Journal of Applied Statistics
Vol. 37, No. 3, March 2010, 515–528

Taylor & Francis
Taylor & Francis Group

Bayesian analysis for bivariate von Mises distributions

Kanti V. Mardia*

Department of Statistics, University of Leeds, Leeds LS2 9JT, UK

(*Received 8 October 2009; final version received 11 December 2009*)

There has been renewed interest in the directional Bayesian analysis for the bivariate case especially in view of its fundamental new and challenging applications to bioinformatics. The previous work had concentrated on Bayesian analysis for univariate von Mises distribution. Here, we give the description of the general bivariate von Mises (BVM) distribution and its properties. There are various submodels of this distribution which have become important and we give a review of these submodels. Also, we derive the normalizing constant for the general BVM distribution in a compact way. Conjugate priors and posteriors for the general case and the submodels are obtained. The conjugate prior for a multivariate von Mises distribution is also examined.

Keywords: bioinformatics; bivariate angular data; conjugate priors; cosine model; directional statistics; distributions on torus; sine model

1. Introduction

In protein bioinformatics, there has been a keen growing interest in bivariate von Mises (BVM) directional distributions. For example, one fundamental application arises from the conformational angles for protein backbones [12, 23]. Another main example is related to Bayesian alignment where the bivariate conditional distributions for rotational matrices (3×3) in the Euler parameterization are some bivariate von Mises distributions [7, 9, 19]. These distributions are particular cases of the 'full' BVM distribution, which was introduced by Mardia [15],

$$\begin{aligned} f(\theta, \phi) = c(\kappa_1, \kappa_2, A) \exp\{&\kappa_1 \cos(\theta - \mu) + \kappa_2 \cos(\phi - \nu) \\ &+ [\cos(\theta - \mu), \sin(\theta - \mu)] A [\cos(\phi - \nu), \sin(\phi - \nu)]^{\mathrm{T}}\}, \end{aligned} \tag{1}$$

where the angles $\theta, \phi \in (-\pi, \pi]$ lie on the torus, a square with opposite sides identified, the matrix $A = (a_{ij})$ is a 2×2 real matrix and $c(\cdot)$ is the normalizing constant. For the circular variables

*Email: k.v.mardia@leeds.ac.uk

ISSN 0266-4763 print/ISSN 1360-0532 online

DOI: 10.1080/02664760903551267
http://www.informaworld.com

(θ, ϕ) distributed as Equation (1), we will write

$$(\theta, \phi) \sim \mathrm{BVM}(\mu, \nu; \kappa_1, \kappa_2, A).$$

Also note that it will be understood that when we refer to BVM, it is the full BVM. The notation FBVM is not used to avoid confusion with the Fisher–Bingham distribution (FB). This model has eight parameters and allows for varied dependence between the two angles. Various submodels with five parameters have appeared [23, 29] to mimic the bivariate normal distribution.

Note that we can rotate the density to the following form. Consider $\exp\{Q\}$ with

$$Q = \kappa_1 \cos(\theta - \mu) + \kappa_2 \cos(\theta - \nu) + (\cos\theta, \sin\theta) B (\cos\phi, \sin\phi)^{\mathrm{T}}. \tag{2}$$

Then the the matrices A in Equation (1) and B in Equation (2) are connected, namely,

$$B = R(\mu)^{\mathrm{T}} A R(\nu), \quad A = R(\mu) B R(\nu)^{\mathrm{T}}, \tag{3}$$

where

$$R(\mu) = \begin{pmatrix} \cos\mu, & \sin\mu \\ -\sin\mu, & \cos\mu \end{pmatrix}. \tag{4}$$

Two important sub-models are the sine and the cosine models. The sine model $\mathrm{BVM}_{S,5}$ has the density [29]

$$f_{\mathrm{s}}(\theta, \phi) \propto \exp\{\kappa_1 \cos(\theta - \mu) + \kappa_2 \cos(\phi - \nu) + \lambda \sin(\theta - \mu) \sin(\phi - \nu)\}. \tag{5}$$

For the cosine model, the density is given by [23]

$$f_{\mathrm{c}}(\theta, \phi) \propto \exp\{\kappa_1 \cos(\theta - \mu) + \kappa_2 \cos(\phi - \nu) - \kappa_3 \cos(\theta - \mu - \phi + \nu)\}. \tag{6}$$

This model with the positive interaction is denoted by $\mathrm{BVM}_{C+,5}$. The cosine model with negative interaction, namely with the exponent in Equation (6) as

$$\kappa_1 \cos(\theta - \mu) + \kappa_2 \cos(\phi - \nu) - \kappa_3 \cos(\theta - \mu + \phi - \nu),$$

will be denoted by $\mathrm{BVM}_{C-,5}$. Some details, notation, and references for these distributions are given in Section 2.

Consider the univariate von Mises distribution with pdf [21].

$$\{2\pi I_0(\kappa)\}^{-1} \exp\{\kappa \cos(\theta - \mu)\},$$

where $-\pi < \theta \leq \pi$, $-\pi < \mu \leq \pi$ and $\kappa \geq 0$, $I_0(\kappa)$ is a Bessel function, μ the mean direction and κ the concentration (precision) parameter. It has been shown in [20] that for given κ, the conjugate prior for μ leads to a posterior distribution which is also von Mises. Guttorp and Lockhart [8] have given the joint conjugate prior for μ and κ, and Mardia [17] has considered a slight variant. However, the posterior distribution for κ is not straightforward. Various suggestions have appeared; for example, take the prior for κ independently as a chi-square distribution, use the noninformative prior and so on.

We will focus here on conjugate priors since these have been popular due to the following properties:

(a) It possesses an algebraic advantage since at least the normalizing constant will also have a simple form in general.
(b) It has numerical advantages.

(c) One can assess its effect on the likelihood.
(d) It is important in directed acyclic graphs (DAG) because the joint probability can be split into the prior part and the likelihood part at different nodes.
(e) The mixture of conjugate priors is again a conjugate prior.

However, when there is very little information, a uniform prior may be adequate.

In this paper, we extend the univariate directional work on conjugate priors to the bivariate case with some multivariate extensions. We discuss various properties of the BVM and submodels in Section 2. In Section 3, we derive a convenient expression for the normalizing constant for the full BVM. We show in Section 4 that the conjugate prior for (μ, ν), given κ_1, κ_2 and A for the full BVM, is also a full BVM and the conjugate priors for submodels are also derived; a historical note is given in Section 4.3. The case for the conjugate prior for κ_1, κ_2, A for the given mean and the full conjugate prior are given in Section 5. In Section 6, we consider the conjugate prior for the multivariate von Mises distribution (sine model). In Section 7, we give an overview of applications of circular distributions in protein bioinformatics. We end with a discussion in Section 8.

