Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Statistical Planning and Inference 136 (2006) 390–400

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Inverse problems with non-compact operators

L. Cavalier*

Université Aix-Marseille I, CMI, 39 rue Joliot-Curie, 1313453 Marseille Cedex, France

Received 2 July 2003; accepted 28 June 2004
Available online 15 September 2004

Abstract

In the recent studies of inverse problems with random noise, one of most common assumptions is that the linear operator A is compact. This hypothesis is natural and has fine statistical properties. However, dealing with compact operators is not necessary. By use of the Spectral Theorem we may extend results for compact operators to any linear continuous operator. Using the method of unbiased risk estimation, we prove some oracle inequality for this estimator. As examples of non-compact operators, the problem of convolution on $\mathbb{R}$ and the estimation of the derivative of some function are studied.

MSC: 62G05; 62G20

Keywords: Inverse problems; Oracle inequalities; Exact minimax constant

1. Introduction

Inverse problems appear in many fields of application, from geophysics to medical image processing. The aim is to reconstruct some unknown object (or function) based on noisy indirect observations. From a mathematical point of view this often corresponds to inverting some operator equation: given g, find f such that $g = Af$. Usually A is some linear operator $A: H \to H$, where H is some Hilbert space. The most interesting cases are ill-posed inverse problems (see Hadamard, 1932) where A is not invertible as an operator, i.e. may be injective but with a non-continuous inverse. These problems are ill-posed in the sense that a small error on g can imply a large one on the “inverse” f.

* Tel.: +4-91-11-3570; fax: +4-91-11-3552
E-mail address: Laurent.Cavalier@cmi.univ-mrs.fr (L. Cavalier).

doi:10.1016/j.jspi.2004.06.063

In this paper, we will consider inverse problems with random noise, i.e. that the error on the observation is random. The model is the following

$$Y = Af + \varepsilon\xi,$$

where ξ is a generalized Gaussian random H-valued noise, $0 < \varepsilon < 1$ is a small parameter (the noise level) and Y is the observation.

The study of inverse problems with random noise is a fastly growing field in statistics. One of the most standard (and sometimes implicit) hypothesis is that the operator A is compact. In this case one often uses Singular Value Decomposition (SVD) in order to project the observation Y on some basis appropriate for the operator. An equivalent model is then obtained in the form of a sequence space model in the coefficients space (see Cavalier et al., 2002 and Cavalier and Tsybakov, 2002). Related works appear in Efromovich (1997) and also in Donoho (1995) and Johnstone (1999) where the Wavelet-Vaguelette Decomposition (WVD), an analog of SVD based on wavelets, is used.

Clearly, dealing with compact operators is an interesting and natural assumption in statistics, in the sense that we decompose the observation Y in some basis, and then handle its coefficients. However, one may ask if this fine hypothesis of compactness is really needed, or is just more easy to deal with. The answer is that it is not necessary to use compact operators and that any continuous linear operator with a known spectral decomposition can be studied, in an analogous way to the compact operator framework. Instead of using the SVD we use the Spectral Theorem. The eigenvalues of the compact operator A are replaced by the continuous spectrum of the linear operator A.

This idea to deal with the Spectral Theorem has already been used in Caroll et al. (1991). This approach is also considered in several papers on statistical inverse problems (see Mair and Ruymgaart, 1996; van Rooij and Ruymgaart, 1996; van Rooij et al., 2000). In these papers, minimax rates of convergence are obtained on some classes of functions. Dey et al. (1996) a data-driven choice of bandwidth is given by use of cross-validation.

The formal model described in (1) is close from previous papers but not exactly the same. In particular, the effect of the Spectral Theorem on the white noise ξ was not really studied. Moreover, our problem is different. In this paper, the aim is rather to choose the best estimator among a given family. Thus, we obtain a completely different kind of result. We prove a precise oracle inequality for our estimator, i.e. that in some sense our choice mimics the best estimator in the family. A main difference with minimax results is that oracle inequalities are non-asymptotic in ε. This oracle approach is important in modern mathematical statistics and could be used in the framework of inverse problems with non-compact operators.

