TSE working paper series

09-030

Research Group: *Econometrics and Statistics*

April 3, 2009

Convergence Rates for Ill-Posed Inverse Problems with an Unknown Operator

JAN JOHANNES, SÉBASTIEN VAN BELLEGEM

AND ANNE VANHEMS

Ecole d'économie de Toulouse (TSE) - Manufacture des Tabacs
Aile Jean-Jacques Laffont - 21, allée de Brienne - 31000 TOULOUSE
Tél : +33 (0) 5 61 12 85 89 - Fax : +33 (0) 5 61 12 86 37 - **www.TSE-fr.eu** - contact@TSE-fr.eu

Toulouse School of Economics

Convergence rates for ill-posed inverse problems with an unknown operator *

Jan Johannes[1] Sébastien Van Bellegem[2] Anne Vanhems[2,3]

This version, April 3, 2009

Abstract

This paper studies the estimation of a nonparametric function φ from the inverse problem $r = T\varphi$ given estimates of the function r and of the linear transform T. The rate of convergence of the estimator is derived under two assumptions expressed in a Hilbert scale. The approach provides a unified framework that allows to compare various sets of structural assumptions used in the econometrics literature. General upper bounds are derived for the risk of the estimator of the structural function φ as well as of its derivatives. It is shown that the bounds cover and extend known results given in the literature. Particularly, they imply new results in two applications. The first application is the blind nonparametric deconvolution on the real line, and the second application is the estimation of the derivatives of the nonparametric instrumental regression function via an iterative Tikhonov regularization scheme.

Keywords: Inverse problem, Hilbert Scale, Blind deconvolution, Instrumental variable, Nonparametric regression, Regularization, Iterative Tikhonov, Spectral cut-off, Thresholding, Sobolev spaces

JEL classifications: Primary C14; secondary C30

*This work was supported by the "Agence National de le Recherche" under contract ANR-09-JCJC-0124-01 and by the IAP research network nr P6/03 of the Belgian Government (Belgian Science Policy). Main results were presented during the Econometric world congress in London in 2005, the Cowles Summer Conferences in Yale in 2008 and the European Meeting of the Econometric Society in Milan in 2008. Comments from the participants of the meeting were greatly appreciated. We gratefully ackowledge the comments and advices from Richard Blundell, Thierry Magnac, Jean-Marc Robin and from the referee which helped to improve the presentation of our results. The usual disclaimer applies.

[1]Université catholique de Louvain

[2]Toulouse School of Economics

[3]Toulouse Business School

1 Introduction

A wide range of econometric problems are related to the identification and the estimation of a nonparametric function φ from a structural model

$$r = T\varphi, \tag{1.1}$$

where r and T are a function and a linear operator that are known or can be estimated from observations.

One important example of such a problem is given when φ is a density function that solves the convolution equation

$$r(y) = \varphi \star f(y) := \int_{-\infty}^{\infty} f(y-u)\varphi(u)du \tag{1.2}$$

where r and f are two other density functions. Such an equation arises when the density φ has to be estimated from a sample that is contaminated by an additive measurement error with probability density f. Here the function r represents the density of the contaminated observation, and the linear operator T is the convolution with f. Example of economic applications can be found e.g. in Horowitz (1998), Postel-Vinay and Robin (2002), Bonhomme and Robin (2006).

Another example of the model (1.1) is given when φ is solution of the moment equation

$$\mathbb{E}[Y|W] = \mathbb{E}[\varphi(Z)|W], \tag{1.3}$$

where Y is a dependent variable, Z is a vector of endogenous explanatory variables and W is a vector of instruments. In the setting of (1.1), the function r is the conditional expectation $\mathbb{E}[Y|W]$ of the response Y given the instruments W and the operator T is the conditional expectation operator, i.e. $T\varphi = \mathbb{E}[\varphi(Z)|W]$. Identification and nonparametric estimation of φ have been the subject of many recent economic studies, see e.g. Darolles, Florens, and Renault (2002), Hall and Horowitz (2005), Blundell, Chen, and Kristensen (2007), Chen and Reiss (2007) and other references below.

A common task in the above problems is to analyse the accuracy of an estimator of φ, for example its mean square convergence. In the literature, rates of convergence are derived under various assumptions on φ, r or T, or on the data generating process involved in the specific problem. To summarize, three kinds of a priori assumptions are usually considered in the literature: (i) a function class on the solution φ, (ii) a smoothness class for the density of the observations, and (iii) a regularity condition on the operator T.

The two first types of assumptions are usual in nonparametric estimation, in which it is assumed that φ or the densities belong to some space of smooth functions, such as a Hölder or a Sobolev space. In the context of inverse problems in econometrics, some works also consider the prior that φ belongs to a compact space, see e.g. Newey and Powell (2003) and the discussion on that assumption in Florens, Johannes, and Van Bellegem (2005). Several conditions of type (iii) were proposed in the literature, among which are the *sieve measure of ill-posedness* (e.g. Blundell, Chen, and Kristensen (2007)), and the *source condition* (e.g. Carrasco, Florens, and Renault (2007)). In Hall and Horowitz (2005), the regularity of the operator, assumed to be compact, is directly expressed by the rate of decreasing of its discrete spectrum.

It is not an easy task to compare the recent results in the econometric literature because they are derived under various combination of those three assumptions. It is a main objective of the present paper to provide a unification of the results that allows this comparison. We show that rates of convergence are essentially driven by two assumptions, both written in a common *Hilbert Scale*. The definition of Hilbert scale and our two structural assumptions are explained in Section 2.

Under those two assumptions, the central results of the paper are risk bounds for the estimation of φ and its derivatives when r and possibly T are unknown and estimated from data. Those results are proved in a setting that is general enough to be applied to a wide range of ill-posed linear problems in econometrics (e.g. as the above deconvolution and instrumental regression problems). They are also proved to be "practical", meaning that, when applied to particular econometric models, they allow to derive new rates of convergence and extend significantly some results of the literature. In the paper, we illustrate that point and we now summarize some of these results.

First of all, we address as an example the nonparametric estimation of the deconvolution density φ from the convolution equation (1.2) when the observation's density r on the left hand side in (1.2) and the error density f are unknown but estimated from data. That setting is sometimes called "blind deconvolution", as the density of the error is not assumed to be known. Although it is a more realistic setting, its study has surprisingly not yet been considered in econometrics, as far as we know. We provide a new estimator (the double-threshold deconvolution estimator) of φ and of the s first derivatives of φ, and derive its rate of convergence under various conditions on f and φ (e.g. under ordinary or super smooth conditions as defined in Fan (1991)). A remarkable result is the rate of convergence of the derivative of φ when both φ and f are super smooth (see Proposition 3.3 below). It is also worth mentioning that we do not assume that the deconvolution density φ or the error density is compactly supported. In particular, the model allows that variables are normally distributed.

Another result is given by the introduction of a new estimator of the derivatives of φ from the instrumental model (1.3). As far as we know, there is no published paper given an estimator of that quantity. To achieve fast rates of convergence, the estimator is regularized by iterative Tikhonov method. Explicit rates of convergence are given under various regularity conditions (see Proposition 4.3 below) and are proved to extend to derivative estimation the rates found e.g. in Hall and Horowitz (2005) and Chen and Reiss (2007).

The paper is organized as follows. In the next section, we define and give examples of Hilbert scale. We state the two central assumptions under which the main results are proved, and make explicit connection with the literature. Then, our study is divided into two cases. In Section 3 we consider the situation where the eigenfunctions of T are known. As we show below, a natural thresholding estimator can be defined. The deconvolution problem (1.2) is an example of that case, where the eigenfunctions of T are given by the Fourier exponentials. Then, in Section 4 we study the most general case where the eigenfunctions of T are not known. In that case, there is no natural estimator and we discuss various regularization schemes. The estimation of a nonparametric instrumental regression from (1.3) is an example of the second case. Technical proofs are in the Appendix.

2 Structural assumptions

2.1 Hilbert scale

The two sufficient assumptions needed in order to derive the mean square convergence of estimators are both written in a given Hilbert scale. As it is a new concept in econometrics,[1] we recall its basic construction and give some intuitive examples below. For a complete exposition we refer e.g. to Krein and Petunin (1966).

A Hilbert scale is a sequence of embedded Hilbert spaces defining the a priori smoothness assumptions of the problem. The simplest example is given by the sequence of spaces of differentiable functions, which constitutes an embedded sequence of regularity spaces indexed by the degree of smoothness. As we will see below, other examples are also useful.

A rigorous construction of a Hilbert scale involves the definition of an operator $B : H \to H$ which is unbounded, self-adjoint and strictly positive, and where H is a Hilbert space (very often $H = L^2(\mathbb{R})$ or $L^2[0,1]$). The *Hilbert scale generated by* B is then the sequence of spaces $(H_q)_{q\in\mathbb{R}}$, where $H_q = \mathcal{D}(B^{q/2})$ is the domain of $B^{q/2}$.

In our example of differentiable functions, B is the second order derivative, and the condition $g \in H_q$ simply means that g has q derivatives. However, as we will clarify below, some econometric papers do not characterize the regularity of φ with respect to the order of derivative. In Section 2.4 we explain why it may also be useful to consider another operator B than the second order derivative to generate the regularity spaces.

It is also worth mentioning that q is not necessarily an integer, and that the above definition holds for $q \in \mathbb{R}$. Note also that the spaces H_q are again Hilbert spaces with the natural inner product $\langle g,h\rangle_q := \langle B^{q/2}g, B^{q/2}h\rangle$ and induced norm $\|g\|_q := \left\|B^{q/2}g\right\|$.

The two following examples expand the case of differentiable functions.

