Multivariate Normal Distributions Parametrized as a Riemannian Symmetric Space

Miroslav Lovrić

Dept. of Mathematics and Statistics, McMaster Univ., Hamilton, Ontario, Canada L8S 4K1

e-mail: lovric@icarus.math.mcmaster.ca

Maung Min-Oo

Dept. of Mathematics and Statistics, McMaster Univ., Hamilton, Ontario, Canada L8S 4K1

e-mail: minoo@mcmaster.ca

Ernst A. Ruh

Dept. of Mathematics, University of Fribourg, Pérolles, CH-1700 Fribourg, Switzerland

e-mail: ruh@unifr.ch

Abstract: The construction of a distance function between probability distributions is of importance in mathematical statistics and its applications. Distance function based on the Fisher information metric has been studied by a number of statisticians, especially in the case of the multivariate normal distribution (Gaussian) on $\mathbb{R}^n$. It turns out that, except in the case $n = 1$, where the Fisher metric describes the hyperbolic plane, it is difficult to obtain an exact formula for the distance function (although this can be achieved for special families with fixed mean or fixed covariance). We propose to study a slightly different metric on the space of multivariate normal distributions on $\mathbb{R}^n$. Our metric is based on the fundamental idea of parametrizing this space as the Riemannian symmetric space $SL(n+1)/SO(n+1)$. Symmetric spaces are well understood in Riemannian geometry, allowing us to compute distance functions and other relevant geometric data.

The first author was supported in part by Canadian NSERC grant OGP-0194144.
The second author was supported in part by Canadian NSERC grant OPG-7873.
The third author was supported in part by Swiss NSF grant # 2000-042054.94/1.
1991 *Mathematics Subject Classification.* Primary 53C35; Secondary 62H30, 62B10.

1. Introduction

The construction of a distance function between probability distributions is of importance in mathematical statistics. Rao [14] first proposed a distance function based on the Fisher information as a Riemannian metric. This metric has been studied by a number of statisticians (see [2], [5], [6] and [15]) in the case of the multivariate normal distribution (Gaussian) on $\mathbb{R}^n$. It turns out that except in the case of $n = 1$, where the Fisher metric describes the hyperbolic plane, it is difficult to obtain an exact formula for the distance function of the Fisher metric, although this can be achieved for special families with fixed mean or fixed covariance, as in [2], [15].

We propose to study a slightly different metric on the space of multivariate normal distributions on $\mathbb{R}^n$. Our metric is based on the fundamental idea of parametrizing this space as the Riemannian symmetric space $SL(n+1)/SO(n+1)$. Symmetric spaces are well understood in Riemannian geometry, allowing us to compute distance functions and other relevant geometric data. We refer to [10], which is a standard introductory textbook on symmetric spaces.

The importance of a convenient metric for probabilistic models is well documented in the literature; compare [6], [15]. The symmetric metric that we introduce does not differ much from the Fisher metric (computed, for example, in [15]), although the extension of the symmetry group to $SL(n+1)$ forces a crucial additional term in one direction. It is, in effect, very close to the metric described in [6] which is based on the Siegel group. The difference here is due to the fact that the embedding of normal distributions into the Siegel space used in [6] is not totally geodesic.

We would like to thank Peter March for explaining the Fisher information metric to us.

2. The $SL(n+1)$–action

Let $\mathcal{N} = \{\gamma|dx|\}$ be the space of normal (Gaussian) distributions on $\mathbb{R}^n$, where $|dx|$ is the Lebesgue measure on $\mathbb{R}^n$.

The affine group of $\mathbb{R}^n$,

$$Aff(n) = \{\Phi_{a,b} : x \longmapsto ax + b \mid a \in GL(n),\ b \in \mathbb{R}^n\},$$

acts transitively on $\mathcal{N}$ by

$$\gamma|dx| \longmapsto \left(\Phi_{a,b}^{-1}\right)^* (\gamma|dx|).$$

This is a left action, with the projection

$$\begin{aligned} \pi : Aff(n) &\longrightarrow \mathcal{N}, \\ \Phi_{a,b} &\longmapsto \left(\Phi_{a,b}^{-1}\right)^* (\gamma_0|dx|), \end{aligned} \tag{2.1}$$

where $\gamma_0|dx| = (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2}|x|^2}|dx|$ is the standard Gaussian distribution on $\mathbb{R}^n$. Moreover, $\pi^{-1}(\gamma_0|dx|) = O(n)$.

