Submitted to the Annals of Statistics

MINIMAX ESTIMATION FOR MIXTURES OF WISHART DISTRIBUTIONS

By L.R. Haff, P.T. Kim*, J.–Y. Koo†, and Donald Richards‡

University of California at San Diego, University of Guelph, Korea University, and Penn State University

The space of positive definite symmetric matrices has been studied extensively as a means of understanding dependence in multivariate data along with the accompanying problems in statistical inference. Many books and papers have been written on this subject, and more recently there has been considerable interest in high-dimensional random matrices with particular emphasis on the distribution of certain eigenvalues. Our present paper is motivated by modern data acquisition technology, particularly, by the availability of diffusion tensor-magnetic resonance data; and, in that context, especially pertinent are the papers by Jian and Vemuri, et al. [15, 16]. With the availability of such data acquisition capabilities, smoothing or nonparametric techniques are required that go beyond those applicable only to data arising in Euclidean spaces. Accordingly, we present a Fourier method of minimax Wishart mixture density estimation on the space of positive definite symmetric matrices.

1. Introduction. The space of positive definite symmetric matrices has been studied extensively in statistics as a means of understanding dependence in multivariate data along with the accompanying problems in statistical inference. Many books and papers [8–10, 26, 28, 32, 33] have been written on this subject, and there has been considerable interest recently in high-dimensional random matrices with particular emphasis on the distribution of certain eigenvalues [17] and on graphical models [24].

While the primary emphasis in statistical research on the space of positive definite matrices arises from investigations of the Wishart distribution,

*Research supported in part by the Natural Sciences and Engineering Research Council of Canada, grant DG 46204.

†Research Support: This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2009-0075827).

‡Supported in part by National Science Foundation grant DMS-0705210.

AMS 2000 subject classifications: Primary 62G20; Secondary 65R32.

Keywords and phrases: Diffusion tensor imaging, deconvolution, Harish-Chandra c-function, Helgason-Fourier transform, Laplace-Beltrami operator, optimal rate, Pinsker constant, Sobolev ellipsoid.

modern data acquisition technologies, such as diffusion tensor-magnetic resonance imaging (DT-MRI), now produce real-world data on this space. The papers [1, 6, 7, 22, 37] analyze various aspects of the general nature of DT-MRI, and the papers by Jian and Vemuri [15] and Jian, *et al.* [16] make explicit applications of Wishart mixture models to DT-MRI.

With such data acquisition capabilities now available, smoothing or nonparametric techniques are in need beyond those generally applicable to data arising in Euclidean spaces or special manifolds [11, 12, 14, 18, 19]. Nonparametric estimation on the space of positive definite matrices was treated in the papers of Kim and Richards [20, 21]. Other than that, we are only aware of [29] and [30].

In this paper we consider the problem of estimating the mixing density of a continuous mixture of Wishart distributions. We construct a nonparametric deconvolution estimator of that density, and obtain minimax rates of convergence for the estimator. Throughout this work, we adopt as a guide results developed for the classical problem of deconvolution density estimation on Euclidean spaces [2, 3, 5, 23, 27, 36]. Much of the difficulty with the space of positive definite symmetric matrices is due to the fact that mathematical analysis on the space is technically demanding. Helgason [13] and Terras [35] provided much insight and technical innovation, however, and we make extensive use of their method of Fourier analysis on $\mathcal{P}_m$, the space of $m \times m$ positive definite symmetric matrices.

We summarize the paper as follows. In Section 2 we discuss and set up the notation for Wishart mixtures. In Section 3 we begin by reviewing the necessary Fourier methods which allow us to construct a nonparametric estimator of the mixing density, and then we provide the estimator. The minimax property of our nonparametric estimator is stated in Section 4 along with supporting results. These results are discussed further in Section 5, especially in connection with DT-MRI technology. Finally, Section 6 presents all of the proofs. Additional technical details are addressed in an appendix.

2. Wishart mixtures. Throughout the paper, for any square matrix y, we denote the trace and determinant of y by $\operatorname{tr} y$ and $|y|$, respectively; further, we denote by $\boldsymbol{I}_m$ the $m \times m$ identity matrix.

For $s = (s_1, \ldots, s_m) \in \mathbb{C}^m$ with $\operatorname{Re}(s_j + \cdots + s_m) > (j-1)/2$, $j = 1, \ldots, m$, the multivariate gamma function is defined as

$$(2.1) \qquad \Gamma_m(s_1, \ldots, s_m) = \pi^{m(m-1)/4} \prod_{j=1}^{m} \Gamma\big(s_j + \cdots + s_m - \tfrac{1}{2}(j-1)\big),$$

where $\Gamma(\cdot)$ denotes the classical gamma function.

We denote by G the general linear group $\mathrm{GL}(m, \mathbb{R})$ of all $m \times m$ nonsingular real matrices and by K the group $\mathrm{O}(m)$ of $m \times m$ orthogonal matrices. The group G acts transitively on $\mathcal{P}_m$ by the action

$$G \times \mathcal{P}_m \to \mathcal{P}_m, \quad (g, y) \mapsto g'yg, \tag{2.2}$$

$g \in G$, $y \in \mathcal{P}_m$, where g' denotes the transpose of g. Under this group action, the isotropy group of the identity in G is K, hence the homogeneous space $K\backslash G$ can be identified with $\mathcal{P}_m$ by the natural mapping from $K\backslash G \to \mathcal{P}_m$ that sends $Kg \mapsto g'g$. In distinguishing between left and right cosets, we place the quotient operation on the left and right of the group, respectively.

For $y = (y_{ij}) \in \mathcal{P}_m$, define the measure

$$\mathrm{d}_* y = |y|^{-(m+1)/2} \prod_{1 \le i \le j \le m} \mathrm{d}y_{ij}.$$

It is well-known that the measure $\mathrm{d}_* y$ is invariant under the action (2.2). Relative to the dominating measure $\mathrm{d}_* y$, the probability density function of the standard Wishart distribution with N degrees of freedom is

$$w(y) = \frac{1}{2^{Nm/2}\Gamma_m(0, \ldots, 0, N/2)} \, |y|^{N/2} \exp\big(- \tfrac{1}{2} \mathrm{tr}\, y\big), \tag{2.3}$$

$y \in \mathcal{P}_m$. Consequently, for $\sigma \in \mathcal{P}_m$, we note that $\mathrm{tr}(\sigma^{-1/2} y \sigma^{-1/2}) = \mathrm{tr}(\sigma^{-1} y)$ and $|\sigma^{-1/2} y \sigma^{-1/2}| = |\sigma^{-1} y|$. It then follows that, relative to the dominating measure $\mathrm{d}_* y$, the density of the general Wishart distribution, with covariance parameter σ, is $w(\sigma^{-1} y)$, $y \in \mathcal{P}_m$.

Suppose next that σ is a random matrix and, relative to the dominating measure $\mathrm{d}_* \sigma$, has a continuous mixing density, f, that is invariant under the action (2.2). By integration with respect to σ, the continuous Wishart mixture density is given by

$$r(y) = \int_{\mathcal{P}_m} f(\sigma) \, w(\sigma^{-1} y) \, \mathrm{d}_* \sigma, \tag{2.4}$$

$y \in \mathcal{P}_m$. For the case in which $m = 1$, the standard Wishart density is essentially a chi-square density, in which case (2.4) is a continuous mixture of chi-square densities. In general, (2.4) is a convolution operation for functions on $\mathcal{P}_m$; by applying Fourier analysis on $\mathcal{P}_m$, we will transform that convolution into a scalar multiplication.

