Riemannian connection

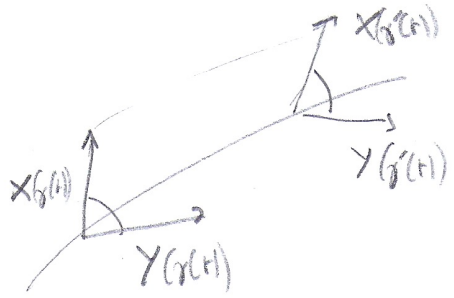
There are many possible choice for affine connections. If one has a notion of metric, a natural choice is to choose the Riemannian connection.

* symmetry: $\Gamma^k_{ij} = \Gamma^k_{ji}$ (torsion free)

* preserve inner product by parallel transport:

? $\frac{d}{dt}\langle X(\gamma(t)), Y(\gamma(t))\rangle = \langle \nabla_{\dot\gamma(t)} X, Y\rangle + \langle X, \nabla_{\dot\gamma(t)} Y\rangle$

↑ covariant derivative ↑ covariant derivative



parallel transport:

$$\nabla_{\dot\gamma} X = \underbrace{\left\{\dot x^k + \dot\gamma^i X^j \Gamma^k_{ij}\right\}}_{0} \partial_k|_{\gamma(t)}$$

X, Y parallel translated $\Rightarrow$ $\frac{d}{dt}\langle X(\gamma(t)), Y(\gamma(t))\rangle = 0$

Requirement to be true for any path and X, Y, $\dot\gamma(t) = z$

$$\partial_k \langle \partial_i, \partial_j\rangle = \langle \nabla_{\partial_k}\partial_i, \partial_j\rangle + \langle \partial_i, \nabla_{\partial_k}\partial_j\rangle$$

$\hookrightarrow$ local:

$$\partial_k g_{ij} = \Gamma_{ki,j} + \Gamma_{kj,i}$$

$$\Gamma_{ij,k} = \frac{1}{2}\left(\partial_i g_{jk} + \partial_j g_{ki} - \partial_k g_{ij}\right)$$

Information-based connection

Natural notion Kullback Leibler divergence

$$D[p\|q] = \int dx\, p(x) \log \frac{p(x)}{q(x)}$$

Problem: not symmetric, do not satisfy triangular inequality

However: $D(p\|q) \geq 0$ (smooth) $\qquad D(p\|q) \Leftrightarrow p = q$

Notations $\quad D(\partial_i|_p \| p') = \partial_i D(p\|p')$

$$D(\partial_i|_p \| \partial_j|_{p'}) = \partial_i \partial'_j (p\|p') \dots$$

Diagonal constraint $\Rightarrow \quad D[\partial_i \| \,] = D[\, \| \partial_j] = 0$

$$\Rightarrow \quad D[\partial_i \partial_j \| \,] + D[\partial_i \| \partial_j] = 0$$

$$\Rightarrow \quad D[\partial_i \partial_j \| \,] = \underbrace{-D[\partial_i \| \partial_j]}_{g_{ij}(p)} = D[\,\| \partial_i \partial_j]$$

($D[\partial_i\partial_j\|\,] = 0$, $D[\,\|\partial_i\partial_j] = 0$)

$$D(p\|q) = \frac{1}{2} g_{ij}(q)\, \Delta\xi^i \Delta\xi^j + O\left(\|\Delta\xi\|^3\right) \qquad \Delta\xi^i = \xi^i(p) - \xi^i(q)$$

($g_{ij}(q)$: metric !)

Connection $\rightarrow$ 3rd order $\qquad \Gamma_{ij,k} = -D[\partial_i\partial_j \| \partial_k]$

$$D[p\|q] = \frac{1}{2} g_{ij}(q)\,\Delta\xi^i\Delta\xi^j + \frac{1}{6} h_{ijk}(q)\Delta\xi^i\Delta\xi^j\Delta\xi^k + o\left(\|\Delta\xi\|^3\right)$$

$$h_{ijk} = D[\partial_i\partial_j\partial_k\|\,] = \partial_i[D[\partial_j\partial_k\|\,\,]] - D[\partial_j\partial_k\|\partial_i]$$

