



# On the derivatives of the normalising constant of the Bingham distribution

A. Kume, Andrew T.A. Wood

## ► To cite this version:

A. Kume, Andrew T.A. Wood. On the derivatives of the normalising constant of the Bingham distribution. *Statistics and Probability Letters*, 2009, 77 (8), pp.832. 10.1016/j.spl.2006.12.003 . hal-00556855

**HAL Id: hal-00556855**

**<https://hal.science/hal-00556855>**

Submitted on 18 Jan 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Author's Accepted Manuscript

On the derivatives of the normalising constant of the Bingham distribution

A. Kume, Andrew T.A. Wood

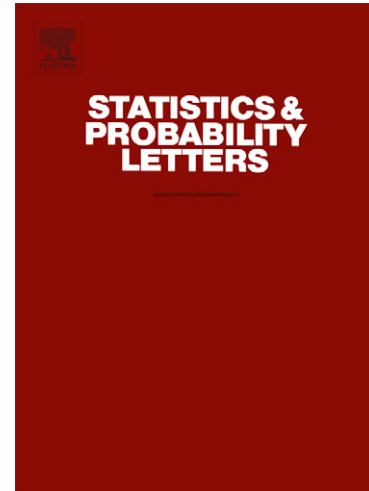
PII: S0167-7152(07)00017-X  
DOI: doi:10.1016/j.spl.2006.12.003  
Reference: STAPRO 4550

To appear in: *Statistics & Probability Letters*

Received date: 9 February 2005  
Revised date: 27 October 2006  
Accepted date: 19 December 2006

Cite this article as: A. Kume and Andrew T.A. Wood, On the derivatives of the normalising constant of the Bingham distribution, *Statistics & Probability Letters* (2007), doi:10.1016/j.spl.2006.12.003

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



[www.elsevier.com/locate/stapro](http://www.elsevier.com/locate/stapro)

# On the derivatives of the normalising constant of the Bingham distribution

A. Kume

*Institute of Mathematics, Statistics and Actuarial Science, University of Kent,  
Canterbury, CT2 7NF, UK  
a.kume@kent.ac.uk*

Andrew T.A. Wood

*School of Mathematical Sciences, University of Nottingham, University Park,  
Nottingham, NG7 2RD, UK  
atw@maths.nott.ac.uk*

## Abstract

It is shown that an arbitrary joint partial derivative of the Bingham normalising constant, expressed in standard form, is proportional to the normalising constant of a Bingham distribution of higher dimension. Two consequences of this result are noted.

*Keywords:* complex Bingham distribution; directional statistics; shape analysis.

## 1 Introduction

The Bingham distribution has proved useful as a model in Directional Statistics and, more recently, the complex Bingham distribution has been widely used in planar Shape Analysis. See, for example, Bingham (1974), Kent (1994), Dryden and Mardia (1998) and Mardia and Jupp (2000). The Bingham distribution is defined as that of a zero-mean multivariate normal random vector in  $\mathbb{R}^p$  conditioned to lie on the unit sphere  $\mathcal{S}^{p-1} \subset \mathbb{R}^p$ . For a given symmetric  $p \times p$  matrix  $\Sigma$ , the Bingham density with respect to the uniform measure  $d\mathcal{S}^{p-1}(x)$  on  $\mathcal{S}^{p-1}$  is given by

$$f(x|\Sigma) = \mathcal{C}_0(\Sigma)^{-1} \exp(-x^\top \Sigma x), \quad x^\top x = 1, \quad (1)$$

where  $\mathcal{C}_0(\Sigma)$  is a normalizing constant. If  $\Sigma = Q\Lambda Q^\top$  is the spectral decomposition of  $\Sigma$  where  $Q \in O(p)$ , the space of  $p \times p$  orthogonal matrices, and  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$ , then it easy to see that  $\mathcal{C}_0(\Sigma) = \mathcal{C}_0(\Lambda)$ , because the orthogonal transformation generated by left

multiplication by  $Q$  is an isometry in  $\mathcal{S}^{p-1}$ . Note that, from (1),

$$\mathcal{C}_0(\Lambda) = \int_{x \in \mathcal{S}^{p-1}} \exp \left( - \sum_{i=1}^p \lambda_i x_i^2 \right) d\mathcal{S}^{p-1}(x). \quad (2)$$

Moreover, as a consequence of the constraint  $x^\top x = 1$ ,  $\mathcal{C}_0(\Lambda + tI_p) = \exp(-t)\mathcal{C}_0(\Lambda)$ , where  $I_p$  is the  $p \times p$  identity matrix.