3. Fourier analysis on $\mathcal{P}_m$ and estimation of the mixing density. In this section we review the Fourier methods needed to transform the convolution product (2.4) and to construct a nonparametric estimator of the mixing density f.

3.1. *The Helgason-Fourier transform.* For $y \in \mathcal{P}_m$, denote by $|y_j|$ the principal minor of order j, $j = 1, \ldots, m$. For $s \in \mathbb{C}^m$, the *power function* $p_s : \mathcal{P}_m \to \mathbb{C}$ is

$$p_s(y) = \prod_{j=1}^{m} |y_j|^{s_j}, \tag{3.1}$$

$y \in \mathcal{P}_m$. Let $\mathrm{d}k$ denote the Haar measure on K, normalized to have total volume equal to one; then

$$h_s(y) = \int_K p_s(k'yk)\, \mathrm{d}k, \tag{3.2}$$

$y \in \mathcal{P}_m$, is the *zonal spherical function* on $\mathcal{P}_m$. It is well-known that the functions h_s are fundamental to harmonic analysis on symmetric spaces [13, 35]. If $s_1, \ldots, s_m$ are nonnegative integers then, up to a constant factor, (3.2) is an integral formula for the zonal polynomials which arise in many aspects of multivariate statistical analysis [28, pp. 231–232].

Let $C_c^\infty(\mathcal{P}_m)$ denote the space of infinitely differentiable, compactly supported, complex-valued, functions f on $\mathcal{P}_m$; also, let

$$C_c^\infty(\mathcal{P}_m/K) = \{f \in C_c^\infty(\mathcal{P}_m) : f(k'yk) = f(y) \text{ for all } k \in K, y \in \mathcal{P}_m\}.$$

For $s \in \mathbb{C}^m$ and $k \in K$, the *Helgason-Fourier transform* [35, p. 87] of a function $f \in C_c^\infty(\mathcal{P}_m)$ is

$$\mathcal{H}f(s,k) = \int_{\mathcal{P}_m} f(y)\, \overline{p_s(k'yk)}\, \mathrm{d}_*y, \tag{3.3}$$

where $\overline{p_s(k'yk)}$ denotes complex conjugation.

For the case in which $f \in C_c^\infty(\mathcal{P}_m/K)$, we make the change of variables $y \mapsto k_1'yk_1$ in (3.3), $k_1 \in K$, and integrate with respect to the Haar measure $\mathrm{d}k_1$. Applying the invariance of f and the formula (3.2), we deduce that $\mathcal{H}f(s,k)$ does not depend on k. Specifically, $\mathcal{H}f(s,k) = \widehat{f}(s)$ where

$$\widehat{f}(s) = \int_{\mathcal{P}_m} f(y)\, \overline{h_s(y)}\, \mathrm{d}_*y, \tag{3.4}$$

$s \in \mathbb{C}^m$, is the *zonal spherical transform* of f.

In the case of the standard Wishart density (2.3), which is a K-invariant function, the zonal spherical transform is well-known (Muirhead [28, p. 248]; Terras [35, pp. 85-86]):

$$\widehat{w}(s) = \frac{\Gamma_m(s_{m-1}, \ldots, s_1, -(s_1 + \cdots + s_m) + N/2)}{\Gamma_m(0, \ldots, 0, N/2)}\, h_s(\tfrac{1}{2}\boldsymbol{I}_m).$$

3.2. *The inversion formula for the Helgason-Fourier transform.* For $a_1, a_2 \in \mathbb{C}$ with $\mathrm{Re}(a_1), \mathrm{Re}(a_2) > 0$, let

$$B(a_1, a_2) = \frac{\Gamma(a_1)\,\Gamma(a_2)}{\Gamma(a_1 + a_2)}$$

denote the classical beta function. For $s \in \mathbb{C}^m$ such that $\mathrm{Re}(s_i + \cdots + s_j) > -\frac{1}{2}(j - i + 1)$ for all $1 \le i < j \le m - 1$, the *Harish-Chandra c-function* is

$$c_m(s) = \prod_{1 \le i < j \le m-1} \frac{B\big(\frac{1}{2}, s_i + \cdots + s_j + \frac{1}{2}(j - i + 1)\big)}{B\big(\frac{1}{2}, \frac{1}{2}(j - i + 1)\big)} \ . \tag{3.5}$$

Let $\rho \equiv \Big(\frac{1}{2}, \ldots, \frac{1}{2}, \frac{1}{4}(1 - m)\Big)$ and set

$$\omega_m = \frac{\prod_{j=1}^{m} \Gamma(j/2)}{(2\pi i)^m \, \pi^{m(m+1)/4} \, m!} \ , \tag{3.6}$$

$$\mathbb{C}^m(\rho) = \{s \in \mathbb{C}^m : \mathrm{Re}(s) = -\rho\}, \tag{3.7}$$

and

$$\mathrm{d}_* s = \omega_m \, |c_m(s)|^{-2} \, \mathrm{d}s_1 \cdots \mathrm{d}s_m.$$

Let $M = \{\mathrm{diag}(\pm 1, \ldots, \pm 1)\}$ be the set of $m \times m$ diagonal matrices with entries ± 1 on the diagonal; then M is a subgroup of K and is of order 2^m. By factorizing the Haar measure $\mathrm{d}k$ on K, it may be shown [35, p. 88] that there exists an invariant measure $\mathrm{d}\bar{k}$ on the coset space K/M such that

$$\int_{\bar{k} \in K/M} \mathrm{d}\bar{k} = 1.$$

The *inversion formula for the Helgason-Fourier transform* $\mathcal{H}$ in (3.4) is that if $f \in C_c^\infty(\mathcal{P}_m)$ then [13, 35]

$$f(y) = \int_{\mathbb{C}^m(\rho)} \int_{\bar{k} \in K/M} \mathcal{H}f(s, k) \, p_s(k' y k) \, \mathrm{d}\bar{k} \, \mathrm{d}_* s. \tag{3.8}$$

$y \in \mathcal{P}_m$. In particular, if $f \in C_c^\infty(\mathcal{P}_m/K)$ then

$$f(y) = \int_{\mathbb{C}^m(\rho)} \widehat{f}(s) \, h_s(y) \, \mathrm{d}_* s,$$

$y \in \mathcal{P}_m$, and there also holds the *Plancherel formula*,

$$\int_{\mathcal{P}_m} |f(y)|^2 \, \mathrm{d}y = \int_{\mathbb{C}^m(\rho)} |\widehat{f}(s)|^2 \, \mathrm{d}_* s. \tag{3.9}$$

We refer to Terras [35, p. 87 ff.] for full details of the inversion formula and for references to the literature.

