The Annals of Statistics
1998, Vol. 26, No. 3, 1083–1102

DECONVOLUTION DENSITY ESTIMATION ON $SO(N)$

BY PETER T. KIM[1]

University of Guelph and Yonsei University

This paper develops nonparametric deconvolution density estimation over $SO(N)$, the group of $N \times N$ orthogonal matrices of determinant 1. The methodology is to use the group and manifold structures to adapt the Euclidean deconvolution techniques to this Lie group environment. This is achieved by employing the theory of group representations explicit to $SO(N)$. General consistency results are obtained with specific rates of convergence achieved under sufficient smoothness conditions. Application to empirical Bayes prior estimation and inference is also discussed.

1. Introduction. In Euclidean nonparametric mixture models, one has

$$\int f(x-\mu) g(\mu)\, d\mu, \tag{1.1}$$

where $f(\cdot)$ is assumed known and the parameter of interest is the unknown mixing density $g(\cdot)$. Estimation of the mixing density can be performed using deconvolution density estimation, which has been studied in depth by several authors; see, for example, Devroye (1989), Zhang (1990), Fan (1991a, b) and Diggle and Hall (1993), and the references therein.

Mixture models in general have been of considerable importance in statistics. Lindsay (1995) provides an excellent account of the subject as well as an extensive bibliography. Although the nonparametric version (1.1) is but one aspect of the entire mixture modelling strategy, it nevertheless provides additional statistical procedures such as nonparametric empirical Bayes estimation; see, for example, Maritz and Lwin (1989), and, for nonparametric errors in variable regression, see Fan and Truong (1993).

Let us now change the discussion and briefly mention some ongoing research on orientation statistics because orientation statistics provides a fundamental rationale for extending the mixture framework into a non-Euclidean environment.

There has been some statistical interest in a situation where one observes three mutually orthogonal unit direction vectors. The data originates from vector cardiogram orientation, which was introduced in Downs (1972) with various authors further developing this area; see Khatri and Mardia (1977) and Prentice (1986, 1989). As one can see, this takes us away from the

Received November 1996; revised November 1997.

[1]Supported in part by a grant from NSERC Canada, OGP46204 and by BC Matthews Alumni Fund, University of Guelph.

AMS 1991 *subject classifications*. Primary 62G05; secondary 58G25.

Key words and phrases. Asymptotic variance, asymptotic bias, consistency, differentiable manifold, irreducible representations, unitary matrices.

Euclidean setting to a non-Euclidean environment where the state space now becomes $SO(N)$, the group of $N \times N$ orthogonal matrices of determinant 1. Mathematically, $SO(N)$ is a compact Lie group and there is a certain appeal to statisticians because $SO(N)$ and compact Lie groups can be realized as the compact space of matrices that are frequently encountered in multivariate analysis; see, for example, Farrell (1985).

A location type model in $SO(N)$ often takes the form

$$f(x\mu^{-1}) = f(\operatorname{tr} x\mu^t), \tag{1.2}$$

where $f(\cdot)$ is a density on $SO(N)$ absolutely continuous with respect to the normalized Haar measure on $SO(N)$, $x, \mu \in SO(N)$ and superscript t denotes matrix transpose; see Khatri and Mardia (1977). If we then extend (1.2) into a nonparametric mixture setting, the analogous representation to (1.1) would be

$$\int f(x\mu^{-1}) g(\mu)\, d\mu, \tag{1.3}$$

where again $f(\cdot)$ is assumed known and the parameter of interest is the unknown mixing density $g(\cdot)$. It turns out that (1.3) is a convolution in the Lie group sense and so, if we wish to estimate the mixing density as in the Euclidean case, one strategy is to develop a deconvolution technique on $SO(N)$. It should be strongly emphasized that if a successful generalization of deconvolution to $SO(N)$ can be made, this fulfills a first, but an important step, in extending the statistical tools associated with mixture models to orientation statistics in general and vector cardiogram orientation in particular. This extension will therefore be the subject of this paper for which we now provide an overview.

In Section 2, we undergo some preparation for Fourier analysis on compact groups specializing down to $SO(N)$. Most of the material is available in the mathematical literature; see, for example, Talman (1968), Vilenkin (1968), Helgason (1978, 1984), Warner (1983), Bröcker and tom Dieck (1985) and Gong (1991).

In Section 3 we tackle the problem of non-Euclidean deconvolution. In the statistical literature, deconvolution methods are mainly done on Euclidean space where the objective is to produce estimators of the measurement density when observations consist of the true measurement plus additive noise. However, as stated at the beginning, deconvolution methodologies for compact Lie groups and homogeneous spaces are also needed. In addition to vector cardiogram orientation, deconvolution would be appropriate for problems associated with errors in variables in spherical regression, as developed by Chang (1989), as well as nonparametric empirical Bayes estimators of prior densities when the parameter space is a compact Lie group; see Kim (1991). We establish L^2 consistent deconvolution density estimators. Rates of convergence are established under sufficient smoothness conditions on the density.

Section 4 deals with applications. We will first examine the case of $SO(3)$, the lowest dimensional non-abelian case. We also discuss a particular error distribution derived from the work of Rosenthal (1994) on random walks on $SO(N)$. An application to nonparametric empirical Bayes estimation and inference for $SO(N)$ parameters is established. This provides (nonparametric) extensions to some of the earlier (parametric) work on orientation statistics; see Downs (1972), Prentice (1986) and Khatri and Mardia (1977).

Some additional comments are made in Section 5, including the relevance of implementing fast algorithms. All proofs are provided in Section 6.

The material in this paper requires some technical knowledge concerning compact Lie groups and their representations. As a minimal requirement, the Appendices as well as Section 2 sketch the relevant material needed to read this paper. Consequently, the reader should review this material first.

Prior to starting the discussion, the following comment should be made. The theory of group representations is a very rich, beautiful and difficult branch of mathematics. Our short account of the topic is included only for the purpose of getting the idea across as needed for the problem at hand. Put differently, we do little justice to portraying the richness of the theory as well as its broad historical evolution. There are numerous books on group representations, and the reader is encouraged to look through them if they find interest in the current paper. A good source for the understanding of this paper is Bröcker and tom Dieck (1985). For general Lie groups, consult Warner (1983) and for finite groups, Serre (1977) or Diaconis (1988). For differential geometry consult Spivak (1973), Helgason (1978) and Warner (1983).

2. Preparation. For a compact Lie group G, Fourier analysis involves expanding functions on G by its irreducible representations. In particular, denote by $\mathrm{Irr}(G, \mathbf{C})$ the collection of inequivalent irreducible representations of G. The definition and some properties are reviewed in Appendix A. For $f \in L^2(G)$, we define the *Fourier transform* with respect to an irreducible representation as

$$\hat{f}(U) = \int_G U(g^{-1}) f(g)\, dg \tag{2.1}$$

for $U \in \mathrm{Irr}(G, \mathbf{C})$, where dg denotes the unit Haar measure on G normalized by the volume of G. The *Fourier inversion* can be written as

$$f(g) = \sum_{U \in \mathrm{Irr}(G, \mathbf{C})} d_U \operatorname{tr} U(g) \hat{f}(U), \tag{2.2}$$

where $g \in G$ and d_U is the dimension of the representation $U \in \mathrm{Irr}(G, \mathbf{C})$. We note that, strictly speaking, (2.2) should be interpreted as in the L^2 sense although with sufficient smoothness, it can hold with equality pointwise almost everywhere.

