Review of rates of convergence and regularity conditions for inverse problems.

J-M. Loubes[1] **and V. Rivoirard**[2]

[1]Institut de Mathématiques de Toulouse, UMR 5219
Université Toulouse 3, Toulouse
loubes@math.univ-toulouse.fr
[2]Laboratoire de Mathématiques, UMR 8628
Université Paris-Sud
Vincent.Rivoirard@math.u-psud.fr

Abstract

The aim of this article is to review the different rates of convergence encountered in inverse problems, with both deterministic and stochastic noise. Indeed, in the litterature, several regularity conditions are often assumed leading to apparently different rates. We point out the different points of view and provide global assumptions that handle most of the cases encountered. Moreover we discuss optimality of some different usual estimators in the minimax but also the maxiset framework.

Keywords: Inverse Problems, Regularization Methods, Maxisets.

2000 Mathematics Subject Classification: 60G17, 62G07.

1 Introduction

Driven by the needs of application in sciences and industry, the field of inverse problem has been one of the fastest growing area over the recent years. Indeed inverse problems are found in a large number of pratical situations, in mathematics, in biology, in seismology or in economy. Hence mathematicians tried to model the different situations encountered using tools from different fields of applied mathematics, more particulary from numerical analysis, econometrics and statistics. We try, in this paper, to review the different sets of results together with the assumptions made for each model. We aim at relating the different results into a more general framework. More precisely we will draw correspondence between the regularity assumptions and the different rates of convergence. A particular attention will be paid to the difference between the deterministic settings and the stochastic framework.

An inverse problem deals with the estimation of a function x_0 which is not observed directly but through an operator A. We assume that we do not observe the image of the function of interest but that our observations are close to this image. More precisely, assume we observe y such that $\|y - Ax\| \leq \delta$. The norm $\|.\|$ depends on the framework as well as the definition of the noise level δ. The noise can either be a deterministic error or a stochastic random noise, which will determine the difficulty of the estimation problem and the asymptotics is given by $\delta \to 0$. The operator is often assumed to be known in a statistical or deterministic framework since the estimation issue is due to observation errors. On the contrary in econometrics, the operator often depends on the law of the data and so has to be estimated. As a result, the noise comes from the estimation of the unknown law of the data, leading to an estimator of the operator, which defines the level of the observation noise.

The problem is ill-posed in the sense that our noise corrupted observations might lead to large deviations when trying to estimate x_0. In the linear case, the best L^2 approximation of x_0 is $x^+ = A^+ Ax$, where A^+ is the Moore-Penrose (generalized) inverse of A. We will say the problem is ill-posed if A^+ is unbounded. This might entail, and is generally the case, that $A^+(y)$ is not close to x^+. Hence, the inverse operator needs to be, in some sense, regularized.

Regularization methods replace an ill-posed problem by a family of well-posed problems. Their solution, called regularized solutions, are used as approximations of the desired solution of the inverse problem. These methods always involve some parameter measuring the closeness of the regularized and the original (unregularized) inverse problem. Rules (and algorithms) for the choice of these regularization parameters as well as convergence properties of the regularized solutions are central points in the theory of these methods, since they allow to find the right balance between stability and accuracy. Generally, one often consider a regularized operator acting over the data $\hat{x}_\delta = R_\alpha(y)$ where α is a smoothing sequence depending on the noise level δ such that $R_\alpha y$ is close to x_0. Regularization techniques are widely described in (Engl, Hanke and Neubauer, 1996). In a deterministic settings, rates of convergence are given in (Tautenhahn and Jin, 2003), (Tikhonov and Arsenin, 1977) or (Engl, 2000) for instance. In a statistical framework, optimal rates are given for linear and non linear procedures in (Cavalier, Golubev, Picard and Tsybakov, 2002), (Cavalier and Tsybakov, 2002) or (Mair and Ruymgaart, 1996) for example and in (Darolles, Florens and Renault, 2004) and (Hall and Horowitz, 2005) in econometrics.

When A is linear, the statistical problem has been extensively studied, although in general efficient parameter choice is still under active research. Two main kinds of estimators have been considered. First regularized estimators such as Tikhonov type estimators, then non linear thresholded estimators. The first approach has been studied in great detail. An interesting early survey of this topic is provided by O'Sullivan in (O'Sullivan, 1986). In this setting, the main issues are what kind of regularizing functional should be considered and closely related what the relative weight of the regularizing functional should be. More recently, Mair and Ruymgaart in (Mair and Ruymgaart, 1996) studied different regularized inverse problems and proved the optimality of the rate of convergence for their estimators. Special attention has been devoted in this setting when considering a Singular value decomposition (SVD) of the operator F. We cite the recent work in this direction developed by Cavalier and Tsybakov in (Cavalier and Tsybakov, 2002) or Cavalier, Golubev, Picard and Tsybakov in (Cavalier et al., 2002). The second approach has its most popular version in the *wavelet-vaguelet* decomposition introduced by Donoho (Donoho, 1995). In this case the main issue is finding an appropriate basis over which F^+, the generalized inverse, is almost diagonal. This idea is further developed by Kalifa

and Mallat (Kalifa and Mallat, 2003) who introduce *mirror wavelets*. Closely related, Cohen, Hoffmann and Reiss in (Cohen and Reiss, 2003) construct an adaptive thresholded estimator based on Galerkin's method.
We point out that scarce statistical literature exists when A is non linear. Among the few papers available, we point out the works (O'Sullivan, 1990) or (Snieder, 1991) where some rates are given, and that of Bissantz et al. in (Bissantz, Hohage, Munk and Ruymgaart, 2007) where they discuss a nonlinear version of the method of regularization (MOR). A different type of approach is developed in Chow and Khasminskii (Chow and Khasminskii, 1997) for dynamical inverse problems. Finally in (Loubes and Ludeña, 2008) are handled a particular class of non linear operators. But such problems are not fully studied. Hence, in this work, we will restrict ourselves to linear operators with linear or non linear estimation procedures.

So, in all these papers, rates are given according to the regularity of the operator and the function to be estimated. First, the difficulty of the inverse problem is measured through the ill-posedness of the operator. Then some regularity is assumed for the function to be estimated but the sets of assumptions differ according to the 3 frameworks. In statistics, one often assume smoothness assumptions such Sobolev or Besov imbeddings. In econometrics, saturation sets are defined to measure the bias of an estimation procedure and will be shown to be equivalent to maxisets (see (Cohen, DeVore, Kerkyacharian and Picard, 2001) for the definition of maxisets). They are also closely related to the source sets, widely used in numerical analysis litterature. All these assumptions lead to different rates of converge depending on the different regularity indexes defined within all frameworks. Hence, we aim at drawing lines between the different settings and show the common points as well as the differences between all ! the different models.

The paper falls into the following sections. Section 2 introduces the different models used in the three field studied here, numerical analysis, econometry and statistics, as well as the common assumptions over the operator and the parameter of interest. In Section 4, we study the different rates of convergence obtained in the different frameworks. Section 5 is devoted to a comparison between stochastic and deterministic noise. The differences between the different regularity sets in econometrics and analysis are highlighted in Section 3. Finally in Section 6, we give some results about optimaly of thresholding procedures as regards the maxiset theory.