3.3. *Eigenvalues, the Laplacian, and Sobolev spaces.* For $y = (y_{ij}) \in \mathcal{P}_m$, we define the $m \times m$ matrix of partial derivatives,

$$\frac{\partial}{\partial y} = \Big(\tfrac{1}{2}\,(1+\delta_{ij})\,\frac{\partial}{\partial y_{ij}}\Big),$$

where δ_{ij} denotes Kronecker's delta. The Laplacian, Δ, on $\mathcal{P}_m$ can be written [35, p. 106] in terms of the local coordinates y_{ij} as

$$\Delta = -\mathrm{tr}\Big(\big(y\frac{\partial}{\partial y}\big)^2\Big).$$

The power function p_s in (3.1) is an eigenfunction of Δ (see [28, p. 229], [31, p. 283], [35, p. 49]). Indeed, let $r_j = s_j + s_{j+1} + \cdots + s_m + \frac{1}{4}(m - 2j + 1)$, $j = 1, \ldots, m$ and define

$$\lambda_s = -(r_1^2 + \cdots + r_m^2) + \tfrac{1}{48} m(m^2 - 1); \tag{3.10}$$

then $\Delta p_s(Y) = \lambda_s p_s(Y)$. Since $\mathrm{Re}(s) = -\rho$ then each r_j, $j = 1, \ldots, m$ is purely imaginary; hence, $\lambda_s > 0$, $s \in C_m(\rho)$.

The operator $\mathcal{H}$ changes the effect of invariant differential operators on functions to pointwise multiplication: If $f \in C_c^\infty(\mathcal{P}_m)$ then

$$\mathcal{H}(\Delta f)(s,k) = \lambda_s \mathcal{H} f(s,k),$$

$s \in \mathbb{C}^m$, $k \in K$ [35, p. 88]. For $\varphi > 0$, we therefore define the fractional power, $\Delta^{\varphi/2}$, of Δ, as the operator such that

$$\mathcal{H}(\Delta^{\varphi/2} f)(s,k) = \lambda_s^{\varphi/2} \mathcal{H} f(s,k),$$

$f \in C_c^\infty(\mathcal{P}_m)$. Having constructed $\Delta^{\varphi/2}$, we define the *Sobolev class*,

$$\mathcal{F}_\varphi = \{f \in C^\infty(\mathcal{P}_m) : \|\Delta^{\varphi/2} f\|^2 < \infty\},$$

where $2\varphi > \dim \mathcal{P}_m = m(m+1)/2$ and, for $f \in C^\infty(\mathcal{P}_m)$,

$$\|f\| = \Big(\int_{\mathcal{P}_m} |f(y)|^2 \,\mathrm{d}_* y\Big)^{1/2}$$

denotes the $L^2(\mathcal{P}_m)$-norm with respect to the measure $\mathrm{d}_* y$. For $Q > 0$, we also define the *bounded Sobolev class*,

$$\mathcal{F}_\varphi(Q) = \{f \in C^\infty(\mathcal{P}_m) : \|\Delta^{\varphi/2} f\|^2 < Q\},$$

where $2\varphi > \dim \mathcal{P}_m$.

4. Main result. We presently state a minimax estimator for the Wishart mixing density (2.4). First, the result depends on the following lower bound:

THEOREM 4.1. *Let $N > (m-1)/2$ and $\varphi > \dim \mathcal{P}_m/2$. Then, for some $C > 0$ and the Wishart mixture (2.4),*

$$\inf_{\tilde{f}} \sup_{f \in \mathcal{F}_\varphi(Q)} \mathbb{E}\|\tilde{f} - f\|^2 \geq C\ (\log n)^{-2\varphi} \tag{4.1}$$

as $n \to \infty$, where the infimum is taken over all estimators $\tilde{f}$ of f.

This result provides a lower bound for any estimator, and so naturally, one wishes to derive an explicit estimator that achieves this bound. We do so by applying the Helgason-Fourier transform to the mixture density (2.4), so that

$$\mathcal{H}r(s,k) = \mathcal{H}f(s,k)\widehat{w}(s), \tag{4.2}$$

$s \in \mathbb{C}^m$, $k \in K$. Having observed a random sample $Y_1, \ldots, Y_n$ from the density (2.4), we estimate $\mathcal{H}r(s,k)$ by its *empirical Helgason-Fourier transform*,

$$\mathcal{H}_n r(s,k) = \frac{1}{n}\sum_{\ell=1}^{n} \overline{p_s(k'Y_\ell k)}. \tag{4.3}$$

On substituting (4.3) in (4.2), together with the assumption that $\widehat{w}(s) \neq 0$, $s \in \mathbb{C}^m$, we obtain

$$\mathcal{H}_n f(s,k) = \frac{\mathcal{H}_n r(s,k)}{\widehat{w}(s)},$$

$s \in \mathbb{C}^m$, $k \in K$.

Analogous with classical Euclidean deconvolution, we introduce a smoothing parameter $T = T(n)$ where $T(n) \to \infty$ as $n \to \infty$, and then we apply the inversion formula (3.8) using a spectral cut-off based on the eigenvalues of Δ. First, we introduce the notation

$$\mathbb{C}^m(\rho, T) = \{s \in \mathbb{C}^m(\rho) : \lambda_s < T\}$$

where $\mathbb{C}^m(\rho)$ is defined in (3.7). We now define

$$f_n(y) = \int_{\mathbb{C}^m(\rho,T)} \int_{\bar{k} \in K/M} \frac{\mathcal{H}_n r(s,\bar{k})}{\widehat{w}(s)}\, p_s(\bar{k}' y \bar{k})\, \mathrm{d}\bar{k}\, \mathrm{d}_* s, \tag{4.4}$$

$y \in \mathcal{P}_m$, and take this as our nonparametric estimator of f.

For technical reasons, explained in Section 6, we assume that the density function r satisfies the following moment condition on the principal minors

$$\int_{\mathcal{P}_m} |y_1|^{-1} \cdots |y_{m-1}|^{-1} \, |y_m|^{(m-1)/2} \, r(y) \, \mathrm{d}_* y < \infty. \tag{4.5}$$

We now state the minimax property for the estimator (4.4), where for two sequences of real numbers $\{a_n\}$ and $\{b_n\}$ we use the notation $a_n \asymp b_n$ to mean that there exist constants $0 < c \le C < \infty$ so that $cb_n \le a_n \le Cb_n$ as $n \to \infty$.

THEOREM 4.2. *Let $N > (m-1)/2$, $\varphi > \dim \mathcal{P}_m/2$ and suppose (4.5) is true. Then, for the Wishart mixture (2.4),*

$$\sup_{f \in \mathcal{F}_\varphi(Q)} \mathbb{E}\|f_n - f\|^2 \asymp (\log n)^{-2\varphi} \tag{4.6}$$

as $n \to \infty$.

5. Discussion. In this section let us briefly state the meaning of the above minimax result in the context of covariance estimation as well as provide a synopsis of recent interest in this area in the medical imaging literature.

5.1. *Covariance matrix estimation.* The profound influence of Charles Stein on covariance estimation originates largely from his Rietz lecture - see [32]. The idea is that for certain loss functions over $\mathcal{P}_m$, the usual estimator of the covariance matrix parameter is inadmissible. Through an unbiased estimation of the risk function over covariance matrices, Stein was able to improve upon the usual estimator by pooling the observed eigenvalues of the sample covariance matrix. Subsequent to this, through a series of papers, improvements were obtained in Haff [8–10]. Other related works include Takemura [34], Lin and Perlman [25], and Loh [26], to name a few.