Given two functions $f, h \in L^2(G)$, define the *convolution* by

$$f * h(g) = \int_G f(x^{-1} g) h(x)\, dx. \tag{2.3}$$

We note the similarity of the above to convolution on Euclidean space when we express $x^{-1} = -x$. The following is a key result.

LEMMA 2.1. *For* $f, h \in L^2(G)$,

$$\widehat{(f * h)}(U) = \hat{f}(U)\hat{h}(U),$$

where $U \in \mathrm{Irr}(G, \mathbf{C})$.

The proof is straightforward.

2.1. *Specialization to* $SO(N)$. We now specialize the above discussion to $G = SO(N)$. First, the dimension of $SO(N)$ as a manifold is

$$\dim SO(N) = \sum_{l=1}^{N-1} l = \frac{N(N-1)}{2}. \tag{2.4}$$

This comes from the fact that the Lie algebra of $SO(N)$ (the tangent space at the unit element) is $so(N)$, the space of $N \times N$ skew symmetric matrices.

For $SO(N)$, the indexing of the irreducible representations is fundamental. Each (inequivalent) element of $\mathrm{Irr}(SO(N), \mathbf{C})$ is characterized by a k-tuple of integers $j = (j_1, \dots, j_k)$ called the *signature*. Now this signature varies depending on whether N is even or odd and so let us make the following notation. For $N = 2k + 1$ odd, let

$$J_m = \{ j \in \mathbf{Z}^k : m \geq j_1 \geq j_2 \geq \cdots \geq j_k \geq 0 \}, \tag{2.5}$$

where $\mathbf{Z}$ denotes the set of all integers. On the other hand for $N = 2k$ even, let

$$J_m = \{ j \in \mathbf{Z}^k : m \geq j_1 \geq j_2 \geq \cdots \geq |j_k| \geq 0 \}. \tag{2.6}$$

One notices that in the even case, an extra set of indices come out from the relation $|j_k|$. This is explained in more detail in Appendix B. To get all of the irreducible representations, let $m \to \infty$ for both the even and odd cases and define

$$J = \lim_{m \to \infty} J_m. \tag{2.7}$$

Consequently, each $U \in \mathrm{Irr}(SO(N), \mathbf{C})$ can be indexed by its signature U_j, along with $\chi_{U_j} = \chi_j$ and $d_{U_j} = d_j$ for all j; see Appendix A for the appropriate definitions. This means that for $f \in L^2(SO(N))$, we can express (2.1) by

$$\hat{f}(j) = \int_{SO(N)} U_j(g^{-1}) f(g)\, dg \tag{2.8}$$

and (2.2) by

$$f(g) = \sum_{j \in J} d_j \operatorname{tr} U_j(g) \hat{f}(j). \tag{2.9}$$

We should point out that the characterization of elements of $\mathrm{Irr}(SO(N), \mathbf{C})$ is unique only up to conjugation.

Consider Δ the Laplace–Beltrami operator on $SO(N)$. Then the components of the irreducible representations are the eigenfunctions of Δ so that

$$\left\{d_j^{1/2}U_j\colon j \in J\right\}$$

is a complete orthonormal basis of $L^2(SO(N))$. For $N = 2k + 1$, the corresponding eigenvalue is

$$\lambda_j = j_1^2 + \cdots + j_k^2 + (2k-1)j_1 + (2k-3)j_2 + \cdots + j_k, \tag{2.10}$$

while for $N = 2k$,

$$\lambda_j = j_1^2 + \cdots + j_k^2 + (2k-2)j_1 + (2k-4)j_2 + \cdots + 2j_{k-1}. \tag{2.11}$$

More explicit descriptions are provided in Appendix B.

3. The deconvolution problem and main results. Suppose the observation Y is over $SO(N)$ and is made up of the true measurement X composed with noise ε. The true measurement can then be viewed as some random quantity on $SO(N)$ along with the error being some random quantity also on $SO(N)$. Consequently, the observations consist of

$$Y = X\varepsilon,$$

where the multiplication is with respect to the group action $SO(N) \times SO(N) \to SO(N)$.

The density of Y is then the convolution of the densities of ε and X, that is,

$$f_Y(u) = f_X * f_\varepsilon(u) = \int_{SO(N)} f_X(v^{-1}u) f_\varepsilon(v)\, dv.$$

By Lemma 2.1, we can write

$$\hat{f}_X(j) = \hat{f}_Y(j)\left[\hat{f}_\varepsilon(j)\right]^{-1},$$

provided that $\hat{f}_\varepsilon(j)$ is invertible. For ease of notation, henceforth we will define

$$\left[\hat{f}_\varepsilon(j)\right]^{-1} = \hat{f}_{\varepsilon^{-1}}(j). \tag{3.1}$$

In general f_Y is assumed to be unknown, hence $\hat{f}_Y(j)$ is unknown. Suppose we have a random sample $Y_1, \dots, Y_n$. Then we form the empirical characteristic function

$$\hat{f}_Y^n(j) = \frac{1}{n}\sum_{l=1}^{n} U_j\left(Y_l^{-1}\right), \tag{3.2}$$

similar to the empirical characteristic function on Euclidean space; see Feuerverger and Murieka (1977). Following this by using (3.2) in the Fourier inversion formula (2.9), we can obtain a nonparametric deconvolution density estimator for f_X by

$$f_X^n(g) = \sum_{j \in J_m} d_j \operatorname{tr}\left\{U_j(g)\hat{f}_Y^n(j)\hat{f}_{\varepsilon^{-1}}(j)\right\}, \tag{3.3}$$

where $m = m(n) \to \infty$ as $n \to \infty$.

Alternatively, define

$$K_n^{\varepsilon}(g) = \sum_{j \in J_m} d_j \operatorname{tr}\left\{U_j(g) \hat{f}_{\varepsilon^{-1}}(j)\right\}.$$

Then (3.3) can be written in the more familiar kernel form,

$$f_X^n(g) = \frac{1}{n} \sum_{l=1}^{n} K_n^{\varepsilon}\left(Y_l^{-1} g\right) \tag{3.4}$$

for $g \in SO(N)$.

3.1. *Consistency results*. The following notation will be used. For two sequences $\{a_n\}$ and $\{b_n\}$, we will denote $a_n = O(b_n)$ by $a_n \ll b_n$. Furthermore, $\|\cdot\|_2$ will denote the usual L^2-norm while $\|\cdot\|_{\mathrm{op}}$ will denote the usual operator norm.

We now state the main results where the meaning of differentiability is with respect to $SO(N)$ being a differentiable manifold in addition to being a group.

