Information Geometry and Its Applications

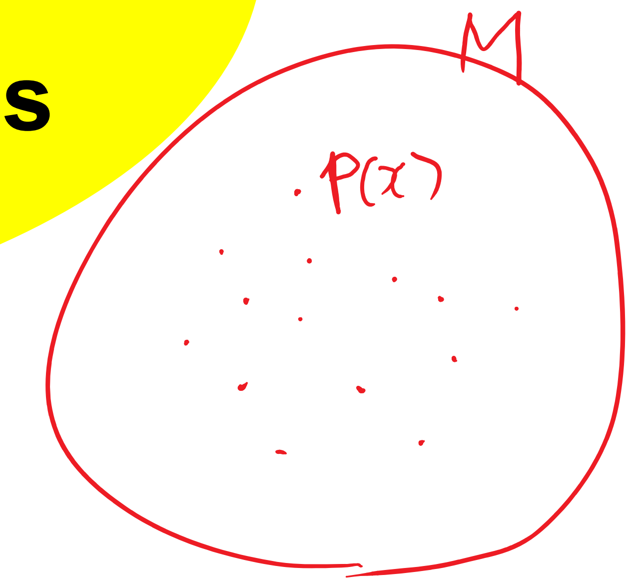
Shun-ichi Amari   RIKEN Brain Science Institute

1. Divergence Function and Dually Flat Riemannian Structure
2. Invariant Geometry on Manifold of Probability Distributions
3. Geometry and Statistical Inference
   semi-parametrics
4. Applications to Machine Learning and Signal Processing

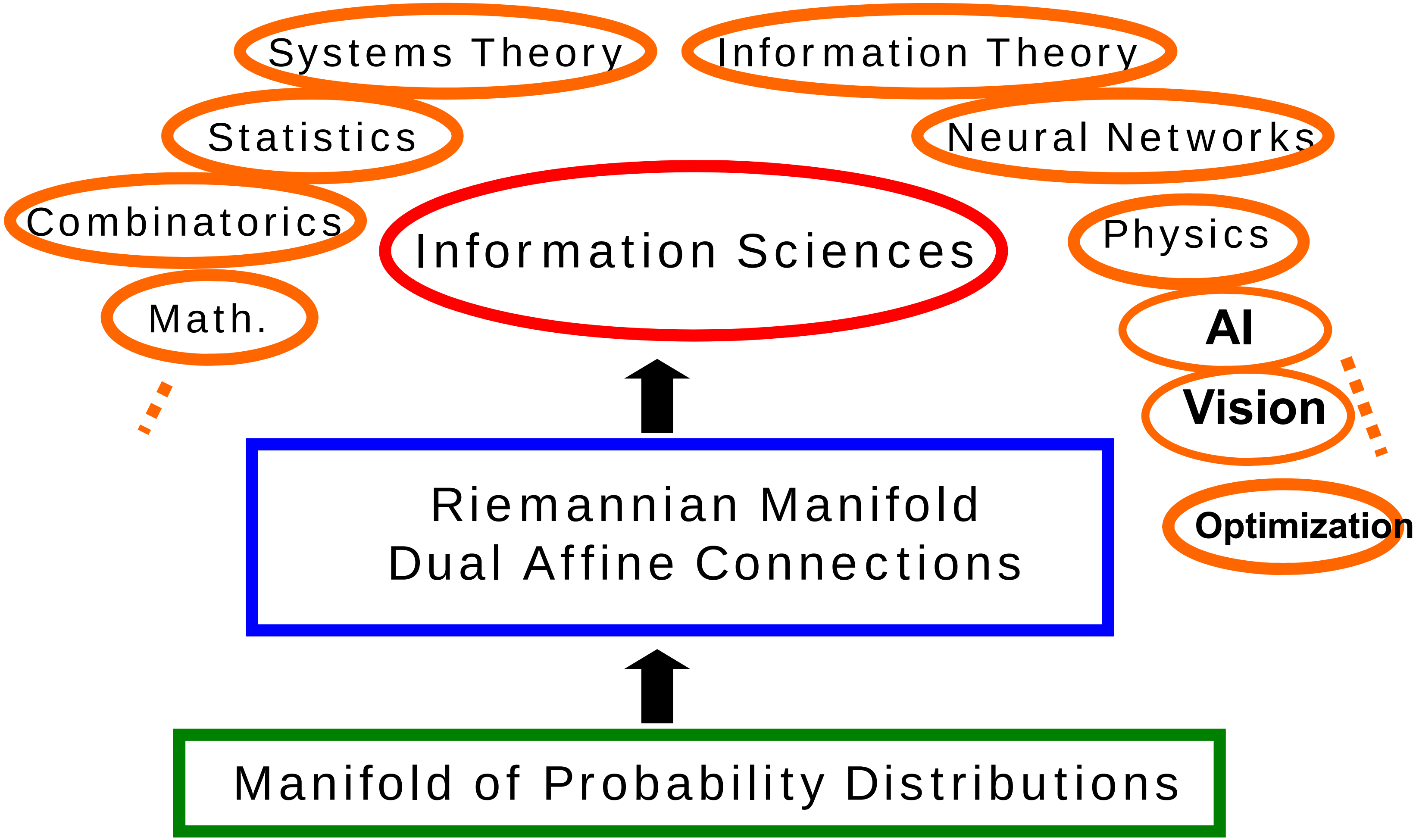
Information Geometry

-- Manifolds of Probability Distributions

$$M = \{ p(\mathbf{x}) \}$$



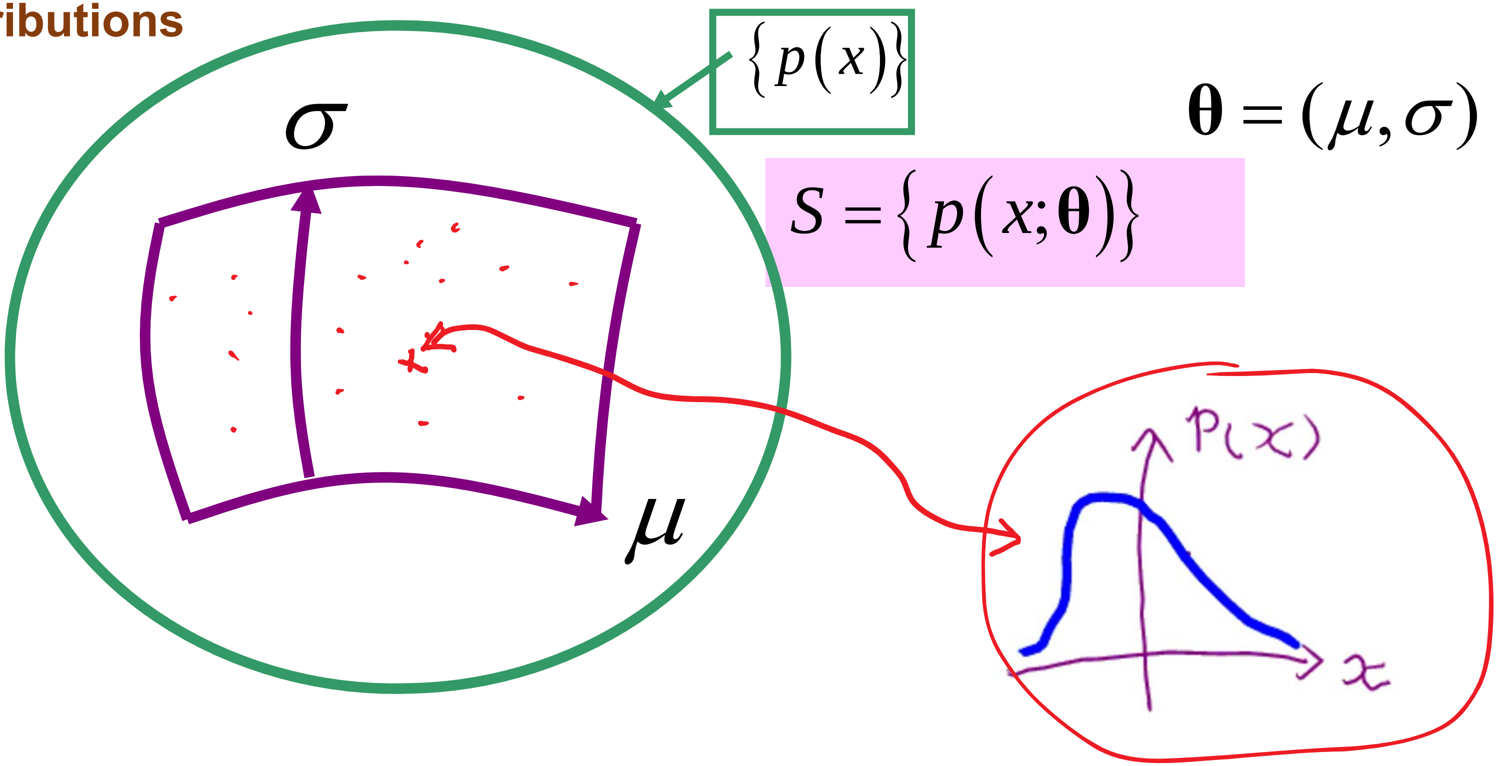
Information Geometry



Information Geometry ?

