On bounds for the Fisher-Rao distance between multivariate normal distributions

João E. Strapasson, Julianna P. S. Porto† and Sueli I. R. Costa†*

INTRODUCTION

Information geometry is approached here by considering **the statistical model of multivariate normal distributions $\mathcal{M}$ as a Riemannian manifold with the natural metric provided by the Fisher information matrix.** This differential geometric approach to probability theory, introduced by C. Rao in 1945, have been recently applied to different domains such as statistical inference, mathematical programming, image processing and radar signal processing [4]. Explicit forms for the Fisher-Rao distance associated to this metric and for the geodesics of general distribution models are usually very hard to determine. In the case of general multivariate normal distributions **lower and upper bounds** have been derived. We approach here some of these bounds and introduce a new one discussing their tightness in specific cases.

PRELIMINARES

For multivariate normal distributions,

$$p(\mathbf{x};\boldsymbol{\mu},\Sigma)=\frac{(2\pi)^{-\left(\frac{n}{2}\right)}}{\sqrt{\mathrm{Det}(\Sigma)}}e^{\left(-\frac{(\mathbf{x}-\boldsymbol{\mu})^t\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}{2}\right)},$$

where $\mathbf{x}=(x_1,\cdots,x_n)^t$, $\boldsymbol{\mu}=(\mu_1,\cdots,\mu_n)^t$ is the mean vector and Σ is the covariance matrix, we consider the statistical model $\mathcal{M}=\{p_{\boldsymbol{\theta}};\boldsymbol{\theta}=(\boldsymbol{\mu},\Sigma)\in\Theta=\mathbb{R}^n\times P_n(\mathbb{R})\}$, where $P_n(\mathbb{R})$ is the space of positive-definite order n symmetric matrices, is $\left(n+\frac{n(n+1)}{2}\right)$-dimensional manifold. A natural Riemannian structure [1] can be provided by the *Fisher information matrix*:

$$g_{ij}(\boldsymbol{\theta})=E_{\boldsymbol{\theta}}\left(\frac{\partial}{\partial\theta_i}\Big(\log p(x;\boldsymbol{\theta})\Big)\frac{\partial}{\partial\theta_j}\Big(\log p(x;\boldsymbol{\theta})\Big)\right),$$

where $E_{\boldsymbol{\theta}}$ is the expected value with respect to the distribution $p_{\boldsymbol{\theta}}$.

The Fisher-Rao distance between two distributions $p_{\boldsymbol{\theta}_1}$ and $p_{\boldsymbol{\theta}_2}$ is then given by the shortest length, $\ell(\alpha)=\int_{t_1}^{t_2}(\langle\alpha'(t),\alpha'(t)\rangle_G)^{\frac{1}{2}}\,dt$, where $\langle\mathbf{x},\mathbf{y}\rangle_G=\mathbf{x}^t[g_{ij}]\mathbf{y}$ is the inner product defined G. The Fisher-Rao infinitesimal arc-length can be expressed as

$$ds^2=d\boldsymbol{\mu}^t\Sigma^{-1}d\boldsymbol{\mu}+\frac{1}{2}\mathrm{tr}[(\Sigma^{-1}d\Sigma)^2]. \qquad (1)$$

An explicit expression for the distance between two general normal multivariate distributions is very hard obtain.

Closed forms:

- In the case $n=1$, a closed form for the Fsher-Rao distance is known via an association with the classical model of the hyperbolic half-plane.
- The submanifold $\mathcal{M}_D$ where the covariance matrix is diagonal, $\Sigma=\mathrm{diag}(\sigma_1^2,\sigma_2^2,\cdots,\sigma_n^2)$. The Fisher matrix [3]:

$$g_{ij}(\boldsymbol{\theta})=\begin{pmatrix}\frac{1}{\sigma_1^2} & 0 & \cdots & 0 & 0\\ 0 & \frac{2}{\sigma_1^2} & \cdots & 0 & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & 0 & \cdots & \frac{1}{\sigma_n^2} & 0\\ 0 & 0 & \cdots & 0 & \frac{2}{\sigma_n^2}\end{pmatrix}.$$

$$d_D(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)=\sqrt{2\sum_{i=1}^{n}(\log B_i)^2},$$