2. Some properties of BVM and sub-models

In Table 1, we give a summary of various submodels for the different choice of parameters of BVM, references and notation. The most central distributions are the sine and the cosine models, each with five parameters.

First, we note that the conditional distributions of BVM are von Mises, and Mardia [15] has proved that within this exponential family we cannot have both marginals as well as conditionals to be von Mises. Further, Mardia [15] has shown that the marginal pdf of ϕ is proportional to

$$\exp\{\kappa_2 \cos(\phi - \nu)\} I_0(Q^{1/2}), \tag{7}$$

where Q is given by

$$\begin{aligned}&\kappa_1^2 + 2\kappa_1(a_{11}\cos\mu + a_{12}\sin\mu)\cos\phi + 2\kappa_1(a_{21}\cos\mu + a_{22}\sin\mu)\sin\phi\\ &\quad + (a_{11}^2 + a_{12}^2)\cos^2\phi + (a_{21}^2 + a_{22}^2)\sin^2\phi + (a_{11}a_{21} + a_{12}a_{22})\cos\phi\sin\phi.\end{aligned}$$

Various sub-cases can be considered. For the sine model, Mardia et al. [25] have shown that the marginals are 'nearly' von Mises.

The distribution can be bimodal, see particular cases examined in [29] and [23] for the sine and the cosine models, respectively. Also, the conditional distribution of $u = \theta + \phi$, given ϕ,

Table 1. Submodels of the general BVM distribution with parameters $\mu = \nu = 0, \kappa_1 \neq 0, \kappa_2 \neq 0$ (unless stated); $(A)_{ij} = a_{ij}, i, j = 1, 2$. $R(\psi) =$ rotation.

	A	Name	Notation	Reference
0.	$A = \kappa_3 R(\psi)$	General cosine	$\mathrm{BVM}_{GC,6}$	[16]
1.	$\alpha \neq 0, \beta \neq 0$	Rivest	$\mathrm{BVM}_{R,6}$	[28]
2.	$\alpha = \beta = -\kappa_3$	Cosine +ve interaction	$\mathrm{BVM}_{C+,5}$	[23]
3.	$\alpha = -\beta = -\kappa_3$	Cosine -ve interaction	$\mathrm{BVM}_{C-,5}$	[23]
4.	$\alpha = 0, \beta = \lambda$	Sine model	$\mathrm{BVM}_{S,5}$	[29]
5.	$\alpha = \cosh\lambda - 1, \beta = \sinh\lambda$	Hybrid	$\mathrm{BVM}_{H,5}$	[10]
6.	$\kappa_1 = \kappa_2 = 0, \alpha \neq 0, \beta \neq 0$	Diffused model	$\mathrm{BVM}_{D,4}$	[9]

Notes: Exponent term $= a_{11}\cos\theta\cos\phi + a_{12}\cos\theta\sin\phi + a_{21}\sin\theta\cos\phi + a_{22}\sin\theta\sin\phi$. For the Case 0, $a_{11} = a_{22}, a_{21} = -a_{12}$. For Cases 1–6, $a_{12} = a_{21} = 0, a_{11} = \alpha, a_{22} = \beta, \kappa_1 \neq 0, \kappa \neq 0$ unless stated.

is the full circular von Mises Bingham distribution. Thus, one should be aware of its complex behaviour [10]. Let $\mu = \nu = 0$ without any loss of generality. It can be shown that for concentrated data, we have

$$\begin{pmatrix}\theta\\ \phi\end{pmatrix} \sim N\left[\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \Sigma\right],$$

where

$$\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix} = \Sigma^{-1}\begin{pmatrix}a_{21}\\ a_{12}\end{pmatrix} \quad \text{and} \quad \Sigma^{-1} = \begin{pmatrix}a_{11}+\kappa_1 & -a_{22}\\ -a_{22} & a_{11}+\kappa_2\end{pmatrix}.$$

Thus, the six parameters (or the eight parameters) collapse to five parameters in a particular way.

One can simulate the full distribution by using the conditional von Mises of $\theta|\phi$, and then the marginal distribution of ϕ using Equation (7). For the sine and cosine case, rejection methods have been developed in [23]. Habeck [9] has given an efficient method to simulate the diffused distribution since it can be transformed into two independent univariate von Mises distributions. Simulating both conditional distributions is not difficult since these are both univariate von Mises; for the univariate von Mises, there is an efficient algorithm of Best and Fisher [2]. This form of sampling has been used in [7] for simulating the Fisher matrix distribution [11].

3. Normalizing constant for the full BVM

The normalizing constant in a compact form for the full BVM has been elusive since [15], but surprisingly we can obtain it in a compact form. Let us write the singular value decomposition of A as

$$A = R(\gamma)\Lambda R(\delta)^{\mathrm{T}}, \tag{8}$$

where $R(\gamma)$ and $R(\delta)$ are two orthogonal matrices and $\Lambda = \mathrm{diag}(\alpha, \beta)$. Thus, the normalizing constant is given by

$$I = \int_0^{2\pi}\int_0^{2\pi} \exp\{\kappa_1\cos(\theta-\gamma) + \kappa_2\cos(\phi-\delta) + \alpha\cos\theta\cos\phi + \beta\sin\theta\sin\phi\}\,\mathrm{d}\theta\,\mathrm{d}\phi. \tag{9}$$

Let

$$\alpha+\beta = 2t, \quad t = (\alpha+\beta)/2; \quad \alpha-\beta = 2u, \quad u = (\alpha-\beta)/2, \quad \alpha = t+u, \quad \beta = t-u.$$

The exponent in Equation (9) can be written as

$$\kappa_1\cos(\theta-\gamma) + \kappa_2\cos(\phi-\delta) + t\cos(\theta-\phi) + u\cos(\theta+\phi),$$

and we can write each of the four terms in the exponent by the expansion of the form

$$\mathrm{e}^{\kappa_1\cos(\theta-\gamma)} = \sum_k \epsilon_k I_k(\kappa_1)\cos k(\theta-\gamma),$$

where $\epsilon_k = 1$ for $k = 0$ and $\epsilon_k = 2$ for $k > 0$. On substituting these in Equation (9) and changing the order of integration, we get the expression for I as follows:

$$\sum_k\sum_l\sum_m\sum_n I_k(\kappa_1)I_l(\kappa_2)I_m(t)I_n(u)\int_0^{2\pi}\int_0^{2\pi}\cos k(\theta-\gamma)\cos l(\phi-\delta)$$
$$\times\cos m(\theta-\phi)\cos n(\theta-\phi)\,\mathrm{d}\theta\,\mathrm{d}\phi. \tag{10}$$