$$S = \{ p(x; \mu, \sigma) \} \qquad p(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\}$$

Gaussian distributions



$$\boldsymbol{\theta} = (\mu, \sigma)$$

$$S = \{ p(x; \boldsymbol{\theta}) \}$$

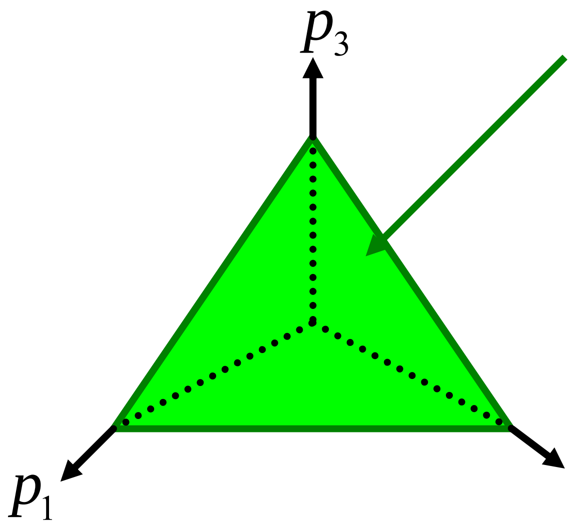
Manifold of Probability Distributions

$$x = 1, 2, 3 \qquad S_n = \{p(x)\}$$

$n = 3$

$$\boldsymbol{p} = (p_1, p_2, p_3) \qquad \boxed{p_1 + p_2 + p_3 = 1}$$

$M = \{p(x; \boldsymbol{\theta})\}$

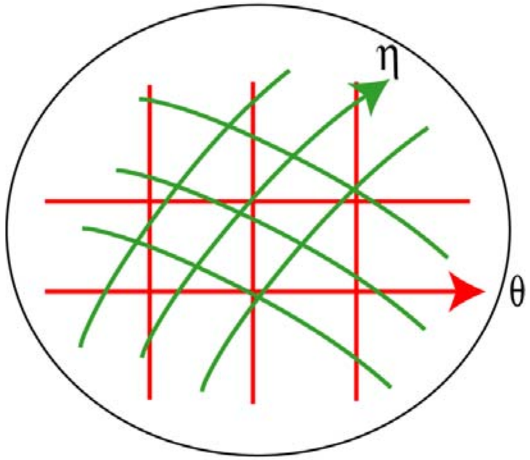


Manifold and Coordinate System

coordinate transformation

$\theta = (\theta_1, \cdots, \theta_n)$

$\eta = (\eta_1, \cdots, \eta_n)$



$\eta = \eta(\theta) \longleftrightarrow \theta = \theta(\eta)$

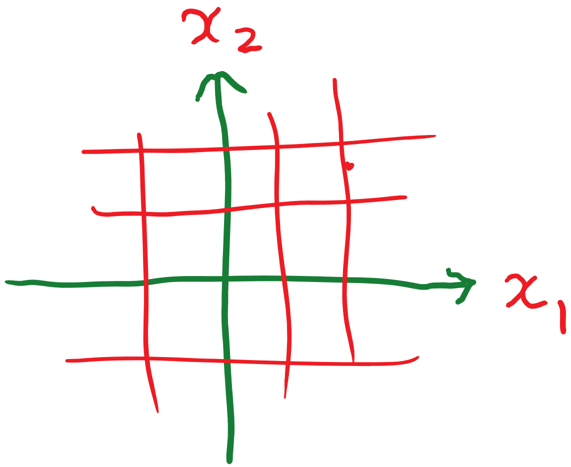
one-to-one

differentiable

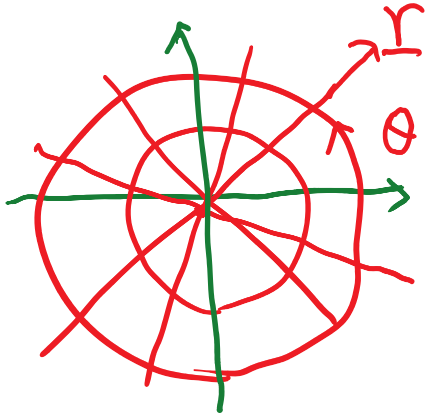
Examples of Coordinate systems

Euclidean space

$$\begin{cases} x_1 = r\cos\theta \\ x_2 = r\sin\theta \end{cases}$$



$$x = (x_1, x_2)$$



$$P = (r, \theta)$$

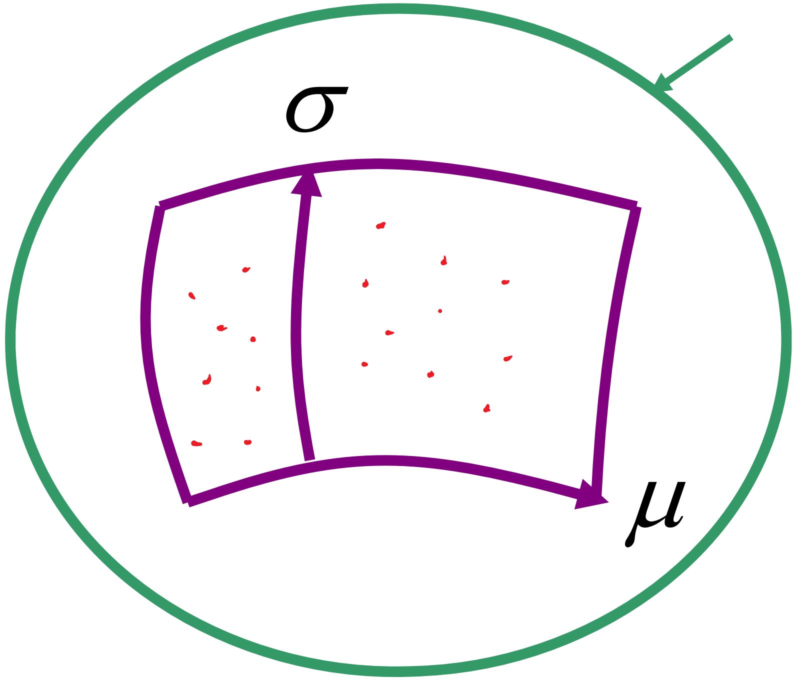
$$\begin{cases} r = \sqrt{x_1^2 + x_2^2} \\ \theta = \tan^{-1} \dfrac{x_2}{x_1} \end{cases}$$

Gaussian distributions

$\xi = (\mu, \sigma^2)$,

$\theta = \left(-\frac{1}{2\sigma^2}, \frac{\mu^2}{\sigma^2}\right)$,

$\eta = (\mu, \mu^2 + \sigma^2)$



$$S = \{p(x;\mu,\sigma)\} \qquad p(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