2 Models for inverse problems

We first recall the basic properties of operators in Hilbert spaces. Let X and Y be Hilbert spaces. Let $A: X \to Y$ be a compact linear operator, one to one and bounded. The adjoint operator A^* is hence defined everywhere from Y to X such that

$$\forall x \in X,\ y \in Y, \quad < Ax, y >=< x, A^* y > .$$

Then the operator A^*A is self-adjoint and non negative. This operator admits a unique spectral measure E_λ such that

$$A^*A = \int \lambda dE_\lambda(\lambda).$$

So we will use the following standard notations

$$\forall f, \quad f(A^*A) = \int f(\lambda) dE_\lambda(\lambda).$$

Moreover, this operator admits a SVD decomposition, $(\sigma_k^2 = \lambda_k, \phi_k, \psi_k),\ k \geq 1$. So we get

$$\forall x \in X,\ A^*Ax = \sum_{k=1}^{\infty} \lambda_k < x, \phi_k > \phi_k.$$
$$Ax = \sum_{k=1}^{\infty} \sigma_k < x, \phi_k > \psi_k.$$

Ill-posedness means using this decomposition that the eigenvalues goes to zero when the resolution level k increases. If Ax were observable, the solution would be given by

$$x^+ = \sum_{k=1}^{\infty} \frac{< Ax, \psi_k >}{\lambda_k} \phi_k.$$

But the asymptotic behaviour of the eigenvalues implies that a small perturbation of the data lead to a large deviation of the estimate. The faster is the decay, the more difficult the estimation becomes, since inverting the operator is a more challenging issue.

As explained previously, we do not observe the image of the unknown function but noisy data. So in statistics or in numerical analysis, the model is defined as

$$y = Ax_0 + \delta\xi, \tag{2.1}$$

where x_0 is a function which belongs to a subset of a Hilbert space X. A is a compact ill-posed operator from X to Y, ξ is a noise and δ the level of the noise.
The following assumption is crucial in inverse problems with deterministic noise.

Assumption D1 The exact solution when $\delta = 0$ is assumed to exist and to be such that there exist s a regularity index and R,

$$||x_0||_s \leq R.$$

The definition of the noise change according to the framework. In numerical analysis, it corresponds to measurement errors so it is assumed that $\xi \in Y$ and is such that $||\xi|| \leq 1$. Moreover, it is assumed that the functions are observed at all points.

In statistics, the noise is a Gaussian white noise $\xi \sim \mathcal{N}(0, 1)$ and does not belong necessarily to Y. In Section 5, we study the main differences between these two definitions and their consequences.
Moreover, the asymptotics is given by the number of observations and not the noise level. For this, Model (2.1) is discretized in n points and can be written as follows

$$y_i = Ax_0(t_i) + \delta_n \xi_i,\ i = 1, \ldots, n\ . \tag{2.2}$$

$t_i,\ i = 1, \ldots, n$ are given observation times. The noise is defined as a sequence of random Gaussian variables

$$\forall i = 1, \ldots, n, \quad \delta_n \xi_i \overset{\text{i.i.d}}{\sim} \mathcal{N}\left(0, \frac{1}{n}\right).$$

So, the noise level is related to the number of observations in the following way $\delta_n = 1/\sqrt{n}$. For sake of simplicity we restric ourselves to the case of Gaussian noise but other kinds of noise can be studied provided concentration bounds exist over the random variables. This model is the more complicated one since it mixes several issues, both the inverse problem and the discretization in a regression model issue. It is tackled in (Bissantz et al., 2007), (Loubes and Ludeña, 2008) for instance.
Dealing with (2.2) requires introducing an empirical norm based on this design. Set Q_n to be the empirical measure of the covariables:

$$Q_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{t_i}.$$

Here we have set δ the Dirac function. The $L_2(Q_n)$-norm of a function $y \in Y$ is then given by

$$\|y\|_n = \left(\int y^2 dQ_n \right)^{1/2},$$

and the empirical scalar product by

$$< y, \xi >_n = \frac{1}{n} \sum_{i=1}^{n} \xi_i y(t_i).$$

Remark this empirical norm is defined over the observation space Y. Over the solution space X we will consider the norm given by the Hilbert space structure. Hence we deal with two norms, one is discretized and the other is continuous, which implies that the adjoint of the operator A with such topology depends also on the discretization scheme. Hence extra assumptions are needed to control the amplification of the error with respect to the choice of the discretization sequence as done in (Neubauer, 1992), (Kaltenbacher, 2000) or (Loubes and Ludeña, 2008).

In the particular case where $Y = X$, the functions ϕ_k, ψ_k are orthonormal families of X. Hence we can associate to model (2.1) the sequence model

$$y_k = \lambda_k x_k + \delta \xi_k, \ \forall k \in \mathbb{N}^*. \tag{2.3}$$

Hence multiplying by λ_k^{-1} gives rise to the following sequence model for inverse problems

$$z_k = x_k + \delta \lambda_k^{-1} \xi_k, \ \forall k \in \mathbb{N}^*. \tag{2.4}$$

In this settings, the inverse problem is more easily studied as a particular case of a sequential heteriscedastic regression model with variance growing to infinity as the resolution level k increases. A large amount of litterature in statistical estimation exists for this model. We refer to (Cavalier et al., 2002), (Mair and Ruymgaart, 1996) for general references.
The main advantage of this model is that the estimation issue as well as the ill-posedness are well defined in terms of decay of coefficients. However this model does not correspond to the real observation due to the bias induced by the discretization. Even if under regularity conditions, discrete coefficients are good approximations of real coefficients, this drawback may induce a bias when working with real data as well as a loss of efficiency as quoted in (Donoho and Johnstone, 1999) or (Autin, Le Pennec, Loubes and Rivoirard, 2008).

Inverse problems in economy appear in particular when studying inference with instrumental variables. Consider a random observable vector $(\tilde{y}, z, w)$ with distribution F such that there is an unknown funtion x_0

$$\begin{cases} \tilde{y} &= x_0(z) + U \\ \mathbf{E}(U|w) &= 0. \end{cases} \tag{2.5}$$

U is a random noise, which is identified by the mean of the instrumental variable w such that $\mathbf{E}(U|w) = 0$. This framework is explained in (Hall and Horowitz, 2005) or (Darolles et al., 2004). So consider the unknown operator $A := x \to \mathbf{E}[x(z)|w]$ and define $y := \mathbf{E}(\tilde{y}|w)$. Hence we can write the regression model with instrumental variables as

$$Ax_0 = y. \tag{2.6}$$

So define the The operator depends on the unknown law F of the data and can be written $A = A_F$. A nonparametric estimation of the distribution provides an estimation of the operator $\hat{A} = A_{\hat{F}}$. Hence Model (2.5) can be rewritten as follows

$$y = \hat{A}x_0 + \left[Ax_0 - \hat{A}x_0\right]. \tag{2.7}$$

In order to compare this model with the classical model for inverse problems (2.1) by considering the estimation of the operator as an error term in the regression model, $\delta\xi = \left[Ax_0 - \hat{A}x_0\right]$, with known operator $\hat{A}$. Then the model (2.7) has the classical form

$$y = \hat{A}x_0 + \delta\xi,$$

where $\delta\xi$ depends on the number of observations through the rate of convergence of the nonparametric estimate of the density of the law of the data.

Finally, all three main models are similar in the sense that the inverse regression model is the main framework for studying the efficiency of estimation procedures. The main differences lay in the definition of the noise level.

3 Regularity conditions for inverse problems

In the litterature of inverse problems, the rate of convergence depends on two parameters

- The ill-posedness of the operator A. It is stated using Hilbert scales or Sobolev imbeddings.
  More precisely ill-posedness can be defined considering the Hilbert scale framework. A Hilbert scale is defined with a family of Hilbert spaces $\{X_\mu,\ \mu \in \mathbb{R}\}$. Consider the inner product in X_μ: $< x, y >_\mu = < L^\mu x, L^\mu y >$, where L is an unbounded self-adjoint strictly positive operator in X and $< ., . >$ is the inner product in X. More precisely X_μ stands for the completion of the intersection of domains of operators $L^\mu,\ \mu \geq 0$. So we get that $\|.\|_\mu =< ., . >_\mu^{1/2}$ and by definition we set $\|.\| = \|.\|_0$

The degree of ill-posedness is a such that

$$\forall x \in X_0, \quad c\|x\|_{-a} \leq \|Ax\| \leq C\|x\|_{-a}, \tag{3.1}$$

for two given positive constants c, C.