In this paper, we contribute to the case in which one observes data from a continuous mixture of Wishart distributions, not merely a sample from a single distribution. Therefore, the parameter of interest would be the mixing density of the covariance parameters. And the nonparametric estimator of the mixing density (4.4) is an attractive candidate because of its minimax property. Based on this procedure, one could consider the moment, or mode, of f_n, as a possible estimator of the corresponding population parameters. Alternatively, one could take a nonparametric empirical Bayes approach as in Pensky [29]

5.2. *Application to brain imaging.* In addition to statistical interest in high-dimensional random matrices [17], there has been much interest recently in the medical imaging literature on DT-MRI; an extensive survey of this area is provided by Alexander [1]. Now that the notation is in place, we can provide a motivating practical application. In brain imaging research associated with DT-MRI, each image is represented as a positive definite symmetric matrix, or, diffusion tensor (DT). From a given sample of images, each as positive definite symmetric matrices, the problem is to reconstruct the overall brain shape.

When a DT-MRI image of the brain is collected, the recorded image generally is an imperfect representation of the physical object under study, i.e., brain tissue. Denoting by Y the recorded image and by X the true tissue shape, then Y is a distorted measurement on X. We refer to Fletcher and Joshi [6] for an elementary account of the relationship between covariance matrices and brain images, and to [7] for an account as to the natural relationship of the problem as the statistical analysis of observations on a Riemannian symmetric space.

As stated before, both Y and X are represented as positive definite matrices. Since $\mathcal{P}_m$ is a Riemannian manifold then for each pair $X, Y \in \mathcal{P}_m$, there is a unique geodesic joining X and Y. Since $\mathcal{P}_m$ is also a homogeneous space, *viz.*, $\mathcal{P}_m = \mathrm{GL}(m, \mathbb{R})/\mathrm{O}(m)$, there exists a unique matrix $V \in \mathrm{GL}(m, \mathbb{R})$ such that $Y = VXV'$.

We can also think of this in terms of the log-map on $\mathcal{P}_m$. Whenever we observe Y, we can think of this as observing y where $\exp(y) = Y$. Similarly, we can represent the true X by $\exp(x) = X$; thus, y and x are in the tangent space of $\mathcal{P}_m$. Then there exists a "small" v in the tangent space such that $y = x + v$, and by exponentiating this relationship between y and x, we are led to $Y = V'XV$ see Terras [35], section 4.1-4.2 for technical details.

As with most spatial measurement errors, distortions are often isotropic which in matrix language would mean V has a $K-$invariant distribution under K, i.e., measurement errors have no preferred orientation, or, equivalently, v has a distribution which depends only on its (Euclidean) norm.

The conclusion is the model $Y = V'XV$, where the errors V are K-invariant, is the appropriate one for modeling measurement errors on $\mathcal{P}_m$ such as the one we have for brain imaging. To further motivate the discussion, since V is invariant then we can replace V by $Z^{1/2}$ where $Z \in \mathcal{P}_m$ by using the fact that every nonsingular V has the polar coordinates representation, $V = kZ^{1/2}$ where $k \in K$. So V has the same distribution as $k'V = k'kZ^{1/2} = Z^{1/2}$; hence Y has the same distribution as $Z^{1/2}XZ^{1/2}$ which is (2.4) in the case where Z follows a Wishart distribution. Further, this Z is invariant under K; in fact, $Z = V'V$, which immediately proves the K invariance. The object now is to nonparametrically estimate the density of X based on observations Y.

The papers [15]and [16], depict the above situation in terms of a finite mixture of Wishart distributions. Thus the contents of this paper generalizes this finite mixture situation to the case of continuous mixtures of Wishart distributions.

6. The proof of Theorem 4.1. The upper bound property of (4.6) is established in [21, Theorem 3.3] with $\beta = 1/2$, and that result is the more straightforward calculation of the bounds provided by (4.6). We note that the technical condition (4.5) is required for the upper bound only. All that remains to be proved is that the upper and lower bounds agree.

To derive the lower bound for estimating f in the $L^2(\mathcal{P}_m)$-norm, we shall follow the standard Euclidean approach. Thus, we choose a pair of functions $f^0, f^n \in \mathcal{F}_\varphi(Q)$ and, with w denoting the Wishart density (2.3), we shall show that, for some constants $C_1, C_2 > 0$,

$$\|f^n - f^0\|^2 \geq C_1 \ (\log n)^{-2\varphi}$$

and

$$\int_{\mathcal{P}_m} \frac{\left(f^0 * w(y) - f^n * w(y)\right)^2}{f^0 * w(y)} \, \mathrm{d}_* y \leq \frac{C_2}{n}, \tag{6.1}$$

as $n \to \infty$.

Precisely, let us suppose we can choose $f^0 \in \mathcal{F}_\varphi(Q)$ and a perturbation $\psi \in \mathcal{F}_\varphi(Q)$, and, for $\delta = \delta_n > 0$, let ψ^δ be a scaling of ψ such that $\|\psi^\delta\| \asymp \delta^{-1/2}\|\psi\|$ as $n \to \infty$. Define

$$f^n = f^0 + C_\psi \delta^{-\varphi+1/2} \psi^\delta,$$

and denote (6.1) by $\chi^2(f^n * w, f^0 * w)$. If δ can be chosen so that

$$\chi^2(f^n * w, f^0 * w) \leq C n^{-1},$$

as $n \to \infty$, then the lower bound rate of convergence is determined by $\delta^{-2\varphi}$. We shall develop such a construction and, moreover, do so in a way such that $\delta \asymp \log n$ as $n \to \infty$.

In the following, we let $\mathcal{SP}_m = \{y \in \mathcal{P}_m : |y| = 1\}$ denote the space of positive definite matrices of determinant one. For $y \in \mathcal{P}_m$, let $y = e^{v/m} z$ where $v = \log|y| \in \mathbb{R}$ and $z \in \mathcal{SP}_m$; then, by [35, p. 23], $\mathrm{d}_* y = \mathrm{d}v \, \mathrm{d}_* z$ and, up to a constant factor, $\mathrm{d}_* z$ is the unique K-invariant measure on $\mathcal{SP}_m$.

6.1. *The function ψ.* For a function $\psi : \mathcal{P}_m \to \mathbb{R}$ and $\delta > 0$, define

$$\psi^\delta(y) = \psi(|y|^{(\delta-1)/m} y), \tag{6.2}$$

$y \in \mathcal{P}_m$. Since $y = |y|^{1/m}(|y|^{-1/m} y)$ and $|y|^{-1/m} y \in \mathcal{SP}_m$, it follows that (6.2) is equivalent to

$$\psi^\delta(e^{v/m} z) = \psi(e^{\delta v/m} z),$$

$v \in \mathbb{R}$, $z \in \mathcal{SP}_m$. If ψ is K-invariant, then ψ^δ also is K-invariant. From now on, we suppose that ψ is K-invariant.