THEOREM 3.1. *Suppose $\|\hat{f}_{\varepsilon^{-1}}(j)\|_{\mathrm{op}} \ll d_j^u$ for some $u \geq 0$. If f_Y is bounded and f_X is the pointwise limit of its Fourier series, then*

$$E|f_X^n(g) - f_X(g)|^2 \to 0$$

as $n \to \infty$ for all $g \in SO(N)$ provided $m^{[(\dim SO(N) - k)u + \dim SO(N)]} = o(n)$.

To obtain rates of convergence, smoothness conditions need to be imposed on f_X.

THEOREM 3.2. *Suppose $\|\hat{f}_{\varepsilon^{-1}}(j)\|_{\mathrm{op}} \ll d_j^u$ for some $u \geq 0$. If f_Y is bounded, f_X is $s \geq 1$ times differentiable and square-integrable, then*

$$E\|f_X^n - f_X\|_2^2 \ll n^{-2s/[2s + (\dim SO(N) - k)u + \dim SO(N)]}$$

as $n \to \infty$.

The question that naturally arises concerns the distribution of the errors ε. At one extreme is the Haar measure (uniform distribution) on $SO(N)$. In this case deconvolution is not possible since $\hat{f}_{\varepsilon} = 0$. One can see that in this case the true measurements are uniformly perturbed according to the group action, thus resulting in no hope of being able to recover f_X.

The other extreme would be point mass at the unit element of $SO(N)$. Denote by δ_e the density concentrated at the unit element $e \in SO(N)$. Then

$$\hat{f}_{\varepsilon}(j) = \int_{SO(N)} U_j(g^{-1}) \delta_e(g)\, dg = U_j(e) = \mathbf{I}_{d_j},$$

where $\mathbf{I}_{d_j}$ is the $d_j \times d_j$ identity matrix; therefore $\|\hat{f}_{\varepsilon^{-1}}(j)\|_{\mathrm{op}} \leq 1$. This corresponds to the case $u = 0$ in the above results and would be the ordinary density estimation on $SO(N)$. In fact, we get the following, which is Theorem 2.1 of Hendriks [(1990), page 834].

COROLLARY 3.3 [Hendriks (1990)]. *Suppose* $f_\varepsilon = \delta_e$. *If* f_X *is* $s \geq 1$ *times differentiable and square-integrable*, *then*

$$E\|f_X^n - f_X\|_2^2 \ll n^{-2s/[2s+\dim SO(N)]}$$

as $n \to \infty$.

Therefore, in order for deconvolution to work and at the same time be meaningful, the situation would have to be somewhere between the above two extremes. In the following section, we look at such an example.

4. Applications and examples. In this section, we will examine some special cases. In addition, application of the methodology to empirical Bayes estimation and inference will be discussed.

4.1. *Application to* $SO(3)$. As described in Section 2.1, define the empirical transform on $SO(3)$ by

$$\hat{f}_Y^n(j) = \frac{1}{n}\sum_{l=1}^{n} D_j\left(Y_l^{-1}\right),$$

where $j = 0, 1, \ldots$ and the (inequivalent) irreducible representations D are explicitly written out in Appendix A. Then

$$\hat{f}_X^n(j) = \frac{1}{n}\sum_{l=1}^{n} D_j\left(Y_l^{-1}\right)\hat{f}_{\varepsilon^{-1}}(j),$$

for $j = 0, 1, \ldots$ and the nonparametric deconvolution density estimator of f_X on $SO(3)$ will be

$$f_X^n(g) = \sum_{j=0}^{m}(2j+1)\operatorname{tr}\left\{D_j(g)\left[n^{-1}\sum_{l=1}^{n} D_j\left(Y_l^{-1}\right)\right]\hat{f}_{\varepsilon^{-1}}(j)\right\}, \tag{4.1}$$

for $g \in SO(3)$. Special cases of (4.1) have been considered in Healy, Hendriks and Kim (1995) and Healy and Kim (1996).

4.2. *An example inspired by Rosenthal*. Although some parametric estimation on $SO(N)$ has appeared in the statistical literature [see, e.g., Chang (1986) and Prentice and Mardia (1995)], a general deconvolution estimation problem on $SO(N)$ has not appeared. Consequently, there is in general a lack of models for errors on $SO(N)$ with well-understood spectral properties.

There has, however, appeared a somewhat related problem in probability associated with random walks on groups; see Diaconis (1988). Here one is interested in performing random walks on groups according to the group structure, 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 converge to the uniform measure is analytically studied using Fourier methods on the group.

Thus one can see the similarity of random walks on groups with deconvolution. 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 uniformity.

A very useful simplification for conjugate invariant functions, that is, $f_\varepsilon(g^{-1}xg) = f_\varepsilon(x)$, $x, g \in SO(N)$ is Schur's lemma; see, for example, Bröcker and tom Dieck (1985). In our case, this amounts to the following:

$$\hat{f}_\varepsilon(j) = \gamma_j \mathbf{I}_{d_j} \quad \text{where } \gamma_j = d_j^{-1} \int_{SO(N)} f_\varepsilon(g)\, \chi_j(g^{-1})\, dg. \tag{4.2}$$

To be concrete, consider the case of $SO(5)$ (although this argument should work for all N) and take the conjugacy class of

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

for $\theta \in (0, \pi]$. Setting $\theta = \pi$ and taking the uniform measure over the conjugacy class of R_π, let f_ε be the p-fold convolution product. Rosenthal [(1994), page 407], shows that

$$\hat{f}_\varepsilon(j) = \left[\frac{c_j}{d_j}\right]^p \mathbf{I}_{d_j},$$

where d_j is defined in Section 6 and c_j is formally the evaluation of the integral in (4.2) for this particular case. The particular evaluation is not of concern for us but rather that $c_j^2 \geq 1$, which can be established by consulting Proposition 3.1 [Rosenthal (1994), page 406]. Therefore,

$$\|\hat{f}_{\varepsilon^{-1}}(j)\|_{\mathrm{op}} \leq d_j^p,$$

for some fixed finite constant $p > 0$.

For $SO(5)$ or any fixed $SO(N)$, as the convolution product index $p \to \infty$ [in Rosenthal (1994), he uses k instead of p], then

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

in various metrics including L^2. Consequently, given such an error structure, under the conditions of Theorem 3.2, convergence occurs at a rate of

$$n^{-s/[s+4p+5]}, \qquad \text{as } n \to \infty.$$

4.3. *Empirical Bayes application.* Deconvolution methods can be used in an empirical Bayes setting. Let the sampling density be of the form

$$f(x \mid \mu) = f(\mu^{-1}x) \tag{4.3}$$

for x, $\mu \in SO(N)$. Let $\Pi(\cdot)$ be the prior density on $SO(N)$. Then the marginal density is

$$M(x) = \int_{SO(N)} \Pi(\mu) f(\mu^{-1} x)\, d\mu, \tag{4.4}$$

$x \in SO(N)$.