This expresses a smoothing action. Indeed the operator maps any space of smoothness s into a space of smoothness $s+a$, for instance $H^s \to H^{a+s}$. In an equivalent way, ill posedness of order a is equivalent with the ellipticity property

$$\forall x \in X, \quad < Ax, x > \sim \|x\|_{H^{-a/2}},$$

where $H^{-a/2}$ stands for the dual space of the Sobolev space $H^{a/2}$.

Ill-posedness can be seen through the decay of the eigenvalues as follows. Under previous assumptions (3.1), we obtain the following equivalence

$$\lambda_k \sim k^{-a}.$$

If we do not consider natural assumptions over the operators but only the decay of the eigenvalues, we can imagine a larger class of inverse problems with other kinds of decay, for instance exponential decay.

- The regularity of the function x_0 to be estimated. This regularity can be expressed in several ways, corresponding to the various frameworks.

In statistical estimation, when considering the sequence model (2.4), regularity is expressed through sparsity constraints. This assumption is standard in non parametric estimation and can be expressed by the decay of the coefficients of the unknown function in a good basis. This decay may be polynomail or exponential. So, there is a regularity parameter s an increasing sequence a_k and a constant $C > 0$ such that

$$x_0 \in W_s(C) := \{x = \sum_{k=1}^{\infty} x_k \phi_k, \ \sum_{k=1}^{\infty} a_k^{2s} x_k^2 \leq C\}.$$

The sequence can be chosen exponential $a_k \asymp e^k$ or polynomial $a_k \asymp k$. It is the case in most of the papers in statistical analysis of inverse problems, see for instance (Donoho, 1995), (Mair and Ruymgaart, 1996) or (Cohen and Reiss, 2003) for instance. Such assumption corresponds to well known regularity spaces, such as Sobolev or Besov spaces. For more references for such functional spaces, we refer to (Besov, Ilin and Nikolskiĭ, 1978). Moreover for pratical use, such set of assumptions can be easily checked.

In econometrics, the situation is different since the operator is unknown and has to be estimated. In this case, the authors often separate the problem into a bias-variance trade-off and write the following decomposition

$$\hat{x}_n^\alpha = R_\alpha^n y = x_0 + \underbrace{\hat{x}_\alpha^n - x_\alpha}_{\text{variance}} + \underbrace{x_\alpha - x_0}_{\text{bias}}.$$

R_α is the regularization operator chosen to build the estimator and α is the smoothing sequence. x_α is obtained by smoothing the data without noise by the smoothing operator. The bias term measures in a way the efficiency of the regularization procedure to handle a deterministic ill-posed problem for a given class of functions. Hence the regularity of the function of interest is assumed to undergo the following condition: $\exists 0 \leq \beta \leq 1$,

$$x_0 \in \Phi_\beta := \{x, \ \|x - x_\alpha\| = O(\alpha^\beta)\}. \tag{3.2}$$

Such set is called a saturation space, which is often used in econometrics paper. It is the case when studying inverse problems due to instrumental variables, the consumer surplus variations or generalized moments methods, see for instance (Darolles et al., 2004), (Carasco, Florens and Renault, 2004) or (Loubes and Vanhems, 2004). Such spaces are defined for a given regularization scheme. They define the regularity of a function as the necessary smoothness to achieve, without noise, an estimation of the inverse problem at a given rate of convergence $O(\alpha^\beta)$. Hence such definition can also be interprated as a deterministic maxiset associated to the estimation procedure R_α with rate of convergence α^β, provided α is a given sequence. Hence we should assume that these sets are properly defined, so assume that

Assumption A1 For every $0 < t < s < 1$, we have $\Phi_s \subset \Phi_t$.

In a deterministic framework, or in a statistical framework where the sequence model can not be used, authors define the following spaces

$$X_\beta = \{x \in X, \, \exists \omega \in L^2, \, x = (A^*A)^\beta \omega\} = \mathcal{R}\left[(A^*A)^\beta\right],$$

where $\mathcal{R}$ stands for the range of an operator. We refer to (Engl et al., 1996) for the complete definition of such spaces in a deterministic settings and to (Bissantz et al., 2007) or (Loubes and Ludeña, 2008) with a random noise regression framework.
Such spaces aim at relating the decay of the coefficients of the function to be estimated, with the decay of the eigenvalues of the operator. So the spaces X_β compares the regularity of the unknown function with the ill-posedness of the operator. These spaces are usually called source sets.

For compact operators, these spaces can be characterized via the singular values a follows

$$X_\beta = \{x, \sum_{k=1}^{\infty} \frac{| < x, \phi_k > |^2}{\sigma_k^{4\beta}} < +\infty\}. \tag{3.3}$$

This way of characterizing the regularity in ill-posed inverse problem appears naturally when considering the bias term in the estimation. However, such assumption is difficult to check in practice since the two characteristics of the issue (the ill-posedness and the regularity) are mixed in one main condition.

There is a close connection between source sets and maxisets of a particular regularization procedure. Assume that the regularization procedure assumes the following assumptions

Assumptions A2 The regularization scheme R_α defined on $[0, 1]$ is such that there are a constant α_0 and constants $c, \, C, K$ such that for all sequence $\alpha = (\alpha_k)_{k \geq 1}, \ 0 < \alpha_k < \alpha_0, \ \forall k \geq 1$

$$c \leq |1 - \lambda R_\alpha(\lambda)| \leq C, \, \forall 0 \leq \lambda < \alpha,$$

$$|1 - \lambda R_\alpha(\lambda)| \frac{\lambda}{\alpha} \leq K, \, \forall \alpha \leq \lambda \leq 1.$$

Moreover, it is possible to give another interpretation of such maxisets for well chosen regularization sequence.

Proposition 3.1 (Spectral and sequential definitions of maxisets for regularization methods). *Under Assumtions* **A2**, *we obtain the following equivalent definitions*

$$\Phi_s = \{x \in X, \int_0^t \|E_\lambda d\lambda\, x\|^2 = 0\left(t^{2s}\right)\}$$
$$= \{x = \sum_{k=1}^{\infty} x_k \phi_k, \sum_{k \geq n} x_k^2 = O\left(\lambda_n^{2s}\right)\}.$$

Proof. Recall here the guidelines in (Neubauer, 1997) or (Engl et al., 1996), we get clearly that

$$\|x^\alpha - x_0\|^2 = \int_0^\infty (1-\lambda R_\alpha(\lambda))^2 d\|E_\lambda x_0\|^2$$
$$= \alpha^{2s}\left(\alpha^{-2s}\int_0^\alpha (1-\lambda R_\alpha(\lambda))^2 d\|E_\lambda x_0\|^2\right)$$
$$+\alpha^{2s}\left(\alpha^{2-2s}\int_\alpha^\infty (1-\lambda R_\alpha(\lambda))^2 \frac{\lambda^2}{\alpha^2}\lambda^{-2} d\|E_\lambda x_0\|^2\right).$$