Discrete Distributions

$$S_n = \{ p(x) \}, \quad x = 0, 1, \cdots, n$$

$$p(x) = \sum_{i=0}^{n} p_i \, \delta_i(x)$$

$$= \sum_{i=1}^{n} p_i \, \delta_i(x) - p_0$$

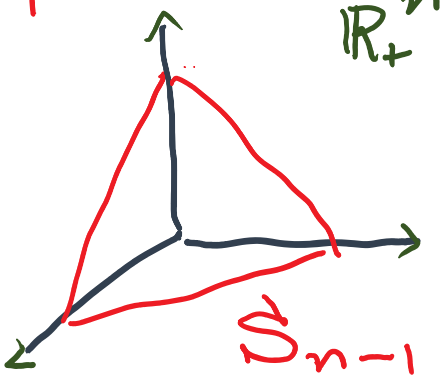
$$\eta = (p_1, \cdots, p_n)$$

$$\theta = \left(\log \frac{p_1}{p_0}, \cdots, \log \frac{p_n}{p_0} \right)$$

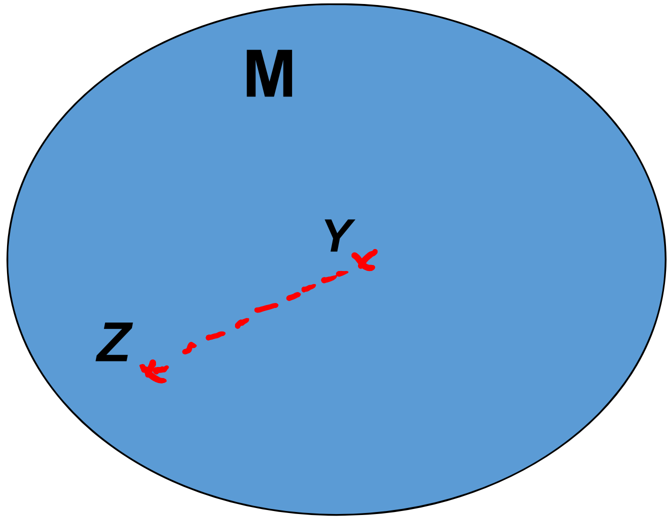
Positive measures

$$M_n : \; m(x) = \sum m_i \, \delta_i(x)$$

$$M_n = \mathbb{R}_+^n$$



Divergence: $D[z:y]$



$$D[\mathbf{z}:\mathbf{y}] \geq 0$$

$$D[\mathbf{z}:\mathbf{y}] = 0, \quad \text{iff } \mathbf{z}=\mathbf{y}$$

Not necessarily symmetric

D[z : y] = D[y : z]

$$D[\mathbf{z}:\mathbf{z}+d\mathbf{z}] = \frac{1}{2}\sum g_{ij}\, dz_i\, dz_j$$

positive-definite $G = (g_{ij})$

Taylor expansion

$$D(z : z+dz) = \frac{1}{2}\sum g_{ij}\, dz_i\, dz_j + \frac{1}{6}\sum K_{ijk}\, dz_i\, dz_j\, dz_k + \cdots$$

Various Divergences

Euclidean

$$D[x : y] = \frac{1}{2} \sum (x_i - y_i)^2$$

: symmetric

f-divergence

$$D[p(x) : q(x)] = \int p(x) f\left\{ \frac{q(x)}{p(x)} \right\} dx$$

f : convex function,

$f(1) = 0$

KL-divergence

(α-β)-divergence

Kullback-Leibler Divergence

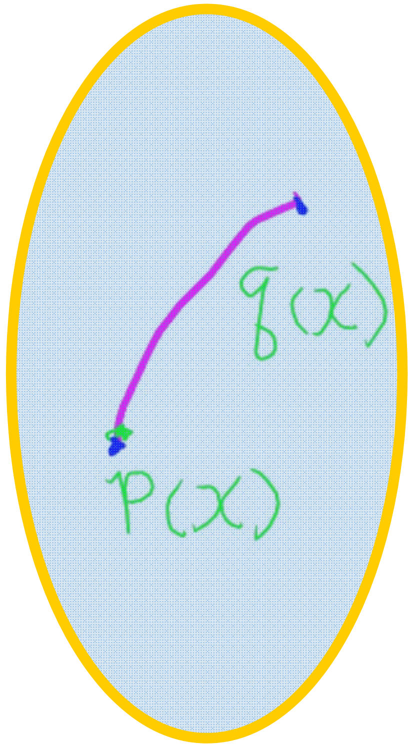
quasi-distance

$f(u) = -\log(u)$

$$D[p(x):q(x)] = \sum_x p(x) \log \frac{p(x)}{q(x)}$$

$$D[p(x):q(x)] \geq 0 \quad =0 \text{ iff } p(x) = q(x)$$

$$D[p:q] \neq D[q:p]$$



$(\alpha,\beta)-divergence$ $\quad$ $p,\ q \in S_n$

$$D_{\alpha,\beta}[p:q] = \sum_{i} \{ \frac{\alpha}{\alpha+\beta} p_i^{\alpha+\beta} + \frac{\beta}{\alpha+\beta} q_i^{\alpha+\beta} - p^{\alpha}{}_i q_i^{\beta} \}$$

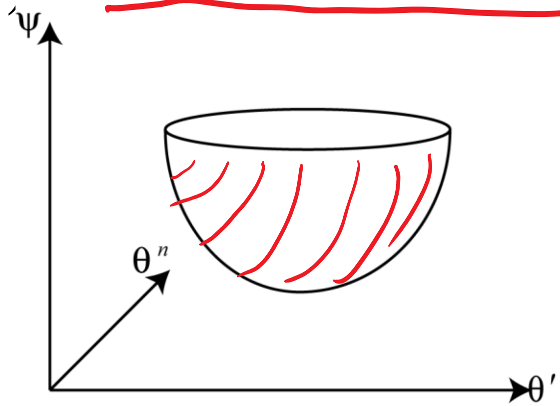
$\beta = -\alpha:\quad \alpha-$divergence

$\alpha = 1:\quad \beta$-divergence

Manifold with Convex Function

S : **coordinates** $\boldsymbol{\theta} = \left(\theta^1, \theta^2, \cdots, \theta^n\right)$

$\psi\left(\boldsymbol{\theta}\right)$: **convex function**



$$\psi\left(\boldsymbol{\theta}\right) = \frac{1}{2}\sum\left(\theta^i\right)^2$$

negative entropy

$$\varphi\left(p\right) = \int p\left(x\right)\log p\left(x\right)dx$$

energy

mathematical programming, control systems
physics, engineering, vision, economics

Riemannian metric and flatness (affine structure)

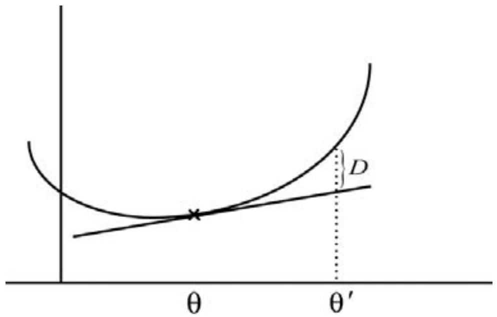
$$\{S,\ \psi(\theta),\ \theta\}$$

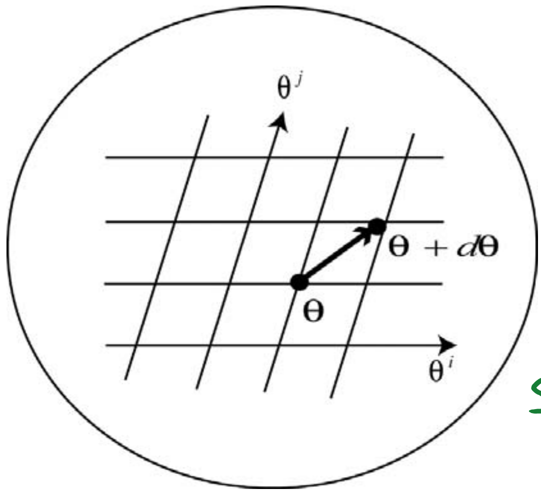
convex : $\tilde{\theta} = A\theta + c$

affine structure

Bregman divergence

$$D(\boldsymbol{\theta}',\boldsymbol{\theta}) = \psi(\boldsymbol{\theta}') - \psi(\boldsymbol{\theta}) - (\boldsymbol{\theta}' - \boldsymbol{\theta}) \cdot \operatorname{grad} \psi(\boldsymbol{\theta})$$





straight line

$$D(\boldsymbol{\theta},\boldsymbol{\theta}+d\boldsymbol{\theta}) = \frac{1}{2}\sum g_{ij}(\boldsymbol{\theta})\, d\theta^i d\theta^j$$

$$g_{ij} = \partial_i \partial_j \psi(\boldsymbol{\theta}), \quad \partial_i = \frac{\partial}{\partial \theta^i}$$