It follows from (A.1), see Appendix, that, for $s \in \mathbb{C}^m$,

$$\begin{aligned} \widehat{\psi^\delta}(s) &= \int_{v\in\mathbb{R}} \exp(\bar{\tilde{s}} v) \int_{\mathcal{SP}_m} \psi^\delta(e^{v/m} z) \overline{h_s(z)}\, \mathrm{d}_* z\, \mathrm{d}v \\ &= \int_{v\in\mathbb{R}} \exp(\bar{\tilde{s}} v) \int_{\mathcal{SP}_m} \psi(e^{\delta v/m} z) \overline{h_s(z)}\, \mathrm{d}_* z\, \mathrm{d}v \\ &= \delta^{-1} \int_{u\in\mathbb{R}} \exp(\delta^{-1}\bar{\tilde{s}} u) \int_{\mathcal{SP}_m} \psi(e^{u/m} z) \overline{h_s(z)}\, \mathrm{d}_* z\, \mathrm{d}u. \end{aligned} \tag{6.3}$$

Define $q = (q_1, \ldots, q_m) \in \mathbb{C}^m$ by $q_j = s_j$, $1 \le j \le m-1$, and $q_m = \tilde{s}$, so that

$$\widehat{\psi}(q) = \int_{v\in\mathbb{R}} \exp(\overline{q_m} v) \int_{\mathcal{SP}_m} \psi(e^{v/m} z) \overline{h_q(z)}\, \mathrm{d}_* z\, \mathrm{d}v.$$

Since $h_s(z) = h_q(z)$ for $z \in \mathcal{SP}_m$ then $\widehat{\psi}(q) = \widehat{\psi}(s)$. From (6.3), we have

$$\widehat{\psi^\delta}(q) = \delta^{-1} \widehat{\psi}(q_\delta), \tag{6.4}$$

where $q_\delta = (q_1, \ldots, q_{m-1}, q_m/\delta)$ and $q^\delta = (q_1, \ldots, q_{m-1}, \delta q_m)$. If $s \in \mathbb{C}^m(\rho)$ then

$$\mathrm{Re}(q_m) = \mathrm{Re}\Big(m^{-1} \sum_{j=1}^m j s_j\Big) = 0,$$

which implies that $q \in \mathbb{C}^m(\rho_0) \equiv \{a \in \mathbb{C}^m : \mathrm{Re}(a) = -\rho_0\}$, where $\rho_0 = (1/2, \ldots, 1/2, 0)$.

Let $\mathrm{d}_* q = \omega_m c_m(q) \mathrm{d}q$ where ω_m and $c_m(q)$ are defined in (3.6) and (3.5), respectively. Note that $\mathrm{d}_* q = \mathrm{d}_* s$ and $\mathrm{d}_* q = \delta \mathrm{d}_* q_\delta$. Because the Jacobian of the change of variables from s to q is one, it follows from the Plancherel formula (3.9) for the Helgason-Fourier transform that

$$\|\psi^\delta\|^2 = \int_{\mathcal{P}_m} |\psi^\delta(y)|^2\, \mathrm{d}_* y = \int_{\mathbb{C}^m(\rho)} |\widehat{\psi^\delta}(s)|^2\, \mathrm{d}_* s.$$

Consequently, by (6.4), we have

$$\|\psi^\delta\|^2 = \delta^{-1} \|\psi\|^2. \tag{6.5}$$

6.2. *The convolution $\psi * w$.* For $s \in \mathbb{C}^m$, let $a_j = \mathrm{Re}(s_j)$ and $b_j = \mathrm{Im}(s_j)$, $1 \le j \le m$. Consider the rectangle $\mathcal{S} = \{s \in \mathbb{C}^m : 1 \le b_j \le 2, 1 \le j \le m\}$. Recalling the definition of q, *infra* Equation (6.3), it follows that if $s \in \mathcal{S}$ then $q \in \mathcal{Q}$, where

$$\mathcal{Q} = \Big\{ q \in \mathbb{C}^m : 1 \le \mathrm{Im}(q_j) \le 2, 1 \le j \le m-1, \tfrac{1}{2}(m+1) \le \mathrm{Im}(q_m) \le m+1 \Big\}.$$

For $j = 1, \dots, m$, define

$$r_j(q) = q_j + q_{j+1} + \cdots + q_{m-1} + q_m - m^{-1} \sum_{l=1}^{m-1} l q_l + \tfrac{1}{4}(m - 2j + 1).$$

If $q \in \mathbb{C}^m(\rho_0) \cap \mathcal{Q}$, then

$$\begin{aligned} -[r_j(q^\delta)]^2 &= \Big[b_j + \cdots + b_{m-1} + \Big(\delta m^{-1} \sum_{l=1}^{m} l b_l - m^{-1} \sum_{l=1}^{m-1} l b_l \Big) \Big]^2 \\ &\le -C\delta^2 r_j(q), \end{aligned}$$

and thus

$$\lambda_{q^\delta} \le C\delta^2 \lambda_q = C\delta^2 \lambda_s.$$

Suppose $\widehat{\psi}$ is supported on $\mathcal{Q}$. Then, by (3.9) and (6.3)–(6.5),

$$\|\Delta^{\varphi/2} \psi^\delta\|^2 \le \delta^{2\varphi - 1} \|\Delta^{\varphi/2} \psi\|^2.$$

Hence, $\delta^{-\varphi + 1/2} \psi^\delta \in \mathcal{F}_\varphi(Q)$ when $\psi \in \mathcal{F}_\varphi(Q)$.

For $q \in \mathbb{C}^m(\rho_0) \cap \mathcal{Q}$, we have

$$|\widehat{w}(q)|^2 \le C \exp\Big(-\pi \sum_{j=1}^{m} j b_j \Big) = C \exp\big(-\pi m q_m \big),$$

hence

$$|\widehat{w}(q^\delta)|^2 \le C \exp\big(-\pi m q_m \delta \big) \le C \exp\big(-\tfrac{1}{2}\pi m (m+1)\delta \big).$$

6.3. *The function f^0.* At this stage, we need to provide some detailed calculations for $m = 2$. These calculations being quite extensive, some of the necessary technical details are provided in the appendix. We also remark that generalizations of these calculations to $m > 2$ may be obtained by using higher order hyperbolic spherical coordinates.

Let p_b be a density on $\mathbb{R}$ such that p_b is sufficiently smooth and satisfies

$$p_b(u) = c_b \exp\left(-b|u|/2 \right),$$

$|u| \geq c_0$, where c_b is a normalizing constant. Consider a natural spline, $\texttt{ns}$, defined by

$$\texttt{ns}(u) = u + \sum_{j=1}^{4} c_j (u - t_j)_+^3,$$

where $(u)_+ = \max(u, 0)$, and $c_1 = c_4 = -1/3$, $c_2 = c_3 = 1/3$, $t_1 = -2$, $t_2 = -1$, $t_3 = 1$ and $t_4 = 2$. Then, $\texttt{ns}(u) = u$ for $u < -2$ and, for $u > 2$,

$$\texttt{ns}(u) = u + \sum_{j=1}^{4} c_j \Big(u^3 - 3t_j u^2 + 3t_j^2 u - 3t_j^3 \Big) = -u.$$

Define the twice continuously differentiable density,

$$p_b(u) = c_b \exp\left(-b\, \texttt{ns}(u)/2\right),$$

$u \in \mathbb{R}$. Then $p_b(u) = c_b \exp(-b|u|/2)$ for $|u| \geq c_0 = 2$. We have the following straightforward result.