As a result, if $\int_0^t \|E_\lambda d\lambda\, x\|^2 = 0\left(t^{2s}\right)$, then since

$$c^2\int_0^\alpha d\|E_\lambda x_0\|^2 \leq \int_0^\alpha (1-\lambda R_\alpha(\lambda))^2 d\|E_\lambda x_0\|^2 \leq C^2\int_0^\alpha d\|E_\lambda x_0\|^2, \tag{3.4}$$

we obtain that $\|x_0 - x^\alpha\| = O(\alpha^s)$.
Conversely, if $\|x_0 - x^\alpha\| = O(\alpha^s)$, then bound (3.4), as well as

$$\int_\alpha^\infty (1-\lambda R_\alpha(\lambda))^2 \frac{\lambda^2}{\alpha^2}\lambda^{-2} d\|E_\lambda x_0\|^2 \leq D^2\int_\alpha^\infty \lambda^{-2} d\|E_\lambda x_0\|^2$$

enable to conclude the proof. □

Such equivalent definitions enable to characterize the maxisets for particular class of inverse problems. Once the degree of ill-posed of the operator A is fixed, say for instance a, the rate of decay of the eigenvalues is fixed. More precisely we get $\lambda_k \sim k^{-a}$. This implies that

$$\forall x = \sum_{k=1}^{\infty} x_k \phi_k \in \Phi_s \to \sum_{k \geq n} x_k^2 = O\left(n^{-2as}\right).$$

Hence in problem (2.1), the quadratic bias estimation error in Φ_s corresponds to an approximation, by its projection onto a finite dimensional space with dimension n, of order n^{-2as}. Hence we obtain clearly that for $\beta = as$, we get sets which are related to Sobolev spaces with degree of regularity β. Indeed, define as in (Mair and Ruymgaart, 1996) the following Sobolev spaces

$$W^\beta := \{x = \sum_{k=1}^{\infty} < x, \phi_k > \phi_k \in L^2([0,1]), \sum_{k=1}^{\infty} k^{2\beta} | < x, \phi_k > |^2 < +\infty\}.$$

Remark 3.1. In this case (compact operator with a its degree of ill-posedness), we can also interpret definition of the source sets to interpret them in terms of Sobolev regularity, leading to the following result.

$$X_s = W^{\frac{as}{2}}.$$

This remark will be extended in Theorem 4.2 in order to compare rates of convergence in the source sets framework and in the statistical smoothness framework.

The following theorem draws some relation between maxisets in econometrics and the source sets used in a deterministic settings.

Theorem 3.2 (Deterministic maxisets and source sets). *Under assumptions* **A1** *and* **A2***, we obtain the following inclusions.*

1. *For* $0 < \beta < 1$*,*
$$X_\beta \subset \Phi_\beta, \quad X_\beta \nsubseteq \Phi_t, \, \forall t > \beta.$$

2. $$X_1 = \Phi_1.$$

This theorem is gathered from the results in (Engl et al., 1996).

We recall here the main regularization scheme used in the statistical litterature

- Tikhonov regularization: $\lambda R_\alpha(\lambda) := \frac{\lambda}{\lambda+\alpha}$.
- Landweber regularization: $\lambda R_\alpha(\lambda) := 1 - (1 - c\lambda)^{\frac{1}{\alpha}}$ for some $c > 0$,
- Spectral cut-off: $\lambda R_\alpha(\lambda) := \mathbf{1}_{[\alpha,1]}(\lambda) + \frac{\lambda}{c\alpha}\mathbf{1}_{[0,\alpha)}$.

These usual regularization operators satisfy clearly the assumptions **A2**.

4 Rates of convergence for inverse problems

Define for a deterministic noise and respectively a random noise, the optimal (minimax) rate of convergence for recovering an unknown signal $x_0 \in X$ as

$$e^\delta_{\text{det}} := \inf_{\hat{x}} \sup_{x\in X} \sup_{||\xi|| \leq 1} ||x - \hat{x}(\xi)|| \tag{4.1}$$
$$e^\delta_{\text{sto}} := \inf_{\hat{x}} \sup_{x\in X} \left(\mathbf{E}||x - \hat{x}(\xi)||^2\right). \tag{4.2}$$

Regularization of ill-posed inverse problems in Hiblert scales lead to the following result

$$e^\delta_{\text{det}}(A, X_\beta(R)) \asymp \delta^{\frac{\beta}{a+\beta}}, \tag{4.3}$$

and this rate of convergence is achieved. Such results have been introduced by Natterer in (Natterer, 1984) and are also given in (Engl et al., 1996).

In a statistical setting, consider first the sequence model (2.4) with Gaussian noise. We now obtain the following theorem

Theorem 4.1 (Stochastic convergence in sequence model). *Suppose we are given in a sequence space*

$$y_k = x_{0k} + \delta\lambda_k^{-1}\xi_k, \; k = 1, \ldots$$

where ξ_k *are i.i.d* $\mathcal{N}(0,1)$ *and* $\lambda_k \asymp k^{-a}$ *and*

$$x_0 \in \{x, \sum_{k\geq 1} k^{2s} x_k^2 \leq R^2\}.$$

Hence

$$\inf_{\hat{x}} \sup_{x,\, \sum_{k\geq 1} k^{2s} x_k^2 \leq R^2} (\mathbf{E}\|x_0 - \hat{x}\|_2^2)^{1/2} \asymp \delta^{\frac{s}{s+b+1/2}}.$$

This result is taken from (Donoho, 1995) or (Cavalier and Tsybakov, 2002). Here also the optimal rate is achieved for a particular class of estimators.

If we do not want to consider this sequential model but try to work with a more realistic model, we need other types of assumptions more adapted to operators. These assumptions are given by the deterministic settings by the sets X_β. In this framework, the following theorem provides a rate of convergence.

Theorem 4.2 (Stochastic convergence in operator model). *Suppose we are given the following observations*

$$y_i = Ax_0(t_i) + \delta\xi_i,\ i = 1, \ldots, n,\ \xi_i \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, 1).$$

Assume that the source condition holds $x_0 \in \mathcal{R}(A^*A)^\beta$ *and that* A *is ill-posed with degree* a*, then we obtain*

$$e^\delta_{\text{sto}} \asymp \delta^{\frac{4\beta a}{4\beta a+2a+1}}.$$

This result is obtained in (Bissantz et al., 2007), (Loubes and Ludeña, 2008) using respectively regularized estimators and either projection estimators or adaptive regularized estimators.
This rate can be compared to the rate obtain in Theorem 4.1 with the following correspondence $s = 2\beta a$. This relation enlights that in the definition of the spaces X_β, the degree of ill-posedness and the Hilbert type regularity of the unknown function are put together while all the sequence model enables to separate the two kind of assumptions.

To sum up, the different rates obtained in Hilbert scales for linear ill-posed problems with known operators are given in the following table. We separated the different cases of degree of ill-posedness together with the two frameworks with either deterministic noise or random noise.

Rates of convergence	Deterministic noise	Stochastic noise
Polynomial ill-posedness	$\frac{s}{s+a}$	$\frac{2s}{2s+2a+1}$
Exponential ill-posedness	$\frac{s}{s+a}$	$\frac{s}{s+a}$

Finally, we proved that the two sets of assumptions (Hilbert scales of source sets) in a statistical framework lead to the same rate of convergence, provided the indexes of regularity are properly defined. The correspondence $s = 2\beta a$ relates the smoothness regularity assumed over the function x_0 and the smoothness of the function with respect to the operator, defined through the source sets. If the operator maps the Hilbert space onto itself, then model (2.4) is well appropriated and various smoothness assumptions are clearly settled for each parameter of the estimation problem. This also explains the particular interest of this model to analyze inverse problems. Otherwise, the action of the operator is more difficult to understand. So it is not possible to separate anymore the two effects and the source sets are an esay tool to provide regularity assumptions to functions studied in an ill-posed framework. Hopefully, the rates of convergence are the same.