In Section 2 the model is given. The use of the Spectral Theorem, in the case of a self-adjoint operator A, is explained in order to obtain a problem analogous to the case of compact operators. As an example, the main oracle inequality from Cavalier et al. (2002) is proved. Section 3 generalizes the results to the case of non-self-adjoint operators. Section 4 consists in examples of some inverse problems with non-compact linear operators. The main point is to understand that this generalization to non-compact operators is not only a theoretical exercise, but that in fact it helps to really understand the problem. The examples deal with the operator of convolution on $\mathbb{R}$ and with the estimation of the derivative of some function. The difference from convolution on $\mathbb{R}$ with circular convolution, studied for example in

Efromovich (1997) and Cavalier and Tsybakov (2002), is that the operator is not any more compact and the Fourier basis are not the eigenfunctions. However, from a heuristic point of view, it is easy to understand that in this framework the Fourier transform will play the role of the Fourier basis. This idea is true and the Fourier coefficients which are the eigenvalues of the circular convolution are replaced here by the Fourier transform on $L^2(\mathbb{R})$ of the convolution kernel r. Thus, we generalize the results to this natural extension.

2. Model

Let H be a separable Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. Consider the operator equation $g = Af$ where A is a known linear continuous operator from H into $\mathrm{Range}(A) \subseteq H$. The inverse problem with random noise consists of statistical estimation of f from noisy observations of g.

The statistical model can be written in the form

$$Y = Af + \varepsilon \xi, \tag{1}$$

where Y is the observation and ξ is a generalized random H-valued noise, i.e. for any $h \in H$, $\langle \xi, h \rangle$ is a random variable, $0 < \varepsilon < 1$ is a small parameter (the noise level). Note that sometimes ξ does not belong to H.

In this paper ξ is Gaussian, and Eq. (1) is understood in the sense that for any $u \in H$, the random variable

$$\langle Y, u \rangle = \langle Af, u \rangle + \varepsilon \langle \xi, u \rangle,$$

is observable, where $\langle \xi, u \rangle$ is a Gaussian random variable on a probability space $(\Omega, \mathscr{A}, \mathbf{P})$, with mean 0 and variance $\|u\|^2$. We also assume that $\mathbf{E}(\langle \xi, u \rangle \langle \xi, u \rangle) = \langle u, v \rangle$, for any $u, v \in H$, where $\mathbf{E}$ is the expectation w.r.t. $\mathbf{P}$. This definition may correspond to the white noise (see Hida, 1980) or to the usual Gaussian white noise model in statistics where ξ plays the role of $\mathrm{d}W(t)$.

In Cavalier et al. (2002) and Cavalier and Tsybakov (2002), the case of compact operators was studied. Here we are going to consider general linear operators, and in particular non-compact ones. The results of the compact operators case will be shown to be valid.

Suppose that A is a Hermitian, i.e. self-adjoint and strictly positive. An operator is strictly positive if $\langle Af, f \rangle > 0$, for any $f \in H - \{0\}$, which is equivalent to A being injective, i.e. with $Ker(A) = \{0\}$.

The *Spectral Theorem* (more precisely a version which may be found in Halmos (1963)) implies that A is unitarily equivalent to a multiplication operator: there exists a σ-finite measure space $(S, \mathscr{F}, \mu)$, a real valued $b \in L^\infty(S, \mathscr{F}, \mu)$ with $b > 0$, and a unitary $U : H \to L^2(S, \mathscr{F}, \mu)$, such that

$$A = U^{-1} M_b U. \tag{2}$$

Recall that unitary means that U preserves the inner product

$$\langle f, h \rangle = \langle Uf, Uh \rangle, \quad \text{for all } f, h \in H \tag{3}$$

and the multiplication operator is such that

$$M_b\theta = b \cdot \theta, \quad b \in L^{\infty}(\mu), \theta \in L^2(\mu). \tag{4}$$

Remark. The framework of compact operators is a special case where $S = \mathbb{N}$, μ is the counting measure and $L^2(\mu) = \ell^2(\mathbb{N})$.