Now,

$$2\cos m(\theta-\phi)\cos n(\theta+\phi) = \cos(m-n)\theta\cos(m+n)\phi + \sin(m-n)\theta\sin(m+n)\phi + \cos(m+n)\theta\cos(m-n)\phi + \sin(m+n)\theta\sin(m-n)\phi. \tag{11}$$

Substituting this expression in Equation (10), we can write the first term as

$$\int_0^{2\pi} \cos k\theta\cos k\gamma + \sin k\theta\sin k\gamma\cos(m-n)\theta\,\mathrm{d}\theta \times \int_0^{2\pi} \cos l\phi\cos l\delta + \sin l\phi\sin l\delta\cos(m+n)\phi\,\mathrm{d}\phi.$$

Using the standard orthogonality results, we find that the last term is given by

$$\pi^2\cos(k\gamma - l\delta),$$

only if $k+l=$ even or $k-l=$ even, zero otherwise. Further,

$$\epsilon_{(k+l)/2}\epsilon_{(k-l)/2} = 4 \quad \text{for } k\neq l,\ \ k\neq 0,\ \ l\neq 0.$$

The same argument applies to the other terms in Equation (10), and we get the following result.

THEOREM 3.1 *The normalization constant for the full BVM* $(\mu,\nu;\kappa_1,\kappa_2,A)$ *at Equation* (1) *is given by*

$$c(\kappa_1,\kappa_2,A)^{-1} = I = 4\pi^2\sum_k\sum_l\{\cos(k\gamma-l\delta)\}I_k(\kappa_1)I_l(\kappa_2)I_{(k+l)/2}(u)I_{(k-l)/2}(v), \tag{12}$$

where the summation is taken over the pairs (k,l) *such that* $k,\ l=-\infty$ *to* ∞, $k+l$ *is even, and* $u=(\alpha+\beta)/2$, $v=(\alpha-\beta)/2$, *and* A *is expresed in terms of* α,β *and* γ *and* δ *given by Equation* (8).

From Equation (12), we can deduce the normalizing constants for various submodels by noting that

$$I_0(0)=1,\quad I_k(0)=0,\quad \text{for } k\neq 0,$$

and some constants are tabulated in Table 2. However, how the constant for the sine model collapses to a single series is not obvious, and a proof for the first time is given in Appendix 1 but this type of proof must have been known to Singh et al. [29], though not published.

We can also deduce other cases. For example, consider the density with exponent

$$\kappa_1\cos\theta + \alpha\cos\theta\sin\phi + \beta\sin\theta\cos\phi.$$

Here, we can rotate θ by $\pi/2$ so that we have $\gamma=\pi/2$ and $\delta=0$. Hence from Equation (12), we have

$$I = 4\pi^2\sum_k\sum_\ell\left(\cos\frac{\pi k}{2}\right)I_k(\kappa_1)I_\ell(\kappa_2)I_{(k+\ell)/2}(u)I_{(k-\ell)/2}(v).$$

Since

$$\cos\frac{\pi k}{2} = (-1)^{k/2},\quad \text{for } k \text{ even};=0 \text{ otherwise},$$

we can simplify the expression further. In the cosine model with the exponent in the pdf as

$$\kappa_1\cos\theta + \alpha\sin\theta\cos\phi + \beta\cos\theta\sin\phi,$$

the role of α and β is changed so $\gamma=0$ and $\delta=\pi/2$ in Equation (12).

Table 2. Normalizing constants $\{c(\kappa_1, \kappa_2, A)\}^{-1}$ of submodels (see for notation Table 1).

Model	Normalizing constant
General cosine	$I_0(\kappa_1)I_0(\kappa_2)I_0(\kappa_3) + 2\sum_{k=1}^{\infty}(\cos k\psi)I_k(\kappa_1)I_k(\kappa_2)I_k(\kappa_3)$
Rivest	$4\pi^2\Sigma_k\Sigma_\ell I_k(\kappa_1)I_\ell(\kappa_2)I_{(k+\ell)/2}(u)I_{(k-\ell)/2}(v)$ and the sums are over $k+\ell$ even, $u=(\alpha+\beta)/2$, $v=(\alpha-\beta)/2$
Cosine +ve/cosine −ve	$(2\pi)^2\{I_0(\kappa_1)I_0(\kappa_2)I_0(\kappa_3) + 2\sum_{k=1}^{\infty} I_k(\kappa_1)I_k(\kappa_2)I_k(\kappa_3)\}$
Sine	$4\pi^2\sum_{k=0}^{\infty}\binom{2k}{k}\left(\frac{\lambda^2}{4\kappa_1\kappa_2}\right)^k I_k(\kappa_1)I_k(\kappa_2)$
Diffused	$4\pi^2 I_0\left(\frac{1}{2}(\alpha+\beta)\right)I_0\left(\frac{1}{2}(\alpha-\beta)\right)$

Note that for I given by Equation (12), we have

$$|I| \le 4\pi^2\Sigma_k\Sigma_\ell I_k(\kappa_1)I_\ell(\kappa_2)I_{(k+\ell)/2}(u)I_{(k-\ell)/2}(v),$$

where now the RHS is the normalizing constant for the Rivest distribution. Rivest [28] has given a numerical algorithm for calculating the estimates from this distribution (see Section 3.3, entitled ‘an algorithm for efficient estimation of the parameters’) and in view of this bound, his method will be applicable to Equation (12). This point will be examined more fully in future work.