$$= \partial_i g_{jk} + \Gamma_{jk,i}$$

$D^*[q\|p] = D[p\|q] \rightarrow D^*$ natural dual Fonction

$$= \frac{1}{2} g_{ij}(p)\,\Delta\xi^i\Delta\xi^j + \frac{1}{6} h^*_{ijk}(p)\,\Delta\xi^i\Delta\xi^j\Delta\xi^k + o\left(\|\Delta\xi\|^3\right)$$

$$\frac{1}{2}\left(g_{ij}(q)\,\Delta\xi^i\Delta\xi^j + \partial_k g_{ij}(p)\,\Delta\xi^i\Delta\xi^j\Delta\xi^k\right)$$

$$h^*_{ijk} = D[\,\|\partial_i\partial_j\partial_k] = \partial_i g_{jk} + \Gamma^*_{jk,i}, \quad \Gamma^*_{jk,i} = -D[\partial_i\|\partial_k\partial_j]$$

$\Rightarrow$ <u>Dual relation</u> : $\partial_k g_{ij} = \Gamma_{ki,j} + \Gamma^*_{jk,i}$

$$\partial_k g_{ij} = -\partial_k D[\partial_i\|\partial_j] = -D[\partial_k\partial_i\|\partial_j] - D[\partial_i\|\partial_k\partial_j]$$

<u>Almost Riemannian</u> : $\Gamma_{ki,j} = \Gamma^*_{ki,j}$ Self-dual

f - divergence

convex, smooth

$$D_f(p\|q) = \int p(x) f\left(\frac{q(x)}{p(x)}\right) dx$$

$D_f(p\|q) = 0 \iff p = q$ if f strictly convex at 1

$\Uparrow$

Jensen inequality: $D_f(p\|q) \geq f\left(\int p(x) \frac{q(x)}{p(x)} dx\right) \overset{\text{def}}{=} f(1) = 0$

Monotonicity: $K(y|x) \rightarrow$ $p_K(y) = \int K(y|x) p(x) dx$, $q_K(y) = \int K(y|x) q(x) dx$

$\Downarrow$

Convexity

$$D_f(p\|q) = \int dx \int dy\, K(y|x) p(x) f\left(\frac{q(x)}{p(x)}\right) dx$$

$p_K(x|y) = \frac{K(y|x)p(x)}{p_K(y)}$

$$= \int dy\, p_K(y) \int dx\, p_K(x|y) f\left(\frac{q(x)}{p(x)}\right)$$

Jensen $\rightarrow$

$$\geq \int dy\, p_K(y) f\left(\int dx\, p_K(x|y) \frac{q(x)}{p(x)}\right)$$

$$= \int dy\, p_K(y) f\left(\frac{q_K(y)}{p_K(y)}\right) = D_f(p_K\|q_K)$$

$D_f(p\|q) = D_f(p_K\|q_K) \iff K$ sufficient statistics p, q

$\iff p_K(x|y) = q_K(x|y)$ (strictly convex)

$K(y|x) = \delta_{F(x)}(y)$

$$f^{\alpha}(u) = \begin{cases} \frac{4}{1-\alpha^2}\left\{1 - u^{(1+\alpha)/2}\right\} & \alpha \neq \pm 1 \\ u \log u \\ -\log u \end{cases} \quad \left.\begin{matrix} \\ \\ \end{matrix}\right\} \text{KL}$$

α-divergence

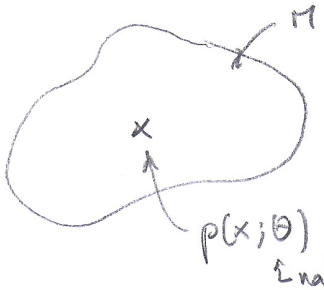
$$D^{(0)}(p\|q) = 2 \int \left(\sqrt{p(x)} - \sqrt{q(x)}\right)^2 dx$$

$\sqrt{D^{(0)}(p\|q)}$

Hellinger distance

Fisher metric

θ^i : coordinates.



