Statistica Sinica **19** (2009), 785-800

OPTIMAL BOUNDS FOR INVERSE PROBLEMS WITH JACOBI-TYPE EIGENFUNCTIONS

Thomas Willer

Université de Provence

Abstract: We consider inverse problems where one wishes to recover an unknown function from the observation of a transformation of it by a linear operator, corrupted by an additive Gaussian white noise perturbation. We assume that the operator admits a singular value decomposition where the eigenvalues decay in a polynomial way, and where Jacobi polynomials appear as eigenfunctions. This includes, as an application, the well known Wicksell's problem. We establish asymptotic lower bounds for the minimax risk in a wide framework (i.e., with $(L^p)_{1<p<\infty}$ losses and Besov-like regularity spaces), which shows that the estimator of Kerkyacharian, Picard, Petrushev, and Willer (2007) is quasi-optimal, and thus yields the minimax rates. We also establish some new results on the needlets introduced by Petrushev and Xu (2005) which appear as essential tools in this setting. Lastly we discuss the interest of the results concerning the treatment of inverse problems by wavelet procedures.

Key words and phrases: Minimax estimation, second-generation wavelets, statistical inverse problems.

1. Motivation

We consider the problem of recovering a function f from a blurred and noisy version Y:

$$\forall v \in V, \quad Y(v) = (Kf, v)_V + \epsilon\xi(v),$$

where K is a linear operator between two Hilbert spaces: $K : U \mapsto V$, ξ is a Gaussian white noise on V, and for H a Hilbert space and $h_1, h_2 \in H$, $(h_1, h_2)_H$ denotes the scalar product in H of h_1 and h_2. We assume that f belongs to $U = L^2([-1, 1], \mu(x)dx)$, with $\mu(x) = (1-x)^\alpha(1+x)^\beta$, $\alpha, \beta > -1/2$, and that K admits a singular value decomposition (SVD), i.e., there exists an orthonormal basis (called the SVD basis) formed by the eigenfunctions of the self-adjoint operator K^*K (where K^* is the adjoint of K). Moreover we assume that this SVD basis consists of the classical Jacobi polynomials of type (α, β), and that the corresponding sequence of eigenvalues tends to zero at a polynomial rate. We name such problems "Jacobi-type inverse problems".

The main motivation of this article is to establish asymptotic lower bounds for the minimax risk in a wide framework, considering $L^p([-1, 1], \mu)$ losses, for all $1 < p < \infty$, and a Besov-like regularity space. This, combined with the result of Kerkyacharian, Picard, Petrushev and Willer (2007) (where upper bounds are provided), shows some new rate phenomenom for inverse problems.

1.1. The results

The most popular technique for the treatment of inverse problems is probably singular value decomposition estimation, where the unknown function is expanded in the SVD basis, and the corresponding coefficients are estimated thanks to Y. Such techniques are very attractive theoretically and can be shown to be asymptotically minimax in many situations (see e.g. Mathé and Pereverzev (2003), Cavalier and Tsybakov (2002), Cavalier, Golubev, Picard and Tsybakov (2002), Tsybakov (2000), Goldenshluger and Pereverzev (2003)). However there are limitations to the minimax framework, in particular such estimators generally cannot estimate functions exhibiting inhomogeneous regularity. To avoid this problem, several wavelet methods have been introduced during the last decade (for example Donoho (1995) and Abramovich and Silverman (1998)), which are minimax over wide sets of target functions, for example Besov spaces. Nevertheless such methods apply only to a category of inverse problems where the operator is well-adapted to the structure of "first generation" wavelets, which are built from a Fourier analysis perspective. Thus many wavelet estimators are available whenever the operator displays some convolution structure (see for instance Pensky and Vidakovic (1999), Fan and Koo (2002) and Kalifa and Mallat (2003)).

The main interest here is to grapple with quite different inverse problems, where the operator displays a polynomial structure. Classical wavelets cannot be used, and new estimation techniques have been given by Kerkyacharian, Picard, Petrushev and Willer (2007); one uses new wavelets built upon polynomials (termed needlets, and introduced by Petrushev and Xu (2005)) to develop the "NEEDD" estimator, and new spaces (which appear as adaptations of the classical Besov spaces) to assess its performances. Here we establish a lower bound for the minimax risk that matches the rate of convergence of NEEDD (up to log factors). Consequently we obtain the minimax rates in all the Jacobi-type inverse problems, and we prove the quasi-optimality of NEEDD. Note also that the results are established for all $L^p([-1,1],\mu)$ losses, whereas in most works cited previously, only the case $p = 2$ is considered, with one exception: for the deconvolution problem in a periodic setting, Johnstone, Kerkyacharian, Picard and Raimondo (2004), combined with Willer (2005), established the minimax rates for all $L^p([0,1],dx)$ losses over Besov spaces. We draw a parallel between those rates and the ones obtained here: we exhibit elbow effects, and we show that the rates in the deconvolution model appear as a critical case of the rates in the Jacobi-type model. Moreover, we give an application of our results to Wicksell's problem, which satisfies the required assumptions on the operator. This problem concerns the recovery of the density of the radii of spherical particles when a sample of planar cuts is given, and has many applications in medecine and in biology.

In this paper, we only consider standard inverse problems where the operator is known. Recently, SVD or wavelet estimators have also been developed for noisy operators (see e.g. Efromovich and Koltchinskii (2002), Cavalier and Hengartner (2005), Cavalier and Raimondo (2007), or Hoffmann and Reiss (2008)), and it may be interesting in the future to expand our results to that setting.

1.2. The idea

The main idea behind NEEDD is to decompose the problem by using a family of functions (the needlets) which in some sense "both quasi-diagonalizes the operator K and the prior information on f" (to use the terms in Donoho (1995)). In the lower bound problem here, a similar problem arises, as we need a family of functions $\{f_\lambda,\ \lambda \in \Lambda\} \subset U$ representative of the difficulties of estimation inside the regularity space considered for the risk. This means that the functions f_λ must be chosen such that:

- they are distant from one another in $L^p(\mu)$ norm,
- at the same time the distributions of the associated processes Y are close to one another (in a Kullback sense, for example).