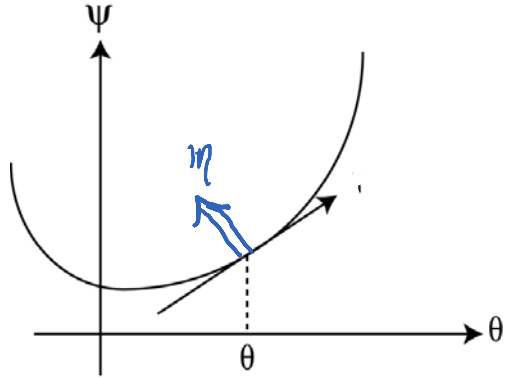
$\theta(t) = ta + b$

Flatness (affine) θ : **geodesic** **(not Levi-Civita)**

Legendre Transformation

dual coordinates θ, η

$$\eta_i = \partial_i \psi(\boldsymbol{\theta}), \quad \partial_i = \frac{\partial}{\partial \theta^i}$$



$\psi(\theta)$ $$\boldsymbol{\theta} \leftrightarrow \boldsymbol{\eta}$$ **one-to-one** $\varphi(\eta)$

$$\varphi(\boldsymbol{\eta}) + \psi(\boldsymbol{\theta}) - \theta_i \eta^i = 0$$

$$\theta^i = \partial^i \varphi(\boldsymbol{\eta}), \quad \partial^i = \frac{\partial}{\partial \eta_i}$$

$$\varphi(\eta) = \max_\theta \{\theta^i \eta_i - \psi(\theta)\}$$

$$D(\boldsymbol{\theta}, \boldsymbol{\theta}') = \psi(\boldsymbol{\theta}) + \varphi(\boldsymbol{\eta}') - \boldsymbol{\theta} \cdot \boldsymbol{\eta}'$$

: proof easy

Proof

$$D(\boldsymbol{\theta},\boldsymbol{\theta}')=\psi(\boldsymbol{\theta})+\varphi(\boldsymbol{\eta}')-\boldsymbol{\theta}\cdot\boldsymbol{\eta}'$$

$$D(\boldsymbol{\theta},\boldsymbol{\theta}')=\psi(\boldsymbol{\theta})-\psi(\boldsymbol{\theta}')-(\boldsymbol{\theta}-\boldsymbol{\theta}')\cdot\underbrace{\mathrm{grad}\ \psi(\boldsymbol{\theta}')}_{\eta'}$$

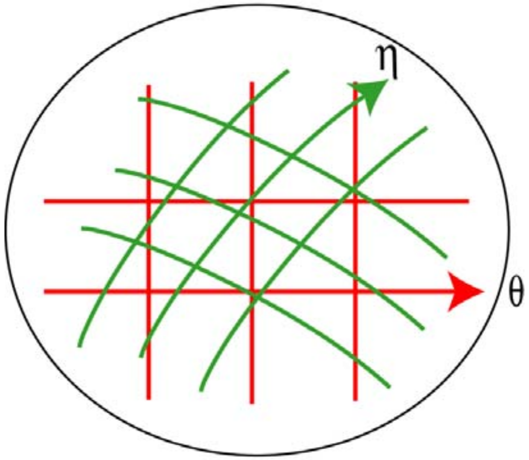
$$\varphi(\boldsymbol{\eta}')+\psi(\boldsymbol{\theta}')-\theta'_i\eta'^i=0$$

$$-\psi(\boldsymbol{\theta}')=\varphi(\boldsymbol{\eta}')-\theta'_i\eta'^i$$

Two affine coordinate systems (θ, η)

θ : geodesic (e-geodesic)

η : dual geodesic (m-geodesic)



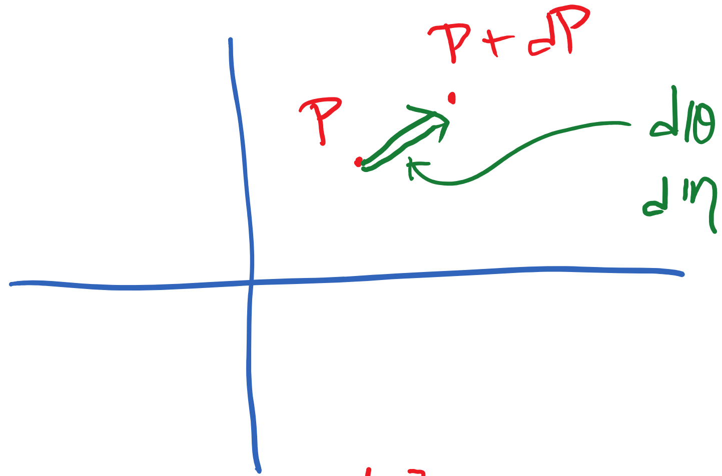
"dually orthogonal"

$$\langle \partial_i, \partial^j \rangle = \delta_i^j \quad - \quad \langle e_i, e_j \rangle$$

$$\partial_i = \frac{\partial}{\partial \theta^i}, \quad \partial^i = \frac{\partial}{\partial \eta_i}$$

$$X < Y, Z > = < \nabla_X Y, Z > + < Y, \nabla^*_X Z >$$

Bi-orthogonality



$$d\theta = \sum d\theta^i e_i$$

$$d\eta = \sum d\eta_i e^i$$

$$\eta_i = \partial_i \psi(\theta)$$

$$d\eta_i = \sum \partial_i \partial_j \psi(\theta)\, d\theta^j = \sum g_{ij}\, d\theta^j$$

$$d\theta^j = \sum g^{ij}\, d\eta_i$$

$$(g_{ij})^{-1} = \partial^i \partial^j \varphi(\eta)$$

$$ds^2 = \langle d\theta, d\theta \rangle$$

$$= \sum \underbrace{\langle e_i, e_j \rangle}_{g_{ij}} d\theta^i d\theta^j$$

$$e_i = \sum g_{ij} e^j$$

$$\langle e_i, e^j \rangle = \delta_i^j$$

Dually flat manifold

θ-coordinates $\leftrightarrow$ η-coordinates

potential functions $\psi(\theta), \varphi(\eta)$

$$g_{ij}(\theta) = \frac{\partial^2}{\partial\theta_i \partial\theta_j}\psi(\theta) \cdots g^{ij} = \frac{\partial^2}{\partial\eta_i \partial\eta_j}\varphi(\theta)$$

$$\psi(\theta) + \varphi(\eta) - \sum \theta_i \eta_i = 0$$

exponential family: $p(x,\theta) = \exp\left\{\sum \theta_i x_i - \psi(\theta)\right\}$

ψ : cumulant generating function

φ : negative entropy

canonical divergence D(P: P')$= \psi(\theta) + \varphi(\eta') - \sum \theta_i \eta_i'$

Exponential Family

$$p(x,\theta) = \exp\{\theta \cdot x - \psi(\theta)\}$$

$\psi(\theta)$: convex function, free-energy

Gaussian:

$e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

$$\left[\begin{array}{l} x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x \\ x^2 \end{pmatrix}, \quad \theta = \begin{pmatrix} \theta^1 \\ \theta^2 \end{pmatrix} = \begin{pmatrix} -\frac{1}{2\sigma^2} \\ \frac{\mu}{\sigma^2} \end{pmatrix} \\ \theta \cdot x = -\frac{(x-\mu)^2}{2\sigma^2} + c \end{array}\right.$$

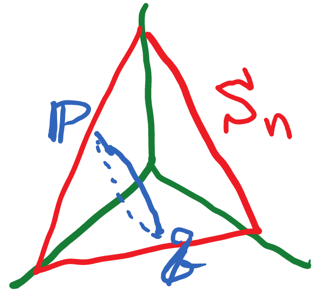
Negative entropy

$$\varphi(\eta) = \int p(x,\theta) \log p(x,\theta)$$

natural parameter : θ

expectation parameter : $\eta = E[x]$

x : discrete   X = {0, 1, …, n}



p

$$S_n = \{ p(x) \mid x \in X \}:$$ **exponential family**

$$p(x) = \sum_{i=0}^{n} p_i \delta_i(x) = \exp[\sum_{i=1}^{n} \theta^i x_i - \psi(\theta)] = \exp\left\{ \boldsymbol{\theta} \cdot \mathbf{x} - \psi\left(\boldsymbol{\theta}\right) \right\}$$

