Journal of Multivariate Analysis
doi:10.1006/jmva.2000.1968, available online at http://www.idealibrary.com on IDEAL®

Optimal Spherical Deconvolution[1]

Peter T. Kim

University of Guelph, Guelph, Ontario, Canada

and

Ja-Yong Koo

Hallym University, Kangwon-Do, Korea

Received January 27, 1999

This paper addresses the issue of optimal deconvolution density estimation on the 2-sphere. Indeed, by using the transitive group action of the rotation matrices on the 2-dimensional unit sphere, rotational errors can be introduced analogous to the Euclidean case. The resulting density turns out to be convolution in the Lie group sense and so the statistical problem is to recover the true underlying density. This recovery can be done by deconvolution; however, as in the Euclidean case, the difficulty of the deconvolution turns out to depend on the spectral properties of the rotational error distribution. This therefore leads us to define smooth and super-smooth classes and optimal rates of convergence are obtained for these smoothness classes. © 2001 Academic Press

AMS 1991 subject classifications: 62G05; 58G25.

Key words and phrases: consistency; density estimation; minimax; rotational harmonics; smooth; super-smooth; Sobolev spaces; spherical harmonics.

1. INTRODUCTION

Deconvolution techniques have been shown to be of practical use in situations where the data is indirectly observed. Indeed, if the underlying density is a mixture of several densities, deconvolution allows one to recover the main components of the mixture; see Efromovich (1997) for a recent example of the benefits of circular deconvolution. Therefore, asymptotic optimally in deconvolution density estimation has been investigated in the statistical literature; see Fan (1991, 1991a, 1993), Efromovich (1997), and Koo and Chung (1998). The main problem involves identifying the smoothness

[1] This research was supported by Korea Research Foundation Grant (KRF-2000-015-DS0006).

0047-259X/01 $35.00

of the characteristic function of the error distribution into ordinary smooth or super-smooth classes for which the resulting convergence turns out to be polynomial or logarithmic, respectively.

The above involve deconvolution in Euclidean space, however, some recent interest in non-Euclidean deconvolution has appeared in the statistical literature. Rooij and Ruymgaart (1991) first motivated deconvolution on the two dimensional unit sphere, S^2. Healy and Kim (1996) and Healy *et al.* (1998) work out the technical details for consistency. Indeed, the problem is as follows. In the case of S^2, measurement error can be modelled analogous to Euclidean error by using the transitive group action $SO(3) \times S^2 \to S^2$, where $SO(3)$ is the space of 3×3 rotation matrices. Then under appropriate smoothness, rates of convergence are obtained. It is then natural to ask whether or not these rates of convergence are optimal as defined in Fan (1993) and Koo (1993). It will be shown that definitions of ordinary smooth and super-smooth can be made through the operator norm of the rotational Fourier transform of the error distribution. These smoothness classes then lead to polynomial or logarithmic rates of convergence, respectively.

We now provide a summary of what is to follow.

In Section 2, we briefly go over the necessary Fourier tools for the 2-dimensional unit sphere and the 3-dimensional rotation matrices, as well as the connections between the two. The latter involves how convolution as well as how Fourier transforms change convolution into individual Fourier products similar to the Euclidean case.

In Section 3, we outline the deconvolution problem of the 2-sphere. In addition, we define smooth and super-smooth densities on the space of 3-dimensional rotation matrices. This is done in the Fourier domain using the operator norm. Following this we state the main results. We also make a connection with some earlier work by Hendriks (1990). The latter obtains upper bound rates of convergence for nonparametric density estimators on compact Riemannian manifolds. It follows as a corollary to one of our main results that in the case of the 2-sphere, this convergence is optimal.

Since this area is relatively new in statistics, in order to motivate the problem further, we provide examples of rotational error densities in Section 4. Two examples of smooth densities as well as an example of a super-smooth density are introduced. The latter involves the rotational version of the Gaussian distribution, while the former involves the rotational version of the Laplace (double exponential) distribution as well as a distribution obtained from the random walks on groups literature. All of these distributions are spectrally defined.

In Section 5 we examine the von Mises–Fisher matrix distribution by calculating it’s rotational Fourier transform. Once the calculations are complete, we notice that although the super-smooth definition appears appropriate, we can almost (but not exactly) get the same power in the exponent on both sides of the inequalities. Consequently, we can almost get the same upper and lower rates of convergence.

All proofs are collected in Section 6. We first establish upper bounds and demonstrate that these upper bounds are also lower bounds by specifying a subproblem. This method follows the outline of Koo (1993) and Koo and Chung (1998), however, one does need to accommodate for the spherical geometry in the construction. It is found that as far as the rates of convergence are concerned, aside from the smoothness class of the underlying rotational error distribution, these rates only depend on the dimension of the 2-sphere.

2. SOME PRELIMINARIES

We will provide a brief overview of Fourier analysis on $SO(3)$ and S^2. Most of the material in expanded form can be found in Talman (1968), Terras (1985), Healy and Kim (1996), and Healy *et al.* (1998). Papers which directly deal with similar issues can be found in Lo and Eshelman (1979) and Wahba (1981).

The well known *Euler angle decomposition* says, any $g \in SO(3)$ can almost surely be uniquely represented by three angles (ϕ, θ, ψ), known collectively as the *Euler angles*, where $\phi \in [0, 2\pi)$, $\theta \in [0, \pi)$, $\psi \in [0, 2\pi)$; see Healy and Kim (1996) and Healy *et al.* (1998) for details. Consider the function,

$$D^l_{q_1 q_2}(\phi, \theta, \psi) = \mathrm{e}^{-iq_1\phi}\, d^l_{q_1 q_2}(\cos\theta)\, \mathrm{e}^{-iq_2\psi}, \tag{2.1}$$

where, $d^l_{q_1 q_2}$ for $-l \leqslant q_1, q_2 \leqslant l$, $l = 0, 1, \ldots$ are related to the Jacobi polynomials; see Lo and Eshelman (1979). The functions $D^l_{q_1 q_2}$, $-l \leqslant q_1, q_2 \leqslant l$, $l = 0, 1, \ldots$, are the eigenfunctions of the Laplace–Beltrami operator on $SO(3)$, hence, $\{\sqrt{2l+1}\, D^l_{q_1 q_2} : -l \leqslant q_1,\ q_2 \leqslant l,\ l = 0, 1, \ldots\}$ is a complete orthonormal basis for $L^2(SO(3))$ with respect to the probability Haar measure and are otherwise known as the *rotational harmonics*. In addition, if we define a $(2l+1) \times (2l+1)$ matrix by

$$D^l(g) = [D^l_{q_1 q_2}(g)], \tag{2.2}$$

where $-l \leqslant q_1, q_2 \leqslant l$, $l \geqslant 0$ and $g \in SO(3)$, these constitute the collection of inequivalent irreducible representations of $SO(3)$.