4. Conjugate priors for the mean vector

4.1 *The main case*

Let (θ_i, ϕ_i), $i = 1, \ldots, n$, be a random sample from BVM $(\mu, \nu; \kappa_1, \kappa_2, A)$. We will use some compact notation for this case. Let $\ell(\theta)^{\mathrm{T}} = (\cos\theta, \sin\theta)$. We have

$$\ell(\theta)^{\mathrm{T}}R(\mu)^{\mathrm{T}} = (\cos(\theta-\mu), \sin(\theta-\mu)) = (\cos(\mu-\theta), -\sin(\mu-\theta)) = \ell(\mu)^{\mathrm{T}}R(\theta)^{\mathrm{T}}I_{-1},$$

where $I_{-1} = \mathrm{diag}(1, -1)$. Thus,

$$\ell(\theta)^{\mathrm{T}}R(\mu)^{\mathrm{T}}AR(\nu)\ell(\phi) = \ell(\mu)^{\mathrm{T}}R(\theta)^{\mathrm{T}}A^{\#}R(\phi)\ell(\nu),$$

where

$$A^{\#} = I_{-1}AI_{-1} = \begin{pmatrix} a_{11} & -a_{12} \\ -a_{21} & a_{22} \end{pmatrix}.$$

Hence the pdf of (θ_i, ϕ_i) is

$$\begin{aligned} &\propto \exp\{\kappa_1\ell(\mu)^{\mathrm{T}}\ell(\theta_i) + \kappa_2\ell(\nu)^{\mathrm{T}}\ell(\phi_i) + \ell(\theta_i)^{\mathrm{T}}R(\mu)^{\mathrm{T}}AR(\nu)\ell(\phi_i)\} \\ &\propto \exp\{\kappa_1\ell(\theta_i)^{\mathrm{T}}\ell(\mu) + \kappa_2\ell(\phi_i)^{\mathrm{T}}\ell(\nu) + \ell(\mu)^{\mathrm{T}}R(\theta_i)^{\mathrm{T}}A^{\#}R(\phi_i)\ell(\nu)\}, \end{aligned}$$

and the likelihood is given by

$$\propto \exp\{\kappa_1\ell_1^{\mathrm{T}}\ell(\mu) + \kappa_2\ell_2^{\mathrm{T}}\ell(\nu) + \ell(\mu)^{\mathrm{T}}B_1\ell(\nu)\}, \tag{13}$$

where

$$\ell_1 = \Sigma_{i=1}^n\ell(\theta_i), \quad \ell_2 = \Sigma_{i=1}^n\ell(\phi_i), \quad B_1 = \Sigma_{i=1}^n R(\theta_i)^{\mathrm{T}}A^{\#}R(\phi_i).$$

Suppose that the prior for (μ, ν) is

$$\propto \exp\{\kappa_{01}\ell(\mu_0)^{\mathrm{T}}\ell(\mu) + \kappa_{02}\ell(\nu_0)^{\mathrm{T}}\ell(\nu) + \ell(\mu-\mu_0)^{\mathrm{T}}A_0\ell(\nu-\nu_0)\}. \tag{14}$$

Hence from Equations (12)–(14), we find that the posterior pdf of (μ, ν) is given by

$$c(\kappa_1^*, \kappa_2^*, A^*)\exp\{\kappa_1^*\ell(\mu^*)^{\mathrm{T}}\ell(\mu) + \kappa_2^*\ell(\nu^*)^{\mathrm{T}}\ell(\nu) + \ell(\mu-\mu^*)^{\mathrm{T}}A^*\ell(\nu-\nu^*)\}, \tag{15}$$

where

$$\kappa_1^*\ell(\mu^*) = \kappa_1\ell_1 + \kappa_{01}\ell(\mu_0), \quad \kappa_2^*\ell(\nu^*) = \kappa_2\ell + \kappa_{02}\ell(\nu_0), \tag{16}$$

$$A^* = R(\mu^*)(B_0+B_1)R(\nu^*)^{\mathrm{T}}, \quad B_0 = R(\mu_0)^{\mathrm{T}}A_0R(\nu_0), \quad B_1 = \sum_{i=1}^{n} R(\theta_i)^{\mathrm{T}}A^{\#}R(\phi_i), \tag{17}$$

with

$$A^{\#} = \begin{pmatrix} a_{11} & -a_{12} \\ -a_{21} & a_{22} \end{pmatrix}.$$

(The notation A^* is not to be confused with the conjugate of A). Hence, we get the following result.

THEOREM 4.1 *If* (θ_i, ϕ_i), $i = 1, \ldots, n$, *are distributed as* $\mathrm{BVM}(\mu, \nu; \kappa_1, \kappa_2, A)$ *and the prior for* (μ, ν) *is* $\mathrm{BVM}(\mu_0, \nu_0; \kappa_{01}, \kappa_{02}, A_0)$, *then the posterior distribution of* (μ, ν) *is also* $\mathrm{BVM}(\mu^*, \nu^*; \kappa_{01}^*, \kappa_{02}^*, A^*)$ *where the parameters are given by Equations* (16) *and* (17). *Further, the normalizing constant* $c(\kappa_1^*, \kappa_2^*, A^*)$ *is given by Equation* (12).

A straight forward extension is as follows.

THEOREM 4.2 *Let* (θ_i, ϕ_i), $i = 0, 1, \ldots, n$, *be distributd as* $\mathrm{BVM}(\mu_i, \nu_i; \kappa_{1i}, \kappa_{2i}, A_i)$, *where* $\theta_0 = \mu$ *and* $\phi_0 = \nu$ *denote the random variables for the prior distribution of* (μ, ν). *Then the posterior distribution of* (μ, ν) *is also* $\mathrm{BVM}(\mu^*, \nu^*; \kappa_1^*, \kappa_2^*, A^*)$, *where*

$$\kappa_1^*\ell(\mu^*) = \sum_{i=0}^{n}\kappa_{1i}\ell(\theta_i), \quad \kappa_2^*\ell(\nu^*) = \sum_{i=0}^{n}\kappa_{2i}\ell(\phi_i)$$

and

$$A^* = \sum_{i=1}^{n} R(\theta_i-\mu^*)^{\mathrm{T}}A_i^{\#}R(\phi_i-\nu^*) + R(\mu_0-\mu^*)^{\mathrm{T}}A_0R(\nu_0-\nu^*)^{\mathrm{T}}. \tag{18}$$

We can write Equation (18) in the following two forms:

(1)

$$A^* = R(\mu^*)BR(\nu^*)^{\mathrm{T}},$$

where

$$B = B_0 + B_1, \quad B_0 = R(\mu_0)^{\mathrm{T}}A_0R(\nu_0), \quad B_1 = \sum_{i=1}^{n} R(\theta_i)^{\mathrm{T}}A_i^{\#}R(\phi_i).$$

(2)

$$A^* = \sum_{i=1}^{n} R^-(\theta_i - \mu^*) A_i R^-(\phi_i - \nu^*) + R^-(\mu_0 - \mu^*) A_0^\# R^-(\nu_0 - \nu^*), \tag{19}$$

but the form (18) is preferred in the discussion below.