One can see that (4.4) is a convolution on $SO(N)$. Let us assume $f(\mu^{-1}x)$ is known; consequently, $\hat{f}(j)$ is known. The statistical analysis comes in with respect to prior uncertainty, that is, an unknown $\Pi(\cdot)$, which of course implies an unknown $M(\cdot)$ as defined in (4.4). From a Bayesian point of view, we can regard the observations $X_1, \dots, X_n$ as unconditionally coming from (4.4). This of course can then be used to construct an unbiased estimator of $\hat{M}(j)$. Indeed, define

$$\hat{M}^n(j) = \frac{1}{n} \sum_{l=1}^{n} U_j(X_l^{-1}).$$

Assuming that $\|[\hat{f}(j)]^{-1}\|_{\mathrm{op}} \ll d_j^u$ for some $u \geq 0$, a logical estimator for $\Pi(\cdot)$ would be

$$\Pi^n(g) = \sum_{j \in J_m} d_j \operatorname{tr}\left\{U_j(g) \hat{M}^n(j) \left[\hat{f}(j)\right]^{-1}\right\}, \tag{4.5}$$

where $g \in SO(N)$. Consistency results will follow by applying Theorems 3.1 or 3.2.

One can use this result for point estimation of μ. Suppose we want to make inference about μ based on the observation X. We note that in terms of squared error loss, if μ^* is an estimator of μ, then

$$L(\mu, \mu^*) = N - \operatorname{tr}\mu^{*-1}\mu$$

for μ, $\mu^* \in SO(N)$. Consequently, if $\Pi(\cdot)$ is the prior density, then the Bayes risk of μ^* is

$$r(\mu^*) = N - \int_{SO(N)\times SO(N)} \operatorname{tr}\{\mu^{*-1}\mu\} f(\mu^{-1}x) \Pi(\mu)\, dx\, d\mu.$$

Now in terms of the usual Fubini argument, we have

$$\begin{aligned}
&\int_{\mu \in SO(N)} \int_{x \in SO(N)} \operatorname{tr}\{\mu^{*-1}\mu\} f(x^{-1}\mu) \Pi(\mu)\, dx\, d\mu \\
&\quad = \int_{x \in SO(N)} \operatorname{tr}\left\{\int_{\mu \in SO(N)} \mu^{*-1}\mu \Pi(\mu \mid x) M(x)\, d\mu\right\} dx \\
&\quad = \int_{x \in SO(N)} \operatorname{tr} \mu^{*-1} \left\{\int_{\mu \in SO(N)} \mu \Pi(\mu \mid x)\, d\mu\right\} M(x)\, dx \\
&\quad = \int_{x \in SO(N)} \operatorname{tr} \mu^{*-1} \{E^{\Pi(\mu \mid x)} \mu\} M(x)\, dx,
\end{aligned}$$

where $\Pi(\mu \mid x)$ is the posterior density. Thus for each $x \in SO(N)$, the solution to

$$\max_{\mu^* \in SO(N)} \operatorname{tr} \mu^{*-1}\{E^{\Pi(\mu \mid x)}\mu\}$$

is the Bayes estimator. One can solve this problem by using a modified singular value decomposition similar to Chang (1986). Consider the modified singular value decomposition

$$E^{\Pi(\mu \mid x)}\mu = O\Gamma Q^t, \tag{4.6}$$

where $O, Q \in SO(N)$ and Γ is a diagonal matrix of singular values. Then the Bayes estimator is

$$\mu_b = OQ^t. \tag{4.7}$$

We are assuming that the prior density $\Pi(\cdot)$ is unknown; however, suppose we have observations $X_1, \dots, X_{n+1}$. Let $X = X_{n+1}$ and use $X_1, \dots, X_n$ to form a consistent estimator of $\Pi(\cdot)$ as in (4.5). An empirical Bayes estimator of μ can be formulated by

$$\mu_{eb} = O^n(Q^n)^t, \tag{4.8}$$

where $Q^n, O^n \in SO(N)$ are elements of the empirical singular value decomposition

$$E^{\Pi^n(\mu \mid x)}\mu = O^n\Gamma^n(Q^n)^t. \tag{4.9}$$

Under consistency of Π^n along with the continuous mapping theorem, we can show that $\mu_{eb} \to \mu_b$ as $n \to \infty$.

5. Discussion. An enormous amount of statistical literature is available on nonparametric density estimation in Euclidean space. The contributions are cited in several monographs; see, for example, Prakasa Rao (1983), Devroye and Györfi (1985) and Silverman (1985). For an important extension of the above to deconvolution density estimation, see, for example, Devroye (1989), Fan (1991a, b) and Diggle and Hall (1993) and the references therein.

Although theoretical work in non-Euclidean statistical methodologies is abundant [see, for example, Giné (1975), Jupp and Spurr (1983), Naiman (1990) and Prentice and Mardia (1995)], more recently, practical statistical methodology beyond the Euclidean space is gaining momentum. In part this is due to current computing capabilities in addition to statistical problems that are genuinely non-Euclidean. Several examples of such in addition to vector cardiogram orientation are plate tectonic issues studied by Chang (1986), statistical classification of macroscopic folds [Kelker and Langenberg (1988)] as well as problems in geometric quality assurance by Chapman, Chen and Kim (1995).

Therefore, in light of the general statistical interest in non-Euclidean spaces along with the popularity of nonparametric density estimation on Euclidean space, it is only natural to attempt the generalization of these methods to non-Euclidean spaces, which this paper explores. This generalization, aside from theoretical interests, can prove to be very valuable from a practical point of view, particularly with respect to vector cardiogram orientation where the practical benefits of mixture modelling can be extended.

Some works on nonparametric density estimation on non-Euclidean spaces are available, although the number is miniscule in comparison to the Euclidean counterpart; see Beran (1979), Hall, Watson, and Cabrera (1987), Bai, Rao and Zhao (1988) and Hendriks (1990). To date, contributions on non-Euclidean deconvolution density estimation are restricted to Healy, Hendriks and Kim (1995) and Healy and Kim (1996), as far as this author is aware, and each are special cases of the contents of this paper. Further, the methods of this paper should easily extend to all of the classical compact Lie groups.

Finally, some comments on computational considerations should be made. In Healy and Kim (1996), computational consideration using a fast Fourier transform now available on S^2, the unit 2-sphere is given explicit attention. The idea comes from applying the fast algorithm on S^2, as developed in Driscoll and Healy (1994), in a format similar to the idea of Silverman (1985) for the case of the circle S^1. We note that S^1 and S^2 are not only different in dimension, they are quite different topologically so the generalization is not necessarily straightforward. Now it is a mathematical fact that S^2 can be realized as a homogeneous space of $SO(3)$, consequently, the computational discussion in Healy and Kim (1996) can be carried over to $SO(3)$. In fact, a generalization of Driscoll and Healy (1994) has been made in a Harvard Ph.D. dissertation [Maslan (1993)] to compact groups of which $SO(N)$ is an example. Therefore, computational considerations for efficiently implementing the ideas of this paper can be formatted to $SO(N)$ according to Silverman (1985) and Healy and Kim (1996).

6. Proofs. We will work out the odd case, that is, $N = 2k + 1$. The even case can be worked out using similar arguments.