A natural way to build such hypotheses is to use functions that enjoy localization properties, and whose images by K can be easily studied; here again needlets are an essential tool. The hypotheses are built as linear combinations of such functions, with some parameters left free, which we adjust optimally with respect to the two constraints cited above. Then the minimal $L^p(\mu)$ distance between the hypotheses yields the lower bound on the entire regularity space. This approach of combining wavelets and lower bound techniques is classical, but the main tool used here - the needlets - is not: properties of needlets are still not thoroughly known, and in several ways they do not behave like classical wavelets. Thus, in Section 5.4, we give a brief list of needlet properties and establish some of them. We show that, in particular, the non-orthogonality of the needlets and the heterogeneity of their $L^p(\mu)$ norms make the lower bound problem more difficult than in other inverse problems, such as deconvolution for example (for which a proof using the classical Meyer wavelets can be found in Willer (2005)).

The paper is organized as follows. In Section 2 we describe the model and state the main result, in Section 3 we give an application to Wicksell's problem, and in Section 4 we discuss the interest of the results within the literature on inverse problems. Lastly, in Section 5, we give the proof of the main theorem.

2. Main Result

2.1. Model and assumptions

We are interested in nonparametric inverse problems in white noise, with a polynomial structure of the operator. We define this framework as follows. Let f be an unknown function belonging to the Hilbert space $U = L^2([-1,1], \mu(x)dx)$, with $\mu(x) = (1-x)^\alpha(1+x)^\beta$, $\alpha, \beta > -1/2$. The estimation problem consists of recovering a good approximation to the function f from the observation of the random variable Y corresponding to a blurred and noisy version of f:

$$\forall v \in V, \quad Y(v) = (Kf, v)_V + \epsilon\xi(v). \tag{2.1}$$

Blurring effect: Let $I = [a,b]$ or $I = [a,b[$, with $-\infty < a < b \leq \infty$, and $\lambda : I \mapsto \mathbb{R}_+^*$ a continuous function. We set $V = L^2(I, \lambda(x)dx)$. Let $K : U \mapsto V$ be a linear operator satisfying the two following conditions. First assume K^*K (where K^* denotes the adjoint of K) is diagonalizable, with a countable set of eigenvalues (denoted $(b_k^2)_{k\in\mathbb{N}}$) that are strictly positive and decrease at a polynomial rate for some ill-posedness coefficient $\nu > 0$ (for two positive sequences (u_k) and (v_k), the notation $u_k \asymp v_k$ means that there exist $0 < c_1 \leq c_2 < \infty$ such that $c_1 v_k \leq u_k \leq c_2 v_k$): $\forall k \in \mathbb{N}^*, \quad b_k \asymp k^{-\nu}$. Second, assume that the classical Jacobi polynomials normalized in U (we denote by $P_k^{\alpha,\beta}$, or simply P_k, the polynomial of degree k) appear as an orthonormal basis of eigenfunctions of K^*K. So P_k is the polynomial of degree k such that $\int_{-1}^1 P_k P_l d\mu = \delta_{k,l}$, and $\forall k \in \mathbb{N}$, $K^*KP_k = b_k^2 P_k$.

Noise effect: $\epsilon > 0$ is deterministic, and ξ is a Gaussian white noise on V:

$$\forall v, w \in V, \begin{cases} \xi(v) \sim \mathcal{N}(0, \|v\|_V^2), \\ E[\xi(v)\xi(w)] = (v,w)_V. \end{cases}$$

2.2. Minimax rates

The aim of the paper is to establish asymptotic minimax rates (when $\epsilon \to 0$) for the inverse problems described above, in a wide framework, i.e., for numerous choices of functions f and of measures of estimation errors. For the latter, we consider all $\mathbb{L}^p(\mu)$ losses (for any $1 < p < +\infty$) defined by, $\forall u \in U$, $\|u\|_{\mathbb{L}^p(\mu)} = [\int_{-1}^1 |u(x)|^p d\mu(x)]^{1/p}$. Concerning the target functions, we introduce spaces $B_{\pi,r}^s(M)$ below, which appear as an adaptation of the classical Besov spaces. Let $(\psi_{j,\eta})_{j\geq 0,\eta\in\mathbb{Z}_j}$ denote the tight frame of needlets described in Section 5.4. For any $f \in U$, we have the following decomposition:

$$f = \sum_{j\geq 0}\sum_{\eta\in\mathbb{Z}_j} \beta_{j\eta}\psi_{j\eta}, \qquad \text{where } \beta_{j\eta} = (f, \psi_{j\eta})_U.$$

Then for $\pi \geq 1$, $s \geq 1/\pi$, $r \geq 1$, $M > 0$ we define

$$B^s_{\pi,r}(M) = \left\{ f \in U \;\Big|\; \left\| \left(2^{js}\Big(\sum_{\eta\in\mathbb{Z}_j} |\beta_{j,\eta}|^\pi \|\psi_{j,\eta}\|^\pi_\pi \Big)^{\frac{1}{\pi}} \right)_{j\geq -1} \right\|_{l^r} \leq M \right\}.$$

If $\psi_{j,\eta}$ were a classical wavelet, then $B^s_{\pi,r}$ would correspond to a Besov space (see e.g. Härdle, Kerkyacharian, Picard and Tsybakov (1998)), which is a very general regularity space including as particular cases Sobolev and Holder spaces, and which can be described very simply, thanks to any regular enough wavelet basis. Such spaces are widely used to study the theoretical performance of wavelet estimators in appropriate inverse problems. However, here the $B^s_{\pi,r}$ correspond to new spaces, characterized by needlets, and appear as a natural alternative to the classical Besov spaces when the inverse problem no longer possesses a convolution structure, but a polynomial structure. Details on this can be found in Narcowich, Petrushev and Ward (2006), and in the appendix of Kerkyacharian, Picard, Petrushev and Willer (2007).

We are interested in the minimax risk

$$R_\epsilon\Big(B^s_{\pi,r}(M), \mathbb{L}^p(\mu)\Big) := \inf_{\hat f} \sup_{f\in B^s_{\pi,r}(M)} \mathbb{E}_f\Big(\|\hat f - f\|^p_{L^p(\mu)}\Big),$$

where the infimum is taken over all $\sigma(Y(t))_{t\geq 0}-$measurable estimators $\hat f$. The results of Kerkyacharian, Picard, Petrushev and Willer (2007), concerning the rates of convergence of the NEEDD estimator, give an immediate upper bound for the risk. This is Theorem 1, where we recall that $\nu > 0$ is a rate of decay of the eigenvalues of the operator ($b_k \asymp k^{-\nu}$), and that $\alpha, \beta > -1/2$ are parameters characterizing U.