The exact formula for the action is

$$\left(\Phi_{a,b}^{-1}\right)^* (\gamma_0|dx|) = (2\pi)^{-\frac{n}{2}} (\det a)^{-1} e^{-\frac{1}{2}|a^{-1}(x-b)|^2}|dx|. \tag{2.2}$$

We can restrict the action to the identity component $Aff^+(n) = \{\Phi_{a,b} \mid \det a > 0\}$. This action is still transitive on $\mathcal{N}$, and so $\mathcal{N} = Aff^+(n)/SO(n)$.

To extend this action to the larger group $SL(n+1)$, we use the following embedding of $Aff^+(n)$ as a subgroup of $SL(n+1)$:

$$\begin{aligned} j : Aff^+(n) &\hookrightarrow SL(n+1), \\ (a,b) &\longmapsto (\det a)^{-\frac{1}{n+1}} \begin{bmatrix} a & b \\ 0 & 1 \end{bmatrix} \end{aligned} \tag{2.3}$$

The simple Lie group $SL(n+1)$ acts canonically on the space $\tilde{\mathcal{S}}_+$ of all positive definite symmetric $(n+1) \times (n+1)$–matrices with determinant 1 via

$$\tilde{s} \longmapsto \tilde{a}\tilde{s}\tilde{a}^t, \qquad \text{for} \quad \tilde{s} \in \tilde{\mathcal{S}}_+ \quad \text{and} \quad \tilde{a} \in SL(n+1), \tag{2.4}$$

where $\tilde{a}^t$ denotes the transpose of $\tilde{a}$. This is a left action which represents $\tilde{\mathcal{S}}_+$ as the Riemannian symmetric space $SL(n+1)/SO(n+1)$ with $SO(n+1)$–principal bundle

$$\tilde{\pi} : SL(n+1) \longrightarrow \tilde{\mathcal{S}}_+, \quad \tilde{a} \mapsto \tilde{a}\tilde{a}^t.$$

The restriction of $\tilde{\pi}$ to the subgroup $\mathit{Aff}^+(n)$ is given by

$$\begin{aligned} \mathit{Aff}^+(n) &\longrightarrow \tilde{\mathcal{S}}_+, \\ (a,b) &\longmapsto (\det a)^{-\frac{2}{n+1}} \begin{bmatrix} aa^t + bb^t & b \\ b^t & 1 \end{bmatrix}, \end{aligned} \tag{2.5}$$

where bb^t denotes the symmetric matrix $[b_i b_j]_{1 \le i,j \le n}$.

This is still a surjective mapping with $\tilde{\pi}^{-1}(Id) = SO(n)$. The standard embedding

$$\begin{aligned} SO(n) &\hookrightarrow SO(n+1), \\ a &\longmapsto \begin{bmatrix} a & 0 \\ 0 & 1 \end{bmatrix}. \end{aligned}$$

can then be extended by the following commutative diagram:

$$\begin{array}{ccc} SO(n) & \hookrightarrow & SO(n+1) \\ \downarrow & & \downarrow \\ \mathit{Aff}^+(n) & \xrightarrow{j} & SL(n+1) \\ \pi \downarrow & \searrow & \downarrow \tilde{\pi} \\ \mathcal{N} & \cong & \tilde{\mathcal{S}}_+ \end{array} \tag{2.6}$$

This gives a canonical identification of $\mathcal{N}$, the space of normal Gaussians on $\mathbb{R}^n$, with the symmetric space $\tilde{\mathcal{S}}_+ = SL(n+1)/SO(n+1)$, which is equivariant with respect to the natural actions of $\mathit{Aff}^+(n)$ on $\mathcal{N}$ and $SL(n+1)$ on $\tilde{\mathcal{S}}_+$ via the embedding j. Both spaces, of course, have the same dimension, equal to $\frac{1}{2}n(n+3)$.