↑ natural parameter :

$$g_{ij}(\theta) = \mathbb{E}_\theta\left[\partial_i \log p\, \partial_j \log p\right] \leftarrow \text{covariant 2.}$$

$$= \int dx\, \frac{\partial_i p}{\sqrt{p}}\, \frac{\partial_j p}{\sqrt{p}}$$

Statistical model: $p(x,\theta) = \begin{cases} \theta_i\,, & x = i \\ \theta_0 = 1 - \sum_{i=1}^{n} \theta_i & \end{cases}$

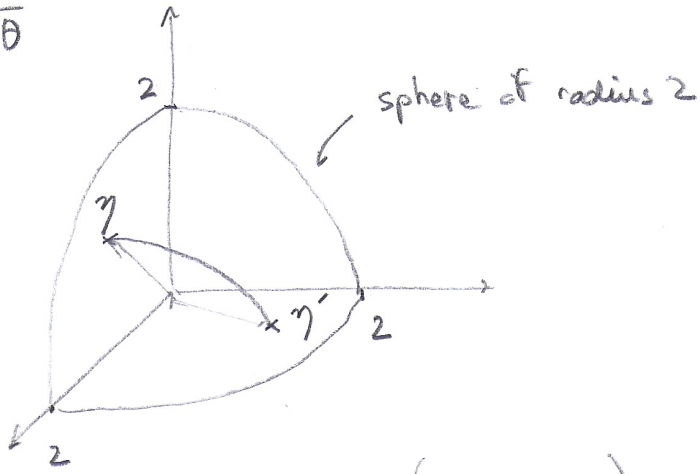
$$g_{ij}(\theta) = \sum_k \theta_k\, \partial_i \log\theta_k\, \partial_j \log\theta_k + \theta_0\, \partial_i \log\theta_0\, \partial_j \log\theta_0$$

$$= \sum_k \theta_k \frac{\delta_{ik}}{\theta_k}\frac{\delta_{jk}}{\theta_k} + \theta_0 \frac{(-1)}{\theta_0}\frac{(-1)}{\theta_0} = \frac{\delta_{ij}}{\theta_i} + \frac{1}{1-\sum\theta_i} \leftarrow n \text{ variables}$$

g_{ij} can be seen as an induced metric :

$$g_{ij}(\theta) = \sum_{k=1}^{n} \frac{\partial_i \theta_k}{\sqrt{\theta_k}}\frac{\partial_j \theta_k}{\sqrt{\theta_k}} + \frac{\partial_i \theta_0}{\sqrt{\theta_0}}\frac{\partial_j \theta_0}{\sqrt{\theta_0}} = \left\langle 2\partial_i \sqrt{\theta}\,,\, 2\partial_j \sqrt{\theta} \right\rangle_{n+1}$$

$$\sqrt{\theta} = \begin{vmatrix} \sqrt{\theta_0} \\ \sqrt{\theta_1} \\ \\ \sqrt{\theta_N} \end{vmatrix}\,, \qquad \eta = 2\sqrt{\theta}$$



$$\langle \eta \cdot \eta' \rangle_{n+1} = 2 \sum_{k=0}^{n} \sqrt{\theta_k}\sqrt{\theta'_k} \Rightarrow L(\theta,\theta') = 2\cos^{-1}\left(\sum_{k=0}^{n+1} \sqrt{\theta_k \theta'_k}\right)$$

Hellinger distance: $\sum_k \left(\sqrt{\theta_k} - \sqrt{\theta'_k}\right)^2 = 2\left(1 - \sum_{k=0} \sqrt{\theta_k \theta'_k}\right) = 2\sin\left(L(\theta,\theta')/4\right)$

Property of Fisher metric

Sufficient statistics : model $p(x|\theta)$

$F: X \to Y$, $x \mapsto F(x)=y$ | $r(x|\theta) = \dfrac{p(x|\theta)}{q(F(x)|\theta)}$ → does not depend on θ for all x

$$p(x|y,\theta) = r(x|\theta)\,\delta_{F(x)}(y)$$

$k(y|x) \to g_k$, $\quad g(\theta) - g_k(\theta) \leftarrow$ positive semi definite

$$\| \\ \Delta g(\theta)$$

$$\Delta g(\theta) = \mathbb{E}_\theta\left[\partial_i \log r(x|\theta)\, \partial_j \log r(x|\theta)\right] \text{ if deterministic}$$

$$= \mathbb{E}_\theta\left[\mathrm{Cov}\left[\partial_i \log p(x|\theta)\, \partial_j \log p(x|\theta)\right] | Y\right]$$

Monotonicity / Chain rule

Additivity : $g(\theta) = g_1(\theta) + g_2(\theta)$ if $p(x|\theta) = p(x_1|\theta)p(x_2|\theta)$