Example 2.1 (Sobolev spaces $\mathcal{W}_q(\mathbb{R})$). Let $H = L^2(\mathbb{R})$ be the space of square integrable functions defined on $\mathbb{R}$. Consider the subset $\mathcal{W}_q(\mathbb{R})$ of H containing only functions g

[1] Assumptions on Hilbert scales were also considered in Chen and Reiss (2007) and Florens, Johannes, and Van Bellegem (2009).

with square integrable q-th derivative $g^{(q)}$. This subset is called a Sobolev space, and the $(\mathcal{W}_q(\mathbb{R}))_q$ is a Hilbert scale. A rigorous definition of that scale necessitates the corresponding operator B in the Fourier domain. Denote by $\mathcal{F}$ the Fourier transform on $\mathbb{R}$, and define B as $Bg(\omega) = \mathcal{F}^{-1}\beta(\omega)\mathcal{F}g(\omega)$ where $\mathcal{F}^{-1}$ is the inverse Fourier transform, and $\beta(\omega) = (1+\omega^2)$. One can show that this operator B generates the above sequence of Sobolev spaces (cf. Mair and Ruymgaart (1996) for a complete exposition). The induced norm is equivalent to the usual Sobolev norm $\|g\|^2 + \|g^{(q)}\|^2 = \int(1+|\omega|^{2q})|\mathcal{F}g(\omega)|^2 d\omega$. $\square$

Example 2.2 (Sobolev spaces of periodic functions $\mathcal{W}_q([0,1])$). In analogy to the previous example consider the space $\mathcal{W}_q[0,1]$ of periodic functions on the interval $[0,1]$ with square integrable q-th derivative . The sequence $(\mathcal{W}_q[0,1])_q$ is also a Hilbert scale generated now by the operator $Bg := \sum_j \beta_j \langle g, \phi_j\rangle \phi_j$ where $\beta_j := (1+(2j)^2)$ and $\{\phi_1 \equiv 1, \phi_{2k}(x) = \sqrt{2}\cos(2\pi kx), \phi_{2k+1}(x) = \sqrt{2}\sin(2\pi kx), k \in \mathbb{N}\}$ is the trigonometric basis (c.f. Neubauer (1988)). $\square$

One advantage of using Hilbert scales is the simplicity by which we derive the rate of convergence of the derivative of the estimator. To illustrate that point, let $g \in \mathcal{W}_q(\mathbb{R})$ and suppose we want to estimate its s-th derivative $g^{(s)}$. From standard theory, it holds $[\mathcal{F}g^{(s)}](\omega) = (\iota\omega)^s[\mathcal{F}g](t)$, $t \in \mathbb{R}$, where $\mathcal{F}$ denotes the Fourier transform and $\iota^2 = -1$. Now we introduce the function $g_s = B^{s/2}g$, where B is defined in Example 2.1. By simple calculation, it holds $\|g^{(s)}\| = \|\mathcal{F}g^{(s)}\| \leqslant \|g_s\|$. Therefore, if $\hat{g}$ denotes some estimator of g, then the risk of the estimator $\widehat{g_s} = B^{s/2}\hat{g}$ provides an upper bound for the risk of $(\hat{g})^{(s)}$.

Remark 2.1. Examples 2.1 and 2.2 can be extended in order to characterize also subsets of analytic functions. One possibility is to set $\beta(\omega) := \exp(|\omega|^{2\gamma})$ in the first example or $b_j = \exp(|j|^{2\gamma})$ in the second, for some$\gamma > 1/2$ (cfr. Kawata (1972)). Subsets of non differentiable functions can also be generated using e.g. the Haar-basis $\{\phi_j\}$ in the second example instead of the trigonometric basis.

2.2 A priori regularity of φ

In non and semiparametric econometrics, it is standard to assume that the function to be estimated belongs to some known space of regularity. The idea is similar for inverse problems, but the difficulty in econometrics is that it is not natural to impose conditions on φ independently of the behavior of T. That is the reason why the conditions on φ and T will be formulated with respect to a common Hilbert scale. Below we show that our conditions, expressed in Hilbert scale, contain (and extend) many of the various assumptions that are used in the literature.

Consider a Hilbert scale $(H_q)_{q\in\mathbb{R}}$ generated by an operator B. The following assumption determines our prior on φ.

Assumption 2.1. *The solution φ of the inverse problem (1.1) belongs to H_p for some $p > 0$.*

In the case where (H_q) is the scale of derivable functions (as in Example 2.1), this assumption simply states that the solution has p square integrable derivatives.

2.3 Link condition

Suppose the function φ satisfies Assumption 2.1 for a given $p > 0$ and a Hilbert scale $(H_q)_{q \in \mathbb{R}}$. Our second fundamental assumption makes a link between the prior regularity of φ and the mapping properties of T.

Assumption 2.2. *If $\varphi \in H_p$, $p > 0$, then there exists a continuous, strictly increasing function $\kappa : \mathbb{R}^+ \to \mathbb{R}^+$ with $\kappa(0+) = 0$ such that, for some finite constant $d \geq 1$, the operator T satisfies*

$$\|g\|_{-p}/d \leq \|\kappa(T^*T)g\| \leq d\,\|g\|_{-p} \quad \textit{for all } g \in H. \tag{2.1}$$

To understand how that assumption quantifies the regularity of T, consider again the case where (H_q) is the scale of differentiable functions (as in Example 2.1) and suppose T is a smoothing operator. For instance, suppose that T is smoothing one time, that is there exists one more derivative of Tg than of g. Another way to express that relation is to assume that the norm of $\|Tg\|$ is equivalent to the norm of $\|B^{-1/2}g\|$ (recall that here B is the second order derivative. Therefore, $B^{-1/2}$ is intuitively speaking the first order integral of g, which indeed has one more derivative than g). In the notations of Assumption 2.2 we equivalently write[2]: $\|(T^\star T)^{1/2}g\| \asymp \|g\|_{-1}$. Now, if we assume that the solution φ has, e.g., 3 derivatives, that is $p = 3$ in Assumption 2.1, then it is easy to verify that condition (2.1) holds with $\kappa(t) = t^{3/2}$.

The last assumption therefore quantifies the relative regularity of T with respect to the regularity of φ. That relative regularity is characterized by the function κ. As we will see in the applications below, the link function κ essentially determines the bound for the risk and hence the rate of convergence.

In the case where we are interested in estimating the derivatives of φ, the above assumption has to be slightly modified. Consider again the scale of Sobolev spaces introduced in Example 2.1. Our target function now is $\varphi_s = B^{s/2}\varphi$ for some $0 \leqslant s < p$, where p determines the prior number of existing derivatives of φ by Assumption 2.1. If we rewrite the initial problem as follows

$$r = T\varphi = TB^{-s/2}B^{s/2}\varphi = T_s\varphi_s \tag{2.2}$$

with obvious definition for T_s, we again see how Hilbert scales easily handle the problem of estimating the derivative. Assumption 2.2 is thus extented to the following assumption:

[2] The following convention is used: $\|u\| \asymp \|v\|$ means that there exists a constant $d \geq 1$ such that $\|v\|/d \leqslant \|u\| \leqslant d\,\|v\|$.

Assumption 2.3. *If $\varphi \in H_p$, $p > 0$, then, for $0 \leq s < p$, there exists a continuous, strictly increasing function $\kappa : \mathbb{R}^+ \to \mathbb{R}^+$ with $\kappa(0+) = 0$ such that, for some finite constant $d \geq 1$, the operator T satisfies*

$$\|g\|_{s-p} / d \leq \|\kappa(T_s^* T_s) g\| \leq d \|g\|_{s-p} \quad \text{for all } g \in H. \tag{2.3}$$

Once again, we have a look at Example 2.1 above to illustrate that assumption. If the true function φ is three times differentiable, that is $p = 3$ in Assumption 2.1, if we are interested by the first derivative of φ, that is $s = 1$, and if the operator T is smoothing one times, then it is easy to check that T_s is smoothing two times and condition (2.1) holds with $\kappa(t) = t^{1/2}$.

2.4 Connection with the literature

Before stating our main results, we link the above conditions with the more standard notions and assumptions that have been used before in the econometric literature on inverse problems. As far as we know, the existing literature in econometrics only covers the case $s = 0$, therefore we concentrate our discussion on that situation. In the next section we derive results for $s \geqslant 0$, that is we are able to derive rates of convergence for the derivatives of φ.

If the Hilbert scale is given by the Sobolev spaces we already mentioned that Assumption 2.1 is equivalent to assume that φ has p derivatives. Moreover, if the link function takes the form $\kappa(t) = t^{p/2a}$, we recover the case where T is said to be *finitely smoothing* (as named in Natterer (1984)). In the particular example of the convolution equation (1.2) this case corresponds to the ordinary smoothing condition of Fan (1991), i.e., $|\mathcal{F}\varphi(t)|^2 \asymp |t|^{-2a}$. In the instrumental regression model (1.3) with a compact operator T, it similarly imposes that the eigenvalues of T are decreasing at a polynomial rate. This case was considered in Section 4.2 of Hall and Horowitz (2005).

Assuming a finitely smoothing T is sometimes too restrictive. For example, in the convolution equation (1.2) it does not cover the case where the error density is super smooth, i.e., $|\mathcal{F}f(t)|^2 \asymp \exp(-|t|^{2a})$. We recover the case where T is infinitely smoothing if we set $\kappa(t) = |\log(t)|^{-p/2a}$ in Assumption 2.2 (e.g. Mair (1994)). In the instrumental regression model (1.3), surprisingly there is no published paper considering that case, as far as we know. However that case is very natural since it covers the situation where (Y, Z, W) are jointly normal.

The two assumptions 2.1 and 2.2 also recover the *Source condition* often used in numerical analysis, and that has been considered in the nonparametric instrumental regression problem e.g. in Darolles, Florens, and Renault (2002) and in Section 4.1 of Hall and Horowitz (2005). That setting is covered if the Hilbert scale is generated by the operator $B = (T^\star T)^{-1}$ and the link function is $\kappa = t^{p/2}$. As before, a polynomial link function is restrictive because it assumes strong conditions on the distribution functions involved in the instrumental regression problem. That was the motivation to consider a logarithmic link function in Johannes, Van Bellegem, and Vanhems (2009).

Finally, in recent work, Chen and Reiss (2007) established minimax risk lower bounds for the nonparametric instrumental regression model under similar assumptions than the above assumptions 2.1 and 2.2. We recover their rate of convergence as a particular case (see the discussion of Proposition 4.3 below).

The following table summarizes the above connections.

Hilbert scale H_p	$\kappa(t) = t^{p/2a}$	$\kappa(t) = \lvert\log(t)\rvert^{-p/2a}$
given by Sobolev spaces	φ is p-times differentiable, T is finitely smoothing with degree of smoothness a	φ is p-times differentiable, T is infinitely smoothing for every a
generated by $B = (T^\star T)^{-1}$	Source condition with $a = 1$	Log source condition with $a = 1$

Table 1: This table identifies the connections between particular Hilbert scales and various types of assumptions used in the literature

3 Risk bounds in case of known eigenfunctions

3.1 Eigenfunctions

In this section, we derive a risk bound in the particular case where the operator B defining the Hilbert scale, and the operator T have the same set of eigenfunctions. The general case will be discussed in the next Section. However, the case where those operators share the same eigenfunctions is already of interest in some econometrics problems. The deconvolution problem is one example, as we will see later.

As it was suggested by the referee, since our main example is the deconvolution problem (1.2) on $\mathbb{R}$, we simplify the exposition and assume in this section that the eigenfunctions are given by the Fourier exponentials. The case of general eigenfunctions is a straightforward extension (see Remark 3.1).

The nice point when eigenfunctions are known is that a natural estimator of φ will be defined by a series estimator in the system of eigenfunctions. Let $\mathcal{F}$ be the Fourier transform on $\mathbb{R}$. In this setting, the structural model (1.1) implies that $\mathcal{F}r(\cdot) = \lambda(\cdot)\mathcal{F}\varphi$, where $\lambda(\cdot)$ is the Fourier spectrum of T (e.g. the Fourier transform of the error density, f, in the convolution case (1.2)).