Using this Spectral Theorem, we can apply the unitary operator U to the model (1), and obtain

$$UY = U(U^{-1}M_bU)f + \varepsilon U\xi = bUf + \varepsilon U\xi.$$

Thus, we have the model in $L^2(\mu)$ equivalent to (1)

$$Z = b \cdot \theta + \varepsilon\eta, \tag{5}$$

where $Z = UY$, $\theta = Uf$ and $\eta = U\xi$.

Consider now the noise η.

Lemma 1. *Let $\eta = U\xi$ where U is a unitary operator. We have for any $\theta = Uf$ and $v = Uh$, where $f, h \in H$,*

$$\langle \eta, \theta \rangle \sim \mathcal{N}(0, \|\theta\|^2)$$

and

$$\mathbf{E}(\langle \eta, \theta \rangle \langle \eta, v \rangle) = \langle \theta, v \rangle.$$

Proof. For any $\theta = Uf$ with $f \in H$, we have

$$\langle \eta, \theta \rangle = \langle \xi, f \rangle \sim \mathcal{N}(0, \|f\|^2) = \mathcal{N}(0, \|\theta\|^2), \tag{6}$$

since U is a unitary operator. In the same way if $\theta = Uf$ and $v = Uh$, where $f, h \in H$, we have

$$\mathbf{E}(\langle \eta, \theta \rangle \langle \eta, v \rangle) = \mathbf{E}(\langle \xi, f \rangle \langle \xi, h \rangle) = \langle f, h \rangle = \langle \theta, v \rangle, \tag{7}$$

since U is unitary. Using (6) and (7), we obtain the lemma. □

Remark. One can see that η has the same properties than ξ, but on $L^2(\mu)$. In other words η is a *white noise* in $L^2(\mu)$.

In the same spirit as Cavalier et al. (2002) we want to consider a model equivalent to (5),

$$X = b^{-1}Z = \theta + \varepsilon\sigma\eta, \tag{8}$$

where $\sigma = b^{-1}$.

Define an estimator $\hat{\theta}$ of $\theta = Uf$ based on the data (8). Then $\hat{f} = U^{-1}\hat{\theta}$ is the associated estimator of *f*. One can define the mean integrated squared risk by

$$\mathscr{R}(\hat{f}, f) = \mathbf{E}_f \|\hat{f} - f\|^2 = \mathbf{E}_\theta \|\hat{\theta} - \theta\|^2 = \mathbf{E}_\theta \int_S |\hat{\theta} - \theta|^2 \,\mathrm{d}\mu,$$

where the notation $\|\cdot\|$ means the $L^2(\mu)$-norm or the norm on *H*.

Define now the family of linear estimators $\hat{\theta} = \lambda X$, where $\lambda \in L^2(\mu)$. The risk of a linear estimator is then

$$R_\varepsilon[\lambda, \theta] = \mathbf{E}_\theta \|\hat{\theta}(\lambda) - \theta\|^2 = \int_S |1 - \lambda|^2 |\theta|^2 \,\mathrm{d}\mu + \varepsilon^2 \int_S \sigma^2 |\lambda|^2 \,\mathrm{d}\mu.$$

Remark that all the functions and spaces that we consider are complex-valued. However, as very often in statistics, we will deal with the real-valued case, leading the natural generalization to the reader.

It is thus clear that the results for the case of compact operators can be generalized for non-compact Hermitian one. The difference is that the series with $\sum_{i=1}^{\infty}\cdot$ are replaced by $\int_S \,\mathrm{d}\mu$.

All the results and the proofs may be rewritten with this remark.

Using the notations of Cavalier et al. (2002), consider the problem of selecting the "best" estimator from a given finite family of linear estimators $\Lambda = \{\lambda^1, \ldots, \lambda^N\}$, $\lambda^i \in L^2(\mu)$ for all $i = 1, \ldots, N$, where *N* is an integer.