Let $f \in L^2(SO(3))$. We define the *rotational Fourier transform* on $SO(3)$ by

$$\hat{f}^l_{q_1 q_2} = \int_{SO(3)} f(g)\, D^l_{q_1 q_2}(g)\, dg, \tag{2.3}$$

where again we think of (2.3) as the matrix entries of the $(2l+1)\times(2l+1)$ matrix $\hat{f}^l = [\hat{f}^l_{q_1 q_2}]$, $-l \leqslant q_1, q_2 \leqslant l$, $l = 0, 1, ...$ and dg is the probability Haar measure on $SO(3)$. The *rotational inversion* can be obtained by

$$\begin{aligned} f(g) &= \sum_{l \geqslant 0} \sum_{q_1, q_2 = -l}^{l} (2l+1)\, \hat{f}^l_{q_1 q_2}\, \overline{D^l_{q_1 q_2}(g)} \\ &= \sum_{l \geqslant 0} \sum_{q_1, q_2 = -l}^{l} (2l+1)\, \hat{f}^l_{q_1 q_2}\, D^l_{q_2 q_1}(g^{-1}), \end{aligned} \tag{2.4}$$

for $g \in SO(3)$, where the overbar denotes complex conjugation. Strictly speaking, (2.4) should be interpreted in the L^2-sense although with additional smoothness conditions, it can hold pointwise.

Spherical Fourier analysis also has similar results. Any point on S^2 can be represented by

$$\omega = (\cos\phi \sin\theta, \sin\phi \sin\theta, \cos\theta)^t,$$

where $\theta \in [0, \pi)$, $\phi \in [0, 2\pi)$ and superscript t denotes transpose. Let

$$Y^l_q(\omega) = Y^l_q(\theta, \phi) = (-1)^q \sqrt{\frac{(2l+1)(l-q)!}{4\pi(l+q)!}}\, P^l_q(\cos\theta)\, e^{iq\phi}, \tag{2.5}$$

where $\theta \in [0, \pi)$, $\phi \in [0, 2\pi)$, $-l \leqslant q \leqslant l$, $l = 0, 1, ...$ and $P^l_q(\cdot)$ are the Legendre functions. We note that we can think of (2.5) as the vector entries to the $2l+1$ vector $Y^l(\omega) = [Y^l_q(\omega)]$, $l \geqslant 0$. In this situation

$$\{Y^l_q\colon -l \leqslant q \leqslant l, l = 0, 1, ...\}$$

form a complete orthonormal basis over $L^2(S^2)$ and is sometimes referred to as the *spherical harmonics*; see Talman (1968).

Let $f \in L^2(S^2)$. We define the *spherical Fourier transform* on S^2 by

$$\hat{f}^l_q = \int_{S^2} f(\omega)\, \overline{Y^l_q(\omega)}\, d\omega, \tag{2.6}$$

where $d\omega$ is the spherical measure on S^2. Again we think of (2.6) as the vector entries of the $(2l+1)$ vector $\hat{f}^l = [\hat{f}^l_q]$, $-l \leqslant q \leqslant l$, $l = 0, 1, \ldots$. The *spherical inversion* can be obtained by

$$f(\omega) = \sum_{l \geqslant 0} \sum_{q=-l}^{l} \hat{f}^l_q Y^l_q(\omega), \tag{2.7}$$

for $\omega \in S^2$. Again, strictly speaking, (2.7) should be interpreted in the L^2-sense although with additional smoothness conditions, it can hold pointwise.

In terms of the Fourier basis, the relation between $SO(3)$ and S^2 can be described in terms of the Euler angles where

$$Y^l_q(\theta, \phi) = \sqrt{\frac{(2l+1)}{4\pi}}\, \overline{D^l_{q0}}(\phi, \theta, \psi), \tag{2.8}$$

$\phi \in [0, 2\pi)$, $\theta \in [0, \pi)$, $-l \leqslant q \leqslant l$ and $l = 0, 1, \ldots$. We note that although an extra angle ψ appears in the right hand side of (2.8), it is in fact independent of ψ. This follows from (2.1) and observing that when $q_2 = 0$, the expression becomes independent of ψ.

One of the most useful tools of Fourier analysis is the fact that convolution of two functions in the Fourier domain turns out to be ordinary matrix multiplication. Indeed, let $f \in L^2(SO(3))$ and $h \in L^2(S^2)$. Define the convolution,

$$f * h(\omega) = \int_{SO(3)} f(u)\, h(u^{-1}\omega)\, du, \tag{2.9}$$

for $\omega \in S^2$. We have the following *convolution property* for $f \in L^2(SO(3))$ and $h \in L^2(S^2)$

$$\widehat{f * h} = \hat{f}\hat{h}. \tag{2.10}$$

In particular, for each $l = 0, 1, \ldots,$

$$(\widehat{f * h})^l_q = \sum_{j=-l}^{l} \hat{f}^l_{qj} \hat{h}^l_j,$$

for all $-l \leqslant q \leqslant l$, see Lemma 2.1 in Healy *et al.* (1998), or exercise 25 of Terras (1985, p. 106).

3. DECONVOLUTION DENSITY ESTIMATION

Consider the following situation

$$Z = \varepsilon X, \tag{3.1}$$

where ε is an $SO(3)$ random element and Z, X are S^2 random elements, with ε and X assumed independent. We note that (3.1) is describing the transitive group action $SO(3) \times S^2 \to S^2$ which consists of ordinary matrix multiplication, where transitive means that for any two ω, $v \in S^2$, there exists a $g \in SO(3)$ such that $\omega = gv$.

Let f_Z, f_ε, f_X denote the densities of Z, ε, X, respectively. Through (3.1), the relation among the densities can be described by convolution,

$$f_Z = f_\varepsilon * f_X$$

as seen by following the familiar corresponding Euclidean result. We note that since f_ε and f_X are density functions, we have

$$\inf_{\omega \in S^2} f_X(\omega) \leqslant \int_{SO(3)} f_\varepsilon(u)\, f_X(u^{-1}\omega)\, du \leqslant \sup_{\omega \in S^2} f_X(\omega). \tag{3.2}$$

Now consider $\hat{f}^l_X$ and $\hat{f}^l_Z$ for each $l \geqslant 0$ given by $[\hat{f}^l_{X,q}]$ and $[\hat{f}^l_{Z,q}]$, respectively, and $\hat{f}^l_\varepsilon$ as the matrix $[\hat{f}^l_{\varepsilon,qj}]$ for each $l \geqslant 0$. By (2.10) we can write

$$\hat{f}^l_X = (\hat{f}^l_\varepsilon)^{-1}\, \hat{f}^l_Z,$$

provided of course that the matrices $(\hat{f}^l_\varepsilon)^{-1}$ exist for all $l = 0, 1, \ldots$ in a range of interest.