Moreover, since the eigenfunctions of T are the Fourier exponential, it is natural to consider the Hilbert scale of Sobolev spaces which is defined in Example 2.1. Recall that the operator B generating that scale is such that for all g, $\mathcal{F}Bg(\omega) = \beta(\omega)\mathcal{F}g(\omega)$, where

$\beta(\omega) = (1+\omega^2)$. In that setting, the link condition (2.3) can be rewritten as a link between the Fourier transform of T (given by $\lambda(\cdot)$) and the Fourier transform of B:

$$\beta^{s-p}(\cdot) \asymp \kappa(\lambda^2(\cdot)/\beta^s(\cdot))^2 \tag{3.1}$$

almost surely. In order to show how this setting simplifies the estimation problem of φ_s (that is the sth derivative of φ), we go back to the inverse problem $r = T_s\varphi_s$, see (2.2). The decomposition implies after straightforward calculations that

$$\mathcal{F}\varphi_s(\cdot) = \frac{\beta^{s/2}(\cdot)}{\lambda(\cdot)}\mathcal{F}r(\cdot) \tag{3.2}$$

almost surely. This shows that in case of a-priori known eigenfunctions the estimation of φ_s is reduced to the estimation of the Fourier spectrum $\lambda(\cdot)$ of T, and the unknown function $\mathcal{F}r(\cdot)$. That idea is developed next.

3.2 Risk bound when the spectrum is known

First we show a direct proof for a risk bound when the Fourier spectrum λ is known. In deconvolution, it means that the error density, f, is known and only an estimator $\widehat{\mathcal{F}r}$ of the function $\mathcal{F}r$ is required. Due to the illposedness of the inverse problem, it is well-known that replacing the unknown function $\mathcal{F}r$ in (3.2) by its estimator will generally not lead to a consistent estimator of φ_s. Therefore, we introduce a threshold that exploits the link condition (3.1). An estimator $\widetilde{\varphi}_s$ of φ_s is then given by the inverse Fourier transform of

$$\mathcal{F}\widetilde{\varphi}_s(\cdot) = \frac{\beta^{s/2}(\cdot)}{\lambda(\cdot)}\,\widehat{\mathcal{F}r}(\cdot)\,\mathbb{1}\{\beta^{s-p}(\cdot) \geqslant d\kappa(\delta^*)^2\}\,. \tag{3.3}$$

The next proposition shows that the rate of convergence of the estimator is driven by the link function κ.

Proposition 3.1. *Let $\widehat{\mathcal{F}r}$ be an estimator of $\mathcal{F}r$ such that $\mathbb{E}|\widehat{\mathcal{F}r} - \mathcal{F}r|^2 \leqslant c\eta$ uniformly over μ a.e. for some $c > 0$. Under Assumptions 2.1 and 2.3, if the threshold $\delta^* := \delta^*(\eta)$ satisfies*

$$D^{-1} \leqslant \frac{\eta}{\kappa(\delta^*)^2}\left\|\frac{\mathbb{1}\{\beta(\cdot)^{s-p} \geqslant d\,\kappa(\delta^*)^2\}}{\sqrt{\Phi(\beta(\cdot)^{s-p}/d)}}\right\|^2 \leqslant D, \tag{3.4}$$

where Φ is the inverse function of $\kappa(\cdot)^2$, then there exists a constant $C \geqslant 1$ only depending on D, d and c such that $\mathbb{E}\,\|\widetilde{\varphi}_s - \varphi_s\|^2 \leqslant C\cdot\kappa(\delta^*)^2\cdot\max(\|\varphi\|_p^2, 1)$.

Proof. Define the regularized solution $\varphi_s^\alpha := \mathcal{F}^{-1}[\mathcal{F}\varphi_s\;\mathbb{1}\{\beta^{s-p} \geqslant d\,\kappa(\delta^*)^2\}]$ and consider the usual decomposition of the risk in a bias and a variance term:

$$\mathbb{E}\,\|\widetilde{\varphi}_s - \varphi_s\|^2 \leqslant 2\mathbb{E}\|\widetilde{\varphi}_s - \varphi_s^\alpha\|^2 + 2\|\varphi_s^\alpha - \varphi_s\|^2$$

Using that $\mathcal{F}$ and $\mathcal{F}^{-1}$ have norm one, and with equation (3.2), the first term of the right hand side is bounded by $2c\eta\|\mathbb{1}\{\beta^{s-p}/d \geqslant \kappa(\delta^*)^2\}\beta^{s/2}/\lambda\|^2$. Now, due to the link condition (3.1) we have $\beta^{s-p}/d \leqslant \kappa(\lambda^2/\beta^s)^2$, which implies $\beta^s/\lambda^2 \leqslant 1/\Phi\left(\beta^{s-p}/d\right)$ by definition of the function Φ. Therefore, the first term is bounded by $2c\eta\|\mathbb{1}\{\beta^{s-p}/d \geqslant \kappa(\delta^*)^2\}/\Phi^{1/2}\left(\beta^{s-p}/d\right)\|^2$. Since the second term is bounded by

$$2d\,\kappa(\delta^*)^2\left\|\beta^{(p-s)/2}\mathcal{F}\varphi_s\mathbb{1}\{\beta^{s-p}/d < \kappa^2(\delta^*)\}\right\|^2 \leqslant 2d\,\kappa(\delta^*)^2\,\|\varphi\|_p^2$$

we obtain that

$$\mathbb{E}\,\|\widetilde{\varphi}_s - \varphi_s\|^2 \leqslant 2c\,d\,\kappa(\delta^*)^2\Big\{\frac{\eta}{\kappa(\delta^*)^2}\left\|\frac{\mathbb{1}\{\beta(\cdot)^{s-p} \geqslant d\,\kappa(\delta^*)^2\}}{\sqrt{\Phi(\beta(\cdot)^{s-p}/d)}}\right\|^2 + 1\Big\}\max(\|\varphi\|_p^2, 1).$$

Therefore, the constraint (3.4) on δ^* implies the result, which completes the proof. $\square$

We show in the application below that $\kappa(\delta^*)^2$ is the minimax-optimal rate of convergence. Before, we consider the extension of the result to the case where the spectrum λ is unknown and hence has to be estimated.

3.3 Risk bound when the spectrum is unknown

In some situations, the spectrum λ is unknown but can be estimated. In deconvolution, it means that the density of the error f is unknown (see the application section 3.4 below). In that case, there exist estimator $\widehat{\mathcal{F}r}$ and $\widehat{\lambda}$ of $\mathcal{F}r$ and λ, respectively. As in the previous section by replacing in (3.3) the unknown spectrum λ by its estimator does not lead to a consistent estimator of φ_s. Therefore, we introduce a second threshold α in order to control the decay of $\widehat{\lambda}$ to 0 (even if λ is far away from 0, its estimator $\widehat{\lambda}$ can be very small). Thereby a natural estimator $\widehat{\varphi}_s$ of φ_s is given by the inverse Fourier transform of

$$\mathcal{F}\widehat{\varphi}_s(\cdot) := \frac{\beta^{s/2}(\cdot)}{\widehat{\lambda}(\cdot)}\,\widehat{\mathcal{F}r}\;\mathbb{1}\{\beta^{s-p}(\cdot) \geqslant d\kappa^2(\delta^*)\}\;\mathbb{1}\{\widehat{\lambda(\cdot)}^2/\beta^s(\cdot) \geqslant \alpha\}. \tag{3.5}$$

The choice of the thresholds δ^* and α are discussed in the next result.

THEOREM 3.2. *Let $\widehat{\mathcal{F}r}$ and $\widehat{\lambda}$ be estimator of $\mathcal{F}r$ and λ respectively such that the inequalities*

$$\mathbb{E}|\widehat{\mathcal{F}r} - \mathcal{F}r|^4 \leqslant c_1\eta^2 \tag{3.6}$$
$$\mathbb{E}|\widehat{\lambda} - \lambda|^4 \leqslant c_2\tau^2 \tag{3.7}$$

holds true uniformly over μ a.e. for some $\eta, \tau > 0$ and $c_1, c_2 \geqslant 1$. Under Assumptions 2.1 and 2.3, if the estimator $\widehat{\varphi}_s$ is defined with a threshold $\delta^* := \delta^*(\eta)$ such that (3.4) holds for some $D \geqslant 1$, and if the second threshold satisfies $\alpha = \max(\delta^*/4, \tau)$, then

$$\mathbb{E}\,\|\widehat{\varphi}_s - \varphi_s\|^2 \leqslant C\{\kappa(\delta^*)^2 + \max(\kappa(\tau)^2, \tau)\}\cdot\max(\|\varphi\|_p^2, 1)$$

where C is a strictly positive constant only depending on κ, c_1, c_2, d and D.

Because the proof involves some technical parts, it is deferred to Appendix A.1.

We show below that this bound is minimax in an important application. Theorem 3.2 surprisingly shows that the choice of the second threshold parameter α is automatic from the optimal choice of the first threshold δ^*. Moreover, the optimal δ^* in Theorem 3.2 is identical to the one in Proposition 3.1 where the spectrum λ is known.

Remark 3.1. In this technical remark, we argue that, in the most general case where the eigenfunctions are not necessarily the Fourier exponentials, the results are extended by using the spectral decomposition of the operator T^*T. According to the spectral theorem (cf Halmos (1963)), there exists a measurable function λ^2 defined on some measure space $(\Omega, \mathfrak{B}, \mu)$ with values in the spectrum of T^*T and a unitary mapping[3] $U : H \to L^2_\mu(\Omega)$ such that $UT^*TU^{-1}g = \lambda^2 g$, μ-a.e., for all $g \in L^2_\mu(\Omega)$. Moreover, there exists a partial isometry $V : G \to L^2_\mu(\Omega)$ such that $VTU^{-1}g = \lambda g$, μ-a.e., for all $g \in L^2_\mu(\Omega)$. In the deconvolution problem, U and V are the Fourier transform. In that setting, we assuming that B has the same eigenfunctions, writes $UBU^{-1}g = \beta g$, μ-a.e., for every g with the same unitary operator U and some measurable function β with values in the spectrum of B. A strong connection is therefore assumed between the Hilbert scale and the inverse problem, and that connection is modeled by the existence of the common operator U in the above spectral decompositions. □

3.4 Application 1: Deconvolution on $\mathbb{R}$ with unknown error distribution

We illustrate the above theorem by an application to density estimation with measurement error. The problem arises when we want to estimate the density φ of a random variable X that is observed with a contamination by some independent additive noise of density f. The observational model is $Y = X + \varepsilon$, where the density of the observation Y, $r(y)$, satisfies $r = f \star \varphi$, see also (1.2). A majority of papers assume the density f to be known, which is convenient in theory but may be not realistic in practice.

In the following application, we do not assume the density of the error to be known. Instead, we observe an iid sample of the error distribution. That sample allows to estimate the error density, and hence it is of interest to analyze the impact of this estimation on the rate of convergence of the resulting deconvolution estimator. That situation has been considered in Neumann (1997) and Johannes (2009). An application of Theorem 3.2 in that setting will show new optimal results on the rate of convergence that was not covered by this literature.

Suppose we observe an iid sample $Y_1, \dots, Y_n$ generated from the distribution r and another iid sample $\varepsilon_1, \dots, \varepsilon_m$ from f. In the convolution model (1.2), the operator to be inverted is the convolution with f.