Some specific results will be needed with respect to d_j. Indeed, the latter is

$$\frac{2^k}{(2k-1)!\cdots 3!1!} \tag{6.1}$$
$$\times \begin{vmatrix} (j_1+k-1/2)^{2k-1} & (j_2+k-3/2)^{2k-1} & \cdots & (j_k+1/2)^{2k-1} \\ (j_1+k-1/2)^{2k-3} & (j_2+k-3/2)^{2k-3} & \cdots & (j_k+1/2)^{2k-3} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ j_1+k-1/2 & j_2+k-3/2 & \cdots & j_k+1/2 \end{vmatrix}$$

[see Gong (1991), page 123]. The evaluation of the above determinant can be expressed in simpler form due to the structure of the matrix in question and in fact is

$$\frac{2^k}{(2k-1)!\cdots 3!1!} \prod_{l=1}^{k} (j_{k-l+1} + l - 1/2) \prod_{r>s} \Big[(j_{k-r+1} + r - 1/2)^2 - (j_{k-s+1} + s - 1/2)^2 \Big],$$

where $j \in J$ and $l = 1, \dots, k$. The case $N = 2k$ is similar and can be found in Gong [(1991), page 123] and Rosenthal [(1994), page 406].

We will need the following lemma, where for two sequences $\{a_n\}$ and $\{b_n\}$, $a_n \sim b_n$ if $a_n / b_n \to 1$, as $n \to \infty$.

LEMMA 6.1. *There exists a $C > 0$ such that*

$$\sum_{j \in J_m} d_j^{2+2u} \sim C m^{(\dim SO(N) - k)u + \dim SO(N)}, \qquad \text{as } m \to \infty,$$

where $u \geq 0$.

PROOF. Define

$$a_l = j_{k-l+1} + l - 1/2,$$

where $l = 1, \dots, k$ and $j \in J_m$. Note that

$$(6.3) \quad 1/2 \leq a_1 < a_2 < \cdots < a_k \leq m + k - 1/2 \quad \text{and} \quad a_j + 1 \leq a_{j+1}$$

for $j = 1, \dots, k-1$. Letting $a = (a_1, \dots, a_k)$,

$$d_j = d_a = \prod_{l=1}^{k} a_l \prod_{r>s} \left[a_r^2 - a_s^2 \right].$$

Now divide (6.3) by m and consider

$$(6.4) \qquad \sum_j \left[\frac{d_j}{m^{k^2}} \right]^{2+2u} \frac{1}{m^k} = \sum_a \left[\frac{d_a}{m^{k^2}} \right]^{2+2u} \frac{1}{m^k}.$$

Notice that (6.4) is a Riemann sum; consequently, as $m \to \infty$, the domain becomes

$$0 \leq x_1 \leq x_2 \leq \cdots \leq x_k \leq 1,$$

and the right-hand side of (6.4) converges to

$$(6.5) \qquad \int_{0 \leq x_1 \leq \cdots \leq x_k \leq 1} \left[\prod_{l=1}^{k} x_l \prod_{r>s} \left\{ x_r^2 - x_s^2 \right\} \right]^{2+2u} dx_1 \cdots dx_k,$$

as $m \to \infty$. Let x^* be a vector such that $0 < x_1^* < \cdots < x_k^* \leq 1$. Then the integrand is strictly positive at x^*. By continuity, we can find an open neighborhood B containing x^* as a subset of $\{0 \leq x_1 \leq \cdots \leq x_k \leq 1\}$ for which the integrand remains strictly positive. Consequently, by the nonnegativity of the integrand of (6.5), the latter can be bounded below by

$$\int_B \left[\prod_{l=1}^{k} x_l \prod_{r>s} \left\{ x_r^2 - x_s^2 \right\} \right]^{2+2u} dx_1 \cdots dx_k > 0,$$

thus providing a lower positive bound for the limit of the sum in question. Some similarity of (6.4) to Selberg's integral is apparent. In fact, exact evaluation may be possible using the ideas surveyed in Richards (1989). □

LEMMA 6.2. *If $\|\hat{f}_{\varepsilon^{-1}}(j)\|_{\mathrm{op}} \ll d_j^u$ for some $u \geq 0$, then*

$$\int_{SO(N)} |K_n^{\varepsilon}(g)|^2 \, dg \ll m^{(\dim SO(N) - k)u + \dim SO(N)}, \qquad \text{as } m \to \infty.$$

PROOF. We have

$$\int_{SO(N)} K_n^{\varepsilon}(g)\overline{K}_n^{\varepsilon}(g)\,dg = \int_{SO(N)} \sum_{j \in J_m} d_j \operatorname{tr}\Big\{U_j \hat{f}_{\varepsilon^{-1}}(j)\Big\} \sum_{j \in J_m} d_j \operatorname{tr}\Big\{\overline{U_j \hat{f}_{\varepsilon^{-1}}(j)}\Big\}$$
$$= \sum_{j \in J_m} d_j \operatorname{tr}\Big\{|\hat{f}_{\varepsilon^{-1}}(j)|^2\Big\},$$

where the overbar denotes complex conjugation; see Lo and Ng (1988). Now by the assumption $\|\hat{f}_{\varepsilon^{-1}}(j)\|_{\mathrm{op}} \ll d_j^u$ for some $u \geq 0$, we note that

$$\sum_{j \in J_m} d_j \operatorname{tr}\Big\{|\hat{f}_{\varepsilon^{-1}}(j)|^2\Big\} \ll \sum_{j \in J_m} d_j^{2+2u}. \tag{6.6}$$

By Lemma 6.1 we have

$$\sum_{j \in J_m} \left[\frac{d_j}{m^{k^2}}\right]^{2+2u} \frac{1}{m^k} \to C, \qquad \text{as } m \to \infty,$$

where $C > 0$ is some constant. Consequently, we have

$$\sum_{j \in J_m} d_j \operatorname{tr}\Big\{|\hat{f}_{\varepsilon^{-1}}(j)|^2\Big\} \ll m^{(\dim SO(N) - k)u + \dim SO(N)}, \qquad \text{as } m \to \infty. \ \square$$

This leads to the following.

LEMMA 6.3. *If $\|\hat{f}_{\varepsilon^{-1}}(j)\|_{\mathrm{op}} \ll d_j^u$ for some $u \geq 0$ and f_Y is bounded, then*

$$\sup_{g \in SO(N)} \operatorname{Var}\big(f_X^n(g)\big) \ll \frac{m^{(\dim SO(N) - k)u + \dim SO(N)}}{n}, \qquad \textit{as } n \to \infty.$$

PROOF. We note that

$$\operatorname{Var}\big(f_\varepsilon^n(g)\big) = \frac{1}{n}\Big[EK_n^{\varepsilon}(X^{-1}g)\overline{K_n^{\varepsilon}(X^{-1}g)} - EK_n^{\varepsilon}(X^{-1}g)E\overline{K_n^{\varepsilon}(X^{-1}g)}\Big]$$
$$\ll \frac{1}{n}\int_G |K_n^{\varepsilon}(x^{-1}g)|^2 f_Y(x)\,dx$$
$$\leq \frac{1}{n} \sup_{g \in SO(N)} f_Y(g) \int_G |K_n^{\varepsilon}(x)|^2\,dx.$$

The result follows from applying Lemma 6.2. $\square$

PROOF OF THEOREM 3.1. Consider the variance bias decomposition

$$E|f_X^n(g) - f_X(g)|^2 = \operatorname{Var}\big(f_X^n(g)\big) + |Ef_X^n(g) - f_X(g)|^2 \tag{6.7}$$

for $g \in SO(N)$. We note that $\operatorname{Var}(f_X^n(g)) \to 0$ as $n \to \infty$. Consequently, we must show that the bias term goes to zero.