$\log r(t) = (1-t) \log p + t \log q$

$$\theta^i = \log(p_i / p_0); \qquad x_i = \delta_i(x); \qquad \psi(\theta) = -\log p_0$$

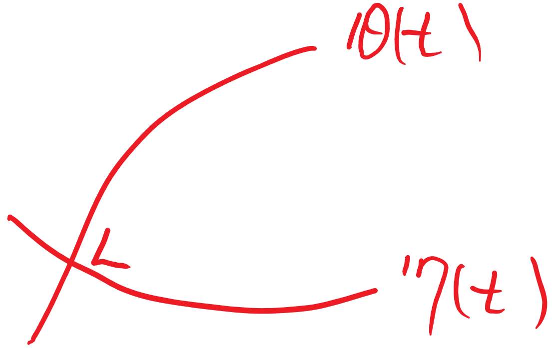
$$\eta_i = E[x_i] = p_i \qquad\qquad \varphi(\boldsymbol{\eta}) = \sum p_i \log p_i$$

$r(t) = (1-t) \cdot p + t q$

Two geodesics

$$\boldsymbol{\theta}(t) = t\,\mathbf{a} + \mathbf{b}$$

$$\boldsymbol{\eta}(t) = t\,\mathbf{a}' + \mathbf{b}$$



Tangent directions

$$\dot{\boldsymbol{\theta}}(t) = \sum \dot{\theta}^i \mathbf{e}_i$$

$$\dot{\boldsymbol{\eta}}(t) = \sum \dot{\eta}_j \mathbf{e}^j$$

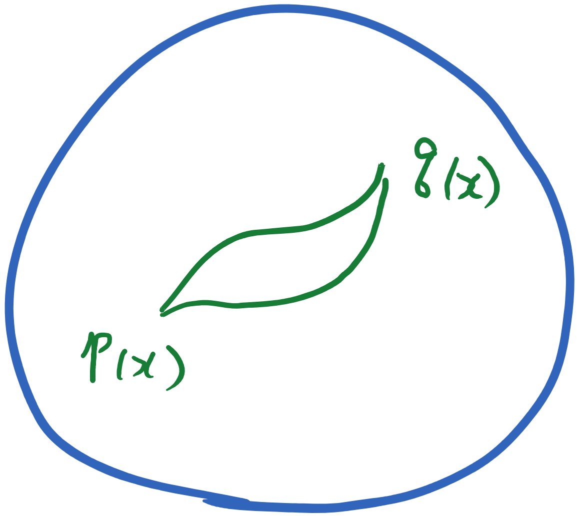
$$\langle \dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\eta}} \rangle = \sum \dot{\theta}^i \dot{\eta}_i = 0$$

orthogonal

Function space of probability distributions: topology

{p(x)}

$$\mathcal{F} = \{ p(x) > 0, \ \int p(x)\,dx = 1 \}$$



m-geodesic

$$r_m(x,t) = (1-t)\,p(x) + t\,q(x)$$

e-geodesic

$$\log r_e(x,t) = (1-t)\log p(x) + t \log q(x) + c(t)$$

Exponential Family

$$\mathcal{F} = \{ p(x) \}$$

$$p(x) = \exp\left\{ \int \theta(s)\, \delta(s - x)\, dx - \psi\{\theta\} \right\}$$

$$\theta(s) = \log p(s), \qquad \psi\{\theta\} = \log \int \exp\{\theta(s)\} ds$$

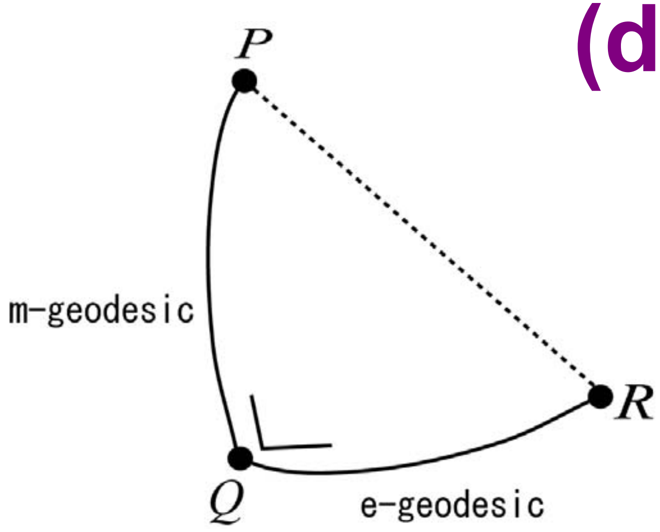
$$\eta(s) = E[\delta(s - x)] = p(s) = \nabla \psi(\theta)$$

$$D_{KL}[p : q] = \int p(x) \log \frac{p(x)}{q(x)}\, dx$$

$$g(s,t) = \nabla\nabla\psi = p(s)\, \delta(s-t)$$

Pythagorean Theorem

(dually flat manifold)



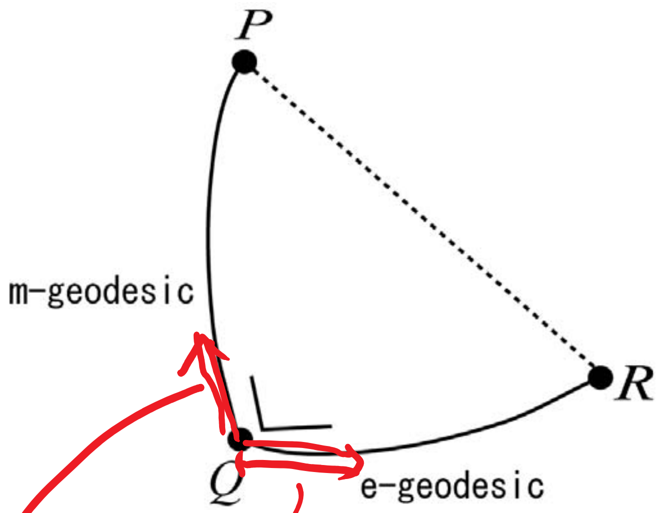
$$D[P:Q]+D[Q:R]=D[P:R]$$

proof

Euclidean space: self-dual $\theta = \eta$

$$\psi(\theta)=\frac{1}{2}\sum(\theta_i)^2$$

Proof



$$D[P:Q] = \psi\left(\boldsymbol{\theta}_P\right) + \varphi\left(\boldsymbol{\eta}_Q\right) - \boldsymbol{\theta}_P \cdot \boldsymbol{\eta}_Q$$

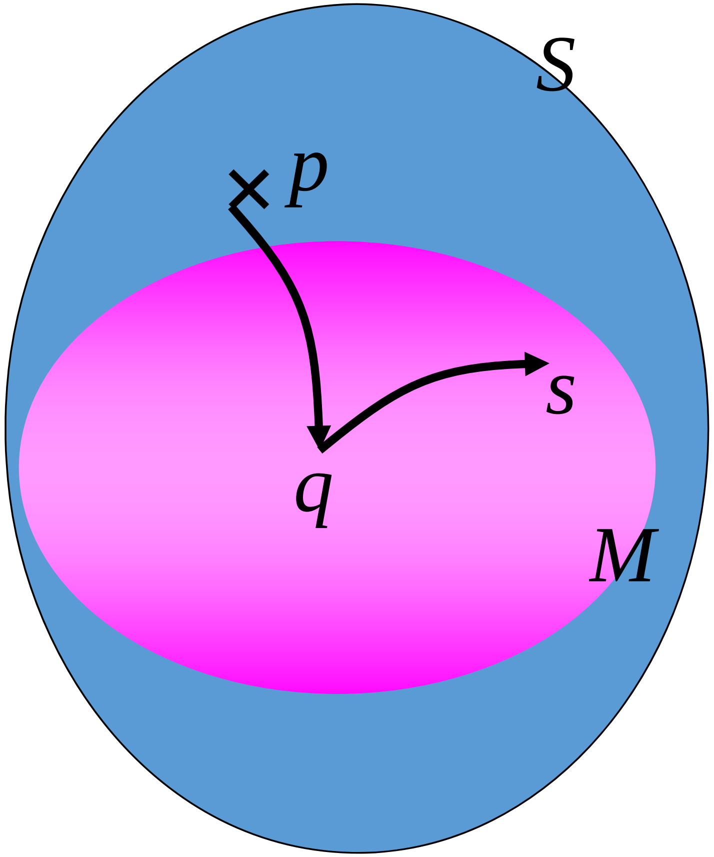
$$D\left[P:Q\right] + D\left[Q:R\right] - D\left[P:R\right] = (\boldsymbol{\eta}_P - \boldsymbol{\eta}_Q) \cdot (\boldsymbol{\theta}_Q - \boldsymbol{\theta}_R)$$

$$(\boldsymbol{\eta}_P - \boldsymbol{\eta}_Q) \cdot (\boldsymbol{\theta}_Q - \boldsymbol{\theta}_R) = 0$$

