A Simple Approximation Method for the Fisher-Rao Distance between Multivariate Normal Distributions

Frank Nielsen

2023

1 The Fisher-Rao distance

Let $\mathbb{P}(d)$ denote the set of symmetric positive?definite (SPD) $d \times d$ matrices and $\mathcal{N}(d)$ denote the set of multivariate normal distributions:

$$\mathcal{N}(d) := \left\{ p_{\mu,\Sigma}(x) = (2\pi)^{\frac{d}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right) \ :\ (\mu,\Sigma) \in \Lambda(d) := \mathbb{R}^d \times \mathbb{P}(d) \right\},$$

The Fisher-Rao distance between two normals $N(\mu_1, \Sigma_1)$ and $N(\mu_2, \Sigma_2)$ is the geodesic Riemannian distance on the manifold $(\mathcal{N}, g^{\mathrm{Fisher}})$ induced by the Fisher information metric:

$$\rho_{\mathcal{N}}(N(\lambda_1), N(\lambda_2)) := \inf_{\substack{c(t) \\ c(0)=p_{\lambda_1} \\ c(1)=p_{\lambda_2}}} \left\{\mathrm{Length}(c)\right\},$$

where

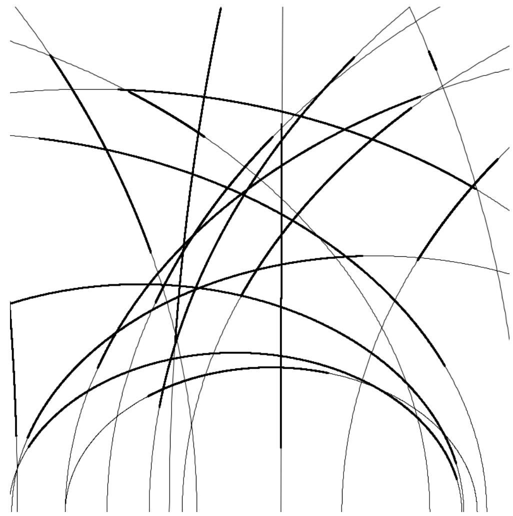
$$\mathrm{Length}(c) := \int_0^1 \mathrm{d}s^{\mathrm{Fisher}}(c(t))\mathrm{d}t,$$

and $\mathrm{d}s^{\mathrm{Fisher}}(t) := \sqrt{\langle \dot{c}(t), \dot{c}(t)\rangle_{c(t)}}$ is the Fisher-Rao length element. The inner product $\langle v_1, v_2\rangle_N$ for $v_1, v_2 \in T_N\mathcal{N}$ at normal N is the called the Fisher-Rao norm (with tangent planes $T_N\mathcal{N}$ is identified to $\mathbb{R}^d \times \mathrm{Sym}(d)$ where $\mathrm{Sym}(d)$ be the set of $d \times d$ symmetric matrices). The statistical model $\mathcal{N}(d)$ is of dimension $m = \dim(\Lambda(d)) = d + \frac{d(d+1)}{2} = \frac{d(d+3)}{2}$ and identifiable: there is a one-to-one correspondence $\lambda \leftrightarrow p_\lambda(x)$ between $\lambda \in \Lambda(d)$ and $N(\mu, \Sigma) \in \mathcal{N}(d)$.

- When $d = 1$, the Fisher-Rao distance is known in closed form:

  $$\rho_{\mathcal{N}}(N_1, N_2) = 2\sqrt{2}\,\mathrm{arctanh}(\Delta(\mu_1, \sigma_1; \mu_2, \sigma_2)),$$

  where $\Delta(a, b; c, d) = \sqrt{\frac{(c-a)^2+2(d-b)^2}{(c-a)^2+2(d+b)^2}}$ is a Möbius distance and $\mathrm{arctanh}(u) := \frac{1}{2}\log\left(\frac{1+u}{1-u}\right)$ for $0 \leq u < 1$. The Fisher?Rao geodesics are semi-ellipses with centers located on the x-axis:



- When the normal distributions belongs to the same submodel $\mathcal{N}_\mu = \{N(\mu,\Sigma) \ : \ \Sigma \in \mathcal{P}(d)\} \subset \mathcal{N}$ of normal distributions sharing the same mean μ, we have [?, ?]:

$$\rho_{\mathcal{N}_\mu}(N_1,N_2) \quad = \quad \sqrt{\frac{1}{2}\sum_{i=1}^{d} \log^2 \lambda_i(\Sigma_1^{-1}\Sigma_2)},$$

  where $\lambda_i(M)$ denotes the i-th generalized largest eigenvalue of matrix M, where the generalized eigenvalues are solutions of the equation $|\Sigma_1 - \lambda\Sigma_2| = 0$. The submanifold $(\mathcal{N}_\mu, g^{\mathrm{Fisher}})$ is totally geodesic in $(\mathcal{N}, g^{\mathrm{Fisher}})$.