Define $U[\lambda, X]$ as an unbiased risk estimator,

$$U[\lambda, X] = \int_S (\lambda^2 - 2\lambda)^2 |X|^2 \,\mathrm{d}\mu + 2\varepsilon^2 \int_S \sigma^2 \lambda \,\mathrm{d}\mu$$

and λ^* the estimator which minimizes this criterion on the set Λ

$$\lambda^* = \arg\min_{\lambda \in \Lambda} U[\lambda, X]. \tag{9}$$

Use the natural generalization (with $\int_S \,\mathrm{d}\mu$ instead of $\sum$) of the technical Assumptions 1 and 2 and notations in Cavalier et al. (2002).

For any $\lambda \in \Lambda$

$$0 < \int_S \sigma^2 \lambda^2 \,\mathrm{d}\mu < \infty \quad \text{and} \quad \max_{\lambda \in \Lambda} \|\lambda\|_\infty \leqslant 1, \tag{10}$$

$$\|\sigma^2 \lambda^2\|_\infty < \infty, \quad \|\sigma^2 \lambda\|_\infty < \infty. \tag{11}$$

There exists a constant $C_1 > 0$ such that, uniformly in $\lambda \in \Lambda$,

$$\int_S \sigma^4 \lambda^2 \,\mathrm{d}\mu \leqslant C_1 \int_S \sigma^4 \lambda^4 \,\mathrm{d}\mu. \tag{12}$$

Denote

$$\rho(\lambda) = \|\sigma^2\lambda\|_\infty \left\{ \int_S \sigma^4\lambda^4 \, \mathrm{d}\mu \right\}^{-1/2}, \quad \rho = \max_{\lambda\in\Lambda} \rho(\lambda),$$

$$T = \max_{\lambda\in\Lambda} \|\sigma^2\lambda^2\|_\infty / \min_{\lambda\in\Lambda} \|\sigma^2\lambda^2\|_\infty,$$

$$M = \sum_{\lambda\in\Lambda} \exp\{-1/\rho(\lambda)\}$$

and

$$L_\Lambda = \log(NT) + \rho^2 \log^2(MT).$$

Now, we obtain an oracle inequality for the estimator θ^*, i.e. we prove that θ^* (almost) mimics the best possible choice in Λ.

Theorem 1. *Let assumptions* (10)–(12) *on the family Λ hold. Then for every $\theta \in L^2(\mu)$, every $B > B_0$ and for the estimator $\theta^* = \lambda^* X$, where X is the data in* (8), *λ^* is defined by* (9), *and $f^* = U^{-1}\theta^*$, we have*

$$\mathbf{E}_f \|f^* - f\|^2 = \mathbf{E}_\theta \|\theta^* - \theta\|^2 \leqslant (1 + \gamma_1 B^{-1}) \min_{\lambda\in\Lambda} R_\varepsilon[\lambda, \theta] + \gamma_2 B \varepsilon^2 L_\Lambda \omega(B^2 L_\Lambda), \tag{13}$$

where

$$\omega(x) = \max_{\lambda\in\Lambda} \sup_{s\in S} \sigma(s)^2\lambda(s)^2 I\left\{ \int_S \sigma^2\lambda^2 \, \mathrm{d}\mu \leqslant x \sup_{s\in S} \sigma(s)^2\lambda(s)^2 \right\}, \quad x > 0$$

and $B_0 > 0$, $\gamma_1 > 0$, $\gamma_2 > 0$ are constants depending only on C_1.

Proof. Using Lemma 1 we have that η is the equivalent of ξ but on $L^2(\mu)$.

Lemma 1 in Cavalier et al. (2002) remains true since here $\mathbf{E}|\sum_{i=1}^\infty v_i \xi_i|$ is replaced by

$$\mathbf{E}|\langle v, \eta\rangle| = \mathbf{E}\left| \int_S v\eta \, \mathrm{d}\mu \right|$$

and the same for Lemma 2. Thus, the whole proof can be written in this way. □

Remark. The results of Cavalier et al. (2002) can be extended to the non-compact case. In the same way, the main remarks of Cavalier et al. (2002) can be made. However, since the model is very general, it is necessary to impose some assumptions before any examples or comments.