More explicitly, we identify

$$\begin{aligned} \mathcal{N} &\longrightarrow \tilde{\mathcal{S}}_+, \\ \left(\Phi_{a,b}^{-1}\right)^* (\gamma_0 |dx|) &\longmapsto (\det a)^{-\frac{2}{n+1}} \begin{bmatrix} aa^t + bb^t & b \\ b^t & 1 \end{bmatrix}. \end{aligned} \tag{2.7}$$

The normal distribution,

$$\frac{1}{(2\pi)^{n/2} \det \sigma} e^{-\frac{1}{2}[(x-\mu)^t \sigma^{-2} (x-\mu)]},$$

is therefore identified with the positive definite symmetric $(n+1)\times(n+1)$–matrix

$$(\det\sigma)^{-\frac{2}{n+1}}\begin{bmatrix}\sigma^2+\mu\mu^t & \mu\\ \mu^t & 1\end{bmatrix}.$$

We now compute the differential of this identification at the standard normal distribution $\gamma_0|dx|$, which is mapped to $\tilde{I}$, the identity matrix in $\tilde{\mathcal{S}}_+$.

The tangent space of $\mathcal{N}$ at $\gamma_0|dx|$ is identified with

$$\{(A,B)\mid A \text{ symmetric } n\times n\text{–matrix}, B\in\mathbb{R}^n\},$$

given by the splitting of the Lie algebra $\mathit{aff}(n)=\{(A,B)\mid A\in gl(n), B\in\mathbb{R}^n\}$ of the affine group as $\mathit{aff}(n)=\mathfrak{o}(n)\oplus T_{\gamma_0|dx|}\mathcal{N}$.

On the other hand, the tangent space at $\tilde{I}$ of the symmetric space $\tilde{\mathcal{S}}_+$ is identified with $\mathfrak{s}_0(n+1)=\{$symmetric $(n+1)\times(n+1)$–matrices of trace zero$\}$ via the Cartan decomposition

$$sl(n+1)=\mathfrak{o}(n+1)\oplus\mathfrak{s}_0(n+1).$$

The differential of the mapping $\mathcal{N}\xrightarrow{\sim}\tilde{\mathcal{S}}_+$ (given in (2.7)) at the identity is then computed to be

$$\begin{aligned} T_{\gamma_O|dx|}\mathcal{N} &\xrightarrow{\sim} T_{\tilde{I}}\tilde{\mathcal{S}}_+=\mathfrak{s}_0(n+1)\\ (A,B) &\longmapsto 2\begin{bmatrix}A-\frac{1}{n+1}\mathrm{tr}(A)\cdot I_n & \frac{1}{2}B\\ \frac{1}{2}B^t & -\frac{1}{n+1}\ \mathrm{tr}(A)\end{bmatrix}.\end{aligned} \tag{2.8}$$

3. Symmetric Riemannian Metric on $\mathcal{N}$

With the explicit identification of $\mathcal{N}$ with $SL(n+1)/SO(n+1)$ given in the previous section, we will now describe the natural symmetric Riemannian metric and the induced distance function.

As is well known in the theory of symmetric spaces, the natural $SL(n+1)$–invariant metric on $\tilde{\mathcal{S}}_+$ is given (up to a positive constant) by restricting the Killing form of the simple Lie algebra $sl(n+1)$ to the subspace $\mathfrak{s}_0(n+1)$ under the Cartan decomposition $sl(n+1)=\mathfrak{o}(n+1)\oplus\mathfrak{s}_0(n+1)$; i.e.,

$$\langle\tilde{s}_1,\tilde{s}_2\rangle=\frac{1}{4}\mathrm{tr}(\tilde{s}_1\tilde{s}_2),\qquad \tilde{s}_1,\tilde{s}_2\in\mathfrak{s}_0(n+1)=T_{\tilde{I}}(\tilde{\mathcal{S}}_+). \tag{3.1}$$

From the identification (2.8) of the tangent spaces $T_{\gamma_0|dx|}\mathcal{N} \cong \mathfrak{s}_0(n+1)$ it follows that the symmetric metric on

$$T_{\gamma_0|dx|}\mathcal{N} \cong \{(A,B) \mid A \text{ symmetric } n \times n\text{–matrix and } B \in \mathbb{R}^n\}$$

is given by

$$\langle (A_1,B_1),(A_2,B_2)\rangle = \mathrm{tr}(A_1A_2) - \frac{1}{n+1}(\mathrm{tr}A_1)(\mathrm{tr}A_2) + \frac{1}{2}\langle B_1,B_2\rangle. \tag{3.2}$$

The Fisher information metric, computed at the standard normal distribution, in contrast, is given by (compare [15])

$$\langle (A_1,B_1),(A_2,B_2)\rangle = 2\ \mathrm{tr}(A_1A_2) + \langle B_1,B_2\rangle. \tag{3.3}$$

Up to scaling, the two metrics (3.2) and (3.3) are the same in the direction of trace zero matrices.