Statistically, (3.1) is describing the non-Euclidean analogue of observations Z made up of the true measurement X, corrupted by noise ε. Our interest is in the unknown f_X. It is assumed that f_ε is known and that $(\hat{f}^l_\varepsilon)^{-1}$ exists for a range of l's that concerns us. Since f_X is unknown, f_Z is also unknown, hence $\hat{f}_Z$ is unknown. Nevertheless, we assume that a random sample $Z_1, \ldots, Z_n$ is available. This will allow us to construct an empirical version $\hat{f}^n_Z$. By (2.10) an estimator for $\hat{f}_X$ is therefore

$$\hat{f}^{n,l}_X = (\hat{f}^l_\varepsilon)^{-1}\, \hat{f}^{n,l}_Z, \tag{3.3}$$

for $l = 0, 1, \ldots$. We can then produce a nonparametric deconvolution density estimator of f_X by (2.7), the spherical inversion.

3.1. *Smooth and Super-Smooth Errors*

Deconvolution density estimation has been investigated for some time now and the degree to which we can recover the density f_X is best characterized in terms of the quality of smoothness of f_ε. Indeed, following Fan (1991a) we will appropriately define the smoothness of f_ε spectrally, with a modification.

The necessary modification required comes from the fact that on $SO(3)$, Fourier transforms are matrices that grow in dimension. Consequently, the quality of smoothness need to be adapted for this change and this can be done by regarding convolution as an operator. Indeed, let $\mathscr{E}_l$ be the $(2l+1)$-dimensional vector space spanned by $\{Y_q^l: -l \leqslant q \leqslant l\}$ for each $l = 0, 1, \ldots$. Thus any $h \in \mathscr{E}_l$ can be written as $h = \sum_{q=-l}^{l} \hat{h}_q^l Y_q^l$ and through Parseval's identity, the usual L^2-norm is $\|h\|_2^2 = \sum_{q=-l}^{l} |\hat{h}_q^l|^2$. Now according to (2.10), $\hat{f}_\varepsilon^l: \mathscr{E}_l \to \mathscr{E}_l$ by $\hat{f}^l h = \sum_{q=-l}^{l} (\sum_{j=-l}^{l} \hat{f}_{\varepsilon, qj}^l \hat{h}_j^l)\, Y_q^l$. Again by Parseval's identity, $\|\hat{f}^l h\|_2 = \sum_{q=-l}^{l} |\sum_{j=l}^{l} \hat{f}_{\varepsilon, qj}^l \hat{h}_j^l|^2$, for all $l \geqslant 0$. Consequently, we have the operator inequality,

$$\|\hat{f}^l h\|_2 \leqslant \|\hat{f}_\varepsilon^l\|_{\text{op}}\, \|h\|_2, \qquad \text{where} \quad \|\hat{f}_\varepsilon^l\|_{\text{op}} = \sup_{\xi \neq 0,\, \xi \in \mathscr{E}_l} \frac{\|\hat{f}_\varepsilon^l \xi\|_2}{\|\xi\|_2}. \tag{3.4}$$

We will say that the distribution of ε is *super-smooth* if the rotational Fourier transform of f_ε satisfies

$$\begin{aligned} &\|(\hat{f}_\varepsilon^l)^{-1}\|_{\text{op}} \leqq d_0^{-1} l^{-\beta_0} \exp(l^\beta/\gamma) \qquad \text{and} \\ &\|(\hat{f}_\varepsilon^l)\|_{\text{op}} \leqq d_1 l^{\beta_1} \exp(-l^\beta/\gamma) \qquad \text{as} \quad l \to \infty, \end{aligned} \tag{3.5}$$

for some positive constants d_0, d_1, β, γ, constants β_0 and β_1. We will say that the distribution of ε is (ordinary) *smooth* if the rotational Fourier transform of f_ε satisfies

$$\|(\hat{f}_\varepsilon^l)^{-1}\|_{\text{op}} \leqq d_0^{-1} l^\beta \quad \text{and} \quad \|(\hat{f}_\varepsilon^l)\|_{\text{op}} \leqq d_1 l^{-\beta} \qquad \text{as} \quad l \to \infty, \tag{3.6}$$

for some positive constants d_0, d_1 and nonnegative constant β. Examples of smooth and super-smooth distributions will be discussed in Section 4.

3.2. *Optimal Estimation*

The empirical Fourier transform on S^2 can be defined by

$$\hat{f}_{Z,q}^{n,l} = \frac{1}{n} \sum_{j=1}^{n} \overline{Y_q^l(Z_j)}, \tag{3.7}$$

which is an unbiased estimator of $\hat{f}^l_{Z,q}$ for $-l \leqslant q \leqslant l$ and $l = 0, 1, ...$ Then by (3.3)

$$\hat{f}^{n,l}_{X,q} = \frac{1}{n} \sum_{j=1}^{n} \sum_{s=-l}^{l} \hat{f}^l_{\varepsilon^{-1},qs} \overline{Y^l_s(Z_j)},$$

where $-l \leqslant q \leqslant l$, $l = 0, 1, ...$ and for ease of notation, we write $\hat{f}^l_{\varepsilon^{-1}} = (\hat{f}^l_{\varepsilon})^{-1}$.

Choosing $m = m(n) \to \infty$ as $n \to \infty$ leads to the following nonparametric deconvolution density estimator of f_X on S^2,

$$f^n_X(\omega) = \sum_{l=0}^{m} \sum_{q=-l}^{l} \left\{ \frac{1}{n} \sum_{j=1}^{n} \sum_{s=-l}^{l} \hat{f}^l_{\varepsilon^{-1},qs} \overline{Y^l_s(Z_j)} \right\} Y^l_q(\omega), \tag{3.8}$$

where $\omega \in S^2$.

For statistical motivation, we can rewrite (3.8) in another way. Define

$$K^{\varepsilon}_n(\omega, v) = \sum_{l=0}^{m} \sum_{q,s=-l}^{l} Y^l_q(\omega)\, \hat{f}^l_{\varepsilon^{-1},qs} \overline{Y^l_s(v)},$$

where $v, \omega \in S^2$. Then an alternative way of writing (3.8) is

$$f^n_X(\omega) = \frac{1}{n} \sum_{j=1}^{n} K^{\varepsilon}_n(\omega, Z_j), \tag{3.9}$$

where $\omega \in S^2$. Note that this resembles an ordinary kernel estimator in Euclidean space.