In an econometrical settings, the problem is quite distinct since the operators are unknown, leading to different rates of convergence. Such rates are given in (Darolles et al., 2004) and (Hall and Horowitz, 2005) but for different assumptions. We obtain the following theorems

Theorem 4.3 (Consistency in Hilbert scale framework). *Consider the following assumptions. A is ill-posed with degree a and $x_0 = \sum_{k\geq 1}^{\infty} x_k \phi_k$ with $x_k \sim k^{-s}$. Under technical additional assumptions and for optimal choices of the smoothing parameter, we can construct an estimate converging at the following rate of convergence*

$$\mathbf{E}||\hat{x} - x_0||_2^2 = O\left(n^{-\frac{2s-1}{2s+a}}\right).$$

This theorem can be found in (Hall and Horowitz, 2005).

Theorem 4.4 (Consistency under determinist maxiset assumption). *Assume that $x \in \Phi_\beta$, then a kernel estimate with an optimal choice of the bandwidth leads to the following rate of convergence.*

$$\mathbf{E}||\hat{x} - x_0||_2^2 = O\left(n^{-\frac{\beta}{2+\beta}}\right).$$

The proof of this result can be found in (Darolles et al., 2004).

Using results from previous section, we can provide an interpretation of the previous rates of convergence. Assume that the operator is ill-posed of order a and that $x_0 \in W^s$ then we obtain that if moreover $x_0 \in \Phi_\beta$, then we can write

$$\beta = \frac{2s}{a}.$$

Theorem 4.4 implies the following rate of convergence

$$\mathbf{E}||\hat{x} - x_0||_2^2 = n^{-\frac{\beta}{2+\beta}}.$$

Using the correspondence, we get that

$$\mathbf{E}||\hat{x} - x_0||_2^2 = n^{-\frac{s}{a+s}},$$

which corresponds to the optimal rate of convergence under source assumptions with a noise in $\delta = \frac{1}{n}$. So the rate in $n^{-\frac{\beta}{2+\beta}}$ can be seen as a rate of converge in a deterministic setting with source sets conditions. The noise is deterministic in the sense that it doest not depend on the level of noise in the observation data but is defined as the noise coming from the estimation of the unknown operator.

The differences between stochastic and deterministic rates of convergence do not depend on the smoothness assumptions but rely on the nature of the observation noise. They are studied in the next section.

5 Comparison between Deteministic and Stochastic settings

Consider the inverse regression framework (2.1). The difference between the deterministic and the stochastic case is not in the inverse problem but in the definition of the observation

noise. Indeed, in the deterministic case, the noise ξ is an element of the Hilbert space Y such that $||\xi|| \leq 1$. One could think the stochastic noise as a random element of Y, but the situation is quite different. The noise is not defined any more as an element of the Hilbert since for a Gaussian white noise, $||\xi|| = \infty$, hence ξ is a distribution over Y.
Let $\phi_k,\ k \geq 1$ be an orthonormal basis of Hilbert space Y, then ξ is defined by the inner products

$$\forall k \geq 1,\ <\xi, \phi_k>= \xi_k \overset{\text{i.i.d}}{\sim} \mathcal{N}(0,1).$$

In the Gaussian white noise, the series $\sum_{k=1}^{\infty} <\xi, \phi_k> \phi_k$ is divergent.

As a result, the stochastic noise is of complete different nature than the deterministic noise, leading to a more difficult estimation issue. It is highlighted by the rates of convergence obtained in Section 4. Recall that for polyniomaly ill-posed problems the determinist rate is given by

$$r_{\text{det}}(\delta) = ||\hat{x} - x_0||^2 = O(\delta^{\frac{4s}{2s+2a}}),$$

while the stochastic rate is given by

$$r_{\text{sto}}(\delta) = \mathbf{E}||\hat{x} - x_0||^2 = O(\delta^{\frac{4s}{2s+2a+1}}).$$

Hence when the inverse problem vanishes as $a \to 0$, when obtain the rates considering to the regression problem

$$r_{\text{det}}(\delta) \to O(\delta^2), \quad r_{\text{sto}}(\delta) \to O(\delta^{\frac{4s}{2s+1}}).$$

Hence denoising in the deterministic noise can be achieved at the parametric rate of convergence. On the contrary the efficiency of a denoising procedure depends on the regularity of the function, here defined by the parameter s. If we consider the sequence model (2.4), the error contribution in the deterministic case is smaller when the resolution level k increases, while the stochastic error is of same order for each coefficient y_k. So in the statistical estimation problem, there is a trade-off between bias and variance, which does not exist in the deterministic case. The whole set of available data can be used, leading to the parametric rate of convergence.

As a conclusion, there is a huge difference between the two situations. To introduce some links between the two cases, let us consider a Y-value Gaussian noise as follows

$$\forall k \geq 1, \quad <\xi^b, \phi_k>= b_k \xi_k, \quad \xi_k \overset{\text{i.i.d}}{\sim} \mathcal{N}(0,1),\ b_k \in l^2(\mathbb{N}). \tag{5.1}$$

In this case the noise is defined as

$$\xi^b = \sum_{k=1}^{\infty} <\xi^b, \phi_k> \phi_k,$$

which converges almost surely in Y. This definition is close to the one used by Munk et al. in (Bissantz et al., 2007). We can associate to ξ^b a linear bounded operator Λ_{ξ^b} acting from Y to the space L^2, such that for any $\phi \in Y$,

$$<\phi, \xi^b>:= \Lambda_{\xi^b} \phi \in L^2.$$

So, the operator can be written as follows

$$\Lambda_{\xi^b} = \sum_{k=1}^{\infty} b_k \xi_k < \phi_k, . > \ .$$

The same definition can be extended for Gaussian white noise with $b_k = 1$.

The following theorem gives the optimal rate of convergence for an ill-posed inversed problem with an adapted stocastic noise.

Theorem 5.1. *Consider the following model*

$$y_k = \lambda_k x_k + \delta b_k \xi_k, \ k \geq 1.$$

Then, with $b_k \asymp k^{-\beta}$ for $\beta \in [0, a]$,

$$e^{\delta}_{\text{stoc}}(A, X^R_s, \xi^{\beta}) \asymp \delta^{\frac{s}{s+a-\beta+1/2}}.$$

Proof. The proof of this theorem implies only rewritting the model with different eigenvalues. Indeed the model can be stated as follows

$$z_k = x_k + \delta k^{a-\beta} \xi_k,$$

leading to the rate of convergence in $\delta^{\frac{s}{s+a-\beta+1/2}}$. □

Remark that for $\beta = \frac{1}{2}$, the optimal rate of convergence obtained in the stochastic framework coincides with the deterministic one. However the noise only belongs to Y for $\beta > 1/2$. To get convergence, we need to consider $b_k \asymp k^{-1/2} \log^{-1}(k+1)$. For this particular choise, we obtain the equivalence between stochastic and determinist framework, up to a logarithmic factor. This particular point is also highlighted in (Cavalier, 2003).

In a maxiset point of view, the situation is also different. Indeed maxisets in a deterministic framework are defined as saturation spaces in Section 3. In a statistical settings and with the model defined in (2.4), maxisets appear as weak Besov spaces, whose nature is far more complicated than previous sets. For more reference about this topic, we refer to (Rivoirard, 2004).