Theorem 1. *For all $1 < p < \infty$, $\pi \geq 1$, $r \geq 1$, and $s > \max_{\gamma\in\{\alpha,\beta\}}\{1/2 - 2(\gamma + 1)(1/2 - 1/\pi) \vee 2(\gamma+1)(1/\pi - 1/p) \vee 0\}$, there exists $C > 0$ such that:*

$$R_\epsilon(B^s_{\pi,r}(M), \mathbb{L}^p(\mu)) \leq C\left[\log\Big(\frac{1}{\epsilon}\Big)\right]^{p+1}\left[\epsilon\sqrt{\log\Big(\frac{1}{\epsilon}\Big)}\right]^{\zeta p},$$

where $\zeta = \min\{\zeta(s),\, \zeta(s,\alpha),\, \zeta(s,\beta)\}$, with

$$\zeta(s) = \frac{s}{s+\nu+1/2}, \quad \zeta(s,\gamma) = \frac{s - 2(1+\gamma)(1/\pi - 1/p)}{s+\nu+2(1+\gamma)(1/2-1/\pi)}.$$

The main purpose of the paper is to prove that these rates coincide with the rates of the minimax risk, up to log factors. We will establish the following.

Theorem 2. *For all* $1 < p < \infty$, $\pi \geq 1$, $r \geq 1$, *and* $s \geq 1/\pi$, *there exists* $C > 0$ *such that* $R_\epsilon(B^s_{\pi,r}(M), \mathbb{L}^p(\mu)) \geq C\epsilon^{\zeta p}$, *where* $\zeta = \min\{\zeta(s),\ \zeta(s,\alpha),\ \zeta(s,\beta)\}$ *with:*

$$\zeta(s) = \frac{s}{s+\nu+1/2}, \quad \zeta(s,\gamma) = \frac{s - 2(1+\gamma)(1/\pi - 1/p)}{s+\nu+2(1+\gamma)(1/2-1/\pi)}.$$

Note that the exact logarithmic factors of the minimax risk are not established yet. We focus here only on the main rate ϵ^ζ, so our results prove that NEEDD is "quasi optimal" in Jacobi-type models.

3. Application to the Wicksell's Problem

The Jacobi-type inverse models considered in this paper find applications in practice, in particular with the well-known Wicksell's problem (Wicksell (1925)) that corresponds to the following situation. Suppose a population of spheres is embedded in a medium, with radii that are assumed to be drawn independently from a density f. A random plane slice is taken through the medium, and some spheres are intersected by it. They furnish circles, the radii of which yield the points of observation $Y_1, \ldots, Y_n$, as illustrated in Figure 3.1. The unfolding problem is to determine the density of the spheres radii from the observed circle radii. This problem arises in medicine, where the spheres might be tumors in an animal's liver (Nychka, Wahba, Goldfarb and Pugh (1984)), as well as in numerous other contexts (biology, engineering, etc.), see for instance Cruz-Orive (1983).

If one uses Lebesgue measure then, by a conditioning argument (see Wicksell (1925)) and under some assumptions, the density of the circles radii is, $\forall y \in [0,1]$, $K_0 f(y) = y\int_y^1 (x^2-y^2)^{-1/2} f(x)dx$ (up to a constant). However few articles use this precise formulation of the problem. In the sequel we adopt the version

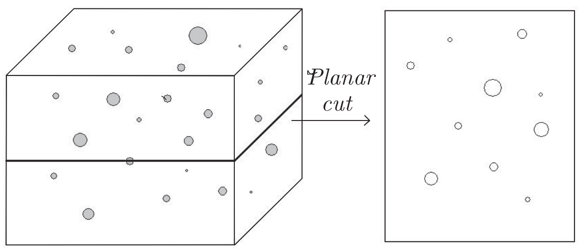

Figure 3.1. Wicksell's problem: observation of radii of disks after a planar cut of spheres.

proposed by Johnstone and Silverman (1991), who replaced Lebesgue measure by two weighted measures. So we observe Y following model (2.1) with $K : \tilde{U} \mapsto V$ given by

$$\begin{cases} \tilde{U} = L^2([0,1], \tilde{\mu}(x)dx), & \tilde{\mu}(x) = (4x)^{-1}, \\ V = L^2([0,1[, \lambda(y)dy), & \lambda(y) = 4\pi^{-1}(1-y^2)^{\frac{1}{2}}, \\ Kf(y) = \frac{\pi}{4} y(1-y^2)^{-\frac{1}{2}} \int_y^1 (x^2-y^2)^{-\frac{1}{2}} f(x) d\tilde{\mu}(x). \end{cases}$$

Johnstone and Silverman (1991) show that K^*K admits the following root eigenvalues and eigenfunctions: $b_k = \pi/16(1+k)^{-1/2}$, $\tilde{P}_k(x) = 4(k+1)^{1/2} x^2 P_k^{0,1}(2x^2 - 1)$. Thus, up to changes in the variables (note $\tilde{U}$ instead of U, and hence the notations $\tilde{P}$ and $\tilde{B}^s_{\pi,r}$ later on), this is a Jacobi-type inverse problem with $(\alpha, \beta, \nu) = (0, 1, 1/2)$. Our results show that NEEDD is a quasi-optimal estimator, and Theorems 1 and 2 establish the rates for the minimax risk R_ϵ^{Wick}. Neglecting $\log(1/\epsilon)$ factors, we have $R_\epsilon^{Wick}[\tilde{B}^s_{\pi,r}(M), \mathbb{L}^p([0,1], x^{3-2p}dx)] \asymp \epsilon^{\zeta p}$, where

$$\zeta = \min\left\{ \frac{s}{s+1}, \frac{s - 2(1/\pi - 1/p)}{s + 3/2 - 2/\pi}, \frac{s - 4(1/\pi - 1/p)}{s + 5/2 - 4/\pi} \right\}.$$

Thus we find rates new to the literature on Wicksell's problem, but of course several comments need to be made. First, we used a transformation, initiated by Johnstone and Silverman (1991), of the original Wicksell problem. Other statistical results are available, but stated in yet another version of the problem, where one considers the *squared* radii of circles and spheres. Then a thorough minimax study can be found in Golubev and Levit (1998) for the estimation of the corresponding distribution function, and in Antoniadis, Fan and Gijbels (2001), convergence rates are established for a wavelet density estimator, but only in $\mathbb{L}^2([0,1], dx)$ norm and over particular Besov spaces. Second, we assumed that the random perturbation is a Gaussian white noise on the space V introduced above, and not a density perturbation as in the original problem. So here we add to the variety of theoretical results on Wicksell: we draw a complete picture of the problem in a minimax perspective, but by using a rather unusual representation. Work still needs to be done to extend our results to a more practical setting; research in that direction is initiated in Chapter 5 of Willer (2006), but a more thorough investigation is under study.