We would like to present our main results in terms of Sobolev spaces. Indeed, on the space $C^{\infty}(S^2)$ of infinitely continuous differentiable functions on S^2, consider the so-called Sobolev norm $\|\cdot\|_{H_s}$ of order s defined in the following way. For any function $h = \sum_{l,q} \hat{h}^l_q Y^l_q$ let

$$\|h\|^2_{H_s} = \sum_{l,q} (1 + l(l+1))^s\, |\hat{h}^l_q|^2. \tag{3.10}$$

One can verify that (3.10) is indeed a norm. Denote by $H_s(S^2)$ the (vector-space) completion of $C^{\infty}(S^2)$ with respect to (3.10), the Sobolev norm of order s. For some fixed constant $M > 0$, let $H_s(S^2, M)$ denote the *smoothness class* of functions $h \in H_s(S^2)$ which satisfy

$$\|h\|_{H_s} \leqslant 1 + M.$$

Consider an unknown distribution P_f depending on the density function $f \in H_s(S^2, M)$ and suppose $\{b_n\}$ is some sequence of positive numbers. This sequence is called a *lower bound* for f if

$$\lim_{c \to 0} \liminf_{n} \inf_{f^n} \sup_{f \in H_s(S^2, M)} P_f(\|f^n - f\|_2 \geqslant cb_n) = 1, \tag{3.11}$$

where the infimum is over all possible estimators f^n based on $Z_1, ..., Z_n$. Alternatively, the sequence in question is said to be an *upper bound* for f if there is a sequence of estimators $\{f^n\}$ such that

$$\lim_{c \to \infty} \limsup_{n} \sup_{f \in H_s(S^2, M)} P_f(\|f^n - f\|_2 \geqslant cb_n) = 0. \tag{3.12}$$

The sequence of numbers $\{b_n\}$ is called the *optimal rate of convergence* for f if it is both a lower bound and an upper bound with the associated estimators $\{f^n, n \geqslant 1\}$, being called *asymptotically optimal*. These definitions are in the sense of Stone (1980).

The following theorems state that the deconvolution density estimators (3.9) are asymptotically optimal, where the minimax rates of convergence depend on the smoothness characteristics of the error distribution.

We will use the following notation. For sequences $\{a_n\}$ and $\{c_n\}$ of positive numbers, let $a_n \ll c_n$ mean that $a_n/c_n \leqslant C$ as $n \to \infty$. When $a_n \ll c_n$ and $c_n \ll a_n$, we write $a_n \asymp c_n$.

THEOREM 3.1. *Suppose f_ε is smooth. If $f_X \in H_s(S^2, M)$ for some $s > 1$, then*

$$n^{-s/(2(s+\beta)+2)}$$

is the optimal rate of convergence, where $m \asymp n^{1/(2(s+\beta)+2)}$.

THEOREM 3.2. *Suppose f_ε is super-smooth. If $f_X \in H_s(S^2, M)$ for some $s > 1$, then*

$$(\log n)^{-s/\beta}$$

is the optimal rate of convergence, where $m \asymp (\log n)^{1/\beta}$.

In Section 4, a discussion of some possible error distributions is presented, however, at this point let us provide some general comments about the extreme cases with respect to distribution of the errors ε.

Indeed, at one extreme is the Haar measure (uniform distribution) on $SO(3)$ in which case deconvolution is not possible since $\hat{f}^l_\varepsilon = 0$ for all $l > 0$.

One can see this by the fact that the true measurements are uniformly corrupted resulting in no hope of being able to recover f_X.

If on the other hand we consider point mass at the unit element of $SO(3)$, i.e., δ_e, where e denotes the unit element in $SO(3)$, then

$$\hat{f}^l_\varepsilon = \int_{SO(3)} D^l(g)\, \delta_e(g)\, dg = D^l(e) = \mathbf{I}_{2l+1},$$

where $\mathbf{I}_v$ is the $v \times v$ identity matrix. Therefore $\|\hat{f}^l_\varepsilon\|_{\mathrm{op}} = 1$ and this corresponds to the smooth case with $\beta = 0$ where

$$K_n^{\delta_e}(\omega, v) = \sum_{l=0}^{m} \sum_{q=-l}^{l} Y_q^l(\omega)\, \overline{Y_q^l(v)},$$

for $v, \omega \in S^2$. Consequently, in this case, (3.9) would be

$$f_X^n(\omega) = \frac{1}{n} \sum_{j=1}^{n} K_n^{\delta_e}(\omega, Z_j), \tag{3.13}$$

where $\omega \in S^2$ and (3.13) would be just ordinary nonparametric density estimation on S^2, since the observations are made without error. This fact along with Theorem 3.1 provides the following corollary which states that for rate of convergence for nonparametric density estimation on S^2, the rate obtained in Theorem 2.1 of Hendriks (1990, p. 834) is optimal.

COROLLARY 3.3. *Suppose* $f_\varepsilon = \delta_e$. *If* $f_X \in H_s(S^2, M)$ *for* $s > 1$ *then*

$$n^{-s/(2s+2)}$$

is the optimal rate of convergence, where $m \asymp n^{1/(2s+2)}$ *as* $n \to \infty$.

4. EXAMPLES OF SMOOTH AND SUPER-SMOOTH DISTRIBUTIONS

In this section we will discuss three different error distributions with two of them being smooth and one of them being super-smooth. All of these distributions are characterized spectrally.

4.1. *Rotational Laplace Distribution*

This distribution is the rotational analogue of the Euclidean Laplace (double exponential) distribution and is discussed in depth in Healy *et al.* (1998). Although a closed form expression for $SO(3)$ is available, see

Theorem 3.5 in Healy *et al.* (1998), it's spectral version is much more informative. Indeed in terms of the rotational harmonics,

$$f_\varepsilon = \sum_{l \geqslant 0} \sum_{q=-l}^{l} (1+\sigma^2 l(l+1))^{-1} (2l+1) \overline{D^l_{qq}}, \tag{4.1}$$

for some $\sigma^2 > 0$. Spectrally,

$$\hat{f}^l_{\varepsilon,\, qj} = (1+\sigma^2 l(l+1))^{-1} \delta_{qj},$$

for $l = 0, 1, \ldots,$ where $\delta_{qj} = 1$ if $q = j$ and is 0 otherwise. As can be seen from (4.1), this is an example of a smooth distribution with $\beta = 2$.

4.2. *The Rosenthal Distribution*

The next distribution comes from a problem in probability associated with random walks on groups; see Diaconis (1988) and Rosenthal (1994). Here one is interested in performing random walks on groups, followed by establishing ways in which the measure converges to the uniform measure, the so-called "mixing". In terms of the mathematical structure, each movement in the random walk is represented by a convolution product. The nature in which finite convolution products converges to the uniform measure is analytically studied using Fourier methods on the group. The case for $SO(N)$ has been studied in Rosenthal (1994). Borrowing from his work, we will consider the situation where f_ε is a p-fold convolution product of conjugate invariant random measures for a fixed axis, where the $p > 0$ measures the degree of uniformly.