We only have $b \in L_\infty(\mu)$, and then the inverse problem will be said to be ill-posed if $b \to 0$ in some sense. In particular, the results have a statistical meaning only for the case of mildly ill-posed inverse problems. If S has a norm, this definition corresponds to $\sigma(s) \sim |s|^\beta$, $\beta > 0$. The case $\beta = 0$ corresponds to the direct model.

The examples of classes of estimators given in Cavalier et al. (2002) are also valid here if S has a norm.

Example 1. *Projection estimators*. In this setting, consider the filters $\lambda(s) = I(|s| \leqslant w)$, where $s \in S,\ w > 0$ and $I(\cdot)$ is the indicator function. In the case of Fourier transform (see Section 4), this family corresponds to band-limited filters.

Example 2. *Tikhonov–Phillips estimators*. Consider the filters $\lambda(s) = (1 + (|s|/w)^{\alpha})^{-1}$, where $s \in S$ and $w > 0$, which corresponds exactly to the Tikhonov–Phillips weight for the non-compact operator case (see Engl et al., 1996).

Theorem 1 is just an example but most of the works on inverse problems using compact operators could be generalized. However, some results may use specific properties of singular values (for example that they are decreasing), and one has to be careful.

The article Cavalier and Tsybakov (2002) could be extended to non-compact operators but since it is based on estimation using block of coefficients, it uses some order on the coefficients numbers. Thus, there must exist some order on S, if we want to generalize the results.

3. Non-self-adjoint operator A

The assumption that A is self-adjoint, can be deleted, and indeed one may consider operators from some Hilbert space H to another Hilbert space G.

Suppose that we have some linear continuous operator $A : H \to G$. Suppose that $Ker(A) = \{0\}$, i.e. that A is injective. Define A^* as the adjoint of A. Then A^*A is an Hermitian operator from H to H. Moreover, A^*A is strictly positive since $\langle A^*Ax, x\rangle = \|Ax\|^2 > 0$ for $x \in H,\ x \neq 0$.

We can apply the Spectral Theorem to A^*A and obtain

$$A^*A = U^{-1}M_{b^2}U, \tag{14}$$

where $b^2 > 0 \in L^{\infty}(\mu)$.

Applying UA^* to the model (1) we have

$$UA^*Y = UA^*Af + \varepsilon UA^*\xi = M_{b^2}Uf + \varepsilon UA^*\xi.$$

Multiplying by b^{-1}, where $b > 0$, the model is then

$$Z = b \cdot \theta + \varepsilon\eta, \tag{15}$$

where $Z = b^{-1}UA^*Y,\ \ \theta = Uf$, and $\eta = b^{-1}UA^*\xi$.

In order to compare to the case where A is self-adjoint, consider the noise η.

Lemma 2. *Let $\eta = b^{-1}UA^*\xi$, where U is a unitary operator. For any $\theta = Uf$ and $v = Uh$, where $f, h \in H$, we have*

$$\langle \eta, \theta\rangle \sim \mathcal{N}(0, \|\theta\|^2)$$

and

$$\mathbf{E}(\langle \eta, \theta \rangle \langle \eta, v \rangle) = \langle \theta, v \rangle.$$

Proof. For any $\theta = Uf$ with $f \in H$, we have

$$\langle \eta, \theta \rangle = \langle \xi, AU^{-1}b^{-1}\theta \rangle \sim \mathcal{N}(0, \|AU^{-1}b^{-1}Uf\|^2). \tag{16}$$

Remark also that

$$\begin{aligned} \|AU^{-1}b^{-1}Uf\|^2 &= \|A(A^*A)^{-1/2}f\|^2 = \langle A(A^*A)^{-1/2}f, A(A^*A)^{-1/2}f \rangle \\ &= \langle (A^*A)^{1/2}f, (A^*A)^{-1/2}f \rangle = \|f\|^2, \end{aligned} \tag{17}$$

since $(A^*A)^{-1/2} = U^{-1}b^{-1}U$ is well-defined (sometimes non-bounded), and is self-adjoint when $b > 0$. In the same way for any $\theta = Uf$ and $v = Uh$, where $f, h \in H$, we have

$$\begin{aligned} \mathbf{E}(\langle \eta, \theta \rangle \langle \eta, v \rangle) &= \mathbf{E}(\langle \xi, AU^{-1}b^{-1}Uf \rangle \langle \xi, AU^{-1}b^{-1}Uh \rangle) \\ &= \langle AU^{-1}b^{-1}Uf, AU^{-1}b^{-1}Uh \rangle = \langle f, h \rangle = \langle \theta, v \rangle. \end{aligned} \tag{18}$$