We have the following:

$$\begin{aligned}
Ef_X^n(g) &= EK_n^{\varepsilon}(Y^{-1}g) \\
&= \int_{SO(N)} \sum_{j \in J_m} d_j \operatorname{tr}\Big\{U_j(y^{-1}g)\hat{f}_{\varepsilon^{-1}}(j)\Big\} f_Y(y)\, dy \\
&= \int_{SO(N)} \sum_{j \in J_m} d_j \operatorname{tr}\Big\{U_j(y^{-1})U_j(g)\hat{f}_{\varepsilon^{-1}}(j)\Big\} f_Y(y)\, dy \\
&= \sum_{j \in J_m} d_j \operatorname{tr}\bigg\{\bigg[\int_{SO(N)} U_j(y^{-1}) f_Y(y)\, dy\bigg] U_j(g)\hat{f}_{\varepsilon^{-1}}(j)\bigg\} \\
&= \sum_{j \in J_m} d_j \operatorname{tr}\Big\{U_j(g)\hat{f}_{Y,j}\hat{f}_{\varepsilon^{-1}}(j)\Big\} \\
&= \sum_{j \in J_m} d_j \operatorname{tr}\Big\{U_j(g)\hat{f}_X(j)\Big\} \\
&\to f_X(g)
\end{aligned}$$

for all $g \in SO(N)$ since f_X is assumed to be the pointwise limit of its Fourier series. Consequently,

$$|Ef_X^n(g) - f_X(g)|^2 \to 0 \qquad \text{as } n \to \infty$$

for all $g \in SO(N)$ as required. □

Proof of Theorem 3.2. We can decompose

$$E\|f_X^n - f_X\|_2^2 = \int_{SO(N)} \operatorname{Var}\big(f_X^n(g)\big)\, dg + \|Ef_X^n - f_X\|^2.$$

By Lemma 6.3,

$$\int_{SO(N)} \operatorname{Var}\big(f_X^n(g)\big)\, dg \ll \frac{m^{(\dim SO(N) - k)u + \dim SO(N)}}{n}, \qquad \text{as } n \to \infty.$$

For the integrated bias, let

$$\begin{aligned}
\Lambda_m &= \Big\{0 \le j_1 \le j_2 \le \cdots \le j_k,\, \lambda_j \le m^2\Big\}, \\
\Lambda'_m &= \Big\{0 \le j_1 \le j_2 \le \cdots \le j_k,\, \lambda_j > m^2\Big\}, \\
J'_m &= \{0 \le j_1 \le j_2 \le \cdots \le j_k, j_k > m\}.
\end{aligned}$$

Clearly $\Lambda_m \subset J_m$ and $J'_m \subset \Lambda'_m$.

Now

$$\begin{aligned}\|Ef_X^n - f_X\|^2 &= \sum_{j\in J'_m} d_j \operatorname{tr}|\hat{f}_X(j)|^2\\ &\le \sum_{j\in \Lambda'_m} d_j \operatorname{tr}|\hat{f}_X(j)|^2\\ &\le \sum_{j\in \Lambda'_m} d_j\lambda_j^s m^{-2s} \operatorname{tr}|\hat{f}_X(j)|^2\\ &\le \|f^{(s)}\|_2^2 m^{-2s}.\end{aligned}$$

The first inequality comes from $J'_m \subset \Lambda'_m$ while the third inequality comes from

$$\int_{SO(N)} |f^{(s)}(g)|^2\,dg = \sum_{j\in J_m} \lambda_j^s d_j \operatorname{tr}|\hat{f}(j)|^2,$$

where $f^{(s)}$ denotes the sth derivative of f for $s \ge 1$; see Lemma 4.1 of Hendriks [(1990), page 842]. Of course the above is also true for $s = 0$, in which case it is the Plancherel Theorem for $SO(N)$; see Helgason (1978, 1984).

Putting the two together, we get that

$$E\|f_X^n - f_X\|_2^2 \ll \frac{m^{(\dim SO(N)-k)u+\dim SO(N)}}{n} + m^{-2s}, \qquad \text{as } n\to\infty.$$

Consequently, this rate is optimized when

$$m \propto n^{1/[2s+(\dim SO(N)-k)u+\dim SO(N)]}. \qquad \square$$

APPENDIX A

Compact Lie groups. A *Lie group* is a differentiable manifold whose group action and the map $g \to g^{-1}$ are continuous. Let G be a Lie group and V a complex vector space. A *representation* of the Lie group G on the vector space V is a continuous mapping

$$U\colon G \to \operatorname{Aut}(V),$$

so that $U(gh) = U(g)U(h)$ and $U(e) = \mathrm{id}_V$, where $\operatorname{Aut}(V)$ is the space of all invertible linear operators on V, e is the identity element in G and id_V is the identity operator on V. The vector space V is known as the *representation space*. If we fix a basis for V, then $\operatorname{Aut}(V) = \mathrm{GL}(n, \mathbf{C})$, the latter being the general linear group of invertible $n \times n$ complex matrices. Consequently, a *matrix representation* of G can be regarded as a group homomorphism $G \to GL(n, \mathbf{C})$.

Let U and W be two representations of G with representation spaces V_U and V_W. Suppose $f\colon V_U \to V_W$ is a linear map between the two representation spaces such that $f(U(g)v) = W(g)f(v)$ for all $g \in G$ and $v \in V_U$. Then f is called an *intertwining operator* and if for a given intertwining operator, a unique inverse intertwining operator exists, then we say that U and W are *equivalent* representations of G.

Let U be a representation of G with representation space V_U and suppose V' is a subspace of V_U such that $U(g)u \in V'$ for all $g \in G$ and for all $u \in V'$, that is, the subspace V' is an invariant subspace of V_U for all operators $U(g)$, $g \in G$. If the only invariant subspaces are $\{0\}$ and V_U, that is, V' is either the trivial subspace or the entire vector space, then the representation U is called an *irreducible representation*.

Denote by $\mathrm{Irr}(G, \mathbf{C})$, the collection of all inequivalent irreducible representations of G. For compact Lie groups, there are countably many. Furthermore, each representation space in $\mathrm{Irr}(G, \mathbf{C})$ is finite dimensional and each representation a *unitary* representation in the sense that there is an inner product $\langle \cdot, \cdot \rangle$ on V such that

$$\langle Uv, Uw \rangle = \langle v, w \rangle$$

for all $v, w \in V$, $U \in \mathrm{Irr}(G, \mathbf{C})$ and $g \in G$.