If  $x_1, \dots, x_n$  is a sample of unit vectors in  $\mathcal{S}^{p-1}$  from the Bingham distribution with density (1), and  $S = \sum_{i=1}^n x_i x_i^\top$ , then the log-likelihood of the data is given by

$$\mathcal{L}(Q, \Lambda) = -\text{tr}(SQ\Lambda Q^\top) - n \log(\mathcal{C}_0(\Lambda)).$$

Moreover, the maximum likelihood estimator (MLE) of  $Q$  is given by  $\hat{Q}$ , where  $S = \hat{Q}\hat{\Lambda}\hat{Q}^\top$  is the spectral decomposition of  $S$ , and

$$\mathcal{L}(\hat{Q}, \Lambda) = - \sum l_i \lambda_i - n \log(\mathcal{C}_0(\Lambda)). \quad (3)$$

See e.g. Mardia and Jupp (2000) for further details. Two points to note are: (i) from standard theory for exponential families, when the MLE of  $\Lambda$  exists, it is unique; and (ii) the MLE of  $\Lambda$  exists with probability 1 when  $n \geq p$ . Broadly, there are two ways to obtain the MLE of  $\Lambda$ : one is to maximise (3); the other, bearing in mind points (i) and (ii) above, is to solve

$$-\mathcal{C}_0(\Lambda)^{-1} \frac{\partial \mathcal{C}_0(\Lambda)}{\partial \lambda_i} = \frac{l_i}{n}, \quad i = 1, 2, \dots, p. \quad (4)$$

Neither  $\mathcal{C}_0(\lambda)$  nor its derivatives have useful closed-form expressions in the general case, but Kume and Wood (2005) have recently proposed saddlepoint approximations to  $\mathcal{C}_0(\lambda)$  which have desirable numerical and theoretical properties. The main purpose of this note is to show that any derivative of (2) is itself proportional to the normalising constant of a higher-dimensional Bingham distribution, with the consequence that these saddlepoint approximations can also be used to approximate the derivatives of (2). As a second application of our main result, we derive an explicit expression for the complex Bingham normalising constant in the case where the relevant eigenvalues have arbitrary multiplicities.

## 2 The main result

A multi-index is defined here as a row vector  $J = (j_1, \dots, j_p)$  whose components  $j_i$  are integers. Define  $|J| = \sum_{i=1}^p |j_i|$ . If all components of  $J$  are non-negative, we say that

$J$  is non-negative. In this paper we only consider non-negative multi-indices. For each non-negative  $J$ , we introduce the partial differential operator  $\mathcal{D}^J$ , where

$$\mathcal{D}^J = \prod_{i=1}^p \mathcal{D}_i^{j_i} = \prod_{i=1}^p \frac{\partial^{j_i}}{\partial \lambda_i^{j_i}} = \frac{\partial^{|J|}}{\partial \lambda_1^{j_1} \dots \partial \lambda_p^{j_p}}.$$

Let  $\mathcal{C}(\lambda; M)$ , where  $\lambda = (\lambda_1, \dots, \lambda_p)^T$  as before and  $M = (m_1, \dots, m_p)$  is a multi-index, denote the Bingham normalising constant  $\mathcal{C}_0(\tilde{\Lambda})$  in (2) where  $\tilde{\Lambda}$  is a diagonal matrix of dimension  $|M| \times |M|$  consisting of diagonal elements  $\lambda_i$  with multiplicity  $m_i$ ,  $i = 1, \dots, p$ . Note that if  $1_p$  is the  $p$ -vector of ones, then  $\mathcal{C}(\lambda; 1_p) = \mathcal{C}_0(\Lambda)$ , where the latter is defined in (2). The main result of this paper is the following.