$$B_i=\frac{\left|\left(\frac{\mu_{1i}}{\sqrt{2}},\sigma_{1i}\right)-\left(\frac{\mu_{2i}}{\sqrt{2}},-\sigma_{2i}\right)\right|+\left|\left(\frac{\mu_{1i}}{\sqrt{2}},\sigma_{1i}\right)-\left(\frac{\mu_{2i}}{\sqrt{2}},\sigma_{2i}\right)\right|^2}{\left|\left(\frac{\mu_{1i}}{\sqrt{2}},\sigma_{1i}\right)-\left(\frac{\mu_{2i}}{\sqrt{2}},-\sigma_{2i}\right)\right|-\left|\left(\frac{\mu_{1i}}{\sqrt{2}},\sigma_{1i}\right)-\left(\frac{\mu_{2i}}{\sqrt{2}},\sigma_{2i}\right)\right|}.$$

- The submanifold $\mathcal{M}_\Sigma$ where Σ is constant,

$$d_\Sigma(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)=\sqrt{(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^t\Sigma^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)}.$$

- The submanifold $\mathcal{M}_\mu$ where $\boldsymbol{\mu}$ is constant,

$$d_\mu^2(\Sigma_1,\Sigma_2)=\frac{1}{2}\sum_{i=1}^{n}[\log(\lambda_i)]^2,$$

λ_i are the eigenvalues of $\Sigma_1^{-1}\Sigma_2$.

$\mathcal{M}_\mu$ is a geodesic submanifold but $\mathcal{M}_D$ and $\mathcal{M}_\Sigma$ are not.

LOWER AND UPPER BOUNDS

A lower bound [2]:

$$d_F(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)\geqslant LB(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)=\sqrt{\frac{1}{2}\sum_{i=1}^{n}[\log(\lambda_i)]^2},$$

where λ_i are the eigenvalues of $S_1^{-1}S_2$,

$$S_i=\begin{pmatrix}\Sigma_i+\boldsymbol{\mu}_i^t\boldsymbol{\mu}_i & \boldsymbol{\mu}_i^t\\ \boldsymbol{\mu}_i & 1\end{pmatrix}.$$

Upper bounds: Upper bounds derived in [2] considering triangular inequalities. For $\boldsymbol{\theta}_1=(\boldsymbol{\mu}_1,\Sigma)$ and $\boldsymbol{\theta}_2=(\boldsymbol{\mu}_2,\alpha\Sigma)$,

$$d_\alpha^2(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)=2\,\mathrm{arccosh}\left(\frac{\sqrt{\alpha}}{2}+\frac{1}{2\sqrt{\alpha}}+\frac{1}{4\sqrt{\alpha}}\boldsymbol{\delta}^t\boldsymbol{\delta}\right)+\frac{n-1}{2}\log^2\alpha,$$

where $\boldsymbol{\delta}=\Sigma^{-\frac{1}{2}}(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1)$.

$$d_F(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)\leqslant d_\alpha(\boldsymbol{\theta}_1,\boldsymbol{\theta}_\alpha)+d_\mu(\boldsymbol{\theta}_\alpha,\boldsymbol{\theta}_2).$$

Upper Bound **U1**: $\alpha=\|\Sigma_1^{-\frac{1}{2}}\Sigma_2\Sigma_1^{-\frac{1}{2}}\|^{\frac{1}{n}}$.
Upper Bound **U2**: $\alpha=\min_\alpha\{d_{\alpha\Sigma}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_\alpha)+d_\mu(\boldsymbol{\theta}_\alpha,\boldsymbol{\theta}_2)\}$.

We proposed a new upper bound, U3, based on special isometries in the manifold $\mathcal{M}$ and on the distance in the submanifold $\mathcal{M}_D$.