Using (16)–(18) we easily obtain the lemma. □

Thus η is a white noise in $L^2(\mu)$ (just as in the self-adjoint case of Section 2).

Finally one can use the model (15) in the non-self-adjoint case just as the model (5) for the case of A adjoint. Thus, we apply the decomposition to A^*A instead of A and use $b = \sqrt{b^2}$. In fact we consider the operator $(A^*A)^{1/2}$.

4. Examples

Example 1. In this section, we are going to present an example of application (given in Ruymgaart (2001)) for which the operator is non-compact. The model considered is the following:

$$\mathrm{d}Y(t) = r * f(t)\,\mathrm{d}t + \varepsilon\,\mathrm{d}W(t), \quad t \in \mathbb{R}. \tag{19}$$

where W is a Wiener process defined on $\mathbb{R}$, $r * f$ denotes the convolution through a known filter $r \in L^1(\mathbb{R})$,

$$Af(t) = r * f(t) = \int_{-\infty}^{\infty} r(t-u)f(u)\,\mathrm{d}u.$$

The aim is to reconstruct the unknown function f based on the observations $\{Y(t), t \in \mathbb{R}\}$.

The problem of convolution is one of the most standard inverse problems. The problem of circular convolution, i.e. periodic on $[a, b]$, appears for example in Efromovich (1997) and Cavalier and Tsybakov (2002). The main difference is that for periodic convolution the operator is compact and the basis of eigenfunctions is the Fourier basis. It seems clear, from a heuristic point of view, that the results could be extended to the case of convolution on $\mathbb{R}$ by using the Fourier transform on $L^2(\mathbb{R})$ instead of the Fourier series. This is the reason why the generalization to non-compact presents some interest, in the sense that it proves that this natural extension is valid.

Suppose that r is symmetric about 0, then

$$\tilde{r}(x)=\int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}tx} r(t)\,\mathrm{d}t=\int_{-\infty}^{\infty} \cos(tx) r(t)\,\mathrm{d}t>0, \quad \forall x \in \mathbb{R}.$$

If A is not self-adjoint, it is easy to understand that one can use Section 3 to obtain the results. For sake of simplicity we deal with the self-adjoint case.

Define the Fourier transform as a unitary operator from $L^2(\mathbb{R})$ into $L^2(\mathbb{R})$ by

$$(Ff)(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}tx} f(t)\,\mathrm{d}t, \quad x \in \mathbb{R},\ f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R}) \tag{20}$$

and its continuous extension on $L^2(\mathbb{R})$.

We have that

$$F(r * f)(x)=\tilde{r}(x).(Ff)(x)$$

and then $A=F^{-1}M_{\tilde{r}}F$. Since $\tilde{r} \in L^2(\mathbb{R})$, the problem is ill-posed.

The last point is to define the Fourier transform applied to the stochastic integral.

The generalized stochastic process

$$\eta(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}tx}\,\mathrm{d}W(t),$$

has by construction a “flat” spectrum and is well-known to be a *white noise* (see Hida, 1980), i.e. such that $\langle \eta, Ff\rangle \sim \mathcal{N}(0, \|f\|^2)$ and

$$\mathbf{E}(\eta(x)\eta(y))=\delta_{x=y}=\begin{cases} 0 & \text{if } x \neq y, \\ +\infty & \text{if } x=y. \end{cases}$$

Example 2. Another example, which does not exactly correspond to our framework, but is very important, is the estimation of a derivative. Suppose that we observe

$$Y=f+\varepsilon\xi, \tag{21}$$

where $H=L_2(\mathbb{R})$, f is a C^β function, $\beta \in \mathbf{N}$, in $L_1(\mathbb{R}) \cap L_2(\mathbb{R})$ and ξ is a white noise. A standard problem in statistics is the estimation of the derivative $D^\beta f=f^{(\beta)}$ of f, or the function f itself when $\beta=0$.