The metric in (3.2) also differs from the metric induced by the Siegel metric on the space of all symmetric, positive $(n+1)\times(n+1)$–matrices used in [6]. In terms of (2.7) and (2.8), Siegel metric is given by

$$\langle (A_1,B_1),(A_2,B_2)\rangle = \frac{1}{4}\left(\mathrm{tr}(A_1A_2) - \frac{1}{(n+1)}(\mathrm{tr}A_1)(\mathrm{tr}A_2)\right) + \langle B_1,B_2\rangle. \tag{3.4}$$

Since $SL(n+1)/SO(n+1)$ is an irreducible symmetric space, the symmetric metric is determined uniquely up to a positive constant factor.

Let (a,b) denote the point $\left(\Phi_{a,b}^{-1}\right)^*(\gamma_0|dx|) \in \mathcal{N}$. The metric g defined in (3.2) can be written as

$$\begin{aligned}
&g((A_1,0),(A_2,0)) = \mathrm{tr}(a^{-1}A_1a^{-1}A_2) - \frac{1}{n+1}\mathrm{tr}(a^{-1}A_1)\mathrm{tr}(a^{-1}A_2),\\
&g((A,0),(0,B)) = 0,\\
&g((0,B_1),(0,B_2)) = \frac{1}{2}\mathrm{tr}(B_1^t a^{-1}B_2),
\end{aligned} \tag{3.5}$$

where $(A,0)$, $(A_1,0)$, $(A_2,0)$, $(0,B)$, $(0,B_1)$ and $(0,B_2)$ are in $T_{(a,b)}\mathcal{N}$.

Usually, one expresses the above metric in terms of multivariate normal parameters. Replacing a by Σ, b by μ, and denoting by X_1 and X_2 vector fields in the Σ-direction and by η_1 and η_2 vector fields in the μ-direction, we write (3.5) as

$$\begin{aligned} g(X_1, X_2) &= \mathrm{tr}(\Sigma^{-1}X_1\Sigma^{-1}X_2) - \frac{1}{n+1}\mathrm{tr}(\Sigma^{-1}X_1)\mathrm{tr}(\Sigma^{-1}X_2), \\ g(X, \eta) &= 0, \\ g(\eta_1, \eta_2) &= \frac{1}{2}\mathrm{tr}(\eta_1^t\Sigma^{-1}\eta_2). \end{aligned} \tag{3.6}$$

Equivalently, the quadratic fundamental form (3.5) or (3.6) can be expressed as

$$ds^2 = \mathrm{tr}(\Sigma^{-1}d\Sigma\Sigma^{-1}d\Sigma) - \frac{1}{n+1}(\mathrm{tr}(\Sigma^{-1}d\Sigma))^2 + \frac{1}{2}d\mu^t\Sigma^{-1}d\mu. \tag{3.7}$$

The Levi-Civita connection of g is computed to be

$$\begin{aligned} \nabla_\eta X &= \nabla_X\eta = -\frac{1}{2}X\Sigma^{-1}\eta, \\ \nabla_{X_1}X_2 &= -\frac{1}{2}(X_1\Sigma^{-1}X_2 + X_2\Sigma^{-1}X_1), \\ \nabla_{\eta_1}\eta_2 &= \frac{1}{8}\left(\eta_1\eta_2^t + \eta_2\eta_1^t + \mathrm{tr}(\Sigma^{-1}(\eta_1\eta_2^t + \eta_2\eta_1^t))\Sigma\right). \end{aligned} \tag{3.8}$$

Finally, the non-zero components of the Riemann curvature tensor of g are

$$\begin{aligned} R(X_1, X_2, \eta_1, \eta_2) &= \frac{1}{8}\mathrm{tr}(\eta_1^t\Sigma^{-1}X_1\Sigma^{-1}X_2\Sigma^{-1}\eta_2 - \eta_1^t\Sigma^{-1}X_2\Sigma^{-1}X_1\Sigma^{-1}\eta_2), \\ R(X_1, \eta_1, X_2, \eta_2) &= \frac{1}{8}\mathrm{tr}(\eta_1^t\Sigma^{-1}X_1\Sigma^{-1}X_2\Sigma^{-1}\eta_2), \\ R(X_1, X_2, X_3, X_4) &= \frac{1}{2}\mathrm{tr}(X_2\Sigma^{-1}X_1\Sigma^{-1}X_3\Sigma^{-1}X_4 - X_1\Sigma^{-1}X_2\Sigma^{-1}X_3\Sigma^{-1}X_4), \\ R(\eta_1, \eta_2, \eta_3, \eta_4) &= -\frac{1}{32}\left(\mathrm{tr}(\eta_1^t\Sigma^{-1}\eta_4)(\eta_2^t\Sigma^{-1}\eta_3) - \mathrm{tr}(\eta_1^t\Sigma^{-1}\eta_3)(\eta_2^t\Sigma^{-1}\eta_4)\right), \end{aligned} \tag{3.9}$$