[3] $L^2_\mu(\Omega)$ is the Hilbert space of square integrable functions defined on $(\Omega, \mathcal{B}, \mu)$ endowed with inner product $\langle f, g\rangle_{L^2_\mu(\Omega)} = \int fgd\mu$

From the observations we consider the estimators $\widehat{\mathcal{F}r}(\omega) = (n\sqrt{2\pi})^{-1}\sum_{j=1}^{n}\exp(-\iota\omega Y_j)$ and $\widehat{\mathcal{F}f}(\omega) = (m\sqrt{2\pi})^{-1}\sum_{j=1}^{m}\exp(-\iota\omega\varepsilon_j)$. In order to apply Theorem 3.2, we first need to check conditions (3.6) and (3.7) on the estimators $\widehat{\mathcal{F}r}$ and $\widehat{\mathcal{F}f}$. These conditions are fulfiled since, by application of Petrov (1995, Theorem 2.11), there exists a positive constant $C > 0$ such that $\sup_{\omega\in\mathbb{R}}\mathbb{E}|[\widehat{\mathcal{F}r}](\omega) - [\mathcal{F}r](\omega)|^4 \leqslant C/n^2$ and $\sup_{\omega\in\mathbb{R}}\mathbb{E}|[\widehat{\mathcal{F}f}](\omega) - [\mathcal{F}f](\omega)|^4 \leqslant C/m^2$.

As recalled in Section 2.4, rates of convergence of the deconvolution estimator is usually derived assuming that φ is ordinary smooth and under two sets of assumptions on $\mathcal{F}f$, namely $\mathcal{F}f$ is ordinary or super smooth. Those conditions are covered in our setting if one consider the scale of Sobolev spaces and various link functions κ (cf Table 1). Therefore, suppose φ belongs to $\mathcal{W}_p(\mathbb{R})$ and we want to estimate the s-th derivative $f_X^{(s)}$ of φ. Following the general approach, we propose a new double threshold deconvolution estimator (DTDE) given by

$$\widehat{\varphi^{(s)}} := \mathcal{F}^{-1}\left[(\iota\omega)^s\frac{\widehat{\mathcal{F}r}(\omega)}{\sqrt{2\pi}\widehat{\mathcal{F}f}(\omega)}\,\mathbb{1}\{(1+\omega^2)^{s-p} \geqslant \kappa^2(\delta^*)\}\,\mathbb{1}\{|\widehat{\mathcal{F}f}(\omega)|^2 \geqslant \alpha(1+\omega^2)^s\}\right], \quad (3.8)$$

where $\delta^* := \delta^*(n)$ and $\alpha^*(n, m)$ are two thresholds that decrease to zero as the samples sizes n and m increase.

Proposition 3.3. *In the convolution model $r = f \star \varphi$ with unknown f, assume $\varphi \in \mathcal{W}_p(\mathbb{R})$ for some $p > 0$ and consider the DTDE estimator (3.8) of the s-th derivative of φ for $0 \leqslant s < p$.*

(i) Let f_ε be ordinary smooth, i.e., $|\mathcal{F}f|^2 \asymp (1+t^2)^{-a}$, $a > 0$. Consider the thresholds $\kappa(\delta^*) \asymp n^{-2(p-s)/(2(p+a)+1)}$ and $\alpha \asymp \max(n^{-2(a+s)/(2(p+a)+1)}, m^{-1})$. Then we have

$$\mathbb{E}\|\widehat{\varphi^{(s)}} - \varphi^{(s)}\|^2 = O\Big(n^{-2(p-s)/(2(p+a)+1)} + m^{-(1\wedge(p-s)/(a+s))}\Big).$$

(ii) Let f_ε be super smooth, i.e., $|\mathcal{F}f|^2 \asymp \exp(-|t|^{2a})$, $a > 0$. Consider the thresholds $\kappa^2(\delta)^* \asymp (\log n)^{-(p-s)/a}$ and $\alpha \asymp \max(n^{-c}, m^{-1})$, $c > 0$. Then we have

$$\mathbb{E}\|\widehat{\varphi^{(s)}} - \varphi^{(s)}\|^2 = O\Big((\log n)^{-(p-s)/a} + (\log m)^{-(p-s)/a}\Big).$$

Proof. Observe that in case (i) Assumption 2.3 is satisfied with $\kappa^2(t) = |t|^{(p-s)/(a+s)}$ and, hence $\Phi(t) = |t|^{(a+s)/(p-s)}$. The well-known approximation $\int_{-T}^{T}|t|^r dt \asymp T^{r+1}$ for $r > 0$ together with the definition of β implies $\int_{-T}^{T} 1/\Phi(\beta(t)^{s-p}/d)dt \asymp |T|^{2(a+s)+1}$. Define T^* by $\delta^* =: \Phi(\beta(T^*)^{s-p}/d)$. It follows that the condition on δ^* given in (3.4) of Proposition 3.1 can be rewritten as

$$1/n \asymp |\beta(T^*)|^{s-p}\Big|\int_{-T^*}^{T^*} 1/\Phi(\beta(t)^{s-p}/d)dt\Big|^{-1} \asymp |T^*|^{-2(p+a)-1}. \quad (3.9)$$

Thereby, we have $\delta^* \asymp n^{-2(a+s)/[2(a+p)+1]}$ and $\kappa(\delta^*) \asymp n^{-2(p-s)/[2(p+a)+1]}$. Consequently, the result follows by applying Theorem 3.2.

In case (ii) Assumption 2.3 is satisfied with $\kappa^2(t) = |\log t|^{-(p-s)/a}$ and, hence $\Phi(t) = \exp(-|t|^{a/(s-p)})$. By applying Laplace's Method (c.f. chapter 3.7 in Olver (1974)) we have $\int_{-T}^{T} 1/\Phi(\beta(t)^{s-p}/d)dt \asymp \exp(|T|^{2a}) \asymp 1/\Phi(\beta(T)^{s-p})$. Hence, the condition on δ^* writes now

$$1/n \asymp |\beta(T^*)|^{s-p}\Phi(\beta(T^*)^{s-p}),$$

which implies $\kappa(\delta^*) \asymp \omega(1/n)$, where ω denotes the inverse function of $\omega^{-1}(t) = t \cdot \Phi(t)$. Since $\omega(t) = |\log t|^{-(p-s)/a}(1 + o(1))$ as $t \to 0$ (c.f. Mair (1994)), we conclude $\kappa^2(\delta^*) \asymp |\log n|^{-(p-s)/a}$ and hence $\delta^* \asymp n^{-c}$ for some sufficiently small c. The bound follows from Theorem 3.2. □

The last proposition establishes the minimax-optimality of the DTDE in case of an ordinary and super smooth error density under much milder assumptions than the one used in Neumann (1997) or Johannes (2009). However, other cases of interest are not considered in the literature, as for example the case where φ and f are both supersmooth (e.g. if they are Gaussian or Cauchy). In this situation, following Assumption 2.1, φ belongs to $H_p(\mathbb{R})$, $p > 0$, with $\beta(t) = \exp(|t|^{2\gamma})$, $2\gamma > 0$ (see Example 2.1), and the DTDE of φ writes

$$\widehat{\varphi} := \mathcal{F}^{-1}\left[\frac{\widehat{\mathcal{F}r}(\omega)}{\sqrt{2\pi}\widehat{\mathcal{F}f}(\omega)}\, \mathbb{1}\{\exp(-p|\omega|^{2\gamma}) \geqslant \kappa^2(\delta^*)\}\, \mathbb{1}\{|\widehat{\mathcal{F}f}(\omega)|^2 \geqslant \alpha\}\right]. \tag{3.10}$$

PROPOSITION 3.4. *Consider the convolution model $r = f \star \varphi$ with super smooth φ, i.e., $\varphi \in H_p(\mathbb{R})$, $p > 0$, with $\beta(t) = \exp(|t|^{2\gamma})$, $2\gamma > 0$, and unknown super smooth f i.e., $|\mathcal{F}f|^2 \asymp \exp(-|t|^{2a})$, with $a > \gamma$. Let DTDE estimator (3.10) of φ be defined by using thresholds $\kappa(\delta^*) \asymp \exp(-p|\log n|^{\gamma/a})$ and $\alpha \asymp \max(n^{-c}, m^{-1})$, $c > 0$, then we have*

$$\mathbb{E}\|\widehat{\varphi}^{(s)} - \varphi^{(s)}\|^2 = O\Big(\exp(-p|\log n|^{\gamma/a}) + \exp(-p|\log m|^{\gamma/a})\Big).$$

Proof. Assumption 2.3 is satisfies with $\kappa^2(t) = \exp(-p|\log t|^{-\gamma/a})$ and, hence $\Phi(t) = \exp(-p^{-a/\gamma}|\log t|^{a/\gamma})$. Then by applying Laplace's Method the condition on δ^* can be rewritten as in (3.9). Consider first the case $\gamma < a$, then from (3.9) we obtain $\kappa(\delta^*) \asymp \exp(-p|\log n|^{\gamma/a})$ and hence $\delta^* \asymp n^{-c}$ for some sufficiently small c. Therefore, the bound follows from Theorem 3.2, which proves the result. □

4 General risk bounds

In the most general case, the eigenfunctions of the operator T in (1.1) are unknown, in contrast to the setting of the previous section. As a consequence, there is no natural orthonormal system that simplifies the problem to equation (3.2), and the threshold estimator is no longer the natural regularized estimator. In order to address the question of deriving rates of convergence, we first clarify the notion of regularized estimator.

4.1 Regularized estimator

In order to estimate the function φ in (1.1), it is first needed to estimate the operator T and the function r. The estimator of course depends on the particular inverse problem we are faced with (deconvolution, nonparametric instrumental regression, etc.). In the following, we do not intend to discuss the quality of these estimators, except that we give some considerations in our applications. Instead, we suppose in the main results that estimators $\hat{T}$, $\hat{T}_s$, $\hat{r}$ of T, T_s, r are given, where $T_s = TB^{-s/2}$ and $\widehat{T}_s = \widehat{T}B^{-s/2}$.

It is well known that the ill-posedness of equation (1.1) implies that a consistent estimator of φ is not found by a simple inversion of the estimated operator $\hat{T}_s$. A modification of the inversion, called regularization, is always necessary. We follow the notations of Tautenhahn (1996) and consider a general continuous regularization scheme in Hilbert scale given by

$$\widehat{\varphi}_s = g_\alpha(\widehat{T}_s^{\star}\widehat{T}_s)\widehat{T}_s^{\star}\widehat{r}. \tag{4.1}$$

where the function $g_\alpha : (0, c] \to \mathbb{R}$ is the regularization scheme that is a piecewise continuous function such that $\lim_{\alpha\to 0+} g_\alpha(t) = 1/t$. Note that, in contrast to Tautenhahn (1996), the regularization scheme (4.1) covers the general case where the operator T is not necessarily known. The following example recalls that the regularization methods usually considered in econometrics are characterized by various functions g_α.