4. Discussion

In the literature on statistical inverse problems, there are few minimax results as general as the one we consider. Usually only the L^2 case is considered, and under the polynomial decay assumption of the eigenvalues, the rate $\zeta = s/(s + \nu + 1/2)$ (named "regular" rate) appears frequently (see Cavalier and Tsybakov

(2002)). For more general L^p losses, only the case of deconvolution in a periodic setting (up to our knowledge) has been studied, in Johnstone, Kerkyacharian, Picard and Raimondo (2004) and in Willer (2005), and elbow effects appear, with a second rate named "sparse". It is interesting to draw a parallel between such a problem, where classical wavelets are widely used tools, and polynomial type problems, which require needlets.

For the deconvolution problem, minimax rates have been established for all $L^p([0,1], dx)$ losses ($1 < p < \infty$), and over balls of a Besov space characterized by parameters $\pi \geq 1$, $s \geq 1/\pi$, $r \geq 1$, as above. Then the rates are given as in Theorem 1 and 2 (up to the logarithmic factors) with ζ replaced by:

$$\zeta = \min\left\{\zeta_{\text{regular}} := \frac{s}{s+\nu+1/2}, \quad \zeta_{\text{sparse}} := \frac{s-1/\pi+1/p}{s+\nu+1/2-1/\pi}\right\}.$$

Then the deconvolution setting appears as a critical case of the Jacobi setting if we set $\alpha = \beta = -1/2$. So the rates in the Jacobi-type models are new, and note that this novelty is not an artifact stemming from the weights on the space, since in the Lebesgue case the rates for the Jacobi scenario (i.e. $\alpha = \beta = 0$) do not coincide with those of the wavelet scenario. Thus the origin of the differences lies in the polynomial structure of the inverse problems, in opposition to the convolution structure of the problems usually treated by first generation wavelet methods.

These results illustrate the fact that the limitations met by classical wavelets in inverse problem theory concerning the type of operators involved, can be circumvented by using new wavelet constructions such as needlets. Similarly other second generation wavelets, meaning wavelets which do not rely on Fourier type constructions, may help to break new ground in statistical inverse problems.

5. Proofs

In this section we give a road map of the proofs of the main results. All the details are given in the on-line version of the paper, available at the following URL: `http://www.stat.sinica.edu.tw/statistica`.

5.1. General scheme of the proof

The proof of Theorem 2 requires well-known methods for minimax lower bounds, as available in Tsybakov (2004), combined with new tools (i.e., needlets). We use Theorem 2.5 in Tsybakov (2004), which involves the Kullback-Leibler divergence $\mathcal{K}(P,Q)$ between two probability measures P and Q, defined by

$$\mathcal{K}(P,Q) = \begin{cases} \int \ln(\frac{dP}{dQ})dP, & \text{if } P \ll Q; \\ +\infty, & \text{otherwise.} \end{cases}$$

Changing notation, and slightly modifying the conditions so as to include the case $m = 1$ (the result remains true using $\tau = 1/\sqrt{m+1}$ instead of $\tau = 1/\sqrt{m}$ in the proof), one has the following.

Theorem 3. *Assume there exist $m+1$ functions $f_0, \dots, f_m$ (with $m \geq 1$) satisfying*

- **Condition (i):** *for all $i \in \{0, \dots, m\}$, $f_i \in B^s_{\pi,r}(M)$,*
- **Condition (ii):** *for all $i \neq j$, $\|f_i - f_j\|_p^p \geq 2\delta$ for some $\delta > 0$,*
- **Condition (iii'):** *for all $i \in \{1, \dots, m\}$, $P_{f_i} \ll P_{f_0}$ and $1/m \sum_{i \geq 1} \mathcal{K}(P_{f_i}, P_{f_0}) \leq \theta \log(m+1)$, where $0 < \theta < 1/8$ and P_f denotes the probability distribution of the process Y under the hypothesis f.*

Then $\inf_{\hat{f}} \sup_{f \in B^s_{\pi,r}(M)} P_f(\|\hat{f} - f\|_p^p \geq \delta) \geq \pi_0$, where π_0 is a positive universal constant.

Consider Condition (iii'). Using Girsanov's formula, one can show that for all $f, g \in U$, P_f is absolutely continuous with respect to P_g and $\mathcal{K}(P_f, P_g) = 1/2\|[K(f-g)]/\epsilon\|_V^2$. Then Condition (iii') can be replaced by

Condition (iii): $f_0 = 0$ and for all $i \in \{1, \dots, m\}$, $\|Kf_i\|_V^2 \leq \theta \log(m+1)\epsilon^2$, where $0 < \theta < 1/4$.

We use Theorem 3 by building several sets of hypotheses $\{f_i,\ i = 0, 1, \dots, m\}$ satisfying the three conditions. Then, using Chebychev's inequality, we have

$$\inf_{\hat{f}} \sup_{f \in B^s_{\pi,r}(M)} E_f \|\hat{f} - f\|_p^p \geq \pi_0 \delta.$$

With an appropriate choice of three sets $\{f_i,\ i = 0, 1, \dots, m\}$, depending on the level of noise ϵ, δ yields the three expected rates. We detail the sparse cases in Section 5.2 and then the regular case in Section 5.3. Throughout these two sections, we use many (old or new) preliminary results on needlets, all of which are given in Section 5.4.

5.2. Sparse cases

The sparse rates $\mu(\alpha)$ and $\mu(\beta)$ are obtained, respectively, by applying Theorem 3 to the following sets of functions: $\{f_0 = 0,\ f_1 = \gamma\psi_{j_0,\eta_1}\}$ and $\{f_0 = 0,\ f_1 = \gamma\psi_{j_1,\eta_{2^{j_1}}}\}$, for some parameters γ, j_0 and j_1 chosen so as to satisfy conditions (i) to (iii). We detail only the proof for $\mu(\alpha)$ (the proof for $\mu(\beta)$ is similar).