Projection Theorem

$$q = \arg\min_{s \in M} D[p : s]$$

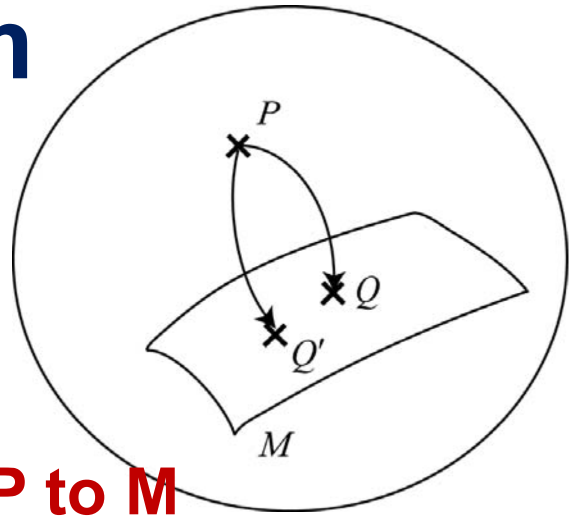
m-geodesic



$$q = \arg\min_{s \in M} D[s : p]$$

e-geodesic

Projection Theorem



$$\min_{Q \in M} D\left[P : Q\right]$$

Q = m-geodesic projection of P to M
unique when M is e-flat

$$\min_{Q \in M} D\left[Q : P\right]$$

Q’ = e-geodesic projection of P to M
unique when M is m-flat

Convex function – Bregman divergence

– Dually flat Riemannian divergence

$$\psi(\theta) \Longrightarrow D_\psi[\theta : \theta'] \Longrightarrow \{\theta, \eta\}, \quad g = \nabla\nabla\psi$$

Dually flat R-manifold – convex function – canonical divergence

KL-divergence

$$\{\theta, \eta\} \qquad \Longrightarrow \qquad \psi(\theta) \Rightarrow D_\psi[\theta : \theta']$$

$$g = \frac{\partial \eta}{\partial \theta} \qquad\qquad D_\psi[\theta : \theta'] = D_{KL}[p(x) : q(x)]$$

Exponential family – Bregman divergence

Banerjee et al

$$p(x, \theta) = \exp\{\theta \cdot x - \psi(\theta)\} \Rightarrow D_\psi(\theta : \theta')$$

$$D_\psi[\theta : \theta'] = \psi(\theta) + \varphi(x') - \theta \cdot x', \quad \eta' = x'$$

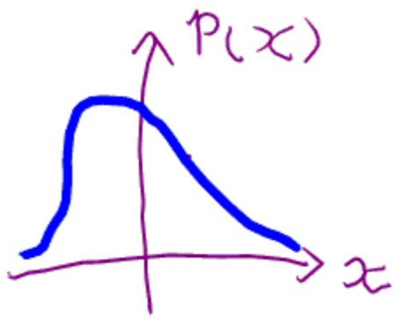
$$\Rightarrow p(x, \theta) = \exp\{-D_\psi[\theta : \theta(x)] + \psi(x)\}$$

Invariance

$$S = \left\{ p\left(x, \boldsymbol{\theta} \right) \right\}$$

Invariant under different representation of x

$$y = y\left(x\right), \qquad \overline{p}\left(y, \boldsymbol{\theta}\right)$$



$$\int \left| p\left(x, \theta_1\right) - p\left(x, \theta_2\right) \right|^2 dx$$

$$\neq \int \left| \overline{p}(y, \theta_1) - \overline{p}(y, \theta_2) \right|^2 dy$$

Invariant divergence

(manifold of probability distributions; $S=\{p(x,\xi)\}$)

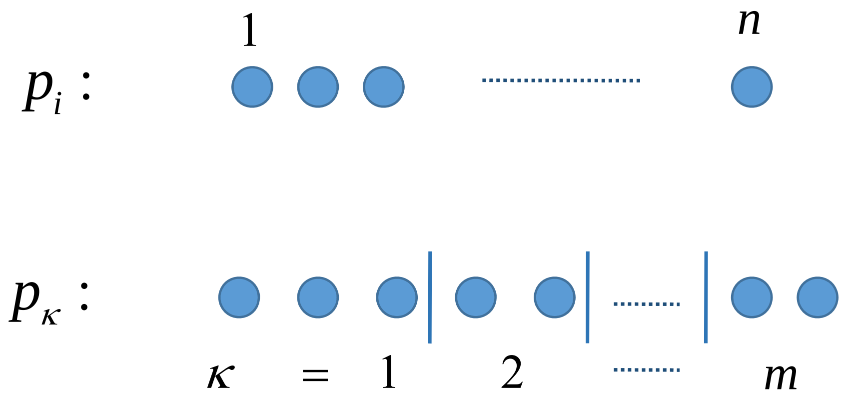
Chentsov
Amari -Nagaoka

$$y=k(x) \quad : \quad \text{sufficient statistics}$$

$$D\left[p_X(x):q_X(x)\right]=D\left[p_Y(y):q_Y(y)\right]$$

Invariance
--- characterization of *f*-divergence

Csiszar



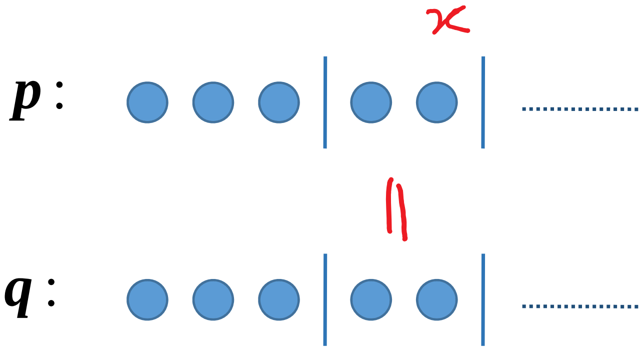
$$p^A = (p_\kappa) \qquad\qquad p_\kappa = \sum_{i \in A_\kappa} p_i$$

$p \to p^A$

$$D[\boldsymbol{p}:\boldsymbol{q}] \ge D\left[\boldsymbol{p}^A : \boldsymbol{q}^A\right]$$

$$D[\boldsymbol{p}:\boldsymbol{q}] = D\left[\boldsymbol{p}^A : \boldsymbol{q}^A\right]$$

$$\Leftrightarrow \ p_i = c_\kappa q_i \ ; i \in A_\kappa$$



Invariance ⇒ *f*-divergence

Csiszar *f*-divergence

Ali-Silvey
Morimoto

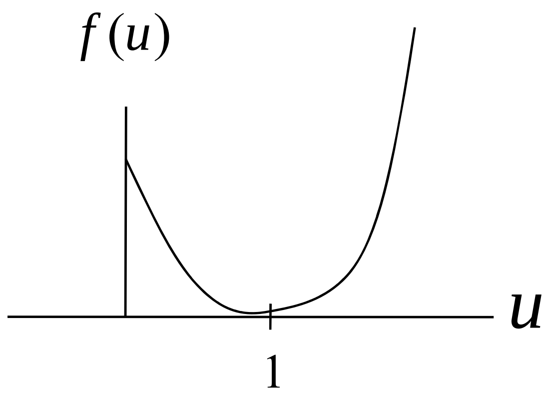
$$D_f[\boldsymbol{p}:\boldsymbol{q}] = \sum p_i f\left(\frac{q_i}{p_i}\right),$$

$$f(u): \text{ convex}, \qquad f(1) = 0,$$

$$D_{cf}[\boldsymbol{p}:\boldsymbol{q}] = cD_f[\boldsymbol{p}:\boldsymbol{q}]$$

$$\tilde{f}(u) = f(u) - c(u-1)$$

$$f(1) = f'(1) = 0\ ;\ f''(1) = 1$$



Theorem

An invariant separable divergence belongs to the class of f-divergence.