4.2 *Particular cases*

We now discuss various particular cases of Theorem 4.2.

4.2.1 *The Rivest model*

For the Rivest model, we have $A_i = \mathrm{diag}(\alpha_i, \beta_i) = A_i^\#, i = 0, 1, \ldots, n,$ so Equation (18) becomes

$$A^* = \sum_{i=0}^{n} (\alpha_i U_i + \beta_i V_i), \tag{20}$$

where

$$U_i = \ell(\theta_i - \mu^*)\ell(\phi_i - \nu^*)^{\mathrm{T}} = \begin{pmatrix} u_{11i} & u_{21i} \\ u_{22i} & u_{22i} \end{pmatrix}, \quad V_i = \begin{pmatrix} u_{22i} & -u_{21i} \\ -u_{12i} & u_{11i} \end{pmatrix}, \quad i = 0, 1, \ldots, n.$$

Hence, the posterior for (μ, ν) is the full BVM. The diffused model can be viewed as its special cases.

4.2.2 *The sine model*

Here $\alpha_i = 0$, $\beta_i = \lambda_i$, so Equation (20) is simply

$$A^* = \sum_{i=0}^{n} \lambda_i V_i. \tag{21}$$

For $\lambda_i = \lambda,\ i = 1, \ldots, n$, the term (21) becomes

$$\lambda \begin{pmatrix} u_{22} & -u_{21} \\ -u_{12} & u_{11} \end{pmatrix} + \lambda_0 \begin{pmatrix} u_{220} & -u_{210} \\ -u_{120} & u_{110} \end{pmatrix},$$

where

$$U = \sum_{i=1}^{n} U_i = \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix}.$$

The key point is that the posterior density for (μ, ν) is not a sine density, but there has been some discussion in recent literature on this point (see [12], cf. [19]). For further details, see also the historical notes in Section 4.3.

4.2.3 *The cosine model*

$$A_i = -\kappa_{3i} I = A_i^\#, \quad i = 0, 1, \ldots, n.$$

It can be shown that the term is of the form

$$A^* = -\sum_{i=0}^{n} \kappa_{3i} R(\theta_i - \nu^* - \phi_i + \mu^* - \psi^*). \tag{22}$$

Thus, it is of the form of general cosine model (see below).

4.2.4 *General cosine model*

We can write here

$$A_i^{\#} = \begin{pmatrix} a_i & -b_i \\ b_i & a_i \end{pmatrix} = \kappa_{3i} R(\psi_i), \quad i = 1, \dots, n; \quad A_0 = \begin{pmatrix} a_0 & -b_0 \\ b_0 & a_0 \end{pmatrix} = \kappa_{30} R(\psi_0).$$

Thus,

$$B = \sum_{i=0}^{n} \kappa_{3i} R(\theta_i - \phi_i - \psi_i) = \kappa_3^* R(\psi^*),$$

where

$$\sum_{i=0}^{n} \kappa_{3i} \cos(\theta_i - \phi_i - \psi_i) = \kappa_3^* \cos\psi^*, \quad \sum_{i=0}^{n} \kappa_{3i} \sin(\theta_i - \phi_i - \psi_i) = \kappa_3^* \sin\psi^*.$$

Now the exponent term $\ell(\mu)^{\mathrm{T}} R(\mu^*) B R(\nu^*)^{\mathrm{T}} \ell(\nu)$ is given by

$$\kappa_3^* \cos(\mu - \nu - \mu^* + \nu^* - \psi^*).$$

Hence the posterior distribution of (μ, ν) for the general cosine case of six parameters is of the same form as the prior, which is a remarkable result. It is unlike the case of the sine model where the posterior has the full BVM.

4.3 *A historical note*

The conjugate prior for the sine model for (μ, ν) first appeared in Lennox et al. [12] but there were errors in the results which were corrected independently in Mardia [19], and in addendum to Lennox et al. [12]. Various results for the conjugate priors for (μ, ν) including Theorems 3.1, 4.1 and 4.2 were sketched in [19]. The normalization constant $c(\mu, \nu; \kappa_1, \kappa_2, A)$ required for the full BVM conjugate prior is new and derived here for the first time so the conjugate prior given here is complete. A remarkable aspect of this constant is that it has a similar form as that of the submodel of [28] except for a cosine multiplier within the summation. For the cosine models, and other models such as the diffused model, the behaviour of conjugate priors is also novel and interesting.

5. Other priors

5.1 *Mean vector given*

We now obtain the conjugate prior for κ_1, κ_2 and A, given the means μ and ν, which is possible knowing the normalizing constant for BVM given in Section 3. Without any loss of generality, we take $\mu = \nu = 0$. We now follow the procedure adopted by Guttorp and Lockhart [8] for the univariate von Mises case and take the conjugate prior density for κ_1, κ_2 and A proportional to

$$\{c(\kappa_1, \kappa_2, A)\}^r \exp\{\kappa_1\kappa_{01} + \kappa_2\kappa_{02} + tr\, B_0 A\}, \tag{23}$$

where $c(\cdot)$ is definted in Equation (12), κ_{01}, κ_{02} and B_0 are the hyperparameters, and we have taken $\mu_0 = \nu_0 = 0$. The normalizing constant requires the value of six variate integrals of Equation (23) over κ_1, κ_2 and A; the problem as expected is harder than in the univariate case; even for the univariate case, the normalizing constant is not simple [8]. But there are now various numerical techniques which bypass the normalizing constants (see Section 5.2 for a strategy). The posterior for (κ_1, κ_2, A) in the previous notation is proportional to

$$\{c(\kappa_1, \kappa_2, A)\}^{n+r} \exp\{\kappa_1 \kappa_1^* \cos\mu + \kappa_2 \kappa_2^* \cos\nu + \mathrm{tr} CA\}, \tag{24}$$

where

$$\kappa_1^* \cos\mu = \kappa_{01} + \Sigma \cos\theta_i, \quad \kappa_2^* \cos\nu = \kappa_{02} + \Sigma \cos\phi_i, \quad C = B_0 + \sum_{i=1}^{n} \ell(\phi_i)\ell(\theta_i)^{\mathrm{T}}. \tag{25}$$

This posterior can be further interpreted as in the univariate case [8].