Example 4.1. **(i)** The Tikhonov regularization is characterized by $g_\alpha(t) = 1/(t + \alpha)$ and the corresponding regularized estimator is the solution of the minimal penalized contrast problem

$$\min_{\phi\in H_s} \left\|\hat{T}_s\phi - \hat{r}\right\|^2 + \alpha \|\phi\|_s^2 \ .$$

(ii) A generalization of the previous example is given by the Tikhonov regularization of order m. It is given by $g_\alpha(t) = (1 - (\alpha/(t + \alpha))^m)/t$ and $m \geqslant 1$. The regularized estimator $\widehat{\varphi}_s := \widehat{\varphi}_{s,m}$ is the solution of the m iterative minimizations

$$\widehat{\varphi}_{s,j} = \arg\min_{\phi\in H_s} \left\|\hat{T}_s\phi - \hat{r}\right\|^2 + \alpha \|\phi - \widehat{\varphi}_{s,j-1}\|_s^2, \quad j = 1, \ldots, m, \quad \widehat{\varphi}_{s,0} = 0.$$

(iii) The spectral cut-off considers $g_\alpha(t) = 1/t$ for $t \geq \alpha$.

(iv) The Landweber iteration procedure takes $g_\alpha(t) = (1 - (1 - t)^{1/\alpha})/t$.

The Tikhonov regularization is the most widely used regularization scheme in econometrics. Iterative Tikhonov is less used, and will be considered in an application below (Section 4.3). In the deconvolution problem presented above we have used the spectral cut-off (see also e.g. Carrasco and Florens (2002), Cavalier and Hengartner (2005) or Bigot and Van Bellegem (2009)). The Landweber iterative regularization scheme is specifically studied in the context of nonparametric instrumental regression in Johannes, Van Bellegem, and Vanhems (2009).

4.2 Main results

In this section, we show that the assumption on the prior assumption on φ and the link condition, both conveniently expressed in a single Hilbert scale, are sufficient to derive an upper bound for the mean square convergence of the regularized estimator. Before stating the result, we introduce the following restriction on the regularization scheme.

Assumption 4.1. *There exist positive constants c_1 and c_2 such that*

$$\sup_{t>0} t^{1/2}|g_\alpha(t)| \leq c_1/\sqrt{\alpha}, \qquad \sup_{t>0} |tg_\alpha(t)| \leq 1,$$
$$\sup_{t>0} t|1-tg_\alpha(t)| \leq c_2\alpha \qquad \sup_{t>0} |1-tg_\alpha(t)| \leq 1.$$

That type of assumption can already be found in Nair, Pereverzev, and Tautenhahn (2005) and leads to substantial simplifications in the proof. Note that the Tikhonov, order m Tikhonov and spectral cut-off regularization fulfil these constraints, but the Landweber iteration procedure does not. Deriving the upper bound for the latter thus needs another proof technique which can be found in Johannes, Van Bellegem, and Vanhems (2009).

Theorem 4.1. *Under Assumptions 2.1 and 2.3 for a given concave link function $\kappa(\cdot)$, if the regularization scheme satisfies Assumption 4.1, then, for all $\alpha > 0$, the risk bound*

$$\mathbb{E}\left\|\hat{\varphi}_s - \varphi_s\right\|^2 \leqslant C\left\{\alpha^{-1}\mathbb{E}\left\|\widehat{r} - \widehat{T}\varphi\right\|^2 + \kappa\left(\alpha + \sqrt{\mathbb{E}\|\hat{T}-T\|^4}\right)\right\},$$

holds true for a strictly positive, finite constant C.

The last upper bound is a sum of two terms. The first term plays the role of a variance term and the second is a bias term. In fact if the approximation error of the operator is sufficiently small (additional restrictions on the class of possible operators are needed for this), the bound becomes $\alpha^{-1}\mathbb{E}\|\hat{r} - T\varphi\|^2 + \kappa(\alpha)$ which is the upper bound that is found when the operator is known and does not have to be estimated. That bound is known to be optimal (Nair, Pereverzev, and Tautenhahn (2005)).

Moreover, under our assumptions the bound of the first term cannot be improved. (To see this, we can consider a deconvolution problem, where it is straightforward to show that an appropriate choice of α the upper bound provides the optimal rate of convergence). However, under more specific assumptions on the stochastic structure of the inverse problem, the bound of the variance can be improved. A detailed discussion with several examples can be found in Bissantz, Hohage, Munk, and Ruymgaart (2007).

The risk bound is derived under the assumption that the function κ is concave. That constraint can be relaxed at the price of a more restrictive connection between κ and the regularization scheme g_α, leading to the following alternative upper bound.

Theorem 4.2. *Under the conditions of Theorem 4.1 where we no longer assume the function κ to be concave, if we assume instead that there exists $c_\kappa > 0$ such that*

$$\sup_{t>0}\left\{\kappa(t)\,(1-tg_\alpha(t))^{1/2}\right\} \leqslant c_\kappa \cdot \kappa(\alpha) \tag{4.2}$$

then, for all $\alpha > 0$, the inequality

$$\mathbb{E}\,\|\hat{\varphi}_s-\varphi_s\|^2 \leqslant C\Big\{\frac{1}{\alpha}\mathbb{E}\left\|\widehat{r}-\widehat{T}\varphi\right\|^2+\kappa(\alpha)+\mathbb{E}\left\|\kappa(\widehat{T}_s^\star\widehat{T}_s)-\kappa(T_s^\star T_s)\right\|^2\Big\},$$

is satisfied for a strictly positive, finite constant C.

The two first terms of the upper bound involve the regularization parameter α. An optimal choice of α depending on $\|\hat{r}-\hat{T}\varphi\|$ allows to balance the two terms. The last term of the inequality involves the deviation between $\kappa(\hat{T}_s^\star\hat{T}_s)$ and $\kappa(T_s^\star T_s)$ which can be simplified in some situations.

To illustrate that simplification, consider the case $s=0$ and the source condition assumption for which Assumption 2.1 reduces to $\|(T^\star T)^{-p/2}\varphi\| < \infty$, and the degree of ill-posedness (Assumption 2.1) is satisfied with $\kappa(t) \equiv t^{p/2}$. In this mildly ill-posed problem, the following inequality due to Egger (2005) can be used:

$$\left\|(\widehat{T}^\star\widehat{T})^{p/2}-(T^\star T)^{p/2}\right\|^2 \leqslant C\left(\|\widehat{T}-T\|^{2\min(1,p)}+\|\widehat{T}-T\|^{2p}\right)\ .$$

Therefore, the last term of the above risk is of order $(\mathbb{E}\|\widehat{T}-T\|^{2p})^{\min(1,1/p)}$.

For severely ill-posed problems where Assumption 2.1 holds with $\kappa(t) \equiv |\log(t)|^{-p/2}$, the following results proved in Hohage (2000) is useful in order to bound the risk:

$$\left\|\kappa(\widehat{T}^\star\widehat{T})-\kappa(T^\star T)\right\|^2 \leqslant C\Big\{\kappa^2\Big(\left\|\widehat{T}^\star\widehat{T}-T^\star T\right\|\Big)+\left\|\widehat{T}^\star\widehat{T}-T^\star T\right\|\Big\}\ .$$

Therefore, the last term of the above risk is of order $\mathbb{E}(|\log(\|\hat{T}-T\|)|^{-p})$.

4.3 Application 2: Nonparametric instrumental regression

We illustrate the main result in the popular model of nonparametric instrumental regression (1.3) in the case $H = L^2[0,1]$ (cf. Hall and Horowitz (2005) Section 4.2 for a detailed exposition). In the following, a new risk bound is derived from the previous results, in the case where r and T are estimated by the method of sieve, and the regularization scheme is the iterative Tikhonov regularization.

From an iid sample (Y_i, Z_i, W_i), $i = 1,\ldots,n$, consider estimators of r and T constructed by projection on the trigonometric basis $\{\phi_j\}_j$ (see Example 2.2) that are not necessarily the eigenfunctions of T. An orthogonal series estimator of r is given by

$$\widehat{r}(\cdot) = \sum_{j=1}^{k}\widehat{r}_j\phi_j(\cdot) \tag{4.3}$$

where $\hat{r}_j = n^{-1}\sum_{i=1}^n Y_i\phi_j(W_i)$, $j = 1,\dots,k$. The number k of estimated coefficients increases as the sample size n increases. In order to derive a series estimator of T, denote $\boldsymbol{\phi}(\cdot) = (\phi_1(\cdot),\dots,\phi_k(\cdot))'$. Then a series estimator of T is given by

$$\widehat{T}g := \boldsymbol{\phi}'\widehat{M}\langle g,\boldsymbol{\phi}\rangle, \tag{4.4}$$

where $\widehat{M} := n^{-1}\sum_{i=1}^n \boldsymbol{\phi}(W_i)\boldsymbol{\phi}(Z_i)'$ and $\langle g,\boldsymbol{\phi}\rangle$ denotes the column vector $(\langle g,\phi_1\rangle,\dots,\langle g,\phi_k\rangle)'$. The estimator of $T^\star$ is the dual of $\widehat{T}$, that is $\widehat{T}^\star h := \boldsymbol{\phi}'\widehat{M}'\langle h,\boldsymbol{\phi}\rangle$ where analogously $\langle h,\boldsymbol{\phi}\rangle := (\langle h,\phi_1\rangle,\dots,\langle h,\phi_k\rangle)'$. We also define the vector $\hat{v} := \frac{1}{n}\sum_{i=1}^n Y_i\boldsymbol{\phi}(W_i)$ such that the series estimator (4.3) of r can be written $\widehat{r} = \boldsymbol{\psi}'\widehat{v}$.

Suppose the prior on φ is described by using the Sobolev spaces $(\mathcal{W}_q[0,1])_q$. For $s \geqslant 0$, consider the Tikhonov regularization scheme of order ℓ (cf. Example 4.1). If we define the diagonal matrix $\nabla^s := \mathrm{Diag}[(\iota)^s,(2\iota)^s\dots,(k\iota)^s]$, the order-$\ell$-iterated-Tikhonov regularized estimator, $\widehat{\varphi}^{(s)} := \boldsymbol{\phi}'\widehat{\varphi}_\ell^{(s)}$, is computed by solving the ℓ linear equations

$$(\widehat{M}'\widehat{M} + \alpha\nabla^s)\widehat{\varphi}_j^{(s)} = \widehat{M}'\widehat{v} + \alpha\nabla^s\widehat{\varphi}_{j-1}^{(s)}, \quad j = 1,\dots,\ell, \quad \widehat{\varphi}_0^{(s)} = 0. \tag{4.5}$$

Proposition 4.3. *Assume $\varphi \in \mathcal{W}_p([0,1])$ for some $p > 0$ and that the link condition (Assumption (2.3)) holds with the scale of Sobolev spaces in $L^2[0,1]$. Consider the estimator (4.5) of the s-th derivative of φ for $0 \leqslant s < p$.*

(i) Suppose the operator T is finitely smoothing, that is $\kappa(t) = t^{(p-s)/2(a+s)}$, and suppose $\hat{T}$ is such that $\mathbb{E}\|\hat{T} - T\|^{2(1\vee(p-s)/(a+s))} = O(n^{-2\tau/(2\tau+1)})$ for some $\tau > (p+a)$. If $\ell \geq (p-s)/(a+s)$, then

$$\mathbb{E}\|\hat{\varphi}^s - \varphi^s\|^2 = O(n^{-2(p-s)/(2(a+p)+1)}).$$

(ii) Suppose the operator T is infinitely smoothing, that is $\kappa(t) = |\log(t)|^{-(p-s)/2a}$. Then

$$\mathbb{E}\|\hat{\varphi}^{(s)} - \varphi^{(s)}\|^2 = O((\log n)^{-(p-s)/a}).$$