**Proposition 1**

For any non-negative multi-indices  $J = (j_1, \dots, j_p)$  and  $M = (m_1, \dots, m_p)$ ,

$$\mathcal{D}^J \mathcal{C}(\lambda; M) = \alpha(J, M) \mathcal{C}(\lambda; M + 2J) \quad (5)$$

where  $\alpha(J, M)$  is the constant given by

$$\alpha(J, M) = (-1)^{|J|} \pi^{-|J|} \prod_{i=1}^p \frac{\Gamma\{(m_i + 2j_i)/2\}}{\Gamma(m_i/2)} \quad (6)$$

and  $\Gamma(\delta) = \int_0^\infty x^{\delta-1} e^{-x} dx$  is the gamma function.

**Proof.** It is convenient to break the proof into two steps.

*Step 1.* Show that

$$\mathcal{C}(\lambda; M) = \frac{2\pi^{|M|/2}}{\Gamma(|M|/2)} \mathcal{E}_p(\lambda; M) \quad (7)$$

where

$$\mathcal{E}_p(\lambda; M) = \frac{\Gamma(|M|/2)}{\prod_{i=1}^p \Gamma(m_i/2)} \int_{\beta \in \mathcal{S}_p} \prod_{i=1}^p \beta_i^{(m_i-2)/2} \exp\left(-\sum_{i=1}^p \lambda_i \beta_i\right) [d\beta], \quad (8)$$

$\mathcal{S}_p$  is the standard simplex in  $\mathbb{R}^p$ , consisting of those vectors with non-negative components which sum to 1, and  $[d\beta]$  is the Lebesgue measure on  $\mathcal{S}_p$ .

*Step 2.* Show that

$$\mathcal{D}^J \mathcal{E}_p(\lambda; M) = \frac{\Gamma(|M|/2)}{\Gamma(|M + 2J|/2)} \pi^{|J|} \alpha(J, M) \mathcal{E}_p(\lambda; M + 2J) \quad (9)$$

where  $\alpha(J, M)$  is defined in (6).

Assuming that Step 1 and Step 2 hold, the proof of Proposition 1 is completed as follows:

$$\begin{aligned}
 \mathcal{D}^J \mathcal{C}(\lambda; M) &= \frac{2\pi^{|M|/2}}{\Gamma(|M|/2)} \mathcal{D}^J \mathcal{E}_p(\lambda; M) \quad [\text{using (7)}] \\
 &= \frac{2\pi^{|M|/2}}{\Gamma(|M|/2)} \frac{\Gamma(|M|/2)}{\Gamma(|M+2J|/2)} \pi^{|J|} \alpha(J, M) \mathcal{E}_p(\lambda; M+2J) \quad [\text{using (9)}] \\
 &= \alpha(J, M) \frac{2\pi^{|M+2J|/2}}{\Gamma(|M+2J|/2)} \mathcal{E}_p(\lambda; M+2J) \\
 &= \alpha(J, M) \mathcal{C}(\lambda; M+2J) \quad [\text{using (7) again}].
 \end{aligned}$$

Step 2 is proved directly by differentiating under the integral sign in (8); we omit the details.

To prove Step 1, let  $\tilde{\lambda}$  denote a vector of length  $|M|$  where  $\lambda_i$  has multiplicity  $m_i$ ,  $i = 1, \dots, p$ . It is clear that

$$\mathcal{C}(\lambda; M) = \mathcal{C}(\tilde{\lambda}; 1_{|M|}). \quad (10)$$

Moreover, it follows from the representation of the Bingham normalising constant given in the proof of Proposition 5 in Kume and Wood (2005), combined with formula (8) in that paper, that

$$\mathcal{C}(\tilde{\lambda}; 1_{|M|}) = \frac{2\pi^{|M|/2}}{\Gamma(|M|/2)} \mathcal{E}_{|M|}(\tilde{\lambda}; 1_{|M|}). \quad (11)$$