Separable divergence:

$$D[\mathbf{p} : \mathbf{q}] = \sum k(p_i, q_i)$$

$$k(p_i, q_i) = p_i f\left(\frac{q_i}{p_i}\right)$$

divergence

(n > 1)

$S = \{\mathbf{p}\}$: space of probability distributions

invariance

dually flat space

invariant divergence

Flat divergence

convex functions
Bregman

F-divergence
Fisher inf metric
Alpha connection

KL-divergence

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

α -Divergence: why?

flat & invariant in $\tilde{S}_{n+1}$ $= M_{n+1}$

$$f_\alpha(u) = \frac{4}{1-\alpha^2}\{1-u^{\frac{1+\alpha}{2}}\} - \frac{2}{1-\alpha}(1-u), \quad \alpha \neq 1$$

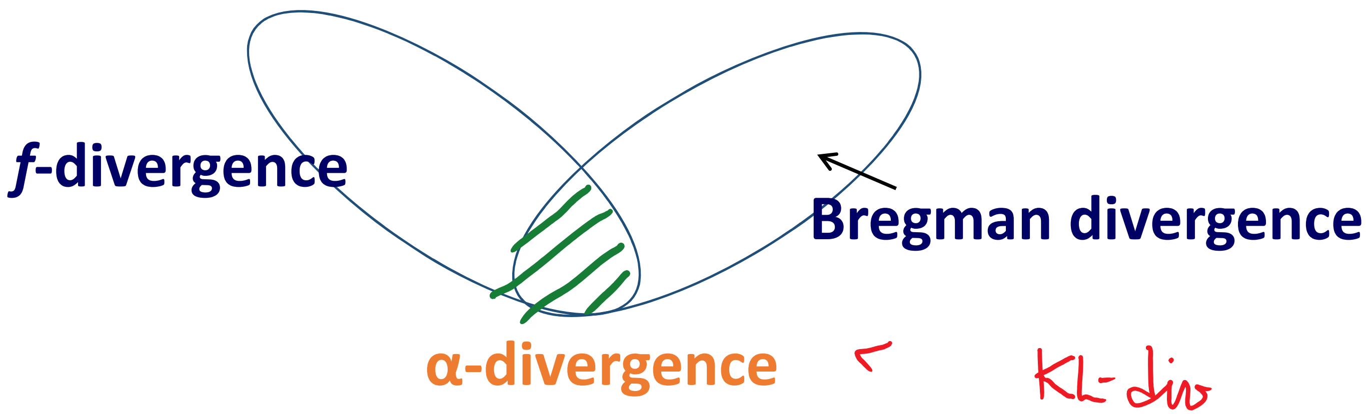
KL-divergence

$$f(u) = u \log u - (u-1)$$

$$D[\tilde{p} : \tilde{q}] = \sum \{\tilde{p}_i \log \frac{\tilde{p}_i}{\tilde{q}_i} + \tilde{p}_i - \tilde{q}_i\}$$

Space of positive measures : vectors, matrices, arrays

$$\tilde{S} = \{\tilde{\boldsymbol{p}}\}, \qquad \tilde{p}_i > 0 : (\sum \tilde{p}_i = 1 \; nn \text{ holds})$$



f divergence of $\tilde{S}$

$$D_f\left[\tilde{\boldsymbol{p}}:\tilde{\boldsymbol{q}}\right]=\sum \tilde{p}_i f\left(\frac{\tilde{q}_i}{\tilde{p}_i}\right)\geq 0$$

$$D_f\left[\tilde{\boldsymbol{p}}:\tilde{\boldsymbol{q}}\right]=0 \Leftrightarrow \tilde{\boldsymbol{p}}=\tilde{\boldsymbol{q}}$$

not invariant under $\tilde{f}(u)=f(u)-c(u-1)$

α divergence

$$D_\alpha[\tilde p : \tilde q] = \sum \left\{ \frac{1-\alpha}{2}\tilde p_i + \frac{1+\alpha}{2}\tilde q_i - \underbrace{\tilde p_i^{\frac{1-\alpha}{2}}}_{=\,\theta_i} \underbrace{\tilde q_i^{\frac{1+\alpha}{2}}}_{=\,\eta_i} \right\}$$

KL-divergence

$$D[\tilde p : \tilde q] = \sum \left\{ \tilde p_i \log \frac{\tilde p_i}{\tilde q_i} + \tilde p_i - \tilde q_i \right\}$$

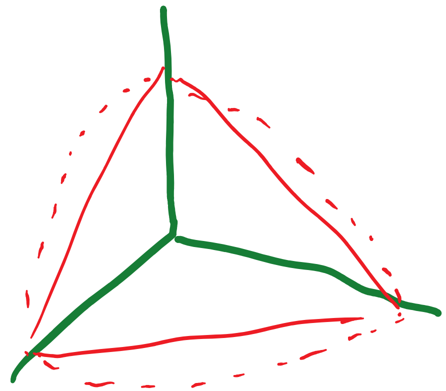
α

$M = \tilde{S}$: dually flat

S : not dually flat (except $\alpha = \pm 1$)

$$\sum p_i = 1$$

$$\sum r_i^{\frac{2}{1-\alpha}} = 1$$



Metric and Connections Induced by Divergence

(Eguchi)

Riemannian metric

$$g_{ij}(\boldsymbol{z}) = \partial_i \partial_j D[\boldsymbol{z} : \boldsymbol{y}]_{|\boldsymbol{y}=\boldsymbol{z}} \quad : D[\boldsymbol{z} : \boldsymbol{y}] = \frac{1}{2} g_{ij}(\boldsymbol{z})(\mathrm{z}_i - \mathrm{y}_i)(\mathrm{z}_j - \mathrm{y}_j)$$

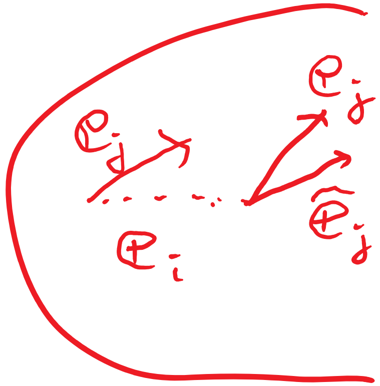
affine connections

$$\{\nabla, \nabla^*\}$$

$$\nabla_{e_i} e_j = \sum \Gamma_{ij}^{k} e_k$$

$$\Gamma_{ijk}(\boldsymbol{z}) = -\partial_i \partial_j \partial'_k D[\boldsymbol{z} : \boldsymbol{y}]_{|\boldsymbol{y}=\boldsymbol{z}}$$

$$\partial_i = \frac{\partial}{\partial z_i}, \quad \partial'_i = \frac{\partial}{\partial y_i}$$



$$\Gamma^*_{ijk}(\boldsymbol{z}) = -\partial'_i \partial'_j \partial_k D[\boldsymbol{z} : \boldsymbol{y}]_{|\boldsymbol{y}=\boldsymbol{z}}$$

Invariant geometrical structure

$$S = \{ p(x, \xi) \}$$

alpha-geometry

(derived from invariant divergence)

$$g_{ij}(\xi) = E\left[\partial_i l \partial_j l \right]$$ **Fisher information**

$$T_{ijk}(\xi) = E\left[\partial_i l \partial_j l \partial_k l \right]$$

$$l = \log p(x, \xi); \qquad \partial_i = \frac{\partial}{\partial \xi^i}$$

α-connection

$$\Gamma^{\alpha}_{ijk} = \{ i, j; k \} - \alpha T_{ijk}$$

Levi-civita:

$$\nabla^{\alpha} \leftrightarrow \nabla^{-\alpha}$$: dually coupled

$$X \langle Y, Z \rangle = \langle \nabla^{\alpha}{}_X Y, Z \rangle + \langle Y, \nabla^{-\alpha}{}_X Z \rangle$$