For $SO(3)$ take the conjugacy class of

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

for $\theta \in (0, \pi]$, followed by taking the uniform measure over the conjugacy class of R_θ. Let f_ε be the p-fold convolution product. Rosenthal (1994, p. 407) shows that

$$\hat{f}^l_{\varepsilon,\, qj} = \left[\frac{\sin(l+1/2)\,\theta}{(2l+1)\sin\theta/2} \right]^p \delta_{qj},$$

for $-l \leqslant q,\ j \leqslant l,\ l = 0, 1, \ldots,$ where $0 < \theta \leqslant \pi$ and $p > 0$. Since

$$\left| \frac{\sin(l+1/2)\,\theta}{(2l+1)\sin\theta/2} \right| \leqslant \frac{1}{(2l+1)\sin\theta/2}$$

TABLE I

Distribution Type and Rates of Convergence

Distribution	Smoothness type	Convergence rate
Laplace	Smooth	$n^{-s/(2s+6)}$
Rosenthal	Smooth	$n^{-s/(2s+2p+2)}$
Gaussian	Super-smooth	$(\log n)^{-s/2}$

whenever $0<\theta\leqslant\pi$ for $l=0, 1, ...,$ then according to (3.6), f_ε is ordinary smooth with $\beta=p$. For $SO(3)$ or any fixed $SO(N)$, as the convolution product index $p\to\infty$, then

$$f_\varepsilon(g)\,dg\to dg$$

in various metrics including L^2.

4.3. *Gaussian Distribution*

The Gaussian distribution can be solved on general Riemannian manifold by solving the appropriate heat equation. Since the $D^l_{q_1, q_2}$ are the eigenfunctions of the Laplacian Δ on $SO(3)$ with eigenvalue $l(l+1)/2$, for $-l\leqslant q\leqslant l$, $l=0, 1, ...,$ after taking account of the rotational symmetries, we can write the distribution as

$$f_\varepsilon=\sum_{l\geqslant 0}\sum_{q=-l}^{l}\exp(-tl(l+1))(2l+1)\,\overline{D^l_{qq}}, \tag{4.2}$$

for $t>0$. Consequently,

$$\hat{f}^{\,l}_{\varepsilon,\,qj}=\exp(-tl(l+1))\,\delta_{qj},$$

so that it is an example of a super-smooth distribution with $\gamma=1/t$ and $\beta=2$.

4.4. *Summary of Distributions*

In Table I we summarize the above distributions along with their smoothness properties. In addition, we state their optimal rate of convergence.

5. THE VON MISES–FISHER DISTRIBUTION

The most widely referenced distribution for directional data is the von Mises–Fisher matrix distribution; see, for example, Khatri and Mardia

(1977). Unlike the distributions of Section 4, this distribution is directly expressed as

$$f_{\varepsilon(g)}(x) = c(\kappa) \exp(\kappa \operatorname{tr} g^{-1}x), \tag{5.1}$$

where $\kappa > 0$ is the concentration parameter around the mean rotation $g \in SO(3)$ and tr denotes the trace operator; see Khatri and Mardia (1977). To understand the nature of the smoothness of (5.1) we need to calculate the rotational Fourier transform; however, prior to doing so, observe that if g and κ are known, then we can assume $g = e$, the unit element of $SO(3)$. This comes from the observation that

$$f_{\varepsilon(g)} * f_X(\omega) = f_{\varepsilon(e)} * f_X(g^{-1}\omega), \tag{5.2}$$

so that we can re-orient the estimation by g^{-1} and estimate the density using $f_{\varepsilon(e)}$ instead of $f_{\varepsilon(g)}$. Using results from the representation theory of $SO(3)$, in particular Schur's lemma and the Clebsch–Gordan formula, the rotational Fourier coefficients of (5.1) can be calculated.

In Section 2, we state that $\{D^l: l = 0, 1, ...\}$ constitute a collection of inequivalent irreducible representations of $SO(3)$. Two consequences of this fact are the following. First, suppose f: $SO(3) \to \mathbb{C}$ is a *class function*, which means that $f(gxg^{-1}) = f(x)$ for all x, $g \in SO(3)$. Then Schur's lemma, see Bröcker and tom Diek (1985), says that

$$\hat{f}^l_{qj} = \frac{\delta_{qj}}{(2l+1)} \int_{SO(3)} f(g) \overline{\chi}_l(g) \, dg, \tag{5.3}$$

where $-l \leqslant q$, $j \leqslant l$ and $\chi_l = \operatorname{tr} D^l$ are the *irreducible characters* of $SO(3)$, $l = 0, 1, 2, ...$, in particular, the rotational Fourier transforms of class functions are a constant multiple of the identity matrix. The second property is that tensor products of irreducible representations, can be decomposed as a direct sum, the so-called Clebsch–Gordan formula; see Bröcker and tom Diek (1985). Indeed,

$$D^k \otimes D^l = D^{|k-l|} \oplus D^{|k-l|+1} \oplus \cdots \oplus D^{k+l}, \tag{5.4}$$

for all k, $l = 0, 1, ...$ Now for the irreducible characters we have

$$\chi_k \cdot \chi_l = \chi_{|k-l|} + \chi_{|k-l|+1} + \cdots + \chi_{k+l}, \tag{5.5}$$

for all k, $l = 0, 1, ...$

We are now ready to discuss the rotational Fourier transform of (5.1). Assume $g=e$ as in (5.2). Note that $f_{\varepsilon(e)}$ is a class function since it depends on $\chi_l(x)=\operatorname{tr} x$. Hence by Schur's lemma we need to calculate

$$\int_{SO(3)} \exp(\kappa \operatorname{tr} x)\, \bar{\chi}_l(x)\, dx = \int_{SO(3)} \exp(\kappa\chi_1(x))\, \bar{\chi}_l(x)\, dx$$

$$= \sum_{k=0}^{\infty} \frac{\kappa^k}{k!} \int_{SO(3)} (\chi_1(x))^k\, \bar{\chi}_l(x)\, dx, \tag{5.6}$$

for all $l=0, 1, \ldots$. By the Cauchy–Schwarz inequality, one can show that (5.6) is bounded, hence taking the Taylor series expansion for the exponential function and interchanging integration with summation is justified.

Now by the Clebsch–Gordan formula, we can write

$$\chi_1^k = \sum_{p=0}^{k} \eta_{p,k}\chi_p, \tag{5.7}$$

where $\eta_{p,k} \geqslant 1$ for $p \geqslant 0$, $k \geqslant 1$. The set of irreducible characters $\{\chi_l: l=0, 1, \ldots\}$, is a complete orthonormal basis for the space of square integrable class functions hence

$$\int \chi_1^k \bar{\chi}_l = \eta_{l,k}, \tag{5.8}$$

where $k \geqslant l$. Now applying Cauchy–Schwarz to (5.8) and noting that $|\chi_1(g)| \leqslant \chi_1(e) = 3$ for all $g \in SO(3)$, we conclude that

$$1 \leqslant \eta_{l,k} \leqslant 3^k, \tag{5.9}$$

for all $k \geqslant l$ and $l=0, 1, 2, \ldots$. The immediate result is that when we apply (5.9) to (5.6), then

$$\frac{c(\kappa)\,\kappa^l}{(2l+1)\, l!} \leqslant \hat{f}^{\,l}_{\varepsilon(e),\, qq} \leqslant \frac{2c(\kappa)(3\kappa)^l}{(2l+1)\, l!}, \tag{5.10}$$

for $-l \leqslant q \leqslant l$ as $l \to \infty$. The upper bound uses the fact that

$$\exp(x) - \sum_{j=0}^{n-1} x^j/j! \leqslant 2x^n/n!$$

for $0 < x < n/2$.