Condition (i) is satisfied if $u_j := 2^{js}(\sum_{\eta \in \mathbb{Z}_j} |\langle f_1, \psi_{j,\eta}\rangle|^\pi \|\psi_{j,\eta}\|_\pi^\pi)^{1/\pi}$ belongs to $l^r(M)$, where $f_1 = \gamma\psi_{j_0,\eta_1}$. Using the first part of Lemma 1, $u_j = 0$ whenever $|j - j_0| \geq 2$. So in the sequel we assume that $j \in \{j_0 - 1, j_0, j_0 + 1\}$, and the

l^r norm of (u_j) is bounded by a constant M (independent of $\gamma > 0$ and j_0), if for instance $u_j \leq 3^{-1/r}M$. Using the second part of Lemma1, we have for any ζ, $|\langle \psi_{j_0,\eta_1}, \psi_{j,\eta_k}\rangle| \leq c_\zeta 1/k^\zeta$. With a large enough ζ, and using the bounds of Theorem 6, one can show that $u_j^\pi \leq c2^{j_0[\pi s+(\pi-2)(\alpha+1)]}\gamma^\pi$. So Condition (i) is satisfied if, for a small enough c depending on M, $\gamma \leq c2^{-j_0[s+(1-2/\pi)(\alpha+1)]}$.

Condition (ii), using Theorem 6, is fulfilled with $\delta \asymp \gamma^p 2^{j_0(p-2)(\alpha+1)}$.

Condition (iii) is satisfied if $\int_I([K(\gamma\psi_{j_0,\eta_1})(t)]/\epsilon)^2 d\lambda(t) \leq C$. We have $\psi_{j_0,\eta}(x) = \sum_{l=2^{j-2}+1}^{2^j-1} c_{j,\eta,l}P_l(x)$ and $K^*KP_l = b_l^2 P_l$, thus

$$\|K(\psi_{j_0,\eta_1})\|_V^2 = \sum_l [b_l c_{j,\eta_1,l}]^2 \asymp 2^{-2\nu j_0}\sum_l [c_{j,\eta_1,l}]^2 = 2^{-2\nu j_0}\|\psi_{j_0,\eta_1}\|_U^2 \leq C2^{-2\nu j_0}.$$

Condition (iii) is then satisfied if $(\gamma 2^{-\nu j_0})/\epsilon \leq c$.

In view of the three conditions, we set $\gamma = c\epsilon 2^{\nu j_0}$ with a small enough c, and $2^{j_0} \asymp \epsilon^{-1/(s+\nu+(1-2/\pi)(\alpha+1))}$. Then $\delta \asymp \epsilon^{(p[s+2(1/p-1/\pi)(\alpha+1)])/(s+\nu+(1-2/\pi)(\alpha+1))}$ gives the sparse lower bound.

5.3. Regular case

Let m be an integer such that $2^m \geq n_2$, where n_2 is the integer from Theorem 7 in the case $p = 2$. For some parameters γ and $j_0 \geq m+1$ chosen further, we consider, for $\varepsilon \in \{0,1\}^{2^{j_0-m-1}}$, the $2^{2^{j_0-m-1}}$ functions

$$f_\varepsilon = \gamma \sum_{k=1}^{2^{j_0-m-1}} \varepsilon_k k^\delta \psi_{j_0,\eta_{2^m k}},$$

for some δ satisfying $\delta > \max[1, \alpha+1/2, (1-2/\pi)(\alpha+1/2)-1/\pi]$. We only keep some of these functions. By the Varshamov-Gilbert Theorem (see for instance Tsybakov (2004)), there exists a subset $E_{j_0} = \{\varepsilon^0, \ldots, \varepsilon^{T_{j_0}}\}$ of $\{0,1\}^{2^{j_0-m-1}}$, and two constants $c > 0$, $\rho > 0$ such that $\forall 0 \leq u < v \leq T_{j_0}$,

$$\sum_{k=1}^{2^{j_0-m-1}} |\varepsilon_k^u - \varepsilon_k^v| \geq c2^{j_0}, \qquad T_{j_0} \geq exp(\rho 2^{j_0}) \quad \text{and} \quad f_{\varepsilon^0} = 0.$$

In the sequel we consider the set $\{f_\varepsilon,\ \varepsilon \in E_{j_0}\}$.

Condition (i): for $\varepsilon \in E_{j_0}$, let $u_j := 2^{js}(\sum_{\eta\in\mathbb{Z}_j} |\langle f_\varepsilon, \psi_{j,\eta}\rangle|^\pi \|\psi_{j,\eta}\|_\pi^\pi)^{1/\pi}$. Once again $u_j = 0$ whenever $|j - j_0| \geq 2$. Now let $j \in \{j_0-1, j_0, j_0+1\}$. Using the same arguments as in the sparse case, but with more technical proofs, one can show that $u_j \leq c2^{j[s+\delta+1/2]}\gamma$. So f_ε belongs to $B_{\pi,r}^s(M)$ if, for a small enough c depending on M, $\gamma \leq c2^{-j_0[s+\delta+1/2]}$.

Condition (ii): for all $\varepsilon^u, \varepsilon^v \in E_{j_0}$ with $u \neq v$, $f_u - f_v = \sum_{k=1}^{2^{j_0-m-1}} \gamma(\varepsilon_k^u - \varepsilon_k^v)k^\delta \psi_{j_0,\eta_{2^m k}}$. So by Theorem 7 and then Theorem 6, $\|f_u - f_v\|_U^2 \geq c\gamma^2 \sum_{k=1}^{2^{j_0-m-1}} (\varepsilon_k^u - \varepsilon_k^v)^2 k^{2\delta} = c\gamma^2 \sum_{\{k \mid \varepsilon_k^u \neq \varepsilon_k^v\}} k^{2\delta}$. Let $N_{u,v}$ denote the cardinality of the set $\{k \in \{1, \ldots, 2^{j_0-m-1}\} \mid \varepsilon_k^u \neq \varepsilon_k^v\}$, then we have $N_{u,v} \geq c2^{j_0}$, and thus one shows that $\|f_u - f_v\|_U^2 \geq c\gamma^2 2^{j_0(1+2\delta)}$. Then distinguishing the cases $2 < p < \infty$ and $1 < p < 2$, one can prove by using the previous inequality, Hölder's inequality, Theorem 5, and Theorem 4, that $\|f_u - f_v\|_{\mathbb{L}^p(\mu)}^p \geq c\gamma^p 2^{j_0 p(1/2+\delta)}$.