where X_i, $i = 1, \dots, 4$, are vector fields in the Σ-direction and η_i, $i = 1, \dots, 4$, are vector fields in the μ-direction.

In the case $n = 1$ (the parameters are σ and $b = \mu$), the only non-zero curvature term is

$$R(\dot\sigma, \dot\mu, \dot\sigma, \dot\mu) = \frac{1}{8}\frac{\dot\sigma^2\dot\mu^2}{\sigma^3},$$

where $\dot\sigma$ is the tangent vector in the σ-direction and $\dot\mu$ is the tangent vector in the μ-direction.

Thus, the sectional curvature is

$$K = -\frac{R(\dot\sigma, \dot\mu, \dot\sigma, \dot\mu)}{g(\dot\sigma, \dot\sigma)g(\dot\mu, \dot\mu) - g(\dot\sigma, \dot\mu)^2} = -\frac{1}{2}.$$

We now compute the induced distance function on the symmetric space $\tilde{\mathcal{S}}_+ = SL(n+1)/SO(n+1)$. As is standard in the theory of symmetric spaces, we reduce the problem to computing the distance function between the "identity" and an element in the "diagonal" maximally flat totally geodesic subspace. We give here the details:

Step 1. Write $\tilde{s}_1 = \tilde{a}_1\tilde{a}_1^t = \tilde{\pi}(\tilde{a}_1)$, with $\tilde{a}_1 \in SL(n+1)$. The element $\tilde{a}_1$ is unique up to right multiplication by an element of the isotropy group $SO(n+1)$. $\tilde{a}_1$ can be thought of as a square root of $\tilde{s}_1$.

Step 2. Let $\tilde{s} = \tilde{a}_1^{-1}\tilde{s}_2(\tilde{a}_1^{-1})^t$. Then $d(\tilde{s}_1, \tilde{s}_2) = d(\tilde{I}, \tilde{s})$, since the distance function is invariant under the action of $SL(n+1)$. Notice that $\tilde{s}$ is well defined only up to the action of the isotropy group $SO(n+1)$, and that it is not the same as $\tilde{s}_1^{-1}\tilde{s}_2$. In fact, $\tilde{s}_1^{-1}\tilde{s}_2$ is, in general, not even symmetric (although one can speak about generalized eigenvalues by solving the equation $\tilde{s}_2(v) = \lambda\tilde{s}_1(v)$).

Step 3. Conjugate $\tilde{s}$ by an element $\tilde{u} \in SO(n+1)$ so that

$$\tilde{u}^{-1}\tilde{s}(\tilde{u}^{-1})^t = \tilde{u}^t\tilde{s}\tilde{u} = \tilde{\delta}$$

is a diagonal matrix. This is equivalent to $\tilde{\delta}$ being in the maximally flat totally geodesic subspace.

Step 4. Take the logarithm of the positive definite diagonal matrix $\tilde{\delta}$; i.e., write

$$\tilde{\delta} = \begin{bmatrix} e^{\mu_1} & & 0 \\ & \ddots & \\ 0 & & e^{\mu_{n+1}} \end{bmatrix} = \exp \begin{bmatrix} \mu_1 & & 0 \\ & \ddots & \\ 0 & & \mu_{n+1} \end{bmatrix},$$

where $\sum_{i=1}^{n+1} \mu_i = 0$.

Step 5. Finally,

$$d(\tilde{s}_1, \tilde{s}_2) = d(\tilde{I}, \tilde{s}) = d(\tilde{I}, \tilde{\delta}) = \left(\sum_{i=1}^{n+1} \mu_i^2 \right)^{1/2}$$

is the distance with respect to the symmetric metric on $\tilde{\mathcal{S}}_+ \cong \mathcal{N}$.