Therefore, using a standard property of the Dirichlet integral,

$$\begin{aligned}
 \mathcal{E}_{|M|}(\tilde{\lambda}; 1_{|M|}) &= \frac{\Gamma(|M|/2)}{\Gamma(1/2)^{|M|}} \int_{\tilde{\beta} \in \mathcal{S}_{|M|}} \left( \prod_{r=1}^{|M|} \tilde{\beta}_r^{-1/2} \right) \exp \left( - \sum_{r=1}^{|M|} \tilde{\lambda}_r \tilde{\beta}_r \right) [d\tilde{\beta}] \\
 &= \frac{\Gamma(|M|/2)}{\Gamma(1/2)^{|M|}} \int_{\tilde{\beta} \in \mathcal{S}_{|M|}} \left( \prod_{r=1}^{|M|} \tilde{\beta}_r^{-1/2} \right) \exp \left( - \sum_{i=1}^p \lambda_i \sum_{r: \tilde{\lambda}_r = \lambda_i} \tilde{\beta}_r \right) [d\tilde{\beta}] \\
 &= \frac{\Gamma(|M|/2)}{\prod_{i=1}^p \Gamma(m_i/2)} \int_{\beta \in \mathcal{S}_p} \left( \prod_{i=1}^p \beta_i^{(m_i-2)/2} \right) \exp \left( - \sum_{i=1}^p \lambda_i \beta_i \right) [d\beta] \\
 &= \mathcal{E}_p(\lambda; M),
 \end{aligned}$$

and so (7) follows from (10) and (11). This completes the proof of Proposition 1.

*Remark 2.1.* Note that  $\mathcal{E}_p(\lambda; M)$  in (8) is the multivariate Laplace transform of a Dirichlet random vector,  $\beta$ , with parameter vector  $\frac{1}{2}M$ .

*Remark 2.2.* Jupp and Mardia (1979), who refer to Bingham (1976), and Muirhead (1982, p. 288, Example 7.8), state that  $\mathcal{C}(\lambda; 1_{|M|}) = {}_1F_1(1/2, |M|/2; \tilde{\Lambda})$ , where  $\tilde{\Lambda}$  is the diagonal matrix whose elements are the components of  $\tilde{\lambda}$ , and  ${}_1F_1$  is the confluent hypergeometric function of matrix argument. We note that (5) does not hold in general for the matrix argument  ${}_1F_1$ ; this can be seen by differentiating the integral representation for  ${}_1F_1$  given in Muirhead (1982, p. 264).

### 3 Numerical approximation of derivatives

There are two main practical motivations for calculating derivatives of the Bingham normalising constant: (i) to calculate the maximum likelihood estimate,  $\hat{\lambda}$  of  $\lambda$ , by solving (4); and (ii) to calculate asymptotic standard errors of  $\hat{\lambda}$  via the Fisher information matrix. In this section we discuss how a saddlepoint approximation proposed by Kume and Wood (2005), combined with Proposition 1, can be used for (i) and (ii).

#### 3.1 Saddlepoint approximations

When specialised to the Bingham normalising constant,  $c(\lambda) = \mathcal{C}_0(\lambda; 1_p)$ , the third and most accurate of the Kume and Wood (2005) approximations,  $\hat{c}_3(\lambda)$ , reduces to the following:

$$\hat{c}_3(\lambda) = 2^{1/2} \pi^{(p-1)/2} \{K_\lambda^{(2)}(\hat{t})\}^{-1/2} \exp\{T(\hat{t}) - \hat{t}\} \prod_{i=1}^p (\lambda_i - \hat{t})^{-1/2},$$

where

$$K_\lambda(t) = -\frac{1}{2} \sum_{i=1}^p \log(1 - t/\lambda_i), \quad K_\lambda^{(j)} = \frac{\partial^j K_\lambda(t)}{\partial t^j} = \sum_{i=1}^p \frac{(j-1)!}{2(\lambda_i - t)^j}, \quad j \geq 1,$$

$\hat{t}$  is the unique solution in  $(-\infty, \min_i \lambda_i)$  to the equation  $K_\lambda^{(1)}(t) = 1$ , and

$$T(t) = \frac{1}{8} \rho_4(t) - \frac{5}{24} \rho_3(t)^2 \quad \text{where} \quad \rho_j(t) = K_\lambda^{(j)}(t) / \{K_\lambda^{(2)}(t)\}^{j/2}. \quad (12)$$

#### 3.2 Approximation of derivatives

Kume and Wood (2005) suggested that, in an approximate likelihood procedure for estimating  $\lambda$ , the derivatives of  $c(\lambda)$  could be approximated using the corresponding derivatives of the approximation. Here we note that an alternative approach is also tractable: to solve (4) approximately using direct approximations to the derivatives obtained using Proposition 1.