LEMMA 6.1. *Define the function $f^0(e^{u/2} D_\alpha) = C_b p_b(u) (\cosh \alpha)^{-b}$, where $C_b = c_b(b-1)/(2\pi)$ for $b > 1$. Then, $g^0 = f^0 * w$ is K-invariant, and*

$$g^0(e^{v/2} D_\beta) \geq C \exp\big(-2b(|v|/2 + \beta)\big)$$

as $|v| \to \infty$ and $\beta \to \infty$.

Let $Y = e^{U/2} Z$ be a random matrix with density $w(e^{u/2} D_\alpha)$, relative to the measure $\mathrm{d}(\alpha, u)$, and distribution function W. Then, $e^{\delta U/2} Z$ has density $\delta^{-1} w(e^{u/(2\delta)} D_\alpha)$, and we denote its probability distribution by W_δ.

LEMMA 6.2. *Let $\mu(u) = e^{-m_2|u|}$, $u \in \mathbb{R}$; $\nu(t) = t^{-m_1}$, $t \in \mathbb{R}$; and define*

$$\eta_\delta(v, \beta) = \int_{\mathcal{D}} \mu(v - u) \nu\Big(\cosh(\alpha - \beta)\Big)\, \mathrm{d}W_\delta(\alpha, u).$$

Then, there exist M and a constant C such that when $\beta \geq M$ and $|v|/\delta \geq M$,

$$\eta_\delta(v, \beta) \leq C \exp\Big\{ -\tfrac{1}{2}(N-1)(1-\xi)(|v|/(2\delta) + \beta)\Big\}$$

for all $\delta \geq 1$ provided that $0 < \xi < 1$, $m_1 \xi > (N-1)(1-\xi)/2$, and $m_2 \xi > (N-1)(1-\xi)/4$.

PROOF. Denote

$$\begin{aligned} J_{11} &= \{(\alpha, u) : |\alpha - \beta| \le \xi\beta, |v - \delta u| \le \xi|v|\} \\ J_{12} &= \{(\alpha, u) : |\alpha - \beta| \le \xi\beta, |v - \delta u| > \xi|v|\} \\ J_{21} &= \{(\alpha, u) : |\alpha - \beta| > \xi\beta, |v - \delta u| \le \xi|v|\} \\ J_{22} &= \{(\alpha, u) : |\alpha - \beta| > \xi\beta, |v - \delta u| > \xi|v|\} \end{aligned}$$

and

$$I_{ij} = \int_{J_{ij}} \mu(v-u)\nu\big(\cosh(\alpha-\beta)\big)\, \mathrm{d}W_\delta(\alpha, u),$$

$i, j = 1, 2$. Using Lemma A.2, one can show that

$$I_{ij} \le C \exp\Big\{ -\tfrac{1}{2}(N-1)(1-\xi)(|v|/(2\delta) + \beta)\Big\}$$

which proves the desired result. □

LEMMA 6.3. *Suppose that* $\big|\psi(e^{v/2}D_\beta)\big| \le \mu(v)\nu(\cosh\beta)$ *for* $(\beta, v) \in \mathcal{D}$. *Then,*

$$\int_{\mathcal{D}} \Big[(\psi^\delta * w)(e^{v/2}D_\beta)\Big]^2 [g^0(e^{v/2}D_\beta)]^{-1}\, \mathrm{d}(\beta, v) \\ \le \frac{1}{\delta} \int_{\mathcal{D}} \eta_\delta(v, \alpha)^2 \big[g^0\big(e^{v/(2\delta)}D_\beta\big)\big]^{-1}\, \mathrm{d}(\beta, v).$$

PROOF. The desired result follows from Lemma A.3, Lemma A.4 and Lemma A.5, see the Appendix. □

6.4. χ^2-*divergence.* In this subsection, for certain densities f and g we obtain bounds on the chi-square divergence,

$$\chi^2(f, g) = \int \frac{(f(y) - g(y))^2}{g(y)}\, \mathrm{d}_* y.$$

LEMMA 6.4. *Suppose that* $\big|\psi(e^{v/2}D_\beta)\big| \le \mu(v)\nu(\cosh\beta)$ *for* $(\beta, v) \in \mathcal{D}$. *For a pair of densities* f^0 *and* $f^n = f^0 + C_\psi \delta^{-\varphi+1/2}\psi^\delta$, *the* χ^2*-divergence between* $g^0 = f^0 * w$ *and* $g^n = f^n * w$ *satisfies*

$$\chi^2(g^0, g^n) = \delta^{-2\varphi+1} \int_{\mathcal{D}} [\psi^\delta * w]^2 (g^0)^{-1} \le \frac{C}{n}$$

provided that $b < \frac{1}{4}\min\big(3\pi, (N-1)(1-\xi) - 1\big)$.

PROOF. Let

$$\begin{aligned} J_{11} &= \{(\beta, v) : |\beta| \le M_n, |v|/(2\delta) \le M_n\}, \\ J_{12} &= \{(\beta, v) : |\beta| \le M_n, |v|/(2\delta) > M_n\}, \\ J_{21} &= \{(\beta, v) : |\beta| > M_n, |v|/(2\delta) \le M_n\}, \\ J_{22} &= \{(\beta, v) : |\beta| > M_n, |v|/(2\delta) > M_n\}, \end{aligned}$$

and

$$\begin{aligned} I_{ij} &= \int_{J_{ij}} \Big[\int_{\mathcal{D}} \Big\{ \frac{1}{2\pi} \int_K \psi(e^{(v-u)/2} D_{R^*(\alpha,\beta,\theta)})\, \mathrm{d}\theta \Big\} \, \mathrm{d}W_\delta(\alpha, u) \Big]^2 \\ &\qquad \cdot [g^0(e^{v/(2\delta)} D_\beta)]^{-1} \, \mathrm{d}(\beta, v), \end{aligned}$$

for $i, j = 1, 2$.

From Lemma 6.1, Lemma 6.2 and Lemma 6.3, we can show that

$$\begin{aligned} I_{11} &\le C\delta^2 e^{4bM_n - 3\pi\delta} \\ I_{12} &\le C\delta e^{-(r-2b)M_n} \\ I_{21} &\le C e^{-(c_1-2b-1)M_n} \\ I_{22} &\le C\delta e^{-(2c_1-1)M_n}, \end{aligned}$$

where $c_1 = (N-1)(1-\xi) - 2b$. We now choose $M_n = \delta$. Also, on letting $\epsilon_1 = 3\pi - 4b > 0$ and $\epsilon_2 = \min\big(\epsilon_1, (N-1)(1-\xi) - 4b - 1\big) > 0$ then we obtain

$$\begin{aligned} \chi^2(g^0, g^n) &= \int_{\mathcal{D}} \Big[C_\psi \delta^{-\varphi+1/2} \psi^\delta * w \Big]^2 (g^0)^{-1} \\ &= C_\psi^2 \delta^{-2\varphi+1} \int_{\mathcal{D}} \big[\psi^\delta * w \big]^2 (g^0)^{-1} \\ &\le C\delta^{-2\varphi} \big(\delta^2 e^{4bM_n - 3\pi\delta} + \delta e^{-(c_1-2b)M_n} + e^{-(c_1-2b-1)M_n} + \delta e^{-(c_1-1)M_n} \big) \\ &\le C\delta^{-2\varphi+2} e^{-\epsilon_2 \delta}. \end{aligned}$$