Remark. Another way of expressing the above computation is as follows:

Write $\tilde{s}_i = \tilde{a}_i \tilde{a}_i^t$ with $\tilde{a}_i \in SL(n+1)$, $i = 1, 2$, and let $\tilde{a} = \tilde{a}_1^{-1} \tilde{a}_2$. Then

$$d(\tilde{s}_1, \tilde{s}_2) = \left(\sum_{i=1}^{n+1} \mu_i^2 \right)^{1/2},$$

where $\{e^{\mu_1}, \ldots, e^{\mu_{n+1}}\}$ are the eigenvalues of the positive definite symmetric matrix $\tilde{a}\tilde{a}^t$. As mentioned above, these eigenvalues can be thought of as generalized eigenvalues of $\tilde{s}_1$ and $\tilde{s}_2$.

We now compute the distance between two normal distributions $\gamma_1 |dx|$ and $\gamma_2 |dx|$ with the *same* variance $\Sigma = aa^t$ but with different means $\mu_i = b_i$, $i = 1, 2$. Consider

$$\gamma_i |dx| = \frac{(\det a)^{-1}}{(2\pi)^{n/2}} e^{-\frac{1}{2}|a^{-1}(x - b_i)|^2} |dx|,$$

$i = 1, 2$. Let

$$\tilde{a}_i = (\det a)^{-\frac{1}{n+1}} \begin{bmatrix} a & b_i \\ 0 & 1 \end{bmatrix} \in SL(n+1).$$

Then $\tilde{s} = (\tilde{a}_1^{-1} \tilde{a}_2)(\tilde{a}_1^{-1} \tilde{a}_2)^t \in \mathcal{S}_+$ is given by

$$\tilde{s} = \begin{bmatrix} I_n + bb^t & b \\ b^t & 1 \end{bmatrix},$$

where $b = a^{-1}(b_2 - b_1)$. The $(n+1)$ positive eigenvalues of $\tilde{s}$ are

$$\left\{ 1, \ldots, 1, 1 + \frac{|b|^2}{2} + \frac{|b|\sqrt{|b|^2 + 4}}{2}, 1 + \frac{|b|^2}{2} - \frac{|b|\sqrt{|b|^2 + 4}}{2} \right\},$$

where the eigenvalue 1 occurs $(n-1)$ times. Hence,

$$d(\gamma_1 |dx|, \gamma_2 |dx|)^2 = 2 \left(\log \left(1 + \frac{|b|^2}{2} + \frac{|b|\sqrt{|b|^2 + 4}}{2} \right) \right)^2 = 2 \left(\operatorname{argcosh}(1 + |b|^2/2) \right)^2.$$

This expression is different from the usual formulas found in the literature. One usually has a Euclidean distance (see [2])

$$d^2 = |b|^2 = |a^{-1}(b_2 - b_1)|^2.$$

Our distance is hyperbolic.

Next, we compute the distance between two normal distributions with the same mean $\mu = b$, but with different variances $\Sigma_i = a_i a_i^t$, $i = 1, 2$.

In this case, $\tilde{s}$ is given by

$$\tilde{s} = (\det a)^{-\frac{2}{n+1}} \begin{bmatrix} aa^t & 0 \\ 0 & 1 \end{bmatrix},$$

where $a = a_1^{-1} a_2$.

We note that $aa^t \neq \sigma_1^{-1}\sigma_2$, unless $a_1 a_2 = a_2 a_1$.

The distance is given by

$$d^2 = \sum_{i=1}^{n} \mu_i^2 - \frac{1}{(n+1)^2} \left(\sum_{i=1}^{n} \mu_i \right)^2 ,$$

where $e^{\mu_1}, \ldots, e^{\mu_n}$ are the eigenvalues of the positive definite symmetric $n \times n$–matrix $(a_1^{-1}a_2)(a_1^{-1}a_2)^t = aa^t$. These eigenvalues are the same as the generalized eigenvalues of the pair of symmetric matrices Σ_1, Σ_2.

Remark. We have computed the distance function of the ambient space $\tilde{\mathcal{S}}_+$, restricted to families of distributions with the same mean or the same variance, which is of course not the same as the distance function computed with respect to the *induced* Riemannian metric on the corresponding submanifolds. In fact, for a fixed variance, the induced Riemannian metric is flat. However, the corresponding submanifold is not totally geodesic.