Proposition 1. *For $\boldsymbol{\theta}_1=(\boldsymbol{\mu}_1,\Sigma_1)$ and $\boldsymbol{\theta}_2=(\boldsymbol{\mu}_2,\Sigma_2)$ we can assert:*

$$d_F(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)=d_F\Big((\mathbf{0},I_n);(\boldsymbol{\mu}_3,D_3)\Big). \qquad (2)$$

where **0** *is the zero mean, I_n is the identity matrix, D_3 is the diagonal matrix given by the eigenvalues of $A=\Sigma_1^{-(1/2)}\Sigma_2\Sigma_1^{-(1/2)}$, Q is the orthogonal matrix whose columns are the respective eigenvectors of A ($A=QD_3Q^t$) and $\boldsymbol{\mu}_3=Q^t\Sigma_1^{-(1/2)}(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1)$.*

Proposition 2. *The Fisher-Rao distance between two multivariate normal distributions $\boldsymbol{\theta}_1=(\boldsymbol{\mu}_1,\Sigma_1)$ and $\boldsymbol{\theta}_2=(\boldsymbol{\mu}_2,\Sigma_2)$ is upper bounded by,*

$$U3(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)=\sum_{i=1}^{n}\sqrt{2(\log C_i)^2}, \qquad (3)$$

$$C_i=\frac{\sqrt{|1-D_{3i}|^2+|\mu_{3i}|^2}+\sqrt{|D_{3i}+1|^2+|\mu_{3i}|^2}}{\sqrt{|D_{3i}+1|^2+|\mu_{3i}|^2}-\sqrt{|1-D_{3i}|^2+|\mu_{3i}|^2}},$$

where μ_{3i} are the coordinates of $\boldsymbol{\mu}_3$ and D_{3i} are the diagonal terms of D_3, as established in the last proposition.

Simulations for $n=2$

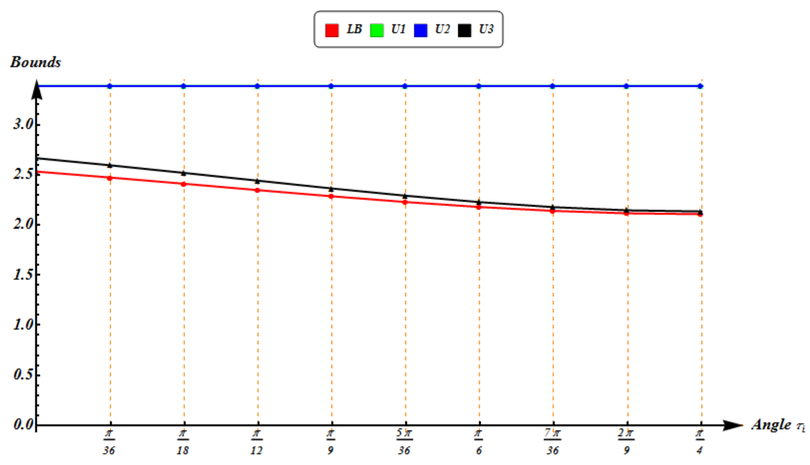

Figure 1: *The mean $\boldsymbol{\mu}_2=(2,2)$ is fixed, $\lambda_1=1$, $\lambda_2=10$ and τ varies from 0 to $\frac{\pi}{4}$.*

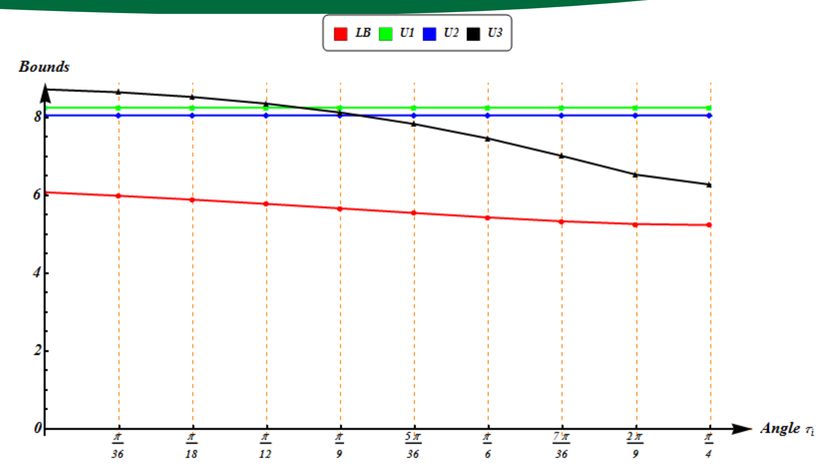

Figure 2: *The mean $\boldsymbol{\mu}_2=(16,16)$ is fixed, $\lambda_1=1$, $\lambda_2=10$ and τ varies from 0 to $\frac{\pi}{4}$.*