For a representation U, define a mapping $\chi: G \to \mathbf{C}$, called the *character* of U by

$$\chi(g) = \mathrm{tr}\, U(g),$$

for all $g \in G$. Note that although we need a matrix to define the character, the *trace* is independent of the basis so that $\chi(\cdot)$ is canonical, that is, basis free. Note also that $\chi(e) = \mathrm{tr}\, \mathrm{id}_{V_U} = \dim V_U$, where the latter denotes the dimension of the representation space V_U. Consequently, $d_U = \chi(e)$ is the dimension for $U \in \mathrm{Irr}(G, \mathbf{C})$.

Some basic examples of representations: the trivial representation is a map $G \to \mathbf{C} - \{0\}$ so that its dimension is 1. Consequently, if we reduce this representation to a unitary representation, $G \to \{1\}$. The standard representation is the matrix form of the group with the group action being matrix multiplication.

Given two representations U, W of a Lie group G, there are two ways we can form new representations. One construction is the direct sum $U \oplus W$ where $(U \oplus W)(gh) = U(gh) \oplus W(gh) = U(g)U(h) \oplus W(g)W(h)$ with $\chi_{U \oplus W}(g) = \chi_U(g) + \chi_W(g)$ for all $g, h \in G$. Thus we have that the $\dim V_{U \oplus W} = \dim V_U + \dim V_W$. A second construction is the direct product $U \otimes W$ where $(U \otimes W)(gh) = U(gh) \otimes W(gh) = U(g)U(h) \otimes W(g)W(h)$ with $\chi_{U \otimes W}(g) = \chi_U(g) \cdot \chi_W(g)$ for all $g, h \in G$. Thus we have that the $\dim V_{U \otimes W} = \dim V_U \dim V_W$.

As an example, we illustrate the situation for $SO(3)$. Let

$$u(\phi) = \begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad a(\theta) = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix},$$

where $\phi \in [0, 2\pi)$, $\theta \in [0, \pi)$. The well-known *Euler angle decomposition* implies that an arbitrary $g \in SO(3)$ can be uniquely written as

$$g = u(\phi)a(\theta)u(\psi),$$

where $\phi \in [0, 2\pi)$, $\theta \in [0, \pi)$ and $\psi \in [0, 2\pi)$ and are known as the *Euler angles*. Consider the function

$$\left[D_j(u(\phi)a(\theta)u(\psi))\right]_{q_1 q_2} = \exp(-iq_1\psi)\, d^j_{q_1 q_2}(\theta) \exp(-iq_2\phi), \quad \text{(A.1)}$$

where $-j \le q_1, q_2 \le j$, $j = 0, 1, \dots$ and $d^j(\cdot)$ are related to the Jacobi polynomials; see Vilenkin (1968). The function (A.1) can be thought of as matrix entries of the $(2j + 1) \times (2j + 1)$ matrix

$$D_j = \left[(D_j)_{q_1 q_2}\right] \quad \text{where } -j \le q_1, q_2 \le j$$

for $j = 0, 1, \dots$. These are the irreducible representations of $SO(3)$.

APPENDIX B

Eigenstructure of $SO(N)$. For $K, L \in so(n)$ consider the invariant inner product

$$\langle K, L \rangle = -\tfrac{1}{2} \operatorname{tr} KL.$$

Then we obtain a left-invariant Riemannian structure $g(\cdot,\cdot)$ on $SO(N)$ satisfying

$$g_e(K, L) = \langle K, L \rangle,$$

for $K, L \in so(n)$.

Now $\dim SO(N) = N(N-1)/2 = q$, hence there exists an orthonormal basis $K_1, \dots, K_q$ on $so(N)$ so that every K_l gives a left-invariant vector field $\tilde{K}_l$ on $SO(N)$ with

$$\left(\tilde{K}_l\right)_e = K_l$$

for $l = 1, \dots, q$.

The Laplace–Beltrami operator related to $g(\cdot,\cdot)$ on $SO(N)$ is

$$\Delta = \tilde{K}_1 + \cdots + \tilde{K}_q.$$

Denote the Cartan subalgebra of $so(N)$ by $\mathscr{H}$ which consists of all the following real matrices:

$$\begin{pmatrix} 0 & \theta_1 \\ -\theta_1 & 0 \end{pmatrix} \oplus \cdots \oplus \begin{pmatrix} 0 & \theta_k \\ -\theta_k & 0 \end{pmatrix}$$

for $N = 2k$,

$$\begin{pmatrix} 0 & \theta_1 \\ -\theta_1 & 0 \end{pmatrix} \oplus \cdots \oplus \begin{pmatrix} 0 & \theta_k \\ -\theta_k & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

for $N = 2k + 1$.

Let $H_l \in \mathscr{H}$ be the above matrix with $\theta_k = 0$ for $k \ne l$ and $\theta_l = 1$ for $l = 1, \dots, k$. Let ν_l be a real linear functional on $\mathscr{H}$ satisfying

$$\nu_l(H) = \langle H_l, H \rangle$$

for any $H \in \mathscr{H}$, $l = 1, \dots, k$. Then every dominant weight ν can be expressed as

$$\nu = j_1 \nu_1 + \cdots + j_k \nu_k,$$

where j_l are integers satisfying

$$j_1 \geq j_2 \geq \cdots \geq |j_k| \geq 0 \quad \text{for } N = 2k,$$
$$j_1 \geq j_2 \geq \cdots \geq j_k \geq 0 \quad \text{for } N = 2k+1.$$

Every such dominant weight determines uniquely an eigenvalue class of irreducible (unitary) representations.

Let U_ν be an irreducible representation of $SO(N)$ with dominant weight ν. Write

$$U_\nu(x) = \left[u_{\nu, ij}(x)\right],$$

$x \in SO(n)$, as a unitary matrix of order d_ν where

$$d_\nu = \prod_{\alpha > 0} \frac{\langle \nu + \delta, \alpha \rangle}{\langle \delta, \alpha \rangle},$$

α is a positive root and $\delta = \frac{1}{2} \sum_{\alpha > 0} \alpha$.

We have

$$\Delta u_{\nu, ij}(x) = -\lambda_\nu u_{\nu, ij}(x),$$

where

$$\lambda_\nu = j_1^2 + \cdots + j_k^2 + (2k-1) j_1 + (2k-3) j_2 + \cdots + j_k \quad \text{for } N = 2k+1$$

and

$$\lambda_\nu = j_1^2 + \cdots + j_k^2 + (2k-2) j_1 + (2k-4) j_2 + \cdots + 2 j_{k-1} \quad \text{for } N = 2k.$$

Acknowledgments. In dealing with a subject as technical as this, one cannot help but draw upon the assistance of various experts. I am greatly indebted to Professor Harrie Hendriks of the University of Nijmegen for his patience and explanations of technical details of spectral geometry. Professor Dennis Healy of Dartmouth College was also of much assistance in dealing with harmonic analysis. I also thank Professor Sheng Gong of the University of Science and Technology for his assistance with the eigenvalues of the Laplace–Beltrami operator on $SO(N)$. I benefitted from comments by an Associate Editor as well as a very insightful referee.