Choosing $\delta = \epsilon_2 / \log n$, we have the desired result. □

APPENDIX A: TECHNICAL MATERIAL

A.1. Power function on $\mathcal{SP}_m$. For $z \in \mathcal{SP}_m$ and $k \in K$, we have for the power function p_s:

$$p_s(k'zk) = |k'zk|^{s_m} \prod_{j=1}^{m-1} |(k'zk)_j|^{s_j} = \prod_{j=1}^{m-1} |(k'zk)_j|^{s_j}$$

and

$$\begin{aligned} p_s(k' e^{v/m} zk) &= \prod_{j=1}^{m} |(e^{v/m} k' zk)_j|^{s_j} = \prod_{j=1}^{m} \Big(e^{vj/m} |(k'zk)_j| \Big)^{s_j} \\ &= e^{vm^{-1}\sum_{j=1}^m js_j} \prod_{j=1}^{m} |(k'zk)_j|^{s_j} = e^{\tilde{s}v} p_s(k'zk), \end{aligned}$$

where $\tilde{s} = m^{-1}\sum_{j=1}^m js_j$. By (3.2) we have, for $z \in \mathcal{SP}_m$,

$$\begin{aligned} h_s(e^{v/m} z) &= \int_K p_s(k' e^{v/m} zk)\, \mathrm{d}k \\ &= e^{\tilde{s}v} \int_K p_s(k'zk)\, \mathrm{d}k \\ &= e^{\tilde{s}v} h_s(z). \end{aligned} \tag{A.1}$$

Since $p_s(k'zk)$ is independent of s_m for any $z \in \mathcal{SP}_m$ and $k \in K$, then $h_s(z)$ is also independent of s_m for any $z \in \mathcal{SP}_m$.

A.2. Integration in hyperbolic polar coordinates. In this setting, we have $m = 2$ and suppose that ψ is K-invariant. For $a \in A^+$, let

$$\gamma(a) = |a|^{-1/2}(a_1 - a_2) \text{ and } \mathrm{d}_* a = \frac{\mathrm{d}a_1}{a_1}\frac{\mathrm{d}a_2}{a_2}.$$

Define

$$\mathcal{D} = \{(\alpha, u) : \alpha \geq 0, u \in \mathbb{R}\} \tag{A.2}$$

and set

$$\mathrm{d}(\alpha, u) = 2\pi \sinh\alpha\, \mathrm{d}\alpha\, \mathrm{d}u, \tag{A.3}$$

$(\alpha, u) \in \mathcal{D}$. We also denote by D_α the matrix $\begin{bmatrix} e^{\alpha} & 0 \\ 0 & e^{-\alpha} \end{bmatrix}$. By changes of variables, we have Lemma A.1.

LEMMA A.1.

$$\int_{\mathcal{P}_2} \psi(y)\, \mathrm{d}_* y = \pi \int_{A^+} \psi(a)\gamma(a) \mathrm{d}_* a = \int_{\mathcal{D}} \psi\Big(e^{u/2} D_\alpha\Big)\, \mathrm{d}(\alpha, u).$$

LEMMA A.2. *For* $c_1 > 0$,

$$\int_{\{u \in \mathbb{R}, \alpha > c_1\}} \mathrm{d}W(\alpha, u) = 4\pi (\cosh c_1)^{-(N-1)}\, \Gamma(N-1), \tag{A.4}$$

(A.5)
$$\int_{\{u>c_2,\alpha>c_1\}} \mathrm{d}W(\alpha,u) = 4\pi(\cosh c_1)^{-(N-1)} \int_{\{t>e^{c_2/2}\cosh c_1\}} t^{N-2}e^{-t}\,\mathrm{d}t,$$

and

(A.6)
$$\int_{\{u<-c_2,\alpha>c_1\}} \mathrm{d}W(\alpha,u) = 4\pi(\cosh c_1)^{-(N-1)} \int_{\{t<e^{-c_2/2}\cosh c_1\}} t^{N-2}e^{-t}\,\mathrm{d}t.$$

PROOF. Observe that

$$\begin{aligned}
&\int_{\{u\in\mathbb{R},\alpha>c_1\}} e^{Nu/2}\exp\big(-e^{u/2}\cosh\alpha\big)\mathrm{d}(\alpha,u)\\
&\quad= 2\pi\int_{\{u\in\mathbb{R},t>\cosh c_1\}} e^{Nu/2}\exp\big(-e^{u/2}t\big)\,\mathrm{d}t\,\mathrm{d}u\\
&\quad= 2\pi\int_{u\in\mathbb{R}} e^{Nu/2}e^{-u/2}\exp\big(-e^{u/2}\cosh c_1\big)\,\mathrm{d}u
\end{aligned}$$

Substituting $t = e^{u/2}$, the integral reduces to a classical gamma integral, and then (A.4) is obtained readily. Next,

$$\begin{aligned}
&\int_{\{u>c_2,\alpha>c_1\}} e^{Nu/2}\exp\big(-e^{u/2}\cosh\alpha\big)\,\mathrm{d}(\alpha,u)\\
&\quad= 2\pi\int_{\{u>c_2,t>\cosh c_1\}} e^{Nu/2}\exp\big(-e^{u/2}t\big)\,\mathrm{d}t\,\mathrm{d}u\\
&\quad= 2\pi\int_{\{u>c_2\}} e^{Nu/2}e^{-u/2}\exp\big(-e^{u/2}\cosh c_1\big)\,\mathrm{d}u
\end{aligned}$$

Substituting $t = e^{u/2}\cosh c_1$, we obtain (A.5). Finally, (A.6) is established analogously. □

In the following results, we let $k_\theta = \begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}$. Lemma A.3 follows from Lemma A.1.

LEMMA A.3.

$$(\psi * w)(e^{v/2}D_\beta) = \frac{1}{2\pi}\int_{\mathcal{D}}\int_0^{2\pi}\psi(e^{(v-u)/2}D_{-\alpha/2}k_\theta D_\beta k'_\theta D_{-\alpha/2})\mathrm{d}\theta \times w(e^{u/2}D_\alpha)\,\mathrm{d}(\alpha,u). \qquad \text{(A.7)}$$

Simple calculations prove Lemma A.4.

LEMMA A.4. *For $\alpha \geq 0$, $\beta \geq 0$, and $\theta \in [0, 2\pi]$, the matrix equation*

$$k_\xi D_R k'_\xi = D_{-\alpha/2} k_\theta D_\beta k'_\theta D_{-\alpha/2} \tag{A.8}$$

has a solution $R^* = R^*(\alpha, \beta, \theta)$ and $\xi^* = \xi^*(\alpha, \beta, \theta)$. Further, $\cosh R^*$ has minimum and maximum values $\cosh(\alpha - \beta)$ and $\cosh(\alpha + \beta)$, respectively, and R^* and ξ^* can be defined uniquely.