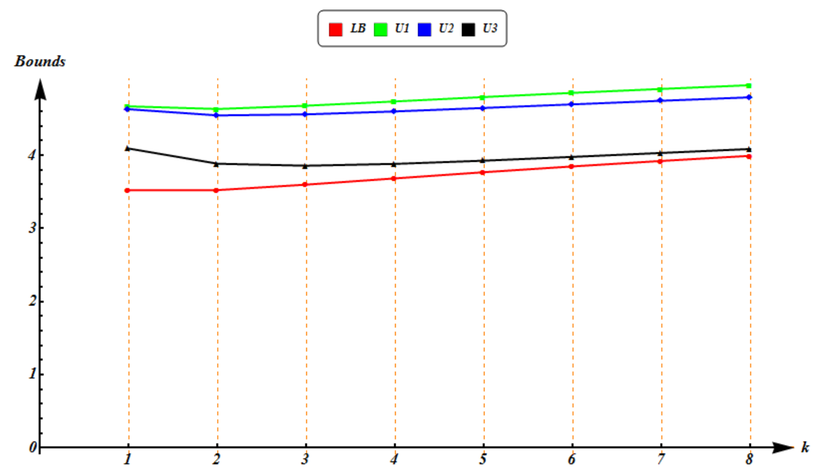

Figure 3: *The mean $\boldsymbol{\mu}_2=(4,4)$ and the angle $\tau=0$ are fixed, $\lambda_1=k$, $\lambda_2=10k$ and k varies from 1 to 8.*

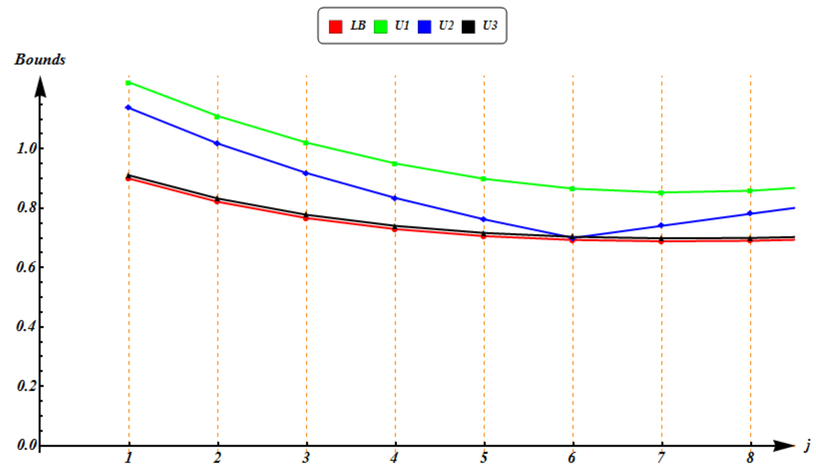

Figure 4: *The mean $\boldsymbol{\mu}_2=(0.5,0.5)$ and the angle $\tau=\frac{\pi}{4}$ are fixed, $\lambda_1=1$ and $\lambda_2=0.5+0.1j$, where j varies from 0 to 10.*

We observe that U3 is sensitively tighter than U_1 and U2 particularly in the cases of discrepancy between eigenvalues and/or rotations of eigenvectors frames.

REFERENCES

[1] S. Amari, and H. Nagaoka, *Methods of Information Geometry*,T. M. M., Vol.191, Am. Math. Soc., 2000.

[2] M. Calvo, and J.M. Oller, Journal of Multivariate Analysis 35.2, 223-242, 1990, Statistics and Decisions 9, 119-138, 1991.

[3] S. I. R. Costa, S. A. Santos, and J. E. Strapasson, *Fisher information distance: a geometrical reading*,arXiv,(to appear)Discrete Applied Mathematics.

[4] F. Nielsen and F. Barbaresco (Eds.), *Geometric Science of Information*, First International Conference, GSI 2013 Paris, France, August 2013, *Lecture Notes in Computer Science* 8085 (2013).

Email address: sueli@ime.unicamp.br (Sueli I. R. Costa).

*School of Applied Sciences, University of Campinas, Brazil
†Institute of Mathematics, University of Campinas, Brazil

Suporte financeiro: FAPESP