- When the normal distributions belongs to the same submodel $\mathcal{N}_\Sigma = \{N(\mu,\Sigma) \ : \ \Sigma \in \mathcal{P}(d)\} \subset \mathcal{N}$ of normal distributions sharing the same covariance matrix Σ we have

$$\sqrt{2}\,\mathrm{arccosh}\left(1 + \frac{1}{4}\Delta_\Sigma^2(\mu_1,\mu_2)\right),$$

  where Δ_Σ is the Mahalanobis distance:

$$\Delta_\Sigma(\mu_1,\mu_2) := \sqrt{(\mu_2-\mu_1)^\top\Sigma^{-1}(\mu_2-\mu_1)}.$$

However, in the general case, the Fisher-Rao distance between normals is not known in closed form [?].

2 Isometric embedding into the higher-dimensional SPD cone

Calvo and Oller [?] show how to embed $N(\mu,\Sigma) \in \mathcal{N}(d) = \left\{\bar{P} = f_\beta(\mu,\Sigma) \ : \ (\mu,\Sigma) \in \mathcal{N}(d) = \mathbb{R}^d \times \mathcal{P}(d)\right\}$ into a SPD matrix of $\mathbb{P}(d+1)$:

$$\bar{P}(N) = f(N) = \begin{bmatrix} \Sigma + \mu\mu^\top & \mu \\ \mu^\top & 1 \end{bmatrix}$$

so that the manifold $(\mathcal{N}(d), g^{\mathrm{Fisher}})$ is isometrically embedded into the submanifold $(\overline{\mathcal{N}}, g^{\mathrm{trace}})$ of the cone equipped with the trace metric

$$g_P^{\mathrm{trace}}(P_1,P_2) := \frac{1}{2}\mathrm{tr}(P^{-1}P_1P^{-1}P_2).$$

However, the submanifold $\overline{\mathcal{N}} \subset \mathbb{P}(d+1)$ is not totally geodesic. Thus Calvo and Oller [?] derived a lower bound on the Fisher-Rao distance:

$$\rho_{\mathrm{CO}}(N_1, N_2) = \sqrt{\frac{1}{2}} \sqrt{\frac{1}{2} \sum_{i=1}^{d} \log^2 \lambda_i(\bar{P}_1^{-1} \bar{P}_2)}$$

which is also metric distance.

3 A simple approximation method

Our method consists in projecting the SPD geodesic $\gamma_{\mathbb{P}(d+1)}(\bar{P}_1^{-1}\bar{P}_2)$ onto $\overline{\mathcal{N}}$ and then maps back the SPD projected curve into $\mathcal{N}$ by using f^{-1}:

$$c_{\mathrm{CO}}(N_1, N_2; t) = f^{-1}\left(\mathrm{proj}_{\overline{\mathcal{N}}}(\gamma_{\mathbb{P}(d+1)}(\bar{P}_1^{-1}\bar{P}_2; t))\right).$$

Indeed, the geodesic $\gamma_{\mathbb{P}(d+1)}(\bar{P}_1^{-1}\bar{P}_2)$ has closed-form equation

$$\gamma_{\mathbb{P}(d+1)}(\bar{P}_1^{-1}\bar{P}_2) = \bar{P}_1^{\frac{1}{2}} \left(\bar{P}_1^{-\frac{1}{2}} \bar{P}_2 \bar{P}_1^{-\frac{1}{2}}\right)^t \bar{P}_1^{\frac{1}{2}}.$$

Now, we need to estimate the Fisher-Rao length of the curve $c_{\mathrm{CO}}(N_1, N_2; t)$ by discretizing the curve at T positions:

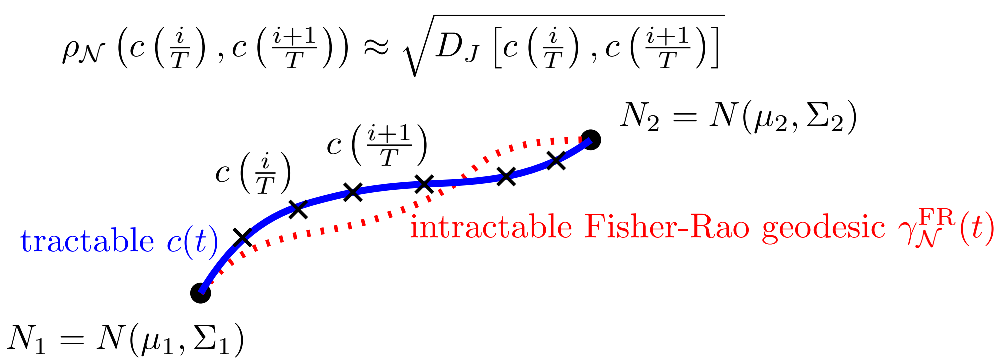
$$\tilde{\rho}_{\mathrm{CO}}(N_1, N_2) \leq \frac{1}{T} \sum_{i=1}^{T-1} \rho_{\mathcal{N}}\left(c\left(\frac{i}{T}\right), c\left(\frac{i+1}{T}\right)\right),$$

and approximate for nearby normals their Fisher-Rao distances by the square root of their Jeffreys divergence:

$$\rho_{\mathcal{N}}\left(c\left(\frac{i}{T}\right), c\left(\frac{i+1}{T}\right)\right) \approx \sqrt{D_J\left[c\left(\frac{i}{T}\right), c\left(\frac{i+1}{T}\right)\right]},$$

where

$$D_J[p_{(\mu_1,\Sigma_1)} : p_{(\mu_2,\Sigma_2)}] = \mathrm{tr}\left(\frac{\Sigma_2^{-1}\Sigma_1 + \Sigma_1^{-1}\Sigma_2}{2} - I\right) + (\mu_2 - \mu_1)^\top \frac{\Sigma_1^{-1} + \Sigma_2^{-1}}{2} (\mu_2 - \mu_1).$$