It is well-known that $D^\beta=F^{-1}M_{b^{-1}}F$, where F is the Fourier transform defined in (20), and $b(s)=(is)^{-\beta}$.

Applying the Fourier transform on (21), one obtains the model

$$Z=FY=Ff+\varepsilon F\xi=\tilde{f}+\varepsilon\eta, \tag{22}$$

where $\tilde{f}$ is the Fourier transform of f and η is a white noise. This model is equivalent to

$$X=b^{-1}Z=b^{-1}\tilde{f}+\varepsilon b^{-1}\eta=\theta+\varepsilon b^{-1}\eta, \tag{23}$$

where θ is the Fourier transform of $f^{(\beta)}$. Thus, this framework corresponds to our problem.

Consider a finite family of projection estimators

$$\hat{\theta}_i = \lambda^i(s)X(s),$$

where $\lambda^i(s) = I(|s| \leqslant w_i)$, $w_i \in \{w_1, \ldots, w_N\}$, $1 \leqslant w_1 < \cdots < w_N$. Suppose also that $N \sim \varepsilon^{-a}$, $w_N/w_1 \sim \varepsilon^{-c}$, where a and c are positive constants, and $\log(1/\varepsilon)/w_1 \to 0$. In this framework the estimator $\hat{f}_\beta^i$ of $f^{(\beta)}$ based on $\hat{\theta}_i$ is usually called a band-limited filter.

It is now possible to obtain more precise computations for the different terms in Theorem 1. Remark for example that

$$\int_S |\sigma\lambda^i|^2 \, \mathrm{d}\mu = \int_{\mathbb{R}} |is|^{2\beta} I(|s| \leqslant w_i) \, \mathrm{d}s = C w_i^{2\beta+1},$$

where C is some generic positive constant. Thus, by a simple algebra one obtains

$$\omega(s) \leqslant C|s|^{2\beta} \tag{24}$$

and

$$L_\Lambda \leqslant C \log\left(\frac{N w_N}{w_1}\right). \tag{25}$$

Using (24) and (25) and Theorem 1 one has,

$$\mathbf{E}_f \|f_\beta^* - f^{(\beta)}\|^2 \leqslant (1 + \gamma_1 B^{-1}) \min_{\hat{f}_\beta^i} \mathbf{E}_f \|\hat{f}_\beta^i - f^{(\beta)}\|^2 + CB\varepsilon^2 \log(1/\varepsilon)^{2\beta+1}. \tag{26}$$

In a nonparametric setting, the remainder term $CB\varepsilon^2 \log(1/\varepsilon)^{2\beta+1}$ will usually be asymptotically small.

Thus, choosing B large enough, we then say that our estimator of the derivative asymptotically mimics the best possible estimator in Λ.

Consider the case when f belongs to some Sobolev ball

$$W(\alpha + \beta, L) = \left\{ f : \int_{\mathbb{R}} (f^{(\alpha+\beta)}(x))^2 \, \mathrm{d}x \leqslant L \right\} = \left\{ f : \int_{\mathbb{R}} |\tilde{f}(s)|^2 |s|^{2\alpha+2\beta} \, \mathrm{d}s \leqslant L \right\},$$

where $\alpha, L > 0$ and $\tilde{f}$ denotes the Fourier transform. The problem is then to estimate, using the observation (23), the function $\theta \in \Theta(\alpha, L)$ where

$$\Theta(\alpha, L) = \left\{ \theta : \int_{\mathbb{R}} |\theta(s)|^2 |s|^{2\alpha} \, \mathrm{d}s \leqslant L \right\}.$$

