylan Togane, and Anna Celler. A new reconstruction methodology for computerized tomography: Frect (fourier regularized computed tomography). *IEEE Transactions on Nuclear Science*, 47(4):1595–1601, 2000.
- [16] Alexander Meister. On the effect of misspecifying the error density in a deconvolution problem. *Canadian Journal of Statistics*, 32(4):439–449, 2004.
- [17] Marianna Pensky and Brani Vidakovic. Adaptive wavelet estimator for nonparametric density deconvolution. *The Annals of Statistics*, 27(6):2033–2053, 1999.
- [18] Leonard A Stefanski and Raymond J Carroll. Deconvolving kernel density estimators. *Statistics*, 21(2):169–184, 1990.

- [19] Andrey Nikolayevich Tikhonov and Vasili Ya Arsenin. *Methods for solving ill-posed problems*. John Wiley and Sons, Inc, 1977.
- [20] Arnoud CM Van Rooij and Frits H Ruymgaart. Regularized deconvolution on the circle and the sphere. In *Nonparametric Functional Estimation and Related Topics*, pages 679–690. Springer, 1991.