Proof. The proof is a straightforward application of Theorem 4.1. When T is finitely smoothing, we use $\mathbb{E}\,\|\hat{r} - r\|^2 = O(n^{-2(p+a)/(2(p+a)+1)})$ and $\mathbb{E}\|\hat{T} - T\|^{2[1\vee(p-s)/(a+s)]} = O(n^{-2\tau/(2\tau+1)})$ for some $\tau > 0$; see Hall and Horowitz (2005). The result follows using $\tau > p+a$. When T is infinitely smoothing, one can show that $\mathbb{E}\|\hat{T}-T\|^4 = O(n^{-2\tau'/(2\tau'+1)})$ and $\mathbb{E}\|\hat{r} - r\|^2 = O(n^{-2\tau/(2\tau+1)})$ for some $\tau > 0$ and $\tau' > 0$. $\square$

The rates given in Proposition 4.3 are new in the context of nonparametric instrumental regression. Case (i) covers Theorem 4.1 of Hall and Horowitz (2005) as a particular case (with $p = \beta - 1/2$ and $\alpha = 2a$ in their notations). Incidentally, observe that it relaxes some of the Hall and Horowitz constraints on the regularity parameters. In case $s = 0$, Theorem 5.3 of Chen and Reiss (2007) derives similar upper bounds of a sieve least square estimator. In addition, Proposition 4.3 establishes the rate of convergence of the derivatives ($s \geqslant 0$).

As a final remark, we recall that the non iterated Tikhonov regularization is known to lead to suboptimal rates of convergence when φ is too regular (it is so-called "saturation effect" of that regularization scheme, see e.g. Florens, Johannes, and Van Bellegem (2009)). As it is highlighted by the last proposition, iterative Tikhonov regularization does not have this limitation. Moreover, when T is infinitely smoothing, there is no constraint on the number of iterations in order to derive the rate of convergence. According to Proposition 4.3, a single iteration is sufficient in order to reach the logarithmic risk.

A Technical appendix

A.1 Proof of Theorem 3.2

We present a proof in the general case where the eigenfunctions are not necessarily the Fourier transform, see Remark 3.1 for notations. In the setting of the Theorem, we consider $U = V = \mathcal{F}$.

The general estimator writes

$$U\widehat{\varphi}_s := \frac{\beta^{s/2}}{\widehat{\lambda}}\,\widehat{Vr}\,\mathbb{1}\{\beta^{s-p}/d \geqslant \kappa^2(\delta^*)\}\,\mathbb{1}\{\widehat{\lambda}^2/\beta^s \geqslant \alpha\}. \tag{A.1}$$

and we work under the assumption that $\widehat{Vr}$ and $\widehat{\lambda}$ are estimators of Vr and λ respectively such that the inequalities

$$\mathbb{E}|\widehat{Vr} - Vr|^4 \leqslant c_1\eta^2 \tag{A.2}$$

$$\mathbb{E}|\widehat{\lambda} - \lambda|^4 \leqslant c_2\tau^2 \tag{A.3}$$

holds true uniformly over $\mu\,a.e.$ for some $\eta, \tau > 0$ and $c_1, c_2 \geqslant 1$.

In the following, we will use the notations

$$\lambda_s := \beta^{-s/2}\lambda, \quad \widehat{\lambda}_s := \beta^{-s/2}\widehat{\lambda}, \quad \text{and} \quad \widehat{\varphi}_s^{\alpha} := U^{-1}\mathbb{1}\{\beta^{s-p} \geqslant d\,\kappa^2(\delta^*)\}\,\mathbb{1}\{\widehat{\lambda}_s^2 \geqslant \alpha\}\,U\varphi_s. \tag{A.4}$$

We also need to introduce the function $\triangle(\tau)$ which is defined for every $\tau > 0$ by

$$\triangle(\tau) := \sup_{t\in\Omega}\left\{\beta^{s-p}(t)\cdot\min\Bigl(\frac{\tau}{\Phi(\beta^{s-p}(t)/d)}, 1\Bigr)\right\}. \tag{A.5}$$

where Φ denotes the inverse function of κ^2.

Lemma A.1. *There exists a constant $C \geqslant 1$ depending only on κ such that $\triangle(\tau)$ for all $\tau \in (0, \Phi(1/d)]$ satisfies*

$$1/(C\,d) \leqslant \max(\kappa^2(\tau), \tau)/\triangle(\tau) \leqslant C\,d.$$

Proof. Consider the upper bound. Taking t_0 such that $\beta^{s-p}(t_0) = d\kappa^2(\tau)$ implies $\triangle(\tau) \geqslant \beta^{s-p}(t_0) = d\kappa^2(\tau)$. On the other hand since $\beta(t) \geqslant 1$ for all $t \in \Omega$, we have

$$\triangle(\tau) = d\,\tau\sup_{u\in(0,1]}\left\{u\cdot\min\Bigl(\frac{1}{\Phi(u)}, \frac{1}{\tau}\Bigr)\right\} \geqslant d\,\tau\,\frac{1}{\Phi(1)}$$

which proves the upper bound. Consider now the lower bound. We distinguish between two cases: there exists a constant $C \geqslant 1$ such that for all $\tau \in (0, \Phi(1)]$, either (i) $\sup_{t\in[\tau,\Phi(1)]}\kappa^2(t)/t \leqslant C$ or

(ii) $\sup_{t\in[\tau,\Phi(1)]}\kappa^2(t)/t \leqslant C\kappa^2(\tau)/\tau$. Note, if $\kappa^2(t)/t = O(1)$ as $t \to 0$ then we are in case (i) and otherwise in (ii). Then due to the monotonicity of κ we have

$$\triangle(\tau) = d\,\tau \max\left(\sup_{u\in(0,\tau]}\frac{\kappa^2(u)}{\tau}, \sup_{u\in[\tau,\Phi(1)])}\frac{\kappa^2(u)}{u}\right) = d\,\tau \sup_{u\in[\tau,\Phi(1)])}\frac{\kappa^2(u)}{u}.$$

Thereby, in case (i) we have $\triangle(\tau) \leqslant C\tau$, while in case (ii) it follows $\triangle(\tau) \leqslant C\kappa^2(\tau)$, which proves the lower bound and hence completes the proof. □

Lemma A.2. *Suppose that $\varphi \in H_p$ for some $p > 0$ and that for some $0 \leqslant s < p$, there exists an index function κ satisfying the link condition (2.3) for some constant $d \geq 1$. Consider $\widehat{\varphi}_s^\alpha$ given in (A.4), where the estimator $\widehat{\lambda}$ of λ satisfies $\mathbb{E}|\widehat{\lambda} - \lambda|^2 \leqslant c\tau$, for some $c > 0$. If in addition the threshold δ^* is defined by (3.4) and $\alpha := \max(\tau, \delta^*/4)$, then there exists a constant $C \geqslant 1$ only depending on d and c such that*

$$\mathbb{E}\,\|\widehat{\varphi}_s^\alpha - \varphi_s\|^2 \leqslant C\cdot\{\kappa(\delta^*)^2 + \triangle(\tau)\}\cdot\|\varphi\|_p^2\,\cdot \tag{A.6}$$

Proof. Let $\varphi_s^\alpha := U^{-1}\mathbb{1}\{\beta^{s-p} \geqslant d\,\kappa(\delta^*)^2\}U\varphi_s$. Then the proof is based on the decomposition $\mathbb{E}\|\widehat{\varphi}_s^\alpha - \varphi_s\|^2 \leqslant 2\|\varphi_s^\alpha - \varphi_s\|^2 + 2\mathbb{E}\|\widehat{\varphi}_s^\alpha - \varphi_s^\alpha\|^2$. We show below the following bound:

$$\mathbb{E}\|\widehat{\varphi}_s^\alpha - \varphi_s^\alpha\|^2 \leqslant C\triangle(\tau)\|\varphi\|_p^2. \tag{A.7}$$

Combining (A.7) with the estimate $\|\varphi_s^\alpha - \varphi_s\|^2 \leqslant d\,\kappa^2(\delta^*)\,\|\varphi\|_p^2$ obtained in the proof of Proposition 3.1 we obtain the result.

Proof of (A.7). Consider the identity

$$\mathbb{E}\,\|\widehat{\varphi}_s^\alpha - \varphi_s^\alpha\|^2 = \int_\Omega |U\varphi_s|^2\mathbb{1}\{\beta^{s-p} \geqslant d\,\kappa(\delta^*)^2\}P(\widehat{\lambda}_s^2 < \alpha)d\mu.$$

Below we show below that for some positive constant C only depending on d and c it holds

$$\mathbb{1}\{\beta^{s-p} \geqslant d\kappa^2(\delta^*)\}P(\widehat{\lambda}_s^2 < \alpha) \leqslant C\cdot\min(\tau/\Phi(b^{s-p}/d), 1). \tag{A.8}$$

Therefore, we can write $\mathbb{E}\|\widehat{\varphi}_s^\alpha - \varphi_s^\alpha\|^2 \leqslant C\cdot\sup_\Omega\{\beta^{s-p}\min(\tau/\Phi(\beta^{s-p}/d), 1)\}\cdot\int_\Omega \beta^{p-s}|U\varphi_s|^2 d\mu$ which implies (A.7) by definition of $\triangle(\tau)$.

Proof of (A.8). We consider two cases.

(i) Let $\alpha = \delta^*/4$. The condition $\beta^{s-p} \geqslant d\kappa^2(\delta^*)$ together with 3.1 implies $\lambda_s^2 \geqslant \delta^* = 4\alpha$ and hence $P(\widehat{\lambda}_s^2 < \alpha) \leqslant P(2|\widehat{\lambda}_s - \lambda_s| \geqslant \lambda_s)$. The Markov inequality implies $P(\widehat{\lambda}_s^2 < \alpha) \leqslant c\,\tau/\Phi(\beta^{s-p}/d)$ under the assumptions of the Lemma on $\hat{\lambda}$. The last bound together with $P(\widehat{\lambda}_s^2 < \alpha) \leqslant 1$ shows (A.8).

(ii) Let $\alpha = \tau$. If $\lambda_s < 4\tau$, then due to the link condition we have $1 \leqslant 4\tau/\Phi(\beta^{s-p}/d)$, which implies (A.8). On the other hand if $\lambda_s \geqslant 4\tau$, then $P(\widehat{\lambda}_s^2 < \alpha) \leqslant P(2|\widehat{\lambda}_s - \lambda_s| \geqslant \lambda_s)$ and (A.8) follows as in case (i), which completes the proof. □

Proof of Theorem 3.2. In Lemma A.2, we show the bound $\mathbb{E}\|\hat{\varphi}_s^\alpha - \varphi_s\|^2 \leqslant C\cdot\{\kappa(\delta^*)^2 + \triangle(\tau)\}\cdot\|\varphi\|_p^2$, for some positive constant C depending on c only. Below we show that it implies

$$\mathbb{E}\|\widehat{\varphi}_s - \widehat{\varphi}_s^\alpha\|^2 \leqslant 2\,c\,(1 + \tau/\alpha)\,D\,\{\kappa(\delta^*)^2 + \triangle(\tau)\,\|\varphi\|_p^2\}, \tag{A.9}$$

The result follows from the conditions imposed on α (i.e., $\alpha \geqslant \tau$), together with the inequality $\triangle(\tau) \leqslant C\,d\,\max(\kappa^2(\tau), \tau)$ (Lemma (A.1)).