Condition (iii): we have $\sqrt{T_{j_0}} \geq exp([\rho/2]2^{j_0})$, so (iii) is satisfied if for all $\varepsilon^u \in E_{j_0}$, $\int_I([K(f_u)(t)]/\epsilon)^2 d\lambda(t) \leq c2^{j_0}$ for a small enough constant c. We have $f_u = \sum_{k=1}^{2^{j_0-m-1}} \beta_{j_0,k}\psi_{j_0,\eta_{2^m k}} = \sum_{k=1}^{2^{j_0-m-1}} \sum_{l\in\mathbb{N}} \beta_{j_0,k} c_{j_0,\eta_k,l} P_l(x)$, with $\beta_{j_0,k} = \gamma\varepsilon_k^u k^\delta$. Then similarly to the sparse case, we have $\|K(f_u)\|_{\mathbb{L}_2(I,\lambda)}^2 \leq c2^{-2\nu j_0}\gamma^2 2^{(2\delta+1)j_0}$. So finally we need $([2^{(\delta-\nu)j_0}\gamma]/\epsilon) \leq C$ with a small enough constant C.

In view of the three conditions, we set $2^{j_0} \asymp \epsilon^{-1/(s+\nu+1/2)}$ and $\gamma \asymp \epsilon^{(s+\delta+1/2)/(s+\nu+1/2)}$, and we obtain the lower bound: $\delta \asymp \epsilon^{ps/(s+\nu+1/2)}$.

5.4. Description of Jacobi needlets

In this section we recall briefly the construction of Jacobi needlets introduced by Petrushev and Xu (2005); for more details we refer the reader to that paper. We recall that (P_k) denote the Jacobi polynomials normalized in U. The first step consists of a Littlewood-Paley decomposition, involving some C^∞ function $a(.)$, supported in $[-2,-1/2] \cup [1/2, 2]$, such that $\sum_{j\geq 0} a^2(x/2^j) = 1$, $\forall |x| \geq 1$. Moreover we add the condition $a(x) > c > 0$ for $3/4 \leq x \leq 7/4$ (so as to use results established in Kerkyacharian, Picard, Petrushev and Willer (2007)). The second step is to use, for each resolution j, a quadrature formula that involves as knots the zeros of the Jacobi polynomial P_{2^j}, denoted by $\mathbb{Z}_j = \{\eta_k : k = 1, \ldots, 2^j\}$, and as coefficients the Christoffel numbers (see Szegö (1975)), denoted by $\{b_{j,\eta_k} : k = 1, \ldots, 2^j\}$. We assume that the $\eta_k = \cos\theta_{j,k}$ are ordered so that $\eta_1 > \eta_2 > \cdots > \eta_{2^j}$, and hence $0 < \theta_{j,1} < \theta_{j,2} < \cdots < \theta_{j,2^j} < \pi$. It is well-known that (cf Szegö (1975)) $\theta_{j,k} \asymp (k\pi/2^j)$ and $b_{j,\eta_k} \asymp 2^{-j}\omega_{\alpha,\beta}(2^j;\eta_k)$ with

$$\omega_{\alpha,\beta}(2^j; x) := (1 - x + 2^{-2j})^{\alpha+\frac{1}{2}}(1 + x + 2^{-2j})^{\beta+\frac{1}{2}}.$$

Then the Jacobi needlets are given by: $\forall j \in \mathbb{N}, k \in \{1, \ldots, 2^j\}$, $\psi_{j,\eta_k}(x) = \sum_{l=2^{j-2}+1}^{2^j-1} c_{j,\eta,l} P_l(x)$, with coefficients $c_{j,\eta,l} = a(l/2^{j-1})P_l(\eta)\sqrt{b_{j,\eta}}$. Some examples of needlets are given at the top of Figure 5.1. Now we give a list of their properties needed to establish Theorem 2.

Similarities with first generation wavelets. First of all, needlets form a tight frame: $\forall f \in \mathbb{H}$, $f = \sum_{j\in\mathbb{N},\eta\in\mathbb{Z}_j} \langle f, \psi_{j,\eta}\rangle\psi_{j,\eta}$ and $\|f\|^2 = \sum_{j\in\mathbb{N},\eta\in\mathbb{Z}_j} |\langle f, \psi_{j,\eta}\rangle|^2$.

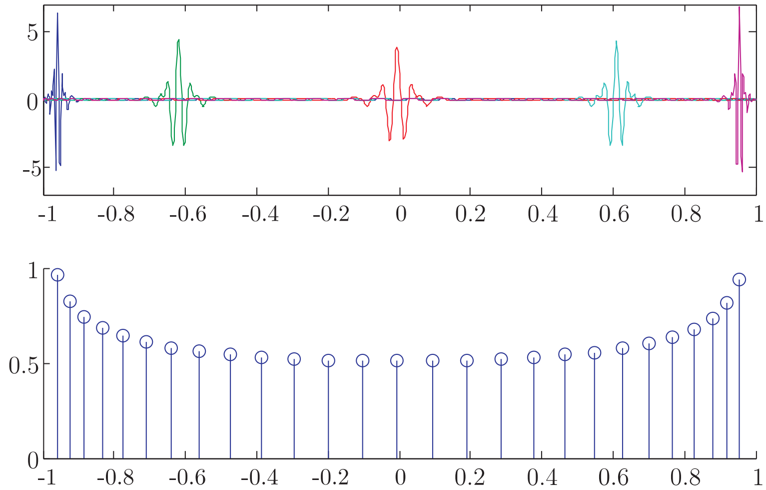

Figure 5.1. For a given resolution j: some of the needlets ψ_{j,η_k} (above), and the values of all the L^3 norms (below) as η_k varies

Second, each needlet ψ_{j,η_k} is concentrated on a small interval centered on η, as established in Petrushev and Xu (2005).