6 Maxisets for Inverse Problems

6.1 Model

In this section, we consider the following heteroscedastic white noise model:

$$y_k = x_k + \delta \mu_k \xi_k, \quad k = 1, 2, \ldots \tag{6.1}$$

where $x = (x_k)_{k \geq 1}$ is an unknown sequence to be estimated by using observations $(y_k)_{k \geq 1}$, $\delta > 0$ is a small parameter and $(\xi_k)_{k \geq 1}$ is an independent and identically distributed (i.i.d.) sequence of Gaussian variables with mean zero and unit variance. Point also that $\mu_k = \lambda_k^{-1}$. Along this paper, we assume that $\mu = (\mu_k)_{k \geq 1}$ is a known sequence of positive real numbers.

This heteroscedastic white noise model, that appears as a generalization of the classical white noise model (for which, we have $\forall\ k \geq 1,\ \mu_k = 1$), is extensively used by statisticians. Let us briefly recall the reasons for this large use and provide references. Given a known linear operator A, we use the heteroscedastic white noise model when we have to estimate the solution f of the linear equation $g = Af$, with noisy observations of g. Most of the time, to deal with such a problem, we exploit the singular value decomposition of A and the sequence $(\mu_k^{-2})_{k\geq 1}$ is then the eigenvalues sequence of the operator A^*A, with A^* the adjoint of A.
In the wavelet context, (Johnstone, 1999) and (Johnstone and Silverman, 1997) explained that the heteroscedastic white noise model can also be used to represent direct observations with correlated structure. More precisely, let us assume that we are given the following non parametric regression model:

$$Y_i = f(\frac{i}{n}) + e_i, \quad i \in \{1, 2, \ldots, n\}, \tag{6.2}$$

where n is an integer, f is the signal to be estimated and the e_i's are drawn from a stationary Gaussian process. By studying the autocorrelation function of the e_i's , (Johnstone, 1999) and (Johnstone and Silverman, 1997) showed that under a good choice of δ and $\mu = (\mu_k)_{k\geq 1}$, the model (6.1) appears as a good approximation of the model (6.2) when n is large.

6.2 The maxiset theory and functional spaces

Let us first motivate the introduction of the maxiset point of view. When non parametric problems are explored, the minimax theory is the most popular point of view: it consists in ensuring that the used procedure $\hat{x} = (\hat{x_k}(y_k))_{k\geq 1}$ achieves the best rate on a given sequence space $\mathcal{S}$. But, at first, the choice of $\mathcal{S}$ is arbitrary (what kind of spaces has to be considered: Sobolev spaces? Besov spaces? why?), secondly, $\mathcal{S}$ could contain sequences very difficult to estimate. Since the unknown quantity $x = (x_k)_{k\geq 1}$ could be easier to estimate, the used procedure could be too pessimistic and not adapted to the data. More embarrassing in practice, several minimax procedures may be proposed and the practitioner has no way to decide but his experiment. To answer these issues, an other point of view has recently appeared: the maxiset point of view introduced by (Cohen et al., 2001) and (Kerkyacharian and Picard, 2000). Given an estimate $\hat{\theta}$, it consists in deciding the accuracy of $\hat{x}$ by fixing a prescribed rate ρ_δ and to point out the set of all the sequences x that can be estimated by the procedure $\hat{\theta}$ at the target rate ρ_δ. So, under the statistical model (6.1), we introduce the following definition.

Definition 6.1. Let $1 \leq p < \infty$ and any estimator $\hat{x} = (\hat{x}_k(y_k))_{k\geq 1}$, the maxiset of $\hat{x}$ associated with the rate ρ_δ and the l_p-loss is

$$MS(\hat{x}, \rho_\delta, p) = \left\{ x = (x_k)_{k\geq 1} : \quad \sup_\delta \left[\left(\mathbb{E} \sum_{k\geq 1} |\hat{x}_k(y_k) - x_k|^p \right)^{\frac{1}{p}} \rho_\delta^{-1} \right] < \infty \right\}.$$

The maxiset point of view brings answers to the previous issues. Indeed, there is no a priori assumption on x and then, the practitioner does not need to restrict his study to an arbitrary sequence space. The practitioner states the desired accuracy and then, knows the quality of the used procedure. Obviously, he chooses the procedure with the largest maxiset. Let us give first examples of maxiset results in the statistical framework of this paper. For this purpose, we need to introduce the following sequence spaces.

Definition 6.2. For all $1 \leq p < \infty$ and $0 < \eta < \infty$, we set:

$$B^{\eta}_{p,\infty} = \left\{ x = (x_k)_{k \geq 1} : \quad \sup_{\lambda > 0} \lambda^{p\eta} \sum_{k \geq \lambda} |x_k|^p < \infty \right\},$$

and if q is a real number such that $0 < q < p$, we set

$$wl_{p,q}(\mu) = \left\{ x = (x_k)_{k \geq 1} : \quad \sup_{\lambda > 0} \lambda^{q} \sum_{k} \mathbf{1}_{|x_k| > \lambda \mu_k} \mu_k^p < \infty \right\}.$$

6.3 Maxisets for thresholding and linear rules

In this section, under the model (6.1), we estimate each x_k by using thresholding rules. More precisely, we focus on thresholding rules associated with the universal threshold $\lambda_{k,\delta} = \mu_k \delta \sqrt{\log{(1/\delta)}}$ (see (Donoho and Johnstone, 1994)) : for all $\delta > 0$, we assume that we are given a real number $\Lambda_\delta > 0$ only depending on δ and tending to $+\infty$ when δ tends to 0, and we set:

$$\hat{x}^t_k(y_k) = \begin{cases} y_k \mathbf{1}_{|y_k| \geq \kappa_* \lambda_{k,\delta}} & \text{if } k < \Lambda_\delta, \\ 0 & \text{otherwise,} \end{cases}$$

where κ_* is a constant. (Kerkyacharian and Picard, 2000) have studied the maxisets for this procedure. They obtained the following result for $\hat{x}^t = (\hat{x}^t_k)_{k \in \mathbb{N}^*}$:

Theorem 6.1. *Let $1 \leq p < \infty$ be a fixed real number and $0 < r < \infty$. We suppose that*

$$\forall\ 0 < \delta \leq \delta_0, \quad \Lambda_\delta = \left(\delta \sqrt{\log{(1/\delta)}}\right)^{-r},$$

where δ_0 is such that $\delta_0 \sqrt{\log(1/\delta_0)} = 1$, and there exists a positive constant T, such that $\forall\ 0 < \delta \leq \delta_0$,

$$\delta^{\frac{\kappa_*^2}{16}} \log{(1/\delta)}^{-\frac{1}{4} - \frac{p}{2}} \sum_{k < \Lambda_\delta} \mu_k^p \leq T.$$

Let q be a fixed positive real number such that $q < p$.
Then, if $\kappa_* \geq \sqrt{2p}$,

$$MS\left(\hat{x}^t, \left(\delta\sqrt{|\log \delta|}\right)^{(1-q/p)}, p\right) = wl_{p,q}(\mu) \cap B^{\frac{1}{r}(1-q/p)}_{p,\infty}.$$

For the same statistical model, (Rivoirard, 2004) proves that under some mild conditions, the maxisets associated with linear estimates of the form $(l_k x_k)_{k \in \mathbb{N}^*}$, where $(l_k)_{k \in \mathbb{N}^*}$ is a non increasing sequence of weights lying in $[0, 1]$, are Besov bodies. These conditions are checked for instance by projection weights, Tikhonov-Phillips weights or Pinsker weights. For further details see (Rivoirard, 2004). We can add that Rivoirard proved that for the rate $\left(\delta\sqrt{\log{(1/\delta)}}\right)^{(1-q/p)}$, the maxisets of linear estimates are strictly included into the maxisets of thresholding rules. It means that from the maxiset point of view, linear estimates are outperformed by thresholding ones, as mentioned in Introduction.