Convexity : $g_\lambda(\theta) \leq \lambda g_1(\theta) + (1-\lambda) g_2(\theta)$

Cramer Rao inequality

unbiased estimator $\mathbb{E}_\theta\left[\hat\theta(x)\right] = \theta \quad \forall \theta$

$$V_\theta^{ij} = \mathbb{E}_\theta\left[\left(\hat\theta^i(x) - \theta^i\right)\left(\hat\theta^j(x) - \theta^j\right)\right]$$

$V_\theta - g^{-1}(\theta) \geq 0$ positive semi definite

α-connection

$$\Gamma^{(\alpha)}_{ij,k} = \mathbb{E}_\theta\left[\left(\partial_i\partial_j \log p_\theta + \frac{1-\alpha}{2}\partial_i \log p_\theta \partial_j \log p_\theta\right)\partial_k \log p_\theta\right]$$

$$\left\langle \nabla^{(\alpha)}_{\partial_i}\partial_j, \partial_k \right\rangle = \Gamma^{(\alpha)}_{ij,k}$$

$$\nabla^{(\alpha)} = \frac{1-\alpha}{2}\nabla^{(0)} + \alpha\nabla^{(1)} + \frac{1+\alpha}{2}\nabla^{(1)} + \frac{1-\alpha}{2}\nabla^{(-1)}$$

$$\partial_k g_{ij} = \Gamma^{(0)}_{ki,j} + \Gamma^{(0)}_{kj,i}$$ → Riemannian metric for the Hellinger distance.

Flat statistical model

Exponential family: $p(x|\theta) = \exp\left(C(x) + \sum \theta^i F_i(x) - \Psi(\theta)\right)$

$$\Psi(\theta) = \log \int \exp\left[C(x) + \sum \theta^i F_i(x)\right] dx$$

Mixture family: $p(x|\theta) = C(x) + \sum \theta^i F_i(x)$

Exp: $C(x) = 0$ $\quad F_i(x) = \begin{cases} 1 & x = i \\ 0 & \end{cases}$, $\quad \theta^i = \log \frac{p_i}{p_0}$

$$\Psi(\theta) = \log\left(1 + \sum_i e^{\theta^i}\right)$$

$$\left.\begin{aligned} \partial_i \log p(x|\theta) &= F^i(x) - \partial_i \Psi(\theta) \\ \partial_i\partial_j \log p(x|\theta) &= -\partial_i\partial_j \Psi(\theta) \end{aligned}\right]$$ Exponential family are 1-flat

θ^i affine coordinate $\quad \nabla^{(1)}$ exponential connection

Property of Fisher metric

Sufficient statistics : model $p(x|\theta)$

$F: X \to Y$, $x \mapsto F(x)=y$ | $r(x|\theta) = \dfrac{p(x|\theta)}{q(F(x)|\theta)}$ → does not depend on θ for all x

$$p(x|y,\theta) = r(x|\theta)\,\delta_{F(x)}(y)$$

$k(y|x) \to g_k$, $g(\theta) - g_k(\theta) \leftarrow$ positive semi definite

$\| \Delta g(\theta)$

$$\Delta g(\theta) = \mathbb{E}_\theta\left[\partial_i \log r(x|\theta)\, \partial_j \log r(x|\theta)\right] \text{ if deterministic}$$

$$= \mathbb{E}_\theta\left[\mathrm{Cov}\left[\partial_i \log p(x|\theta)\, \partial_j \log p(x|\theta)\right] | y\right]$$

Monotonicity / Chain rule

Additivity : $g(\theta) = g_1(\theta) + g_2(\theta)$ if $p(x|\theta) = p(x_1|\theta)p(x_2|\theta)$

Convexity : $g_\lambda(\theta) \leq \lambda g_1(\theta) + (1-\lambda) g_2(\theta)$

Cramer Rao inequality

unbiased estimator $\mathbb{E}_\theta[\hat\theta(x)] = \theta \quad \forall\theta$

$$V_\theta^{ij} = \mathbb{E}_\theta\left[(\hat\theta^i(x) - \theta^i)(\hat\theta^j(x) - \theta^j)\right]$$

$V_\theta - g^{-1}(\theta) \geq 0$ positive semi definite