Proof of (A.9). The term $\mathbb{E}\,\|\widehat{\varphi}_s - \widehat{\varphi}_s^\alpha\|^2$ is bounded by a constant times the sum $I + II$ with

$$I = \mathbb{E}\left\| |\widehat{\lambda}_s|^{-1} |\widehat{Vr} - Vr| \mathbb{1}\{\beta^{s-p} \geqslant d\kappa^2(\delta^*)\} \mathbb{1}\{\widehat{\lambda}_s^2 \geqslant \alpha\} \right\|_{L^2_\mu(\Omega)}^2 ,$$
$$II = \mathbb{E}\left\| |\widehat{\lambda}_s|^{-1} |\widehat{\lambda}_s - \lambda_s| |U\varphi_s| \mathbb{1}\{\beta^{s-p} \geqslant d\kappa^2(\delta^*)\} \mathbb{1}\{\widehat{\lambda}_s^2 \geqslant \alpha\} \right\|_{L^2_\mu(\Omega)}^2 .$$

Consider term I. From the inequality $\widehat{\lambda}_s^2/\lambda_s^2 + |\widehat{\lambda}_s/\lambda_s - 1|^2 \geqslant 1/2$ and under the assumptions of the theorem, we use the Cauchy Schwarz inequality and write

$$I \leqslant 2\,c\,(1+\tau/\alpha)\,\eta\,\left\|\lambda_s^{-1}\mathbb{1}\{\beta^{s-p} \geqslant d\kappa^2(\delta^*)\}\right\|_{L^2_\mu(\Omega)}^2 .$$

Using that $\Phi(\beta^{s-p}/d) \leqslant \lambda_s^2$ (see (3.1)) together with the definition of δ^* given in (3.4), i.e., $\eta\,\|\mathbb{1}\{\beta^{s-p} \geqslant d\kappa^2(\delta^*)\}|\Phi(\beta^{s-p}/d)|^{-1/2}\|_{L^2_\mu(\Omega)}^2 \leqslant D\,\kappa^2(\delta^*)$, we have

$$I \leqslant 2\,c\,(1+\tau/\alpha)\,\eta\,\|\mathbb{1}\{\beta^{s-p} \geqslant d\kappa^2(\delta^*)\}\,|\Phi(\beta^{s-p}/d)|^{-1/2}\|_{L^2_\mu(\Omega)}^2 \leqslant 2\,c\,(1+\tau/\alpha)\,D\,\kappa^2(\delta^*).$$

Term II is handled analogously:

$$II \leqslant 2\int_\Omega |U\varphi_s|^2 \min\Big(\frac{\mathbb{E}|\widehat{\lambda}_s - \lambda_s|^2}{\lambda_s^2} + \frac{\mathbb{E}|\widehat{\lambda}_s - \lambda_s|^4}{\alpha\lambda_s^2}, \frac{\mathbb{E}|\widehat{\lambda}_s - \lambda_s|^2}{\alpha}\Big) d\mu$$

and, under the assumptions of the theorem, we obtain

$$II \leqslant 2\,c\,(1+\tau/\alpha)\,\int_\Omega |U\varphi_s|^2 \min\Big(\frac{\tau}{\Phi(\beta^{s-p}/d)}, 1\Big) d\mu \leqslant 2\,c\,(1+\tau/\alpha)\,\triangle(\tau)\,\|\varphi\|_p^2 .$$

The bound (A.9) follows from the above controls of terms I and II, which completes the proof. □

A.2 Proof of Theorems 4.1 and 4.2

We first recall the notations

$$T_s := TB^{-s/2}, \quad \widehat{T}_s := \widehat{T}B^{-s/2}, \quad \varphi_s := B^{s/2}\varphi \quad \text{and} \quad \widehat{\varphi}_s^\alpha := g_\alpha(\widehat{T}_s^\star\widehat{T}_s)\widehat{T}_s^\star\widehat{T}_s\varphi_s. \tag{A.10}$$

Lemma A.3. *Suppose the assumptions of Theorem 4.1 are satisfied. Then*

$$\mathbb{E}\|\widehat{\varphi}_s^\alpha - \varphi_s\|^2 \leqslant C\,\rho^2\,\kappa\,\Big(C'\{\alpha + (\mathbb{E}\|T - \widehat{T}\|^4)^{1/2}\}\Big) \tag{A.11}$$

where C and C' are positive constants depending only on κ.

Proof. The proof technique is partly inspired from Nair, Pereverzev, and Tautenhahn (2005). Since T satisfies Assumption 2.3 for some index function κ and $\varphi \in H_p$, we conclude $\rho := \left\|\kappa(T_s^\star T_s)^{-1}\varphi_s\right\| < \infty$. Let $\widehat{\psi}_\alpha := \varphi_s - \widehat{\varphi}_s^\alpha$ and $\widehat{R}_\alpha := [I - g_\alpha(\widehat{T}_s^\star\widehat{T}_s)\widehat{T}_s^\star\widehat{T}_s]$, then we have $\widehat{\psi}_\alpha = \widehat{R}_\alpha\varphi_s$. We use $\|\widehat{R}_\alpha^{1/2}\| \leqslant 1$ (Assumption 4.1 (ii)) and obtain due to the Cauchy-Schwarz inequality

$$\begin{aligned}
\|\widehat{\psi}_\alpha\|^2 &= \|\widehat{R}_\alpha\varphi_s\|^2 \\
&\leqslant \|\widehat{R}_\alpha^{1/2}\varphi_s\|^2 = \Big\langle \widehat{R}_\alpha\varphi_s, \varphi_s \Big\rangle = \Big\langle \kappa(T_s^\star T_s)\widehat{\psi}_\alpha, \kappa(T_s^\star T_s)^{-1}\varphi_s \Big\rangle \\
&\leqslant \|\kappa(T_s^\star T_s)\widehat{\psi}_\alpha\|\,\rho.
\end{aligned} \tag{A.12}$$

Thereby we can write

$$\mathbb{E}\|\widehat{\psi}_\alpha\|^2 \leqslant \rho(\mathbb{E}\|\kappa(T_s^\star T_s)\widehat{\psi}_\alpha\|^2)^{1/2} \tag{A.13}$$

which, together with $d\,\alpha \geq \|(\widehat{T}_s^\star \widehat{T}_s)^{1/2}\widehat{R}_\alpha^{1/2}\|^2$ (Assumption 4.1 (ii)), gives

$$\|\widehat{T}_s\widehat{\psi}_\alpha\|^2 = \|\widehat{T}_s\widehat{R}_\alpha\varphi_s\|^2 = \|(\widehat{T}_s^\star\widehat{T}_s)^{1/2}\widehat{R}_\alpha\varphi_s\|^2 \leqslant \alpha d\|\widehat{R}_\alpha^{1/2}\varphi_s\|^2 \leqslant \alpha d\rho\|\kappa(T_s^\star T_s)\widehat{\psi}_\alpha\|$$

and hence,

$$\mathbb{E}\|\widehat{T}\widehat{\psi}_\alpha\|^2 \leqslant \alpha d\rho(\mathbb{E}\|\kappa(T_s^\star T_s)B^{s/2}\widehat{\psi}_\alpha\|^2)^{1/2}. \tag{A.14}$$

Using (A.12) together with $\|B^{-s/2}\| \leqslant c$ and the Cauchy Schwarz inequality:

$$\mathbb{E}\|(T_s-\widehat{T}_s)\widehat{\psi}_\alpha\|^2 \leqslant \mathbb{E}\|(T_s-\widehat{T}_s)\|^2\|\widehat{\psi}_\alpha\|^2 \leqslant c\rho(\mathbb{E}\|T-\widehat{T}\|^4)^{1/2}(\mathbb{E}\|\kappa(T_s^\star T_s)\widehat{\psi}_\alpha\|^2)^{1/2} \tag{A.15}$$

Combining (A.14) and (A.15) we obtain

$$\begin{aligned}\mathbb{E}\|T_s\widehat{\psi}_\alpha\|^2 &\leqslant \mathbb{E}\|(T_s-\widehat{T}_s)\widehat{\psi}_\alpha\|^2 + \mathbb{E}\|\widehat{T}_s\widehat{\psi}_\alpha\|\\ &\leqslant c\rho\Big\{(\mathbb{E}\|T-\widehat{T}\|^4)^{1/2}+\alpha\Big\}(\mathbb{E}\|\kappa(T_s^\star T_s)\widehat{\psi}_\alpha\|^2)^{1/2}.\end{aligned} \tag{A.16}$$

Let Φ be the inverse function of κ^2, which is assumed to be convex on the interval $(0,c^2]$. Define $d^2 = c^2/\|\kappa(T_s^\star T_s)\|^2 \wedge 1$. If $\{\lambda_s^2, U_s : H \to L^2_{\mu_s}(\Omega_s)\}$ denotes the spectral decomposition of $T_s^\star T_s$, then $c^2 \geqslant d^2\kappa(\lambda_s^2)^2$. Hence, using Jensen's inequality we have

$$\Phi\Big(\frac{d^2\mathbb{E}\|\kappa(T_s^\star T_s)\widehat{\psi}_\alpha\|^2}{\mathbb{E}\|\widehat{\psi}_\alpha\|^2}\Big) \leqslant \frac{\mathbb{E}\int_{\Omega_s}\Phi(d^2\kappa(\lambda_s^2(\omega))^2)|U_s\widehat{\psi}_\alpha|^2(\omega)\mu_s(d\omega)}{\mathbb{E}\int_{\Omega_s}|U_s\widehat{\psi}_\alpha|^2(\omega)\mu_s(d\omega)}.$$

Since $\Phi(d^2\kappa(\lambda_s^2)^2) \leqslant \Phi(\kappa(\lambda_s^2)^2) = \lambda_s^2$ we conclude

$$\Phi\Big(\frac{d^2\mathbb{E}\|\kappa(T_s^\star T_s)\widehat{\psi}_\alpha\|^2}{\mathbb{E}\|\widehat{\psi}_\alpha\|^2}\Big) \leqslant \frac{\mathbb{E}\|(T_s^\star T_s)^{1/2}\widehat{\psi}_\alpha\|^2}{\mathbb{E}\|\widehat{\psi}_\alpha\|^2} = \frac{\mathbb{E}\|T_s\widehat{\psi}_\alpha\|^2}{\mathbb{E}\|\widehat{\psi}_\alpha\|^2}. \tag{A.17}$$