Our interest is when $l \to \infty$, hence we can apply Stirlings approximation to (5.10) so that

$$\frac{c(\kappa)(\kappa/\mathrm{e})^l}{\sqrt{2\pi}\,(2l+1)} \exp(-(l+1/2)\log l)$$

$$\leqslant \hat{f}^l_{\varepsilon(e),\,qq} \leqslant \frac{2c(\kappa)(3\kappa/\mathrm{e})^l}{\sqrt{2\pi}\,(2l+1)} \exp(-(l+1/2)\log l), \qquad (5.11)$$

as $l \to \infty$ for $\kappa > 0$.

One can see that with the appearance of the logarithm term in the exponent of (5.11), we cannot get the same value for β on both sides of (5.11). The consequence is that the von Mises–Fisher distribution is somewhat anomalous in that

$$\|(\hat{f}^l_{\varepsilon})^{-1}\|_{\mathrm{op}} \leqq d_0^{-1} l^{3/2} \exp(l^{1+\beta}/\gamma) \qquad \text{and}$$

$$\|(\hat{f}^l_{\varepsilon})\|_{\mathrm{op}} \leqq d_1 l^{-3/2} \exp(-l/\gamma) \qquad \text{as} \quad l \to \infty, \qquad (5.12)$$

for some positive constants β, d_0, d_1, and γ, hence it is smoother than super-smooth.

We have the following result.

COROLLARY 5.1. *Suppose f_ε is distributed according to the von Mises–Fisher distribution. If $f_X \in H_s(S^2, M)$ for some $s > 1$, then*

$$(\log n)^{-s}$$

is a lower bound rate of convergence, while

$$(\log n)^{-s/(1+\beta)}$$

is an upper bound rate of convergence for any $\beta > 0$, as $n \to \infty$.

As an aside, it has long been known in the directional statistics literature that the von Mises–Fisher distribution although close, is not the same as the Gaussian distribution. The calculations of this section along with Section 4.3 show exactly the nature of the difference between the von Mises–Fisher and Gaussian distributions. As $l \to \infty$, the characteristic function of the Gaussian and the von Mises–Fisher distributions behave like

$$\exp(-tl^2) \qquad \text{and} \qquad \exp(-l\log l),$$

respectively for some $t > 0$. Clearly, the Gaussian distribution has slightly smoother tails which therefore accounts for a slower (but not by much) rate of convergence for the deconvolution density estimator relative to the von Mises–Fisher rotational errors.

6. PROOFS

We will prove Theorem 3.1 and Theorem 3.2 by first finding upper bounds for the smooth and super-smooth cases. Following this we will establish lower bounds for these smoothness classes and demonstrate that the upper and lower bounds match so that the resulting bounds are optimal.

The approach of Healy *et al.* (1998) will be used for calculating the upper bounds, while the approach of Koo (1993) and Koo and Chung (1998) will be used to find the lower bounds.

Forthwith, let M, M_1, M_2, ... denote positive constants independent of the sample size n and let C denote a positive constant which may have a different value at each of its appearances.

6.1. *Upper Bounds*

Let $\|f\|_\infty$ denote the L^∞ norm of a function on S^2.

LEMMA 6.1. *If $h \in H_s(S^2, M)$ with $s > 1$, then*

$$\|h\|_\infty \leqslant C(M, s),$$

where $C(M, s)$ is a constant depending only on M and s.

Proof. Write $h = \sum_{l,q} \hat{h}^l_q Y^l_q$. Observe that

$$\begin{aligned} |h(\omega)|^2 &\leqslant \left(\sum_{l,q} (1 + l(l+1))^s \, |\hat{h}^l_q|^2 \right) \left(\sum_{l,q} (1 + l(l+1))^{-s} \, |Y^l_q(\omega)|^2 \right) \\ &\leqslant (1 + M) \sum_l (1 + l(l+1))^{-s} \, (2l+1)/(4\pi). \end{aligned}$$

In the above, we use the addition formula,

$$\sum_{q=-l}^{l} \bar{Y}^l_q(\omega) \; Y^l_q(v) = \frac{2l+1}{4\pi} \, P_l(\cos \gamma(\omega, v)), \tag{6.1}$$

where $\gamma(\omega, v)$ represents the angle between $v, \omega \in S^2$ and $P_l(1) = 1$ are the Legendre polynomials, for all $l \geqslant 0$. Since $s > 1$, the series $\sum_l (1 + l(l+1))^{-s} \, (2l+1)$ is convergent. ∎

LEMMA 6.2. *Suppose* $f_X \in H_s(S^2, M)$ *with* $s > 1$. *Then*

$$\int_{\omega \in S^2} Var(f_X^n(\omega))\, d\omega \ll \begin{cases} m^{2\beta+2} n^{-1} & \textit{smooth} \\ \exp(2m^\beta/\gamma)\, m^{-2\beta_0+2} n^{-1} & \textit{super-smooth} \end{cases}$$

as $n \to \infty$.

Proof. We note that

$$Var(f_X^n(\omega)) = \frac{1}{n}\, Var(K_n^\varepsilon(\omega, Z)) \leqslant \frac{1}{n}\, EK_n^\varepsilon(\omega, Z)\, \overline{K_n^\varepsilon(\omega, Z)},$$

for $\omega \in S^2$, where Z denotes the random S^2 element εX. By Lemma 6.1 and (3.2), f_Z is bounded by a constant C so that

$$\int_{S^2} K_n^\varepsilon(\omega, z)\, \overline{K_n^\varepsilon(\omega, z)}\, f_Z(z)\, dz \leqslant C \int_{S^2} K_n^\varepsilon(\omega, z)\, \overline{K_n^\varepsilon(\omega, z)}\, ds$$
$$= \sum_{l=0}^{m} \sum_{s=-l}^{l} \left| \sum_{q=-l}^{l} Y_q^l(\omega)\, \hat{f}^l_{\varepsilon^{-1}, qs} \right|^2.$$

Define

$$\zeta_\omega(v) = \sum_{q=-l}^{l} Y_q^l(\omega)\, \bar{Y}_q^l(v),$$

where $\omega,\ v \in S^2$. Then

$$\sum_{s=-l}^{l} \left| \sum_{q=-l}^{l} Y_q^l(\omega)\, \hat{f}^l_{\varepsilon^{-1}, qs} \right|^2 = \|\hat{f}^l_{\varepsilon^{-1}} \zeta_\omega\|_2^2$$
$$\leqslant \|\hat{f}^l_{\varepsilon^{-1}}\|_{\text{op}}^2\, \|\zeta_\omega\|_2^2$$
$$= \|\hat{f}^l_{\varepsilon^{-1}}\|_{\text{op}}^2 \sum_{q=-l}^{l} |Y_q^l(\omega)|^2$$