Let  $\hat{t}_0$  and  $\hat{t}_i$  denote the solutions of  $K_\lambda^{(1)}(t) = 1$  and  $(\lambda_i - t)^{-1} + K_\lambda^{(1)}(t) = 1$ , respectively, write  $A_i(\lambda) = -\mathcal{D}_i \log c(\lambda)$ , and let  $T_i(t)$  denote (12), but with  $\rho_j(t)$ ,  $j = 3, 4$ , replaced by

$$\frac{(j-1)!(\lambda_i - t)^{-j} + K_\lambda^{(j)}(t)}{\{(\lambda_i - t)^{-2} + K_\lambda^{(2)}(t)\}^{j/2}}.$$

Then we solve (4) approximately by finding  $\lambda$  to solve  $\hat{A}_i(\lambda) = l_i/n$ , where  $\hat{A}_i(\lambda)$ , the ratio of saddlepoint approximations used to approximate  $A_i(\lambda)$ , is given by

$$\frac{1}{2} \left( \frac{K_\lambda^{(2)}(\hat{t}_0) \prod_{j=1}^p (\lambda_j - \hat{t}_0)}{\{(\lambda_i - \hat{t}_i)^{-1} + K_\lambda^{(2)}(\hat{t}_i)\} (\lambda_i - \hat{t}_i) \prod_{j=1}^p (\lambda_j - \hat{t}_j)} \right)^{1/2} \exp\{\hat{t}_0 - \hat{t}_i + T_i(\hat{t}_i) - T_0(\hat{t}_0)\}. \quad (13)$$

We conclude with two comments. First, if a maximisation procedure which requires the specification of derivatives is to be used to estimate  $\lambda$ , it is better to approximate derivatives by the corresponding derivatives of the approximation, rather than using (13) to approximate the derivatives directly. Second, details for the corresponding approximations for  $A_{ij}(\lambda) = \{c(\lambda)\}^{-1} \mathcal{D}_i \mathcal{D}_j c(\lambda)$ , which are needed to calculate the Fisher information matrix, are similar and we omit the details.

## 4 Complex Bingham normalising constant

In this section we use Proposition 1 to provide an explicit expression for the complex Bingham normalising constant when the Bingham exponent has eigenvalues  $\lambda_1, \dots, \lambda_p$  of multiplicity  $2(1 + m_1), \dots, 2(1 + m_p)$ , respectively. Note that the corresponding normalising constant is  $\mathcal{C}\{\lambda; 2(1_p + M)\}$ . It seems that it would be difficult to calculate this directly, but Proposition 1 gives us a systematic way of doing this. Kent (1994) has shown that

$$\mathcal{C}(\lambda; 2_p) = 2\pi^p \sum_{i=1}^p a_i \exp(-\lambda_i), \quad \text{where} \quad a_i = \prod_{r \neq i} (\lambda_r - \lambda_i)^{-1}, \quad (14)$$

and  $2_p$  is a row vector of 2's. Note that the formula is slightly different to that of Kent (1994) because of the minus sign in the exponent of (1).

**Proposition 2.** *We have*

$$\begin{aligned} \mathcal{C}\{\lambda; 2(1_p + M)\} &= \frac{2\pi^{p+|M|}}{\prod_{i=1}^p m_i!} \sum_{i=1}^p (-1)^{m_i} \exp(-\lambda_i) \\ &\quad \sum_{J_{0(i)} \geq 0, |J_{0(i)}| = m_i} \frac{m_i!}{j_0! \dots j_{i-1}! j_{i+1}! \dots j_p!} (-1)^{j_0} \prod_{r \neq i} \frac{(j_r + m_r)!}{(\lambda_r - \lambda_i)^{1+j_r+m_r}}. \end{aligned}$$



*Remark 4.1.* The second summation in the statement of Proposition 2 is over all non-negative multi-indices  $J_{0(i)} = (j_0, j_1, \dots, j_{i-1}, j_{i+1}, \dots, j_p)$  with  $p$  components which sum to  $m_r$ . When  $m_r = 0$ , the second sum above is interpreted as  $\prod_{r \neq i} j_r! / (\lambda_r - \lambda_i)^{1+j_r}$ . Note that  $j_0$  in the second sum is the number of times that  $\exp(-\lambda_i)$  is differentiated.