In order to combine the three estimates (A.13), (A.16) and (A.17) let us introduce a new function Ψ by $\Psi(\omega) := \Phi(\omega^2)/\omega^2$. Since Φ is convex, we conclude that Ψ is monotonically increasing on $(0,c]$. By (A.13), that is, $(\mathbb{E}\|\kappa(T_s^\star T_s)\widehat{\psi}_\alpha\|^2)^{1/4}/\rho^{1/2} \leqslant (\mathbb{E}\|\kappa(T_s^\star T_s)\widehat{\psi}_\alpha\|^2)^{1/2}/(\mathbb{E}\|\widehat{\psi}_\alpha\|^2)^{1/2}$, the monotonicity of Ψ and (A.17) follows

$$\Psi\Big(\frac{d\cdot(\mathbb{E}\|\kappa(T_s^\star T_s)\widehat{\psi}_\alpha\|^2)^{1/4}}{\rho^{1/2}}\Big) \leqslant \Psi\Big(\frac{d\cdot(\mathbb{E}\|\kappa(T_s^\star T_s)\widehat{\psi}_\alpha\|^2)^{1/2}}{(\mathbb{E}\|\hat{\psi}_\alpha\|^2)^{1/2}}\Big) \leqslant \frac{\mathbb{E}\|T_s\widehat{\psi}_\alpha\|^2}{d^2\ \mathbb{E}\|\kappa(T_s^\star T_s)\widehat{\psi}_\alpha\|^2}.$$

Multiplying by $d^2\ (\mathbb{E}\|\kappa(T_s^\star T_s)\widehat{\psi}_\alpha\|^2)^{1/2}/\rho$ and exploiting (A.16) yields

$$\Phi\Big(\frac{d^2\cdot(\mathbb{E}\|\kappa(T_s^\star T_s)\widehat{\psi}_\alpha\|^2)^{1/2}}{\rho}\Big) \leqslant c\cdot\{\alpha+(\mathbb{E}\|T-\widehat{T}\|^4)^{1/2}\}. \tag{A.18}$$

Thereby the result follows by combining (A.13) and (A.18), which completes the proof. □

Proof of Theorem 4.1. Since T satisfies (2.3) for some index function κ and $\varphi \in H_p$, $p>0$, it follows that $\rho := \|\kappa(T_s^\star T_s)^{-1}\varphi_s\| < \infty$. The proof is based on the decomposition

$$\mathbb{E}\left\|\widehat{\varphi}_s-\varphi_s\right\|^2 \leqslant 2\mathbb{E}\left\|\widehat{\varphi}_s-\widehat{\varphi}_s^\alpha\right\|^2 + 2\mathbb{E}\left\|\widehat{\varphi}_s^\alpha-\varphi_s\right\|^2. \tag{A.19}$$

By definition, together with $\|g_\alpha(\widehat{T}_s^\star\widehat{T}_s)\widehat{T}_s^\star\|^2 \leqslant c/\alpha$ (Assumption 4.1 (i)) we have

$$\mathbb{E}\|\widehat{\varphi}_s-\widehat{\varphi}_s^\alpha\|^2 = \mathbb{E}\|g_\alpha(\widehat{T}_s^\star\widehat{T}_s)\widehat{T}_s^\star(\widehat{r}-\widehat{T}_s\varphi_s)\|^2 \leqslant c\cdot\alpha^{-1}\mathbb{E}\|\widehat{r}-\widehat{T}\varphi\|^2, \tag{A.20}$$

while from Lemma A.3 we obtain

$$\mathbb{E}\|\widehat{\varphi}_s^\alpha-\varphi_s\|^2 \leqslant C\ \rho^2\ \kappa\left(C'[\alpha+(\mathbb{E}\|T-\widehat{T}\|^4)^{1/2}]\right)^2. \tag{A.21}$$

The result follows by combining (A.20) and (A.21) through (A.19) which completes the proof. $\square$

Proof of Theorem 4.2. Since T satisfies (2.3) for some index function κ and $\varphi \in H_p$, $p > 0$, it follows that $\rho := \|\kappa(T_s^\star T_s)^{-1}\varphi_s\| < \infty$. Considering the decomposition (A.19) we bound the first term as in the proof of Theorem 4.1, that is $\mathbb{E}\|\hat{\varphi}_s - \hat{\varphi}_s^\alpha\|^2 \leqslant C\alpha^{-1}\mathbb{E}\|\widehat{r} - \widehat{T}\varphi\|^2$, while we show below that under the assumptions of the theorem the following bound of the second term holds

$$\mathbb{E}\|\hat{\varphi}_s^\alpha - \varphi\|_s^2 \leqslant C\,\rho^2\,\left[\kappa(\alpha)^2 + \mathbb{E}\|\kappa(\widehat{T}_s^\star\widehat{T}_s) - \kappa(T_s^\star T_s)\|^2\right]. \tag{A.22}$$

Thereby, the assertion follows by combining the two bounds.

Proof of (A.22). Let $\widehat{\psi}_\alpha := \varphi_s - \widehat{\varphi}_s^\alpha$ and $\widehat{R}_\alpha := [I - g_\alpha(\widehat{T}_s^\star\widehat{T}_s)\widehat{T}_s^\star\widehat{T}_s]$, then we have

$$\|\widehat{\psi}_\alpha\|^2 = \|\widehat{R}_\alpha\varphi_s\|^2 \leqslant 2\|\widehat{R}_\alpha\kappa(\widehat{T}_s^\star\widehat{T}_s)\kappa(T_s^\star T_s)^{-1}\varphi_s\|^2 + 2\|\widehat{R}_\alpha[\kappa(\widehat{T}_s^\star\widehat{T}_s) - \kappa(T_s^\star T_s)]\kappa(T_s^\star T_s)^{-1}\varphi_s\|^2.$$

Thereby, since $\|\widehat{R}_\alpha\| \leqslant 1$ (Assumption 4.1) and $\|\widehat{R}_\alpha^{1/2}\kappa(\widehat{T}_s^\star\widehat{T}_s)\|^2 \leqslant c_\kappa\kappa(\alpha)^2$ (from (4.2)), it follows (A.22). Indeed, $\|\widehat{\psi}_\alpha\|^2 \leqslant C\rho^2\left(\kappa(\alpha)^2 + \|\kappa(\widehat{T}_s^\star\widehat{T}_s) - \kappa(T_s^\star T_s)\|^2\right)$ and this completes the proof. $\square$

References

Bigot, J., and S. Van Bellegem (2009): "Log-density deconvolution by wavelet thresholding," *Scandinavian Journal of Statistics*, 36, 749–763.

Bissantz, N., T. Hohage, A. Munk, and F. H. Ruymgaart (2007): "Convergence rates of general regularization methods for statistical inverse problems and applications," *SIAM Journal on Numerical Analysis*, 45(6), 2610–2636.

Blundell, R., X. Chen, and D. Kristensen (2007): "Semi-nonparametric IV estimation of shape-invariant Engel curves," *Econometrica*, 75, 1613–1669.

Bonhomme, S., and J.-M. Robin (2006): "Generalized nonparametric deconvolution with an application to earnings dynamics," Working Paper.

Carrasco, M., and J.-P. Florens (2002): "Spectral method for deconvolving a density," Working Paper 138, IDEI, Toulouse.

Carrasco, M., J.-P. Florens, and E. Renault (2007): "Linear inverse problems in structural econometrics: Estimation based on spectral decomposition and regularization," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6B, pp. 5633–5746. Elsevier.

Cavalier, L., and N. Hengartner (2005): "Adaptive estimation for inverse problems with noisy operators," *Inverse Problems*, 21, 1345–1361.

Chen, X., and M. Reiss (2007): "On rate optimality for ill-posed inverse problems in econometrics," Cowles foundation discussion paper No. 1626.

Darolles, S., J.-P. Florens, and E. Renault (2002): "Nonparametric Instrumental Regression," Working Paper # 228, IDEI, Université de Toulouse I.

Egger, H. (2005): "Accelerated Newton-Landweber iterations for regularizing nonlinear inverse problems," SFB-Report 2005-3, Austrian Academy of Sciences, Linz.

Fan, J. (1991): "On the optimal rates of convergence for nonparametric deconvolution problems," *Annals of Statistics*, 19, 1257–1272.

Florens, J.-P., J. Johannes, and S. Van Bellegem (2005): "Instrumental regression in partially linear models," Discussion Paper # 0537, Institut de statistique, Université catholique de Louvain.

——— (2009): "Identification and estimation by penalization in nonparametric instrumental regression," .

Hall, P., and J. L. Horowitz (2005): "Nonparametric methods for inference in the presence of instrumental variables," *Annals of Statistics*, 33, 2904–2929.

Halmos, P. R. (1963): "What does the spectral theorem say?," *Amer. Math. Monthly*, 70, 241–247.

Hohage, T. (2000): "Regularisation of exponentially ill-posed problems," *Numer. Funct. Anal. Optim.*, 21, 439–464.

Horowitz, J. L. (1998): *Semiparametric methods in econometrics.* Springer, New York.

Johannes, J. (2009): "Deconvolution with unknown error density," *The Annals of Statistics*, forthcoming.

Johannes, J., S. Van Bellegem, and A. Vanhems (2009): "Projection estimation in nonparametric instrumental regression," Discussion Paper.

Kawata, T. (1972): *Fourier Analysis in Probability Theory.* Academic Press, New York.

Krein, S., and Y. I. Petunin (1966): "Scales of Banach spaces," in *Russian Math. Surveys*, vol. 21, pp. 85–169.

Mair, B. A. (1994): "Tikhonov regularization for finitely and infinitely smoothing operators," *SIAM Journal on Mathematical Analysis*, 25, 135–147.

Mair, B. A., and F. H. Ruymgaart (1996): "Statistical inverse estimation in Hilbert scales," *SIAM Journal on Applied Mathematics*, 56(5), 1424–1444.

Nair, M., S. V. Pereverzev, and U. Tautenhahn (2005): "Regularization in Hilbert scales under gerneal smoothing conditions," *Inverse Problems*, 21, 1851–1869.

Natterer, F. (1984): "Error bounds for Tikhonov regularization in Hilbert scales," *Applicable Analysis*, 18, 29–37.

Neubauer, A. (1988): "When do Sobolev spaces form a Hilbert scale?," *Proc. Amer. Math. Soc.*, 103(2), 557–562.

Neumann, M. H. (1997): "On the effect of estimating the error density in nonparametric deconvolution," *Journal of Nonparametric Statistics*, 7, 307–330.

Newey, W. K., and J. L. Powell (2003): "Instrumental variable estimation of nonparametric models," *Econometrica*, 71, 1565–1578.

Olver, F. (1974): *Asymptotics and special functions.* Academic Press.

Petrov, V. V. (1995): *Limit theorems of probability theory. Sequences of independent random variables.*, Oxford Studies in Probability. Clarendon Press., Oxford, 4. edn.

Postel-Vinay, F., and J.-M. Robin (2002): "Equilibrium wage dispersion with worker and employer heterogeneity," *Econometrica*, 70, 2295–2350.

Tautenhahn, U. (1996): "Error estimates for regularization methods in Hilbert scales," *SIAM Journal on Numerical Analysis*, 33(6), 2120–2130.