In the next section, we consider a Bayesian model and we evaluate the maxisets respectively for the median and the mean of the posterior distribution.

6.4 Maxisets for Bayes rules

The sequence x to be estimated is supposed to be sparse. With this in mind, we wish to estimate each x_k by using Bayes rules and we consider the following Bayesian model: we suppose that we are given a fixed unimodal density γ, assumed to be positive on $\mathbb{R}$, symmetric about 0 and such that there exist two positive constants M and M_1 such that

$$(H_1) \qquad \sup_{\theta \geq M_1} \left| \frac{d}{dx} \log \gamma(x) \right| = M < \infty.$$

Then, we assume that the x_k's are independent and $\forall\ k \in \mathbb{N}^*$,

$$(M_1) \qquad x_k \sim w_{k,\delta}\gamma_{k,\delta}(x_k) + (1 - w_{k,\delta})\delta_0(x_k),$$

where $\forall\ x \in \mathbb{R}$,

$$\gamma_{k,\delta}(x) = s_{k,\delta}\gamma(s_{k,\delta}x), \quad s_{k,\delta} = (\delta\mu_k)^{-1},$$

and $w_{k,\delta}$ is a real number lying in $(0,1)$. The hypothesis (H_1) will be useful to determine maxisets for Bayes rules. Assumption (H_1) implies that

$$\forall\ u \geq M_1, \quad \gamma(u) \geq \gamma(M_1)\exp(-M(u - M_1)).$$

It means that the tails of γ have to be exponential or heavier. It can be shown (see (Rivoirard, 2005)) that this assumption is essential to get maxisets as large as possible. Furthermore, $w_{k,\delta} = w_\delta$ depends only on δ and we shall assume throughout this paper that $\pi_\varepsilon = (1 - w_\delta)w_\delta^{-1}$ satisfies the following mild assumptions, globally denoted (H_2).

1. $\delta \longrightarrow \pi_\varepsilon$ is continuous,
2. $\inf_{\delta > 0} \pi_\delta > 1$,
3. $\pi_1 = \exp(1)$,
4. $\pi_\varepsilon \overset{\delta \to 0}{\longrightarrow} +\infty$,
5. $\delta\sqrt{\log \pi_\varepsilon} \overset{\delta \to 0}{\longrightarrow} 0$.

In this section, we consider the Bayes rules associated with the l_1-loss (the median of the posterior distribution) and with the l_2-loss (the mean of the posterior distribution). For all $\delta > 0$, we assume that we are given a real number $\Lambda_\delta > 1$ depending only on δ and tending to $+\infty$ when δ tends to 0. We estimate each x_k by $\hat{x}_k^{b_1}(y_k)$ or by $\hat{x}_k^{b_2}(y_k)$ defined by the following procedure.

- If $k < \Lambda_\delta$, $\hat{x}_k^{b_1}(y_k)$ (respectively $\hat{x}_k^{b_2}(y_k)$) is the median (respectively the mean) of the posterior distribution of x_k given y_k.
- If $k \geq \Lambda_\delta$, $\hat{x}_k^{b_1}(y_k) = \hat{x}_k^{b_2}(y_k) = 0$.

We have the following result for the posterior median.

Theorem 6.2. *We assume that (H_1) and (H_2) hold. Let $0 < r < \infty$ and $1 \leq p < \infty$ be two fixed real numbers. We suppose that $\forall\ \delta > 0$,*

$$\Lambda_\delta = (\delta\sqrt{\log \pi_\varepsilon})^{-r},$$

and there exist two positive constants T_1 and T_2, such that $\forall\ \delta > 0$,

$$\delta^{-p}\ \pi_\varepsilon^{-1}(\log \pi_\varepsilon)^{-\frac{1}{2}-\frac{p}{2}} \leq T_1, \tag{6.3}$$

$$\pi_\varepsilon^{-\frac{1}{8}}(\log \pi_\varepsilon)^{-\frac{1}{4}-\frac{p}{2}} \sum_{k<\Lambda_\delta} \mu_k^p \leq T_2. \tag{6.4}$$

Let q be a fixed positive real number such that $q < p$.
Then,

$$MS\left(\hat{x}^{b_1}, (\delta\sqrt{\log \pi_\varepsilon})^{1-q/p}, p\right) = wl_{p,q}(\mu) \cap B_{p,\infty}^{\frac{1}{r}(1-q/p)}.$$

For the posterior mean, we have:

Theorem 6.3. *We assume that (H_1) and (H_2) hold. Let $0 < r < \infty$ and $1 \leq p < \infty$ be two fixed real numbers. We suppose that $\forall\ \delta > 0$,*

$$\Lambda_\delta = (\delta\sqrt{\log \pi_\varepsilon})^{-r},$$

and there exist two positive constants T_1 and T_2, such that $\forall\ \delta > 0$,

$$\delta^{-p}\ \pi_\varepsilon^{-\frac{1}{4}}(\log \pi_\varepsilon)^{-\frac{1}{2}-\frac{p}{2}} \leq T_1, \tag{6.5}$$

$$\pi_\varepsilon^{-\frac{1}{32}}(\log \pi_\varepsilon)^{-\frac{1}{4}-\frac{p}{2}} \sum_{k<\Lambda_\delta} \mu_k^p \leq T_2. \tag{6.6}$$

Let q be a fixed positive real number such that $q < p$. Then,

$$MS\left(\hat{x}^{b_2}, (\delta\sqrt{\log \pi_\varepsilon})^{1-q/p}, p\right) = wl_{p,q}(\mu) \cap B_{p,\infty}^{\frac{1}{r}(1-q/p)}.$$

For the proofs of these results, see (Rivoirard, 2005). When π_ε is a power of δ, then, by using Theorem 6.1, we can compare the Bayesian procedures $\hat{x}^{b_1}$ and $\hat{x}^{b_2}$ with the thresholding one. We can conclude that each of them achieves the same performance as the thresholding one. Finally, since linear estimates are outperformed by thresholding ones, they are also outperformed by $\hat{x}^{b_1}$ and $\hat{x}^{b_2}$.

6.5 Some remarks

Maxiset Procedure

The comparison of procedures using maxisets is not as famous as minimax comparison. However the results that have been obtained up to now are very promising since they generally show that the maxisets of well-known procedures are spaces that are well established and easily interpretable. Indeed, in the maxiset approach, the Besov bodies (the spaces $B_{p,\infty}^{\eta}$)

control the x_k's for the large values of k. As for the spaces $wl_{p,q}(\mu)$, they can be viewed as weighted weak l_q spaces. The weak l_q space is the space $wl_{p,q}(\mu)$ when $\mu_k = 1$ for any $k \geq 1$, so we denote it $wl_{p,q}(1)$. This space was studied in approximation theory and coding by (DeVore, 1989). We easily see that if we order the components of a sequence x according to their size:

$$|x|_{(1)} \geq |x|_{(2)} \geq \cdots \geq |x|_{(n)} \geq \ldots,$$

then

$$x \in wl_{p,q}(1) \iff \sup_n n^{\frac{1}{q}} |x|_{(n)} < \infty.$$

So, $wl_{p,q}(1)$ spaces naturally measure the sparsity of a signal. Of course, the weighted versions of these spaces, the $wl_{p,q}(\mu)$ spaces, share the same property.

Other models

As we pointed out in section 3, the regularity spaces Φ_β can be seen as a very particular class maxisets for the deterministic problem of recovering a signal observed in an inverse framework without noise. They characterize the set of functions that can be recovered by a given regularization procedure at a given rate of convergence. It is stricking to see that such sets can in this setting define regularity spaces in the associated random model.