**Proof.** We prove the result by applying the differential operator  $\mathcal{D}^M$  to Kent's formula (14), multiplying the result by  $\alpha(M, 2_p)^{-1}$ , and then using Proposition 1. Write  $M_{(i)} = (m_1, \dots, m_{i-1}, m_{i+1}, \dots, m_p)$  with the obvious modifications when  $i = 1, p$ . Now

$$\mathcal{D}^{M_{(i)}} a_i = \left( \prod_{r \neq i} \mathcal{D}_r^{m_r} \right) \left( \prod_{r \neq i} \frac{1}{\lambda_r - \lambda_i} \right) = \prod_{r \neq i} \mathcal{D}_r^{m_r} \left( \frac{1}{\lambda_r - \lambda_i} \right) = (-1)^{|M_{(i)}|} \prod_{r \neq i} \frac{m_r!}{(\lambda_r - \lambda_i)^{1+m_r}}.$$

Therefore, using the fact that  $\mathcal{D}^M = \mathcal{D}_i^{m_i} \mathcal{D}^{M_{(i)}}$ ,

$$\begin{aligned} \mathcal{D}^M \mathcal{C}(\lambda; 2_p) &= 2\pi^p \sum_{i=1}^p \mathcal{D}^M [a_i \exp(-\lambda_i)] \\ &= 2\pi^p \sum_{i=1}^p \mathcal{D}_i^{m_i} [\exp(-\lambda_i) \mathcal{D}^{M_{(i)}} a_i] \\ &= 2\pi^p \sum_{i=1}^p (-1)^{|M_{(i)}|} \mathcal{D}_i^{m_i} \left[ \exp(-\lambda_i) \prod_{r \neq i} \frac{m_r!}{(\lambda_r - \lambda_i)^{1+m_r}} \right] \\ &= (-1)^{|M|} 2\pi^p \sum_{i=1}^p (-1)^{m_i} \exp(-\lambda_i) \\ &\quad \sum_{J_{0(i)} \geq 0, |J_{0(i)}| = m_i} \frac{m_i!}{j_0! \dots j_{i-1}! j_{i+1}! \dots j_p!} (-1)^{j_0} \prod_{r \neq i} \frac{(m_r + j_r)!}{(\lambda_r - \lambda_i)^{1+m_r+j_r}}. \end{aligned}$$

The final step above uses the multivariate version of Leibnitz's formula. The proof is completed by multiplying by  $\alpha(M, 2_p)^{-1}$  and then using Proposition 1.

## Acknowledgements

This work was supported by the U.K. Engineering and Physical Sciences Research Council, grant GR/R55757/01.

## References

- [1] Bingham, C. (1974). An antipodally symmetric distribution on the sphere. *Ann. Statist.* **2**, 1201-5.

- [2] Bingham, C. (1976). Expansions related to a hypergeometric function arising in an antipodally symmetric distribution on the sphere. Directional Data Project, Research Report 11, Department of Statistics, Leeds University.
- [3] Dryden, I.L. & Mardia, K.V. (1998). *Statistical Shape Analysis*. John Wiley, Chichester.
- [4] Jupp, P.E. and Mardia, K.V. (1979). Maximum likelihood estimators for the matrix von Mises-Fisher and Bingham distributions. *Ann. Statist.* **7**, 599–606.
- [5] Kent, J.T. (1994). The complex Bingham distribution and shape analysis. *J.R.Statist. Soc. B* **56**, 285–99.
- [6] Kume, A & Wood, A.T.A (2005). Saddlepoint approximations for the Bingham and Fisher-Bingham normalising constants. *Biometrika* **92**, 465–476..
- [7] Mardia, K.V. & Jupp, P.E. (2000). *Directional Statistics*. John Wiley, Chichester.
- [8] Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory*. John Wiley, New York.