4. Expectation and Center of Mass

The expected value of a random variable X with values in a vector space $\mathbb{R}^d$,

$$X : (\Omega, d\mu) \longrightarrow \mathbb{R}^d, \qquad \omega \longmapsto X(\omega),$$

where Ω is a probability space with a measure $d\mu$ and σ–algebra $\mathcal{B}$ of subsets of Ω, is defined by the equation

$$\widehat{X} = E(X) = \int_\Omega X(\omega) d\mu \in \mathbb{R}^d. \tag{4.1}$$

It is characterized by the property that the mean square deviation

$$V : \mathbb{R}^d \longrightarrow \mathbb{R},$$

given by

$$y \longmapsto \int_\Omega |y - X(\omega)|^2 d\mu, \tag{4.2}$$

achieves its unique minimum at $\widehat{X}$. This implies, by differentiation,

$$\int_\Omega \left(\widehat{X} - X(\omega) \right) d\mu = 0,$$

which is the defining property of $\widehat{X} = E(X)$.

The generalization of this definition to the case of a random variable with values in a Riemannian manifold is known as the center of mass construction (see [9], [11] and [13]). This construction has analogous properties to the Euclidean center of mass and is well–defined and unique provided that the random variable takes values in a convex region; in particular, it can be defined for a random variable with values in a simply–connected manifold of non–positive sectional curvature. This fact was exploited by E. Cartan in his work on fixed points of isometries of such manifolds. All Riemannian symmetric spaces of non–compact type, such as $SL(n+1)/SO(n+1)$, are simply connected and have non–positive sectional curvature, so the center of mass construction is defined globally on such spaces.

In this section we review the center of mass construction. It is basically characterized by the same property as in Euclidean space except that the expression $|y - X(\omega)|^2$ is replaced by $d\big(y, X(\omega)\big)^2$, where d is the distance function of the Riemannian metric. So, $\widehat{X}$ is still defined to be the point in the Reimannian manifold where the function

$$y \longmapsto \int_{\Omega} d\big(y, X(\omega)\big)^2 d\mu \tag{4.3}$$

assumes its minimum value.

In order to interpret the defining equation (4.1) suitably on a Riemannian manifold, we need the concept of the exponential map

$$\exp_x : T_x M \longrightarrow M,$$

where x is a point on a Riemannian manifold M and $T_x M$ is the tangent space at x. For $v \in T_x M$, $\exp_x(v)$ is defined to be the point $\gamma_v(1)$, where γ_v is the unique geodesic starting at $\gamma_v(0) = x$ with initial velocity $\dot{\gamma}_v(0) = v$. For convenience, we are assuming that the Riemannian manifold is complete, so that $\exp_x$ is defined for all $v \in T_x M$ and all $x \in M$.

The main technical tool that is now needed is an estimate for the differential of the exponential map. At the origin in $T_x M$, $\mathrm{dexp}_x : T_x M \to T_x M$ is easily seen to be the identity map. Estimates for dexp_x along geodesics are obtained by studying Jacobi fields and Sturm–Liouville–type comparison theorems. One should consult textbooks on Riemannian geometry, compare [7], [3], under the headings: "Jacobi fields, second variation formula, Rauch comparison theorem".

In general, the uniqueness and existence of the center of mass requires injectivity and convexity properties of the exponential map, but the required properties are satisfied for all simply–connected Riemannian manifolds with non–positive sectional curvature (sometimes called Hadamard manifolds in the literature). We refer to [3] for the geometry of such manifolds. In particular, for symmetric spaces of non–compact type the exponential map is a global diffeomorphism and can be expressed explicitly in terms of the Lie groups involved. We will denote the inverse function of $\exp_x$ by

$$\log_x : M \longrightarrow T_x M.$$

(For a construction in general case, see [13].) Let $f : (\Omega, d\mu) \to (M, g)$, $\omega \mapsto f(\omega)$, be a manifold–valued random variable, i.e., a measurable map from a probability space to a Hadamard manifold. The *center of mass* of f, or the expectation value of the random variable f is the unique point $\widehat{f} \in M$ such that

$$\int_\Omega \log_{\widehat{f}}\big(f(\omega)\big) d\mu = 0. \tag{4.4}$$

This is the appropriate Riemannian generalization of the Euclidean definition (4.1); $\widehat{f}$ can also be characterized as the unique point where the mean square deviation

$$V : M \longrightarrow \mathbb{R},$$

given by

$$y \longmapsto \int_\Omega d\big(y, f(\omega)\big)^2 d\mu, \tag{4.5}$$

achieves its minimum value.