The second line uses the operator inequality (3.4), the third line uses Parseval's identity, while the last line uses the addition formula (6.1) along with the fact that $P_l(1) = 1$. Therefore, we have

$$\int_{S^2} Var(f_X^n(\omega))\, d\omega \leqslant \frac{C}{n} \sum_{l=0}^{m} \|\hat{f}^l_{\varepsilon^{-1}}\|_{\text{op}}^2\, (2l+1).$$

Now apply the definitions of smooth and super-smooth to the last expression. ∎

LEMMA 6.3. *Suppose $f_X \in H_s(S^2, M)$, where $s > 1$. Then*

$$\|f_X - Ef^n_X\|_2^2 \ll m^{-2s}.$$

Proof. Observe that

$$F_X(\omega) - Ef^n_X(\omega) = \sum_{l>m} \sum_{q=-l}^{l} \hat{f}^l_{X,q} Y^l_q(\omega).$$

Since

$$m^{2s} \|f_X - Ef^n_X\|_2^2 = m^{2s} \sum_{l>m} \sum_{q=-l}^{l} |\hat{f}^l_{X,q}|^2$$

$$\leqslant \sum_{l>m} \sum_{q=-l}^{l} (1 + l(l+1))^s |\hat{f}^l_{X,q}|^2 \leqslant (1+M)$$

we have the desired result. ∎

By putting together Lemma 6.2 and Lemma 6.3, upper bound estimates can be established as

$$E\|f^n_X - f_X\|_2^2 \ll \begin{cases} m^{2\beta+2} n^{-1} + m^{-2s} & \text{smooth} \\ \exp(2m^\beta/\gamma)\, m^{-2\beta_0+2} n^{-1} + m^{-2s} & \text{super-smooth} \end{cases}$$

as $n \to \infty$. Consequently, choosing

$$m \asymp \begin{cases} n^{1/2(s+\beta+1)} & \text{smooth} \\ (\log n)^{1/\beta} & \text{super-smooth} \end{cases}$$

as $n \to \infty$ optimizes the upper bound rates, respectively.

6.2. *Lower Bounds*

To show that the upper bound rates are optimal rates, we calculate lower bound rates of convergence and show that these are the same as the upper bounds. In calculating the lower bounds we follow the popular approach:

- specify a subproblem;
- use Fano's lemma to calculate the difficulty of the subproblem.

Let N_n be a positive integer depending on n and define

$$V_n = \{(l, q): q = 0, 1, ..., l,\ l = N_n + 1, ..., 2N_n\}.$$

Define

$$\psi^l_q = \begin{cases} (Y^l_q + Y^l_{-q})/\sqrt{2} & \text{if} \quad q>0 \quad \text{is even} \\ Y^l_0 & \text{if} \quad q=0 \\ (Y^l_q - Y^l_{-q})/\sqrt{2} & \text{if} \quad q>0 \quad \text{is odd} \end{cases}$$

for $(l, q) \in V_n$. Since $\overline{Y^l_q} = (-1)^q \, Y^l_{-q}$, ψ^l_q is a real-valued function for each $(l, q) \in V_n$. Let $\tau = \tau(n) = \{\tau^l_q : (l, q) \in V_n\}$ and consider the function

$$f_\tau = (4\pi)^{-1/2} + M_1 N_n^{-s-1} \sum_{l=N_{n+1}}^{2N_n} \sum_{q=-l}^{l} \tau^l_q \psi^l_q, \tag{6.2}$$

where M_1 is a positive constant such that $3(7^s)\, M_1^2 \leqslant M$. Finally, let

$$\mathscr{F}_n = \{f_\tau : \tau \in \{0, 1\}^{|V_n|}\}, \tag{6.3}$$

where for some given finite set, $|\cdot|$ will denote it's cardinality and assume that $N_n \to \infty$ as $n \to \infty$. Under the assumption that $s>1$, we have the following lemma.

LEMMA 6.4. *For n sufficiently large, $\mathscr{F}_n \subset H_s(S^2, M)$ and $M_2^{-1} \leqslant f \leqslant M_2$ for all $f \in \mathscr{F}_n$.*

Proof. Let

$$\tilde{\tau}^l_q = \begin{cases} \tau^l_q/\sqrt{2} & \text{if} \quad q>0 \\ \tau^l_0 & \text{if} \quad q=0 \\ (-1)^q \, \tau^l_q/\sqrt{2} & \text{if} \quad q<0. \end{cases}$$

Then

$$\sum_{q=0}^{l} \tau^l_q \psi^l_q = \sum_{q=-l}^{l} \tilde{\tau}^l_q \, Y^l_q.$$

Using this fact and applying the Sobolev norm to (6.2), we get

$$\begin{aligned} \|f_\tau\|^2_{H_s} &= 1 + M_1^2 N_n^{-2s-2} \sum_{l=N_n+1}^{2N_n} \sum_{q=-l}^{l} (\tilde{\tau}^l_q)^2 \, (1 + l(l+1))^s \\ &\leqslant 1 + M_1^2 N_n^{-2s-2} \sum_{l=N_n+1}^{2N_n} (l+1)(1+l)(1+l(l+1))^s \\ &\leqslant 1 + 7^s M_1^2 N_n^{-2} \sum_{l=N_N+1}^{2N_n} (l+1) \\ &\leqslant 1 + 3(7^s)\, M_1^2. \end{aligned} \tag{6.4}$$

Consequently, we have shown that $f_\tau \in H_s(S^2, M)$. Since $s > 1$, we have the desired result. ∎

Now let $f, g \in \mathscr{F}_n$ with $f \neq g$. By the orthonormality of ψ^l_q, we have

$$\|f - g\|_2 \geqslant M_1 N_n^{-s-1}. \tag{6.5}$$

It follows from (6.5) and Lemma 3.1 of Koo (1993) that there exists an $\mathscr{F}^0_n \subset \mathscr{F}_n$ such that for all $u, v \in \mathscr{F}^0_n$ with $u \neq v$,

$$\|u - v\|_2 \geqslant M_3 N_n^{-s} \qquad \text{and} \qquad \log(|\mathscr{F}^0_n| - 1) \geqslant M_4 N_n^2. \tag{6.6}$$

Define

$$\delta \equiv u - v = M_1 N_n^{-s-1} \sum_{l=N_{n+1}}^{2N_n} \sum_{q=-l}^{l} \tilde{\tau}^l_q Y^l_q.$$