5.2 *The general case*

The prior pdf of $(\mu, \nu, \kappa_1, \kappa_2, A)$ can be taken as proportional to

$$\{c(\kappa_1, \kappa_2, A)\}^{r} \exp\{\kappa_1 \kappa_{01} \cos(\mu - \mu_0) + \kappa_2 \kappa_{02} \cos(\nu - \nu_0) + \mathrm{tr} B_0 R(\mu)^{\mathrm{T}} A R(\nu)\}, \tag{26}$$

where B_0 is a 2×2 matrix of hyperparameter for A. Thus, the posterior for $(\mu, \nu, \kappa_1, \kappa_2, A)$ is proportional to

$$\{c(\kappa_1, \kappa_2, A)\}^{n+r} \exp\{\kappa_1 \kappa_1^* \cos(\mu - \mu_0^*) + \kappa_2 \kappa_2^* \cos(\nu - \nu_0^*) + \mathrm{tr} C R(\mu)^{\mathrm{T}} A R(\nu)\}, \tag{27}$$

where

$$C = B_0 + U, \quad U = \sum_{i=1}^{n} \ell(\phi_i)\ell(\theta_i)^{\mathrm{T}}, \tag{28}$$

$$\kappa_1^* \cos\mu_0^* = \kappa_{01} \cos\mu_0 + \sum \cos\theta_i, \quad \kappa_1^* \sin\mu_0^* = \kappa_{01} \sin\mu_0 + \sum \sin\theta_i,$$

$$\kappa_2^* \cos\nu_0^* = \kappa_{02} \cos\nu_0 + \sum \cos\phi_i, \quad \kappa_2^* \sin\nu_0^* = \kappa_{02} \sin\nu_0 + \sum \sin\phi_i.$$

Let us now discuss the normalizing constant for Equation (27). First, consider the univariate case with $A = 0$ and $\kappa_2 = 0$, so the joint density of μ and κ_1 from Equation (27) is proportional to

$$\{I_0(\kappa_1)\}^{n+r} \exp\{\kappa_1 \kappa_1^* \cos(\mu - \mu_0^*)\}.$$

Even in this case, the normalizing constant is intricate [8]. However, Damien and Walker [5] have bypassed this problem in simulating this distribution by constructing a Gibb sampler where one needs only to sample from uniform random variables. It will be worthwhile to extend their method in simulating Equation (27). In certain cases, we can use a Wishart distribution as prior for the precision parameters in view of the approximate normality [12], whereas the mean vector can have the prior distribution as given in Section 4; both priors being independent.

6. The multivariate von Mises distribution

We now consider the extension of the sine model to p dimensions proposed in [17, 24]. The pdf of $\boldsymbol{\theta}^{\mathrm{T}} = (\theta_1, \ldots, \theta_p)$ can be written as

$$\{T(\boldsymbol{\kappa}, \boldsymbol{\Lambda})\}^{-1} \exp\left\{\boldsymbol{\kappa}^{\mathrm{T}} c(\boldsymbol{\theta}, \boldsymbol{\mu}) + \frac{1}{2} s(\boldsymbol{\theta}, \boldsymbol{\mu})^{\mathrm{T}} \Lambda s(\boldsymbol{\theta}, \boldsymbol{\mu})\right\}, \tag{29}$$

where $-\pi < \theta_i \leq \pi$, $-\pi < \mu_i \leq \pi$, $\kappa_i \geq 0$, $-\infty < \lambda_{ij} < \infty$, $\boldsymbol{\kappa}^{\mathrm{T}} = (\kappa_1, \ldots, \kappa_p)$,

$$c(\boldsymbol{\theta}, \boldsymbol{\mu})^{\mathrm{T}} = (\cos(\theta_1 - \mu_1), \ldots, \cos(\theta_p - \mu_p)), s(\boldsymbol{\theta}, \boldsymbol{\mu})^{\mathrm{T}} = (\sin(\theta_1 - \mu_1), \ldots, \sin(\theta_p - \mu_p)),$$

and $(\boldsymbol{\Lambda})_{ij} = \lambda_{ij} = \lambda_{ji}$, $i \neq j$, $\lambda_{ii} = 0$, with $\{T(\boldsymbol{\kappa}, \boldsymbol{\Lambda})\}^{-1}$ a normalizing constant. We call this the multivariate von Mises distribution. Note that for $p = 1$, this is a univariate von Mises density and for $p = 2$, this density corresponds to the bivariate sine model. For $p > 2$, the normalizing constant is not known in any closed form. For large concentrations, we have ($\boldsymbol{\mu} = \mathbf{0}$ without any loss of generality)

$$\boldsymbol{\theta} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}^{-1}), \quad \text{where } (\boldsymbol{\Sigma}^{-1})_{ii} = \kappa_i, \ (\boldsymbol{\Sigma}^{-1})_{ij} = -\lambda_{ij}, \ i \neq j.$$

Using the conjugate prior for the mean vector $\boldsymbol{\mu}$ for given $\boldsymbol{\kappa}$ and $\boldsymbol{\Lambda}$, the posterior density of $\boldsymbol{\mu}$ is found to belong to an extension of Equation (32) given in [22]. For $\boldsymbol{\kappa}$, $\boldsymbol{\Lambda}$ we can use the independent prior distribution as Wishart for Γ, where $(\Gamma)_{ii} = \kappa_i$ and $(\Gamma)_{ij} = -\lambda_{ij}$.

Here, we have considered only the multivariate sine model but the conjugate prior for the multivariate cosine model as well as the general model given in [22] can be written down for the mean vector. For example, for the general multivariate von Mises model of [22], the conjugate prior has the same form as the original distribution. For a potential application to the RNA of this distribution, see [6].

7. Circular variables and protein bioinformatics

We describe the background to protein bioinformatics where new applications of circular variables have appeared recently. A protein is a sequence of amino acids which, when it folds, has a 3D structure. In this 3D structure, we can describe the protein by the coordinates of the atoms. The main part of the protein is its backbone (main chain) which can be considered as a repeated sequence of three atoms: nitrogen (N), carbon (C_α or C^α) and carbon (C'), so the backbone in 3D has coordinates of the sequences

$$\mathrm{N}_1 - \mathrm{C}_1^\alpha - \mathrm{C}_1' - \mathrm{N}_2 - \mathrm{C}_2^\alpha - \cdots - \mathrm{N}_r - \mathrm{C}_r^\alpha - \mathrm{C}_r',$$

where r denotes the 'length' of the backbone. In view of the physico-chemical properties, three dihedral angles ϕ, ψ, ω are sufficient to summarize a dipeptide of the backbone and Table 3 gives the definition of the three dihedral angles ϕ, ψ and ω. Recall that the dihedral angle θ of the four points $x_1, \ldots, x_4$ in 3D is defined as the angle between the two planes (x_1, x_2, x_3) and (x_2, x_3, x_4).