When considering maxisets with observations drawn from model (2.1), we face the problem of discretization. Indeed, as seen in (Autin et al., 2008), the smoothness assumptions needed to ensure the equivalence between this model and the sequential model are stronger than the minimum assumptions necessary to ensure the convergence of the estimator. As a consequence, the maxiset associated to any estimation procedure comes from the discretization assumptions and not from the inverse problem, leading to weak results preventing any interpretation of maxisets results.

7 Conclusion

In this paper, we have linked the different notions of regularity used in the inverse problem litterature in the three main fields, statistics, analysis and econometry. We have showed that under some regularity assumptions, the three main regularity spaces can be compared to each other. This leads to rates of convergence which are well understood in Hilbert scales and when the operator is known and that can be studied in the three settings. Several main differences also appeared, mainly when dealing with observation noise of different nature, i.e deterministic or random noise.

Several questions still remain unadressed, such as whether it is possible to separate regularity conditions over te operator and the unknown function when the operator maps a space into a different one. Neither did we pay attention to the adaption issue and its different understanding in the different frameworks.

Acknowledegments

We are very gratefull to A. Munk, T. Hohage and J-P. Florens for their helpfull discussions.

References

Autin, F., Le Pennec, E., Loubes, J.-M. and Rivoirard, V. 2008. Maxisets for complexity regularization procedures, *Preprint* .

Besov, O. V., Ilin, V. P. and Nikolskiĭ, S. M. 1978. *Integral representations of functions and imbedding theorems. Vol. I*, V. H. Winston & Sons, Washington, D.C. Translated from the Russian, Scripta Series in Mathematics, Edited by Mitchell H. Taibleson.

Bissantz, N., Hohage, T., Munk, A. and Ruymgaart, F. 2007. Convergence rates of general regularization methods for statistical inverse problems and applications, *SIAM J. Numer. Anal.* **45**(6): 2610–2636 (electronic).

Carasco, M., Florens, J.-P. and Renault, E. 2004. Linear inverse problems in structural econometrics, *Handbook of Econometrics* .

Cavalier, L. 2003. Habilitation à diriger les recherches, *Université Aix-Marseille 1* .

Cavalier, L., Golubev, G. K., Picard, D. and Tsybakov, A. B. 2002. Oracle inequalities for inverse problems, *Ann. Statist.* **30**(3): 843–874. Dedicated to the memory of Lucien Le Cam.

Cavalier, L. and Tsybakov, A. 2002. Sharp adaptation for inverse problems with random noise, *Probab. Theory Related Fields* **123**(3): 323–354.

Chow, P. and Khasminskii, R. 1997. Statistical approach to dynamical inverse problems, *Communications in Mathematical Physics* **189**(2): 521–531.

Cohen, A., DeVore, R., Kerkyacharian, G. and Picard, D. 2001. Maximal spaces with given rate of convergence for thresholding algorithms, *Appl. Comput. Harmon. Anal.* **11**(2): 167–191.

Cohen, H. and Reiss 2003. Adaptive wavelet galerkin methods for linear inverse problems, *SIAM* **1**(3): 323–354.

Darolles, S., Florens, J.-P. and Renault, E. 2004. Nonparametric instrumental regression, *preprint* .

DeVore, R. A. 1989. Degree of nonlinear approximation, *Approximation theory VI, Vol. I (College Station, TX, 1989)*, Academic Press, Boston, MA, pp. 175–201.

Donoho, D. 1995. Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition, *Appl. Comput. Harmon. Anal.* **2**(2): 101–126.

Donoho, D. L. and Johnstone, I. M. 1994. Ideal spatial adaptation by wavelet shrinkage, *Biometrika* **81**(3): 425–455.

Donoho, D. L. and Johnstone, I. M. 1999. Asymptotic minimaxity of wavelet estimators with sampled data., *Stat. Sin.* **9**(1): 1–32.

Engl, H. W. 2000. Regularization methods for solving inverse problems, *ICIAM 99 (Edinburgh)*, Oxford Univ. Press, Oxford, pp. 47–62.

Engl, H. W., Hanke, M. and Neubauer, A. 1996. *Regularization of inverse problems*, Vol. 375 of *Mathematics and its Applications*, Kluwer Academic Publishers Group, Dordrecht.

Hall, P. and Horowitz, J. 2005. Nonparametric methods for inference in the presence of instrumental variables, *Ann. Statist.* **33**(6): 2904–2929.

Johnstone, I. M. 1999. Wavelet shrinkage for correlated data and inverse problems: adaptivity results, *Statist. Sinica* **9**(1): 51–83.

Johnstone, I. M. and Silverman, B. W. 1997. Wavelet threshold estimators for data with correlated noise, *J. Roy. Statist. Soc. Ser. B* **59**(2): 319–351.

Kalifa, J. and Mallat, S. 2003. Thresholding estimators for linear inverse problems and deconvolutions, *The Annals of Statistics* **31**(1): 58–109.

Kaltenbacher, B. 2000. Regularization by projection with a posteriori discretization level choice for linear and nonlinear ill-posed problems., *Inverse Probl.* **16**(5): 1523–1539.

Kerkyacharian, G. and Picard, D. 2000. Thresholding algorithms, maxisets and well-concentrated bases, *Test* **9**(2): 283–344.

Loubes, J.-M. and Ludeña, C. 2008. Choosing a penalty for regularized estimators of inverse problems, *accepted in Electronic Journal of Statistics* .

Loubes, J.-M. and Vanhems, A. 2004. Saturation spaces in inverse problems in econometrics, *Proceedings of the American Econometric Society* .

Mair, B. A. and Ruymgaart, F. H. 1996. Statistical inverse estimation in Hilbert scales, *SIAM J. Appl. Math.* **56**(5): 1424–1444.

Natterer, F. 1984. Error bounds for tikhonov regularization in hilbert scales, *Applic. Analys.* **19**: 29–37.

Neubauer, A. 1992. Tikhonov regularization of nonlinear ill-posed problems in Hilbert scales, *Appl. Anal.* **46**(1-2): 59–72.

Neubauer, A. 1997. On converse and saturation results ofr tikhonov regularization of ill-posed problems, *SIAM J. Numer. Anali* **33**(2): 517–527.

O'Sullivan, F. 1986. A statistical perspective on ill-posed inverse problems, *Statist. Sci.* **1**(4): 502–527. With comments and a rejoinder by the author.

O'Sullivan, F. 1990. Convergence characteristics of methods of regularization estimators for nonlinear operator equations., *SIAM J. Numer. Anal.* **27**(6): 1635–1649.

Rivoirard, V. 2004. Maxisets for linear procedures, *Statist. Probab. Lett.* **67**(3): 267–275.

Rivoirard, V. 2005. Bayesian modelling of sparse sequences and maxisets for bayes rules, *To appear in Mathematical Methods of Statistics.* .

Snieder, R. 1991. An extension of Backus-Gilbert theory to nonlinear inverse problems., *Inverse Probl.* **7**(3): 409–433.

Tautenhahn, U. and Jin, Q.-n. 2003. Tikhonov regularization and a posteriori rules for solving nonlinear ill posed problems, *Inverse Problems* **19**(1): 1–21.

Tikhonov, A. N. and Arsenin, V. Y. 1977. *Solutions of ill-posed problems*, V. H. Winston & Sons, Washington, D.C.: John Wiley & Sons, New York. Translated from the Russian, Preface by translation editor Fritz John, Scripta Series in Mathematics.