Theorem 4. *For any $l \geq 1$ there exists a constant $C_l > 0$ such that*

$$|\psi_{j,\eta_k}(\cos\theta)| \leq C_l \frac{1}{\sqrt{\omega_{\alpha,\beta}(2^j, \cos\theta)}} \frac{2^{\frac{j}{2}}}{(1 + 2^j|\theta - \frac{\pi k}{2^j}|)^l}, \quad 0 \leq \theta \leq \pi.$$

This almost exponential concentration property implies wavelet-like inequalities for the L^p norms of linear combinations of needlets. This has been established in Kerkyacharian, Picard, Petrushev and Willer (2007).

Theorem 5. *If $0 < p < \infty$, there exists a constant $C_p > 0$ such that for any collection of numbers $\{\lambda_k : k = 1, \ldots, 2^j\}$, $j \geq 0$,*

$$\left\| \sum_{k=1}^{2^j} \lambda_k \psi_{j,\eta_k} \right\|_{\mathbb{L}^p(\mu)}^p \leq C_p \sum_{k=1}^{2^j} |\lambda_k|^p \|\psi_{j,\eta_k}\|_{\mathbb{L}^p(\mu)}^p.$$

Differences from first generation wavelets. Needlets do not issue from a translation/dilatation scheme. In particular, for a given resolution j, L^p norms are not constant with respect to η (this is illustrated in Figure 5.1), and this plays an important role in the proofs of Theorems 1 and 2. The following bounds have been established in Petrushev and Xu (2005) (upper bounds), and in Kerkyacharian, Picard, Petrushev and Willer (2007) (lower bounds).

Theorem 6. $\forall\ 0 < p \leq \infty$, $\forall\ j \in \mathbb{N}$, *we have up to scalars depending only on* p,

$$\forall\ 1 \leq k \leq 2^{j-1}, \qquad \|\psi_{j,\eta_k}\|_p \asymp \left(\frac{2^{j(\alpha+1)}}{k^{\alpha+1/2}}\right)^{1-\frac{2}{p}},$$

$$\forall\ 2^{j-1} < k \leq 2^j, \qquad \|\psi_{j,\eta_k}\|_p \asymp \left(\frac{2^{j(\beta+1)}}{(1+(2^j-k))^{\beta+1/2}}\right)^{1-\frac{2}{p}}.$$

Moreover, unlike first generation wavelets, needlets do not form an orthonormal basis, but only a redundant frame. So we need two new results for the proof of the minimax rates. First we need an upper bound for the scalar products between needlets.

Lemma 1. *We have*

1. $\forall j,\ j',\ k,\ l$ *such that* $|j'-j| \geq 2$, $\langle \psi_{j,\eta_k}, \psi_{j',\eta_l}\rangle = 0$,
2. $\forall \zeta > 0$, $\exists c_\zeta$ *such that* $\forall j,\ j',\ k,\ l$ *with* $|j'-j| \leq 1$, $|\langle \psi_{j,\eta_k}, \psi_{j',\eta_l}\rangle| \leq [c_\zeta/(1+|k-2^{j-j'}l|)^\zeta]$.

Second, we need a lower bound for the L^p norm of linear combinations of needlets. A result as general as the upper bound of Theorem 5 is impossible, but we have the following result for needlets with a large enough distance between the indexes of the η's.

Theorem 7. *If* $p \in 2\mathbb{N}^*$, *there exist a constant* $c_p > 0$ *and an integer* n_p *such that, for any collection of numbers* $\{\lambda_k : k \in I_j\}$, $j \geq 0$, *where* $I_j \subset \{1, \ldots, 2^j\}$ *and* $k, l \in I_j$, $k \neq l \Longrightarrow |k-l| \geq n_p$,

$$\Big\| \sum_{k\in I_j} \lambda_k \psi_{j,\eta_k} \Big\|^p_{\mathbb{L}^p(\mu)} \geq c_p \sum_{k\in I_j} |\lambda_k|^p \|\psi_{j,\eta_k}\|^p_{\mathbb{L}^p(\mu)}.$$

Proof of Lemma 1. The needlets are $\psi_{j,\eta} = \sum_{l=2^{j-2}+1}^{2^j-1} c_{j,\eta,l} P_l(x)$, with coefficients $c_{j,\eta,l} = a(l/2^{j-1}) P_l(\eta)\sqrt{b_{j,\eta}}$. So if $|j'-j| \geq 2$ then $\{2^{j-2}+1, \ldots, 2^j-1\} \cap \{2^{j'-2}+1, \ldots, 2^{j'}-1\} = \emptyset$, and $\langle \psi_{j,\eta_k}, \psi_{j',\eta_l}\rangle = 0$, $\forall (k,l)$. For the second part of the lemma, we use Theorem 4. After numerous (but simple) inequalities for the integrand, one obtains the desired upper bound for $|\langle \psi_{j,\eta_k}, \psi_{j',\eta_l}\rangle_U|$.

Proof of Theorem 7. Let $p \in 2\mathbb{N}^*$ and $I_j \subset \{1, \ldots, 2^j\}$. We have the decomposition $\|(\sum_{k\in I_j} \lambda_k \psi_{j,\eta_k})\|^p_{\mathbb{L}^p(\mu)} = A + B$, where

$$A = \sum_{k\in I_j} \lambda_k^p \|\psi_{j,\eta_k}\|^p_{\mathbb{L}^p(\mu)},$$

$$B = \sum_{(p_k)_{k\in I_j}\in\Lambda} \frac{p! \prod_{k\in I_j} \lambda_k^{p_k}}{\prod_{k\in I_j} p_k!} \int_{-1}^{1} \Big(\prod_{k\in I_j} \psi_{j,\eta_k}^{p_k}(x)\Big)\mu(x)dx,$$

and $\Lambda = \{(p_k)_{k\in I_j} \mid p_k \in \mathbb{N}, \sum_{k\in I_j} p_k = p$ and $\exists u \neq v$ such that $p_u > 0$ and $p_v > 0\}$. Let $\varphi_{j,k}(x) = (1/\sqrt{\omega_{\alpha,\beta}(2^j, x)})[2^{j/2}/(1 + 2^j|\arccos x - \pi k/2^j|)^{2/s}]$ for some $0 < s < \min\{1, p/(\alpha \vee \beta + 1)\}$.