LEMMA A.5.

$$(\psi * w)(e^{v/2} D_\beta) = \int_{\mathcal{D}} \Big[\frac{1}{2\pi} \int_K \psi(e^{(v-u)/2} D_{R^*(\alpha,\beta,\theta)})\, \mathrm{d}\theta \Big] w(e^{u/2} D_\alpha)\, \mathrm{d}(\alpha, u).$$

PROOF. From equation (A.7) and Lemma A.5, we obtain the desired result. □

REFERENCES

[1] Alexander, D.C. (2005). Multiple-fiber reconstruction algorithms for diffusion MRI. *Annals of the New York Academy of Sciences*, **1064**, 113–133.

[2] Butucea, C., Tsybakov, A.B. (2008). Sharp optimality in density deconvolution with dominating bias, I, II. *Theory Probab. Appl.*, **52**, 24–39, 237–249.

[3] 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.

[4] Dryden, I. L., Koloydenko, A., and Zhou, D. (2009). Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Ann. Appl. Statist.*, to appear.

[5] Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems *Ann. Statist.*, **19**, 1257-1272.

[6] Fletcher, P. T., and Joshi, S. (2004). Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors. In: *Proceedings of the ECCV 2004 Workshop on Computer Vision Approaches to Medical Image Analysis (CVAMIA)*, Lecture Notes in Computer Science, **3117**, pp. 87–98, Springer, New York.

[7] Fletcher, P. T., and Joshi, S. (2007). Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Processing*, **87**, 250-262.

[8] Haff, L. R. (1979). Estimation of the inverse covariance matrix: Random mixtures of the inverse Wishart matrix and the identity. *Ann. Statist.*, **7**, 1264-1276.

[9] Haff, L. R. (1980). Empirical Bayes estimation of the multivariate normal covariance matrix. *Ann. Statist.*, **8**, 586-597.

[10] Haff, L. R. (1991). The variational form of certain Bayes estimators. *Ann. Statist.*, **19**, 1163-1190.

[11] 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.

[12] Healy, D. M., Hendriks, H., and Kim, P. T. (1998). Spherical deconvolution. *J. Multivariate Anal.*, **67**, 1–22.

[13] Helgason, S. (1978). *Differential Geometry, Lie Groups and Symmetric Spaces*, Academic Press, New York.

[14] Hendriks, H. (1990). Nonparametric estimation of a probability density on a Riemannian manifold using Fourier expansions. *Ann. Statist.*, **18**, 832–849.
[15] Jian, B., and Vemuri, B. C. (2007). Multi-fiber reconstruction from diffusion MRI using mixture of Wisharts and sparse deconvolution. *Inform. Process Med. Imaging*, **20**, 384–395.
[16] Jian, B., Vemuri, B. C., Özarslan, E., Carney, P. R., and Mareci, T. H. (2007). A novel tensor distribution model for the diffusion-weighted MR signal. *NeuroImage*, **37**, 164–176 [Erratum: **42** (2008), 1045–1046].
[17] Johnstone, I. M. (2008). Multivariate analysis and Jacobi ensembles: Largest eigenvalue, Tracy–Widom limits and rates of convergence. *Ann. Statist.*, **36**, 2638–2716.
[18] Kim, P. T. (1998). Deconvolution density estimation on $SO(N)$. *Ann. Statist.*, **26**, 1083–1102.
[19] Kim, P. T., and Richards, D. St. P. (2001). Deconvolution density estimation on compact Lie groups. *Contemporary Mathematics*, **287**, 155–171.
[20] Kim, P. T., and Richards, D. St. P. (2008). Diffusion tensor imaging and deconvolution on the space of positive definite symmetric matrices. In: *Mathematical Foundations of Computational Anatomy: Geometric and Statistical Methods for Biological Shape Variability Modeling*, X. Pennac and S. Joshi, eds., MFCA'08 New York, 140–149.
[21] Kim, P. T., and Richards, D. St. P. (2009). Deconvolution density estimation on spaces of positive definite symmetric matrices. *IMS Lecture Notes-Monograph Series: Festschrift for T. P. Hettmansperger*, to appear.
[22] Koltchinskii, V., Sakhanenko, L., and Cai, S. (2007). Integral curves of noisy vector fields and statistical problems in diffusion tensor imaging: nonparametric kernel estimation and hypotheses testing. *Ann. Statist.*, **35**, 1576–1607.
[23] Koo, J.–Y. (1993). Optimal rates of convergence for nonparametric statistical inverse problems. *Ann. Statist.*, **21**, 590–599.
[24] Letac, G., and Massam, H. (2007). Wishart distributions for decomposable graphs. *Ann. Statist.*, **35**, 1278–1323.
[25] Lin, S.P., and Perlman, M.D. (1985). A Monte Carlo comparison of four estimators for a covariance matrix. In *Multivariate Analysis VI* (P.R. Krishnaiah, ed.), 411-429. North-Holland, Amsterdam.
[26] Loh, W.–L. (1991). Estimating covariance matrices. *Ann. Statist.*, **19**, 283–296.
[27] Mair, B.A., and Ruymgaart, F. H. (1996). Statistical inverse estimation in Hilbert scales. *SIAM J. Appl. Math.*, **56**, 1424–1444.
[28] Muirhead, R. J. (1982). *Aspects of Multivariate Statistics*. Wiley, New York.
[29] Pensky, M. (1998). Nonparametric empirical Bayes estimation of the matrix parameter of the Wishart distribution. *J. Multivariate Anal.*, **69**, 242–260.
[30] Pesenson, I. (2006). Deconvolution of band-limited functions on non-compact symmetric spaces. *Houston J. Math.*, **32**, 183–204.
[31] Richards, D. St. P. (1985). Applications of invariant differential operators to multivariate distribution theory. *SIAM J. Appl. Math.*, **45**, 280–288.
[32] Stein, C. (1975). Estimation of a covariance matrix. Rietz Lecture, 1975 Annual Meeting of the American Statistical Association, Atlanta, Ga.
[33] Stein, C. (1977). Lectures on the theory of estimation of many parameters. In: *Studies in the Statistical Theory of Estimation, Part I* (I. A. Ibragimov and M. S. Nikulin, eds.) *Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov.* (LOMI), **74**, 4–65, 146, 148 (in Russian) [English transl.: *J. Soviet Math.*, **34** (1986), 1373–1403].
[34] Takemura, A. (1984). An orthogonally invariant minimax estimator of the covariance matrix of a multivariate normal population. *Tsukuba J. Math.*, **8**, 367–376.

[35] Terras, A. (1988). *Harmonic Analysis on Symmetric Spaces and Applications, II.* Springer, New York.
[36] Zhang, C.H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.*, **18**, 806–830.
[37] Zhu, H., Zhang, H., Ibrahim, J. G., Peterson, B. S. (2007). Statistical analysis of diffusion tensors in diffusion-weighted magnetic resonance imaging data. *J. Amer. Statist. Assoc.*, **102**, 1085–1102.

Department of Mathematics
University of California at San Diego
La Jolla, California 92093, USA
e-mail: haff@ucsd.edu

Department of Mathematics and Statistics
University of Guelph
Guelph, Ontario N1G 2W1, Canada
e-mail: pkim@uoguelph.ca

Department of Statistics
Korea University
Anam-Dong Sungbuk-Ku
Seoul, 136-701 Korea
e-mail: jykoo@korea.ac.kr

Department of Statistics
Penn State University
University Park, PA 16802, USA
e-mail: richards@stat.psu.edu