Parts of this paper were written while the author was visiting the Department of Applied Statistics, Yonsei University, Seoul, Korea, during the fall semester of 1997. I thank them for their hospitality. Special thanks go to Professor B. Sam Yoo.

REFERENCES

BAI, Z. D., RAO, C. R. and ZHAO, L. C. (1988). Kernel estimators of density function of directional data. *J. Multivariate Anal.* **27** 24–39.

BERAN, R. (1979). Exponential models for directional data. *Ann. Statist.* **7** 1162–1178.

BRÖCKER, T. and TOM DIECK, T. (1985). *Representations of Compact Lie Groups*. Springer, New York.

CHANG, T. (1986). Spherical regression. *Ann. Statist.* **14** 907–924.

CHANG, T. (1989). Spherical regression with errors in variables. *Ann. Statist.* **17** 293–306.

CHAPMAN, G. R., CHEN, G. and KIM, P. T. (1995). Assessing geometric integrity using spherical regression techniques (with discussion) *Statist. Sinica* **5** 173–220.

DEVROYE, L. (1989). Consistent deconvolution in density estimation. *Canad. J. Statist.* **17** 235–239.

DEVROYE, L. and GYÖRFI, L. (1985). *Nonparametric Density Estimation: The L_1 View*. Wiley, New York.

DIACONIS, P. (1988). *Group Representations in Probability and Statistics*. IMS, Hayward, CA.

DIGGLE, P. J. and HALL, P. (1993). A Fourier approach to nonparametric deconvolution of a density estimate. *J. Roy. Statist. Soc. Ser. B* **55** 523–531.

DOWNS, T. D. (1972). Orientation statistics. *Biometrika* **59** 665–676.

DRISCOLL, J. R. and HEALY, D. M., JR. (1994). Computing Fourier transforms and convolutions on the 2-sphere. *Adv. in Appl. Math.* **15** 202–250.

FAN, J. (1991a). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** 1257–1272.

FAN, J (1991b). Global behavior of deconvolution kernel estimates. *Statist. Sinica* **1** 541–551.

FAN, J. and TRUONG, Y. K. (1993). Nonparametric regression with errors in variables. *Ann. Statist.* **21** 1900–1925.

FARRELL, R. H. (1985). *Multivariate Calculation: Use of the Continuous Groups*. Springer, New York.

FEUERVERGER, A. and MURIEKA, R. A. (1977). The empirical characteristic function and its applications. *Ann. Statist.* **5** 88–97.

GINÉ, E. M. (1975). Invariant tests for uniformity on compact Riemannian manifolds based on Sobolev norms. *Ann. Statist.* **3** 1243–1266.

GONG, S. (1991). *Harmonic Analysis on Classical Groups*. Springer, Berlin.

HALL, P., WATSON, G. and CABRERA, J. (1987). Kernel density estimation with spherical data. *Biometrika* **74** 751–762.

HEALY, D. M., HENDRIKS, H. and KIM, P. T. (1995). Spherical deconvolution. Preprint.

HEALY, D. M. and KIM, P. T. (1996). An empirical Bayes approach to directional data and efficient computation on the sphere. *Ann. Statist.* **24** 232–254.

HELGASON, S. (1978). *Differential Geometry, Lie Groups and Symmetric Spaces*. Academic Press, New York.

HELGASON, S. (1984). *Groups and Geometric Analysis*. Academic Press, Orlando.

HENDRIKS, H. (1990). Nonparametric estimation of a probability density on a Riemannian manifold using Fourier expansions. *Ann. Statist.* **18** 832–849.

JUPP, P. E. and SPURR, B. D. (1983). Sobolev tests for symmetry of directional data. *Ann. Statist.* **11** 1225–1231.

KELKER, D. and LANGENBERG, C. W. (1988). Statistical classification of macroscopic folds as cylindrical, circular conical, or elliptical conical. *Math. Geol.* **20** 717–730.

KHATRI, C. G. and MARDIA, K. V. (1977). The von Mises–Fisher matrix distribution in orientation statistics. *J. Roy. Statist. Soc. Ser. B* **39** 95–106.

KIM, P. T. (1991). Decision theoretic analysis of spherical regression. *J. Multivariate Anal.* **38** 233–240.

LINDSAY, B. G. (1995). *Mixture Models: Theory, Geometry and Applications*. IMS, Hayward, CA.

LO, J. T. and NG, S. (1988). Characterizing Fourier series representations of probability distributions on compact Lie groups. *SIAM J. Appl. Math.* **48** 222–228.

MARITZ, J. S. and LWIN, T. (1989). *Empirical Bayes Methods*, 2nd ed. Chapman and Hall, London.

MASLEN, D. (1993). Fast transforms and sampling for compact groups. Ph.D. dissertation, Dept. Mathematics, Harvard Univ.

NAIMAN, D. Q. (1990). Volumes of tubular neighborhoods of spherical polyhedra and statistical inference. *Ann. Statist.* **18** 685–716.

PRAKASHA RAO, B. L. S. (1983). *Nonparametric Functional Estimation*. Academic Press, New York.
PRENTICE, M. J. (1986). Orientation statistics without parametric assumptions. *J. Roy. Statist. Soc. Ser. B* **48** 214–222.
PRENTICE, M. J. (1989). Spherical regression on matched pairs of orientation statistics. *J. Roy. Statist. Soc. Ser. B* **51** 241–248.
PRENTICE, M. J. and MARDIA, K. V. (1995). Shape changes in the plane for landmark data. *Ann. Statist.* **23** 1960–1974.
RICHARDS, D. ST. P. (1989). Analogs and extensions of Selberg's integral. In *q-Series and Combinatorics* (D. Stanton, ed.) 109–137. Springer, New York.
ROSENTHAL, J. S. (1994). Random rotations: characters and random walks on $SO(N)$. *Ann. Probab.* **22** 398–423.
SERRE, J. P. (1977). *Linear Representations of Finite Groups*. Springer, New York.
SILVERMAN, B. (1985). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.
SPIVAK, M. (1979). *A Comprehensive Introduction to Differential Geometry* **I**. Publish or Perish, Wilmington, DE.
TALMAN, J. D. (1968). *Special Functions*: *A Group Theoretic Approach*. Benjamin, New York.
VILENKIN, N. J. (1968). *Special Functions and the Theory of Group Representations*. Amer. Math. Soc., Providence, R.I.
WARNER, F. (1983). *Foundations of Differentiable Manifolds and Lie Groups*. Springer, New York.
ZHANG, C. H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* **18** 806–830.

DEPARTMENT OF MATHEMATICS
AND STATISTICS
UNIVERSITY OF GUELPH
GUELPH, ONTARIO N1G 2W1
CANADA
E-MAIL: pkim@ana.mathstat.uoguelph.ca