Fisher-Rao and pullback Hilbert cone distances on the multivariate Gaussian manifold with applications to simplification and quantization of mixtures

Sony CSL

Frank Nielsen

Sony Computer Science Laboratories

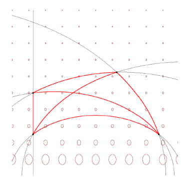
Tokyo, Japan

Problem: Define fast principled distances between multivariate normal distributions

$$N(\mu,\Sigma) \sim p_{\mu,\Sigma}(x) = \frac{(2\pi)^{-\frac{d}{2}}}{\sqrt{\det(\Sigma)}} \exp\left(-\frac{(x-\mu)^\top\Sigma^{-1}(x-\mu)}{2}\right)$$

$$\mathcal{N}(d) = \{N(\lambda) \ : \ \lambda = (\mu,\Sigma) \in \Lambda(d) = \mathbb{R}^d \times \mathrm{Sym}_+(d,\mathbb{R})\}$$

Fisher metric: $ds^2_{\mathrm{Fisher}} = d\mu^\top\Sigma^{-1}d\mu + \frac{1}{2}\mathrm{tr}\left((\Sigma^{-1}d\Sigma)^2\right)$



Some closed form formula for specific cases:

- In 1D $\rho_{\mathcal{N}}(N_0,N_1) = \sqrt{2}\log\left(\frac{1+\Delta(\mu_0,\sigma_0;\mu_1,\sigma_1)}{1-\Delta(\mu_0,\sigma_0;\mu_1,\sigma_1)}\right)$

with Möbius distance $\Delta(a,b;c,d) = \sqrt{\frac{(c-a)^2+2(d-b)^2}{(c-a)^2+2(d+b)^2}}$

- Same mean: $\rho_{\mathcal{N}_\mu}(N(\mu,\Sigma_0),N(\mu,\Sigma_1)) = \sqrt{\frac{1}{2}\sum_{i=1}^d \log^2\lambda_i(\Sigma_0^{-\frac{1}{2}}\Sigma_1\Sigma_0^{-\frac{1}{2}})}$

Fisher-Rao geodesic Riemannian distance:

$$\rho_{\mathrm{FR}}(N_0,N_1) = \inf_{\substack{c(t)\\ c(0)=p_{\mu_0,\Sigma_0}\\ c(1)=p_{\mu_1,\Sigma_1}}} \{\mathrm{Len}(c)\}$$

Geodesic equation:

$$\begin{cases} \ddot{\mu} - \dot{\Sigma}\Sigma^{-1}\dot{\mu} &= 0,\\ \ddot{\Sigma} + \dot{\mu}\dot{\mu}^\top - \dot{\Sigma}\Sigma^{-1}\dot{\Sigma} &= 0.\end{cases}$$

Invariance under the positive affine group:
$\mathrm{Aff}_+(d,\mathbb{R}) := \{(a,A) \ : \ a \in \mathbb{R}^d, A \in \mathrm{GL}_+(d,\mathbb{R})\}$
$\rho_{\mathrm{FR}}((a,A).N_0 : (a,A).N_1) = \rho_{\mathrm{FR}}(N_0,N_1)$

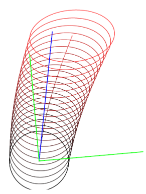
Solving with initial conditions: Use natural parameterization $(\xi = \Sigma^{-1}\mu, \Xi = \Sigma^{-1})$

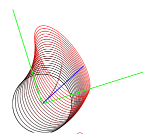
Initial conditions: $(a = \dot{\xi}(0), B = \dot{\Xi}(0)) = \dot{\gamma}^{\mathrm{Fisher}}_{\mathcal{N}}(N_0, v_0; 0)$

Let $B = -\Xi(0)^{-\frac{1}{2}}\,\dot{\Xi}(0)\,\Xi(0)^{-\frac{1}{2}}, a = \Xi(0)^{-\frac{1}{2}}\dot{\xi}(0) + B\Xi_0^{-\frac{1}{2}}\xi(0), G = (B^2 + 2aa^\top)^{\frac{1}{2}}$.

Then we have $\Xi(t) = \Xi(0)^{\frac{1}{2}}\,R(t)R(t)^\top\,\Xi(0)^{\frac{1}{2}}, \xi(t) = 2\Xi(0)^{\frac{1}{2}}\,R(t)\mathrm{Sinh}\left(\frac{1}{2}Gt\right)G^\dagger a + \Xi(t)\Xi^{-1}(0)\xi(0)$

with $R(t) = \mathrm{Cosh}\left(\frac{1}{2}Gt\right) - BG^\dagger\mathrm{Sinh}\left(\frac{1}{2}Gt\right)$



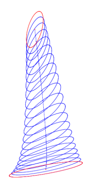


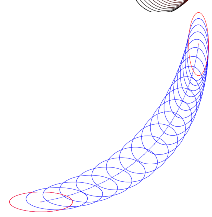
Solving with boundary conditions:

Fisher-Rao geodesic $N_t = N(\mu(t),\Sigma(t)) = \gamma^{\mathcal{N}}_{\mathrm{FR}}(N_0,N_1;t)$:

- For $i \in \{0,1\}$, let $G_i = M_i\,D_i\,M_i^\top$, where

$$M_i = \begin{bmatrix} \Sigma_i^{-1} & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & \Sigma_i \end{bmatrix}, D_i = \begin{bmatrix} I_d & 0 & 0\\ \mu_i^\top & 1 & 0\\ 0 & -\mu_i & I_d \end{bmatrix}$$





- Consider the Riemannian geodesic in $\mathrm{Sym}_+(2d+1,\mathbb{R})$ with respect to the trace metric: $G(t) = G_0^{\frac{1}{2}}\left(G_0^{-\frac{1}{2}}G_1G_0^{-\frac{1}{2}}\right)^t G_0^{\frac{1}{2}}$
- Retrieve $N(t) = \gamma^{\mathcal{N}}_{\mathrm{FR}}(N_0,N_1;t) = N(\mu(t),\Sigma(t))$ from $G(t)$:

$\Sigma(t) = [G(t)]^{-1}_{1:d,1:d}, \mu(t) = \Sigma(t)\,[G(t)]_{1:d,d+1}$ where

$[G]_{1:d,1:d}$ denotes the block matrix with rows and columns ranging from 1 to d extracted from $(2d+1)\times(2d+1)$ matrix G, and $[G]_{1:d,d+1}$ is similarly the column vector of $\mathbb{R}^d$ extracted from G

Approximate Fisher-Rao distance by $\tilde{\rho}_T(N_0,N_1) = \sum_{i=0}^{T-1}\sqrt{D_J\left(N_{\frac{i}{T}}, N_{\frac{i+1}{T}}\right)} \approx \rho_{\mathrm{FR}}(N_0,N_1)$ where D_J is the symmetrized Kullback-Leibler distance: $D_J(N_1,N_2) = I_{f_J}(N_1,N_2) = \mathrm{tr}\left(\frac{\Sigma_2^{-1}\Sigma_1 + \Sigma_1^{-1}\Sigma_2}{2} - I\right) + (\mu_2-\mu_1)^\top\frac{\Sigma_1^{-1}+\Sigma_2^{-1}}{2}(\mu_2-\mu_1)$

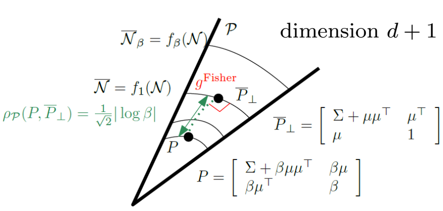
Embed $N(\mu,\Sigma)$ into SPD matrix of dim $d+1$:

$$f_a(N(\mu,\Sigma))^{-1} = \begin{bmatrix} \Sigma^{-1} & -\Sigma^{-1}\mu\\ -\mu^\top\Sigma^{-1} & \mu^\top\Sigma^{-1}\mu + \frac{1}{a}\end{bmatrix}$$

When $a = 1$ isometric embedding:

Totally geodesic submanifold, get lower bound:

$\rho_{\mathrm{CO}}(N_0,N_1) = \rho_{\mathrm{FR}}(N(0,f(N_0)),N(0,f(N_1)))$

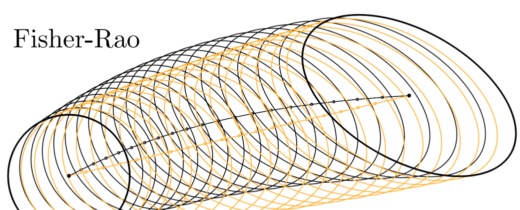


Consider the Hilbert projectice distance on the SPD cone:

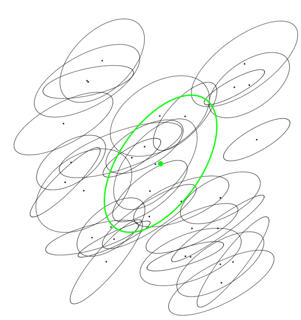
$$\rho_{\mathrm{Hilbert}}(P_0,P_1) = \log\left(\frac{\lambda_{\max}(P_0^{-\frac{1}{2}}P_1P_0^{-\frac{1}{2}})}{\lambda_{\min}(P_0^{-\frac{1}{2}}P_1P_0^{-\frac{1}{2}})}\right),$$

$$= \log\left(\frac{\lambda_{\max}(P_0^{-1}P_1)}{\lambda_{\min}(P_0^{-1}P_1)}\right)$$

→ fast calculations: only extreme eigenvalues



Applications to clustering:



References:

- Calvo & Oller: An explicit solution of information geodesic equations for the multivariate normal model, Statistics and Decision (1991)
- Kobayashi: Geodesics of multivariate normal distributions and a Toda lattice type Lax pair, arXiv:2304.12575 (2023)
- Nielsen: A Simple Approximation Method for the Fisher–Rao Distance between Multivariate Normal Distributions, Entropy (2023)