For $(p_k)_{k\in I_j} \in \Lambda$, by using Theorem 4 with $l = 2/s+1$ for every ψ_{j,η_k}, $k \in I_j$, one can show that there exists C such that

$$\sum_{(p_k)_{k\in I_j}\in\Lambda} \frac{p!\prod_{k\in I_j}|\lambda_k^{p_k}|}{\prod_{k\in I_j} p_k!} \prod_{k\in I_j} |\psi_{j,\eta_k}|^{p_k} \leq C\frac{(\sum_{k\in I_j}|\lambda_k|\varphi_{j,\eta_k})^p}{n_{inf}}.$$

By using similar arguments as in the proof of Theorem 5 of Kerkyacharian, Picard, Petrushev and Willer (2007) (based notably on the Fefferman-Stein maximal inequality, see Fefferman and Stein (1971)) we have

$$\left\|\sum_{k=1}^{2^j} |\lambda_k|\varphi_{j,\eta_k}\right\|_{\mathbb{L}^p(\mu)}^p \leq C' \sum_{k=1}^{2^j} |\lambda_k|^p \|\psi_{j,\eta_k}\|_{\mathbb{L}^p(\mu)}^p.$$

So there exists $C > 0$ such that $|B| \leq C(A/n_{inf})$, and if we impose the condition on I_j that $n_{inf} \geq 2C$, then we obtain $|B| \leq (1/2)A$, and thus

$$\left\|(\sum_{k\in I_j} \lambda_k\psi_{j,\eta_k})\right\|_{\mathbb{L}^p(\mu)}^p \geq \frac{1}{2}\sum_{k\in I_j} \lambda_k^p \|\psi_{j,\eta_k}\|_{\mathbb{L}^p(\mu)}^p.$$

Acknowledgement

The author would like to thank Professor Dominique Picard and the referees for their insightful comments on the paper.

References

Abramovich, F. and Silverman, B. W. (1998). Wavelet decomposition approaches to statistical inverse problems. *Biometrika* **85**, 115-129.

Antoniadis, A., Fan, J. and Gijbels, I. (2001). A wavelet method for unfolding sphere size distributions. *The Canadian Journal of Statistics* **29**, 265-290.

Cavalier, L. and Hengartner, N. W. (2005). Adaptive estimation for inverse problems with noisy operators. *Inverse problems* **21**, 1345-1361.

Cavalier, L. and Raimondo, M. (2007). Wavelet deconvolution with noisy eigen-values. *IEEE Transactions on Signal Processing* **55**, part 1.

Cavalier, L. and 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. and Tsybakov, A. B. (2002). Oracle inequalities for inverse problems. *Ann. Statist.* **30**, 843-874.

Cruz-Orive, L. M. (1983). Distribution-free estimation of sphere size distributions from slabs showing overprojections and truncations, with a review of previous methods. *J. Microscopy* **131**, 265-290.

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

Efromovich, S. and Koltchinskii, V. (2002). On inverse problems with unknown operators. *IEEE Trans. Inform. Theory* **47**, 2876-2894.

Fan, J. and Koo, J. Y. (2002). Wavelet deconvolution. *IEEE Trans. Inform. Theory* **48**, 734-747.

Fefferman, C. and Stein, E. M. (1971). Some maximal inequalities. *Amer. J. Math.* **93**, 107-115.

Goldenshluger, A. and Pereverzev, S. V. (2003). On adaptive inverse estimation of linear functionals in Hilbert scales. *Bernoulli* **9**, 783-807.

Golubev, G. K. and Levit, B. Y. (1998). Asymptotically efficient estimation in the Wicksell problem. *Ann. Statist.* **26**, 2407-2419.

Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. B. (1998). *Wavelets, Approximation and Statistical Applications.* Springer-Verlag.

Hoffmann, M. and Reiss, M. (2008). Nonlinear estimation for linear inverse problems with error in the operator. *Annals of Statistics* **36**, 310-336.

Johnstone, I. M. and Silverman, B. W. (1991). Discretization effects in statistical inverse problems. *J. Complexity* **7**, 1-34.

Johnstone, I. M., Kerkyacharian, G., Picard, D. and Raimondo, M. (2004). Wavelet deconvolution in a periodic setting. *Journal of the Royal Statistical Society* **66**, 1-27.

Kalifa, J. and Mallat, S. (2003). Thresholding estimators for linear inverse problems and deconvolutions. *Ann. Statist.* **31**, 58-109.

Kerkyacharian, G., Picard, D., Petrushev, P. and Willer, T. (2007). Needlet algorithms for estimation in inverse problems. *Electronic Journal of Statistics* **1**, 30-76.

Mathé, P. and Pereverzev, S. V. (2003). Geometry of linear ill-posed problems in variable Hilbert scales. *Inverse Problems* **19**, 789-803.

Narcowich, F. J., Petrushev, P. and Ward, J. M. (2006). Decomposition of besov and triebel-lizorkin spaces on the sphere. *J. Funct. Anal.* **238**, 530-564

Nychka, D., Wahba, G., Goldfarb, S. and Pugh, T. (1984). Cross validated spline methods for the estimation of three-dimensional tumor size distributions from observations on two-dimensional cross sections. *J. Amer. Statist. Assoc.* **79**, 832-846.

Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Annals of Statistics* **27**, 2033-2053.

Petrushev, P. and Xu, Y. (2005). Localized polynomial frames on the interval with Jacobi weights. *J. Fourier Anal. Appl.* **11**, 557-575.

Szegö, G. (1975). Orthogonal Polynomials. *American Mathematical Society, Providence, R.I.*

Tsybakov, A. B. (2000). On the best rate of adaptive estimation in some inverse problems. *C. R. Acad. Sci. Paris Sér. I Math.* **330**, 835-840.

Tsybakov, A. B. (2004). Introduction à l'estimation non-paramétrique. *Volume 41 of Mathématiques and Applications, Springer-Verlag, Berlin.*

Wicksell, S. D. (1925). The corpuscle problem: a mathematical study of a biometric problem. *Biometrika* **17**, 84-99.

Willer, T. (2005). Deconvolution in white noise with a random blurring effect. *preprint LPMA.*

Willer, T. (2006). Non parametric estimation and inverse problems. *Ph.D. thesis, Université Paris* 7.

Laboratoire d'Analyse, Topologie et Probabilités (LATP), Université de Provence, Marseille, France.

E-mail: twiller@cmi.univ-mrs.fr

(Received March 2007; accepted January 2008)