Due to physico-chemistry, the angle ω can only take two values 0 or π (mainly for the proline amino acid but otherwise there is only one value for ω) and therefore only the bivariate angles (ϕ, ψ) play a key role in understanding the backbone, particularly in predicting how the 1D sequences fold into 3D; this is one of the major unsolved problems of biology.

Ramachandran et al. [27] in their fundamental work plotted (ϕ, ψ) to summarize the recurring basic shapes in protein (α helices and β sheets) and found their empirical distribution. These scatter plots of (ϕ, ψ) are now known as Ramachandran plots [4]. The first statistical work in some sense to understand parametrically these plots through bivariate directional distributions

Table 3. Four atoms involved in the definition of dihedral angles (ϕ, ψ, ω) of the backbone.

θ	x_1	x_2	x_3	x_4
ϕ_i	C_i^α	N_i	C_i^α	C_i'
ψ_i	N_i	C_i^α	C_i'	N_{i+1}
ω_i	C_i^α	C_{i-1}'	N_i	C_i^α

(sine and cosine models) has been given by [23]. Boomsma et al. [3] have used these distributions in the local structure prediction through a hidden Markov model.

Recently, Lennox et al. [12] have provided a nonparametric Bayesian model. They use the sine distribution as the centring and component distribution of the Dirichlet process mixture model. In the Dirichlet process, the key steps include a conjugate prior of the sine model [12, Section 3]. Lennox et al. [13] have used again a Dirichlet process mixture for the protein structure prediction with a Dirichlet process which also contains the sine model as the centring distribution. For the nonparametric estimation of density, see [14, 30].

8. Discussion

There are still many challenging questions in this field. One of the major questions is which the BVM model is a 'true' extension of the univariate von Mises distribution. The question has been discussed in [10, 23]. It seems the sine model may be preferred since the multivariate extension is attractive. But note that among other negative properties of the sine model is that the conditional distribution of θ given $\theta - \phi$ is the full von Mises Bingham distribution, whereas for the cosine model, it is a von Mises distribution. Further, we have shown here that the posterior for the cosine model for the mean vector has less parameters (six) than for the sine model (eight parameters). On the other hand, the approximate bivariate normality of cosine may not be as attractive with the somewhat nonmonotone behaviour of limiting precision matrix. On the whole, the situation is somewhat similar to how approximate is the univariate von Mises distribution to the wrapped normal distribution. In practice, there is hardly a difference [26] and one should view the differences in the same light between the sine and the cosine model. Further, work is required to implement the full BVM though various sub-models have been well investigated. The multivariate sine model for high dimensional data also requires further study. Thus the field is still evolving!

Acknowledgements

I wish to express my thanks to David Dahl, Chris Fallaize, Peter Green, John Kent, Charles Taylor and Zhengzheng Zhang for their helpful comments.

References

[1] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions*, Dover Publication, New York, 1972.
[2] D.J. Best and N.I. Fisher, *Efficient simulation method of the von Mises distribution*, Appl. Stat. 24 (1979), pp. 152–157.
[3] W. Boomsma, K.V. Mardia, C.C. Taylor, J. Ferkinghoff-Borg, A. Krogh, and T. Hamelryck, *A generative, probabilistic model of local protein structure*, PNAS 105 (2008), pp. 8932–8937.
[4] C. Branden and J. Tooze, *Introduction to Protein Structure*, Garland, New York, 1998.
[5] P. Damien and S. Walker, *A full Bayesian analysis of circular data using the von Mises distribution*, Can. J. Stat. 27 (2008), pp. 291–298.

[6] J. Frellsen, I. Moltke, M. Thiim, K.V. Mardia, J. Ferkinghoff-Borg, and T. Hamelryck, *A probabilistic model of local RNA 3-D structure*, PLoS Comput. Biol. 5 (2009), pp. 1–11.
[7] P.J. Green and K.V. Mardia, *Bayesian alignment using hierarchical models, with applications in protein bioinformatics*, Biometrika 93 (2006), pp. 235–254.
[8] P. Guttorp and R.A. Lockhart, *Finding the location of a signal: A Bayesian analysis*, J. Am. Stat. Soc. 83 (1988), pp. 322–330.
[9] M. Habeck, *Generation of three dimensional random rotations in fitting and matching problems*, J. Comp. Statist. 24 (2009), pp. 719–731.
[10] J.T. Kent, K.V. Mardia, and C.C. Taylor, *Modelling Strategies for Bivariate Circular Data*, Proceedings of the Leeds Annual Statistical Research Conference, The Art and Science of Statistical Bioinformatics, Leeds University Press, Leeds, 2008.
[11] C.G. Khatri and K.V. Mardia, *The von Mises-Fisher matrix distribution in orientation statistic*, J. R. Stat. Soc. Ser. B 39 (1977), pp. 95–106.
[12] K.P. Lennox, D.B. Dahl, D.B. Vannucci, and J.W. Tsai, *Density estimation for protein conformation angles using a bivariate von Mises distribution and Bayesian nonparametrics*, J. Am. Stat. Soc. 104 (2009a), pp. 586–596, correction to appear.
[13] K.P. Lennox, D.B. Dahl, M. Vannucci, R. Day, and J.W.A. Tsai, *Dirichlet process mixture of hidden Markov models for protein structure prediction*, preprint (2009b), to appear in the Annals of Applied Statistics.
[14] S.C. Lovell, I.W. Davis, B. Arendall, III, et al., *Structure validation by C_α geometry: ϕ, ψ and C_β deviation*, Prot. Struct. Funct. Genet. 50 (2003), pp. 437–450.
[15] K.V. Mardia, *Statistics of directional data (with discussion)*, J. R. Stat. Soc. Ser. B 37 (1975a), pp. 349–393.
[16] K.V. Mardia, *Characterization of directional distributions*, in *Statistical Distributions in Scientific Work 3*, G.P. Patil, S. Kotz, and J.K. Ord, eds., D. Riedel Publishing Co., Dordrecht, Holland, 1975b, pp. 365–386.
[17] K.V. Mardia, *On some recent advancements in applied shape analysis and directional statistics*, in *Systems Biology & Statistical Bioinformatics*, S. Barber, P.D. Baxter, and K.V. Mardia, eds., Leeds University Press, Leeds, 2007, pp. 9–17.
[18] K.V. Mardia, *Holistic statistics and contemporary life sciences*, in *The Art and Science of Statistical Bioinformatics*, S. Barber, P.D. Baxter, A. Gusnanto, and K.V. Mardia, eds., Leeds University Press, Leeds, 2008, pp. 9–17.
[19] K.V. Mardia, *Statistical complexity in protein bioinformatics*, in *Statistical Tools for Challenges in Bioinformatics*, A. Gusnanto, K.V. Mardia and C.J. Fallaize, eds., Leeds University Press, Leeds, 2009, pp. 9–20.
[20] K.V. Mardia and S.A.M. El-Atoum, *Bayesian inference for the von Mises-Fisher distribution*, Biometrika 63 (1976), pp. 203–205.
[21] K.V. Mardia and P.E. Jupp, *Directional Statistics*, Wiley, Chichester, 2000.
[22] K.V. Mardia and V. Patrangenaru, *Directions and projective shapes*, Ann. Stat. 33 (2005), pp. 1666–1699.
[23] K.V. Mardia, C.C. Taylor, and G.K. Subramaniam, *Protein bioinformatics and mixtures of bivariate von Mises distributions for angular data*, Biometrics 63 (2007), pp. 505–512.
[24] K.V. Mardia, G. Hughes, C.C. Taylor, and H. Singh, *A multivariate von Mises distribution with applications to bioinformatics*, Can. J. Stat. 36 (2008), pp. 99–109.
[25] K.V. Mardia, J.T. Kent, G. Hughes, and C.C. Taylor, *Maximum likelihood estimation using composite likelihoods for closed exponential families*, Biometrika 96 (2009), pp. 975–982.
[26] A. Pewsey and M.C. Jones, *Discrimination between the von Misis and wrapped normal distributions: Just how big does the sample size have to be?*, J. Appl. Stat. 39 (2005), pp. 81–89.
[27] G.N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, *Stereochemistry of polypeptide chain configurations*, Mol. Biol. 7 (1963), pp. 95–99.
[28] L.P. Rivest, *A distribution for dependent unit vectors*, Commun. Stat. A 17 (1988), pp. 461–483.
[29] H. Singh, V. Hnizdo, and E. Demchuk, *Probabilistic model for two dependent circular variables*, Biometrika 89 (2002), pp. 719–723.
[30] C.C. Taylor, *Automatic bandwidth selection for circular density estimation*, Comput. Stat. Data Anal. 52 (2008), pp. 3493–3500.