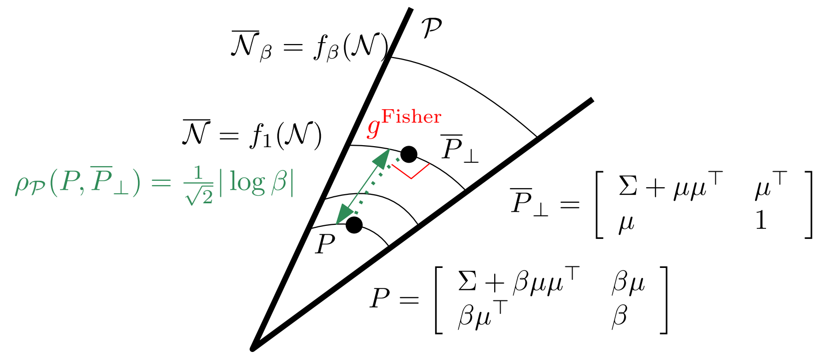


The projection of a SPD matrix $P \in \mathbb{P}(d+1)$ onto $\overline{\mathcal{N}}$ is done as follows: Let $\beta = P_{d+1,d+1}$ and write $P = \begin{bmatrix} \Sigma + \beta\mu\mu^\top & \beta\mu \\ \beta\mu^\top & \beta \end{bmatrix}$. Then the orthogonal projection at $P \in \mathcal{P}$ onto $\overline{\mathcal{N}}$ is:

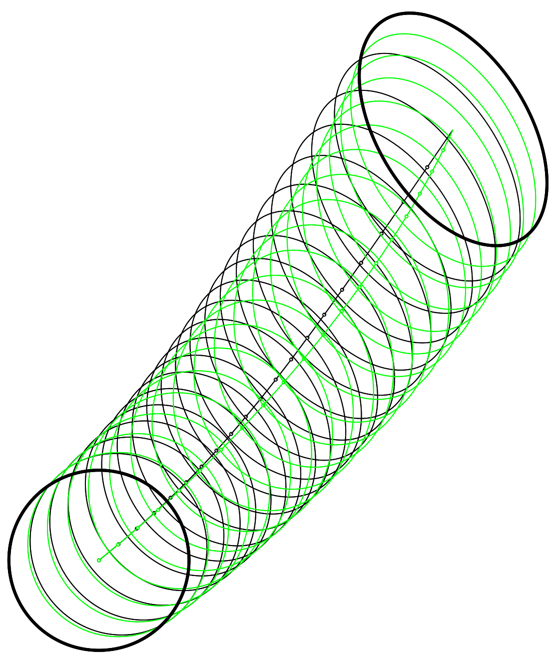
$$\bar{P}_\perp := \mathrm{proj}_{\overline{\mathcal{N}}}(P) = \begin{bmatrix} \Sigma + \mu\mu^\top & \mu^\top \\ \mu & 1 \end{bmatrix},$$

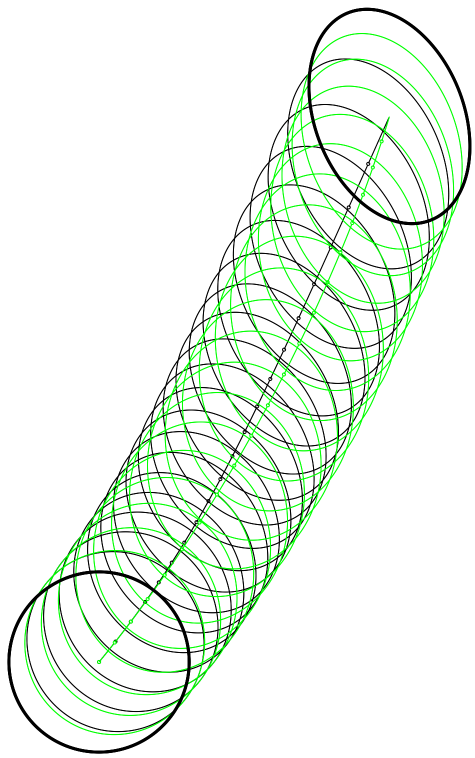
and the SPD distance between P and $\bar{P}_\perp$ is

$$\rho_{\mathcal{P}}(P, \bar{P}_\perp) = \frac{1}{\sqrt{2}} |\log \beta|.$$



Here are some examples of the curves c_{CO} (in green) compared to the Fisher-Rao geodesics (in black):





More details and quantitative analysis: `https://www.mdpi.com/1099-4300/25/4/654`