One can easily verify that for any projection estimator $\hat{\theta}_w$ of θ with $\lambda(|s| \leqslant w)$ we have

$$\begin{aligned} \sup_{\theta \in \Theta(\alpha, L)} \mathbf{E}_f \|\hat{\theta}_w - \theta\|^2 &= \sup_{\theta \in \Theta(\alpha, L)} \left(\int_{|s|>w} |\theta(s)|^2 \, \mathrm{d}s \right) + \varepsilon^2 \int_{|s| \leqslant w} |is|^{2\beta} \, \mathrm{d}s \\ &\geqslant C\varepsilon^{4\alpha/(2\alpha+2\beta+1)}, \end{aligned}$$

as $\varepsilon \to 0$. The risk of the best projection estimator is then asymptotically larger than the remainder term in (26). Suppose that the best projection estimator in the family Λ has a risk which asymptotically attains the order $\varepsilon^{4\alpha/(2\alpha+2\beta+1)}$. Then we have, with $B = \log(1/\varepsilon)$,

$$\sup_{f \in W(\alpha+\beta, L)} \mathbf{E}_f \| f^*_\beta - f^{(\beta)} \|^2 \leqslant (1 + \mathrm{o}(1)) \min_{\hat{f}^i_\beta} \sup_{f \in W(\alpha+\beta, L)} \mathbf{E}_f \| \hat{f}^i_\beta - f^{(\beta)} \|^2$$
$$= \mathrm{O}(\varepsilon^{4\alpha/(2\alpha+2\beta+1)}),$$

as $\varepsilon \to 0$.

It could also be easily proved that $\varepsilon^{4\alpha/(2\alpha+2\beta+1)}$ is the optimal rate of convergence among all the estimators.

Since we have a $(1 + \mathrm{o}(1))$ it is even possible on some special families to attain exactly the minimax risk and not only the rate of convergence.

Acknowledgements

I would like to thank the editor and two referees for comments leading to improvements in the paper.

References

Caroll, R.J., van Rooij, A.C.M., Ruymgaart, F.H., 1991. Theoretical aspects of ill-posed problems in statistics. Acta Appl. Math. 24, 113–140.

Cavalier, L., Tsybakov, A.B., 2002. Sharp adaptation for inverse problems with random noise. Probab. Theory Related Fields 123, 323–354.

Cavalier, L., Golubev, G.K., Picard, D., Tsybakov, A.B., 2002. Oracle inequalities in inverse problems. Ann. Statist. 30, 843–874.

Dey, A.K., Ruymgaart, F.H., Mair, B., 1996. Cross-validation for parameter selection in inverse estimation problems. Scand. J. Statist. 23, 609–620.

Donoho, D.L., 1995. Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. Appl. Comput. Harmonic Anal. 2, 101–126.

Efromovich, S., 1997. Robust and efficient recovery of a signal passed through a filter and then contaminated by non-Gaussian noise. IEEE Trans. Inform. Theory 43, 1184–1191.

Engl, H.W., Hanke, M., Neubauer, A., 1996. Regularization of Inverse Problems, Kluwer Academic Publishers, Dordrecht.

Hadamard, J., 1932. Le problème de Cauchy et les équations aux dérivées partielles linéaires hyperboliques, Herman, Paris.

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

Hida, T., 1980. Brownian Motion, Springer, New York, Berlin.

Johnstone, I.M., 1999. Wavelet shrinkage for correlated data and inverse problems: adaptivity results. Statist. Sinica 9, 51–83.

Mair, B., Ruymgaart, F.H., 1996. Statistical estimation in Hilbert scale. SIAM J. Appl. Math. 56, 1424–1444.

van Rooij, A.C.M., Ruymgaart, F.H., 1996. Asymptotic minimax rates for abstract linear estimators. J. Statist. Plann. Inference 53, 389–402.

van Rooij, A.C.M., Ruymgaart, F.H., van Zwet, W.R., 2000. Asymptotic efficiency for inverse estimators. Theory Probab. Appl. 44, 722–738.

Ruymgaart, F.H., 2001. A short introduction to inverse statistical inference, Lecture in IHP, Paris.