Appendix 1. Normalizing constant for the sine model

Here we assume $\mu = \nu = 0$ without any loss of generality. Using the marginal pdf of θ from Equation (7), we have

$$c^{-1} = 2\pi \int_{-\pi}^{\pi} I_0\{a(\theta)\} \mathrm{e}^{\kappa \cos\theta} \, \mathrm{d}\theta,$$

where $a(\theta) = (\kappa_2^2 + \lambda^2 \sin^2\theta)^{1/2}$. Now, using the series expansion of $I_0(\cdot)$, we find that

$$c^{-1} = 2\pi \sum_{m=0}^{\infty} \int_{-\pi}^{\pi} \frac{(\kappa_2^2 + \lambda^2 \sin^2\theta)^m}{2^{2m}(m!)^2} \mathrm{e}^{\kappa \cos\theta}\, \mathrm{d}\theta.$$

Expanding $(\kappa_2^2 + \lambda^2 \sin^2\theta)^m$ by the binomial expansion, we have

$$c^{-1} = 2\pi \sum_{m=0}^{\infty} \sum_{r=0}^{m} \frac{\lambda^{2r}\kappa_2^{2(m-r)}}{2^{2m} m! r! (m-r)!} \int_{-\pi}^{\pi} \sin^{2r}\theta \mathrm{e}^{\kappa_1 \cos\theta}\, \mathrm{d}\theta.$$

We now change the order of the summation and note that $\sum_{m=0}^{\infty}\sum_{r=0}^{m}$ is equal to $\sum_{r=0}^{\infty}\sum_{m=r}^{\infty}$, and then put $k = m - r$ in the first summation $m = r, \ldots, \infty$, which leads to

$$c^{-1} = 2\pi \sum_{r=0}^{\infty} \frac{1}{r!} \left(\frac{\lambda^2}{2\kappa_2}\right)^r \left(\int_{-\pi}^{+\pi} \sin^{2r}\theta \mathrm{e}^{\kappa\cos\theta}\, \mathrm{d}\theta\right) \sum_{k=0}^{\infty} \frac{\kappa^{2k+r}}{k!(r+k)! 2^{2k+r}}.$$

The summation over k is simply $I_r(\kappa_2)$, whereas the integral is $2\pi(2r!)/\{(2\kappa_1)^r r!\} I_r(\kappa_1)$ by [1, Equation (9.6.18)]. Hence we get the normalizing constant for the sine model given in Table 2.

Appendix 2. Algebraic identities

The following results are used in Section 4 and can be easily verified.

(1) $\ell(\theta)^{\mathrm{T}} A \ell(\phi) = \mathrm{tr} A\{\ell(\phi)\ell(\theta)^{\mathrm{T}}\} = a_{11}\cos\theta\cos\phi + a_{12}\cos\theta\sin\phi + a_{21}\sin\theta\cos\phi + a_{22}\cos\theta\sin\phi$.

(2) $R(\mu)R(\theta)^{\mathrm{T}} = R(\mu - \theta) = R(\theta - \mu)^{\mathrm{T}}$, $R(\mu)^{\mathrm{T}} R(\theta) = R(\theta - \mu)$.

(3) $R(\mu)\ell(\theta) = \ell(\theta - \mu)$, $\ell(\theta - \mu)^{\mathrm{T}} = \ell(\theta)^{\mathrm{T}} R(\mu)^{\mathrm{T}}$.

(4) Let

$$I_{-1} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Then

$$I_{-1} R(\mu)\ell(\theta) = \ell(\mu - \theta).$$

(5) Let

$$R^{-}(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}.$$

Then

$$|R^{-}(\theta)| = -1,\ I_{-1}R(\theta) = R(\theta)^{\mathrm{T}} I_{-1} = R^{-}(\theta);\ R(\theta) = I_{-1}R^{-}(\theta).$$

(6) $A^{\#} = I_{-1} A I_{-1} = \begin{pmatrix} a_{11} & -a_{12} \\ -a_{21} & a_{22} \end{pmatrix}$.