In accordance with the notation in probability and statistics we will denote the center of mass by $\widehat{f} = E(f)$. Since the definition is geometric, the center of mass satisfies the following invariance property under isometries of the target manifold:

Proposition. If $\sigma : M \to M$ is an isometry of a Riemannian manifold M, then

$$E(\sigma \circ f) = \sigma\big(E(f)\big)$$

for any random variable $f : (\Omega, d\mu) \to (M, g)$.

We can also generalize the definition of the mean square deviation of a random variable to the Riemannian case:

Definition. Let $f : (\Omega, d\mu) \to (M, g)$ be a random variable. Then

$$V(f) = \int_\Omega d\big(\widehat{f}, f(\omega)\big)^2 d\mu = \int_\Omega \Big|\log_{\widehat{f}}\big(f(\omega)\big)\Big|^2 d\mu.$$

5. Concluding Remarks

With the extension of the action of the affine group to the special linear group $SL(n+1)$ on the space of multivariate normal distributions in $\mathbb{R}^n$, this set of distributions inherits the structure of a Riemannian space of non–compact type. This point of view plays a useful role in statistics as well, compare [4]. Another important aspect of the geometry of Riemannian symmetric spaces is the role of the ideal boundary, i.e., the points at infinity of a symmetric space. The concept of the center of mass can be extended to measures on the ideal boundary, compare [8]. Perhaps, the points at infinity could be interpreted as the sample of a distribution. In that case, the center of mass, or expectation value, could serve as an estimator. There are other interesting families of statistical distributions (see [12]) which can also be studied using similar geometric methods. We plan to investigate these problems in the future.

6. References

1. S. Amari, "Differential-geometric methods in statistics", (Lecture Notes in Statistics, 28) Springer-Verlag, New York, 1985.

2. C. Atkinson, and A.F.S. Mitchell, Rao's distance measure, *Sankhaya Ser.* **A 43** (1981), 345–365.

3. W. Ballmann, M. Gromov, and V. Schroeder, "Manifolds of nonpositive curvature", Birkhäuser, Boston, 1985.

4. O.E. Barndorff-Nielson, “Parametric statistical models and likelihood”, (Lecture Notes in Statistics, 50) Springer-Verlag, New York, 1988.

5. J. Burbea, and C. R. Rao, Entropy Differential Metric, Distance and Divergence Measures in Probability Spaces: A Unified Approach, *J. of Multivariate Analysis* **12** (1982), 575–596.

6. M. Calvo, and J. M. Oller, A distance between multivariate normal distributions based on an embedding into the Siegel group, *J. of Multivariate Analysis* **35** (1990), 223–242.

7. M. DoCarmo, “Riemannian Geometry”, Birkhäuser, Boston, 1992.

8. H. Fürstenberg, A Poisson formula for semi–simple Lie groups, *Ann. of Math.* **77** (1963), 335–386.

9. K. Grove, and H. Karcher, How to conjugate C^1–close group actions, *Math. Z.* **132** (1973), 11–20.

10. S. Helgason, “Differential Geometry, Lie Groups, and Symmetric Spaces”, Academic Press, 1978.

11. H. Karcher, Riemannian center of mass and mollifier smoothing, *Comm. Pure and Appl. Math.* **30** (1977), 509–541.

12. A. F. S. Mitchell, The Information Matrix, Skewness Tensor and α-connections for General Multivariate Elliptic Distribution, *Ann. Inst. Stat. Math.* **41** No. 2 (1989), 289–304.

13. J. M. Oller, On an intrinsic bias measure, *in* “Proceedings of the Intl. Seminar, Suzdal, Russia”, pp. 134–158, (Lecture Notes in Mathematics, 1546) Springer-Verlag, 1993

14. C. R. Rao, Information and the accuracy attainable in the estimation of statistical parameters, *Bull. Calcutta Math. Soc.* **37** (1945), 81–91.

15. L.T. Skovgaard, A Riemannian geometry of the multivariate normal model, *Scand. J. Statistics* **11** (1984), 211–223.

Multivariate Normal Distributions Parametrized as a Riemannian Symmetric Space

Miroslav Lovrić, Maung Min-Oo and Ernst A. Ruh

Running title head:

Multivariate Normal Distributions

Please send proofs to:

Maung Min-Oo

Dept. of Mathematics and Statistics

McMaster University

Hamilton, Ontario, Canada L8S 4K1

e-mail: minoo@mcmaster.ca

The final manuscrpit will be submitted electronically in TeX