Then

$$\begin{aligned}
\|f_\varepsilon * \delta\|_2^2 &= \sum_{l=N_{n+1}}^{2N_n} \sum_{q=-l}^{l} |\widehat{(f_\varepsilon * \delta)}{}^l_q|^2 \\
&= M_1^2 N^{-2(s+1)} \sum_{l=N_{n+1}}^{2N_n} \sum_{q=-l}^{l} \left| \sum_{j=-1}^{l} \hat{f}^l_{\varepsilon, qj} \tilde{\tau}^l_j \right|^2 \\
&\leqslant M_1^2 N^{-2(s+1)} \sum_{l=N_{n+1}}^{2N_n} \|\hat{f}^l_\varepsilon\|^2_{\text{op}} \sum_{q=-l}^{l} |\tilde{\tau}^l_j|^2 \\
&\leqslant M_1^2 N^{-2(s+1)} \sum_{l=N_n+1}^{2N_n} \|\hat{f}^l_\varepsilon\|^2_{\text{op}} (l+1).
\end{aligned}$$

The first inequality above is obtained by the operator inequality, (3.4), while the last line uses the definition of $\tilde{\tau}$. This therefore implies that

$$\|f_\varepsilon * u - f_\varepsilon * v\|_2^2 \ll \begin{cases} N_n^{-2(s+\beta)} & \text{smooth} \\ \exp(-2N_n^\beta/\gamma)\, N_n^{-2s+2\beta_1} & \text{super-smooth} \end{cases} \tag{6.7}$$

as $n \to \infty$.

By (3.2) and Lemma 6.4, we have

$$M_2^{-1} \leqslant f_\varepsilon * f_\tau \leqslant M_2. \tag{6.8}$$

Now the Kullback–Leibler information divergence $D(f \| g)$ between two densities f and g is defined by $D(f \| g) = \int_{S^2} f \log(f/g)$ and

$$D(f_\varepsilon * f \| f_\varepsilon * g) \leqslant \int_{S^2} \frac{(f_\varepsilon * f - f_\varepsilon * g)^2}{f_\varepsilon * g}. \tag{6.9}$$

Therefore, by (6.7), (6.8) and Jensen's inequality

$$D(f_\varepsilon * u \| f_\varepsilon * v) \ll \begin{cases} N_n^{-2(s+\beta)} & \text{smooth} \\ \exp(-2N_n^{\beta}/\gamma)\, N_n^{-2s+2\beta_1} & \text{super-smooth} \end{cases} \tag{6.10}$$

for all $u, v \in \mathscr{F}_n^0$, as $n \to \infty$.

By Fano's lemma, see, for example, Birgé (1983), Yatrocos (1988), or Koo (1993), if f^n is any estimator of f, then

$$\begin{aligned} \sup_{f \in H_s(S^2, M)} P_f(\|f^n - f\|_2 > cN_n^{-s}) &\geqslant \sup_{f \in \mathscr{F}_n^0} P_f(\|f^n - f\|_2 > cN_n^{-s}) \\ &\geqslant 1 - \frac{nD(f_\varepsilon * u \| f_\varepsilon * v) + \log 2}{\log(|\mathscr{F}_n^0| - 1)}. \end{aligned} \tag{6.11}$$

Apply (6.6) and (6.10) to the last line in (6.11). Finally, let

$$N_n \asymp \begin{cases} n^{1/2(s+\beta+1)} & \text{smooth} \\ (\log n)^{1/\beta} & \text{super-smooth} \end{cases}$$

as $n \to \infty$. Then it follows for the two smoothness classes that

$$\lim_{c \to 0} \liminf_{n} \sup_{f \in H_s(S^2, M)} P_f(\|f^n - f\| > cN_n^{-s}) = 1,$$

thus establishing the lower bound.

We can now use these lower bounds along with the upper bound results which then completes the proofs to Theorem 3.1 and Theorem 3.2.

ACKNOWLEDGMENTS

Parts of this research were completed while the first author was visiting the Department of Statistics, Hallym University, and while the second author was visiting the Department of Mathematics and Statistics, University of Guelph. They thank the two institutions for their hospitality during these visits.

REFERENCES

1. L. Birgé, Approximation dans les espaces metriques et theorie de l'estimation, *Z. Wahrsch. Verw. Gebiete* **65** (1983), 181–237.
2. T. Bröcker and T. tom Diek, "Representations of Compact Lie Groups," Springer-Verlag, New York, 1985.
3. P. Diaconis, "Group Representations in Probability and Statistics," Institute of Mathematical Statistics Lecture Notes, Monograph Series, Hayward, 1988.
4. S. Efromovich, Density estimation for the case of supersmooth measurement error, *J. Amer. Statist. Assoc.* **92** (1997), 526–535.
5. J. Fan, Global behavior of deconvolution kernel estimates, *Statist. Sinica* **1** (1991), 541–551.
6. J. Fan, On the optimal rates of convergence for nonparametric deconvolution problems, *Ann. Statist.* **19** (1991a), 1257–1272.
7. J. Fan, Adaptively local one-dimensional subproblems with application to a deconvolution problem, *Ann. Statist.* **21** (1993), 600–610.
8. D. M. Healy, H. Hendriks, and P. T. Kim, Spherical deconvolution, *J. Multivariate Anal.* **67** (1998), 1–22.
9. D. M. Healy and P. T. Kim, An empirical Bayes approach to directional data and efficient computation on the sphere, *Ann. Statist.* **24** (1996), 232–254.
10. H. Hendriks, Nonparametric estimation of a probability density on a Riemannian manifold using Fourier expansions, *Ann. Statist.* **18** (1990), 832–849.
11. C. G. Khatri and K. F. Mardia, The von Mises–Fisher distribution in orientation statistics, *J. Roy. Statist. Soc. Ser. B* **39** (1977), 95–106.
12. J. Y. Koo, Optimal rates of convergence for nonparametric statistical inverse problems, *Ann. Statist.* **21** (1993), 590–599.
13. J. Y. Koo and H. Y. Chung, Log-density estimation in linear inverse problems, *Ann. Statist.* **26** (1998), 335–362.
14. J. T. Lo and L. R. Eshelman, Exponential Fourier densities on $SO(3)$ and optimal estimation and detection for rotational processes, *SIAM J. Appl. Math.* **36** (1979), 73–82.
15. A. C. M. Rooij and F. H. Ruymgaart, Regularized deconvolution on the circle and the sphere, *in* "Nonparametric Functional Estimation and Relate Topics" (G. Roussas, Ed.), pp. 679–690, Kluwer Academic, Amsterdam, 1991.
16. J. S. Rosenthal, Random rotations: Characters and random walks on $SO(N)$, *Ann. Probab.* **22** (1994), 398–423.
17. C. J. Stone, Optimal rates of convergence for nonparametric estimators, *Ann. Statist.* **8** (1980), 1348–1360.
18. J. D. Talman, "Special Functions: A Group Theoretic Approach," Benjamin, New York, 1968.
19. A. Terras, "Harmonic Analysis on Symmetric Spaces and Applications," Springer-Verlag, New York, 1985.
20. G. Wahba, Spline interpolation and smoothing on the sphere, *SIAM J. Sc. Statist. Comput.* **2** (1981), 5–16.
21. Y. G. Yatracos, A lower bound on the error in nonparametric regression type problems, *Ann. Statist.* **16** (1988), 1180–1187.