Duality: $X < Y, Z > = < \nabla_X Y, Z > + < Y, \nabla^*{}_X Z >$

$$\partial_k g_{ij} = \Gamma_{kij} + \Gamma^*_{kji}$$

$$\Gamma_{ijk} = \Gamma^*_{ijk} - T_{ijk}$$

$$\boxed{\{M, g, T\}}$$

g_{ij} , T_{ijk}

Riemannian Structure

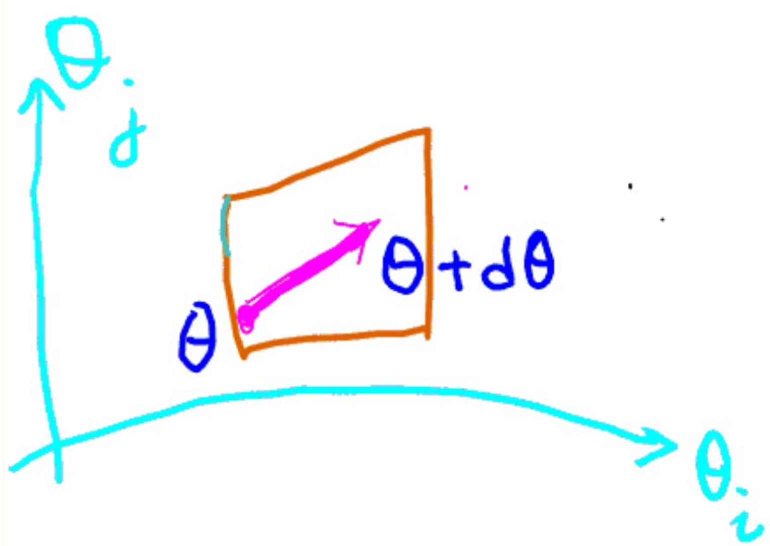
$$ds^2 = \sum g_{ij}(\theta) d\theta^i d\theta^j$$

$$= d\theta^T G(\theta) d\theta$$

$$G(\theta) = (g_{ij})$$

Euclidean $\quad G = E$

Fisher information

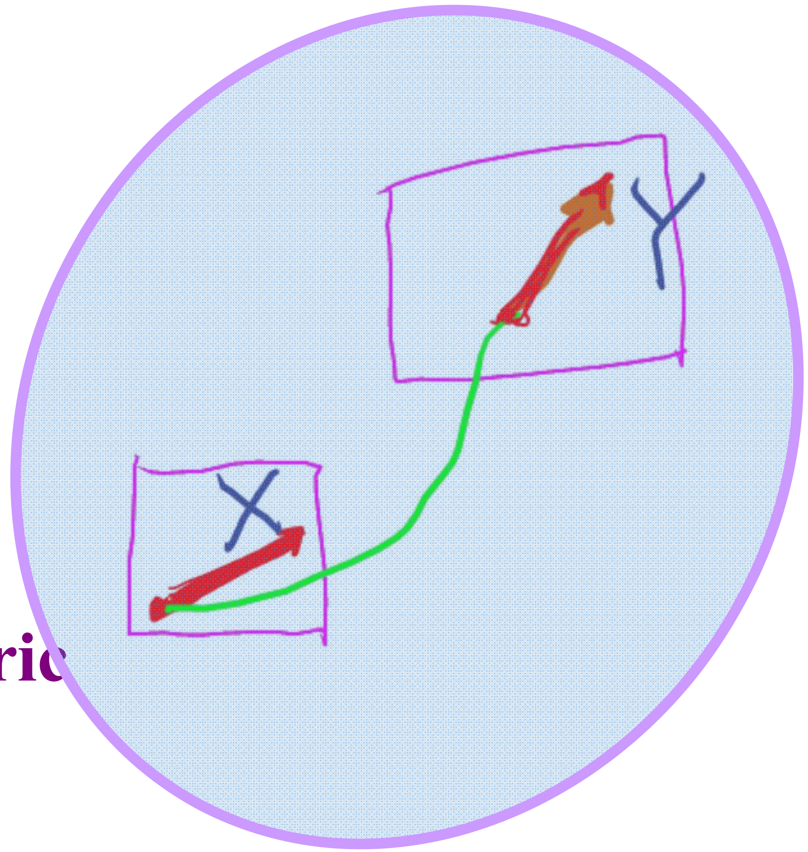


Affine Connection

covariant derivative

$$\nabla_X Y, \quad \Pi_c X = Y$$

geodesic $\nabla_{\dot{X}} \dot{X} = 0, \quad$ X=X(t)



$$s = \int \sqrt{\sum g_{ij}(\theta) d\theta^i d\theta^j}$$

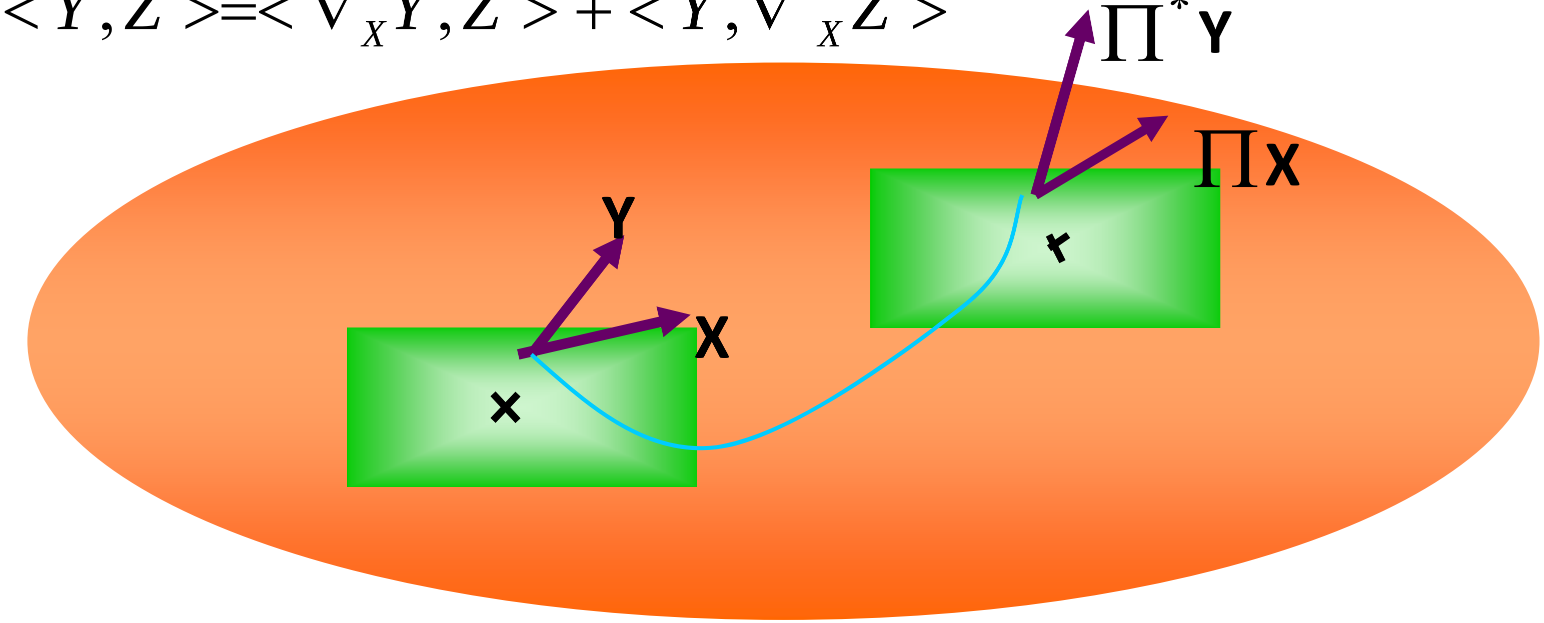
minimal distance $\Rightarrow$ non-metric

straight line

Duality

$$\langle X, Y \rangle = \langle \Pi X, \Pi^* Y \rangle \qquad < X, Y > = \sum g_{ij} X^i Y^j$$

$$X < Y, Z > = < \nabla_X Y, Z > + < Y, \nabla^*_X Z >$$



Riemannian geometry: $\Pi = \Pi^*$

Dual Affine Connections

$$(\nabla, \nabla^*)$$

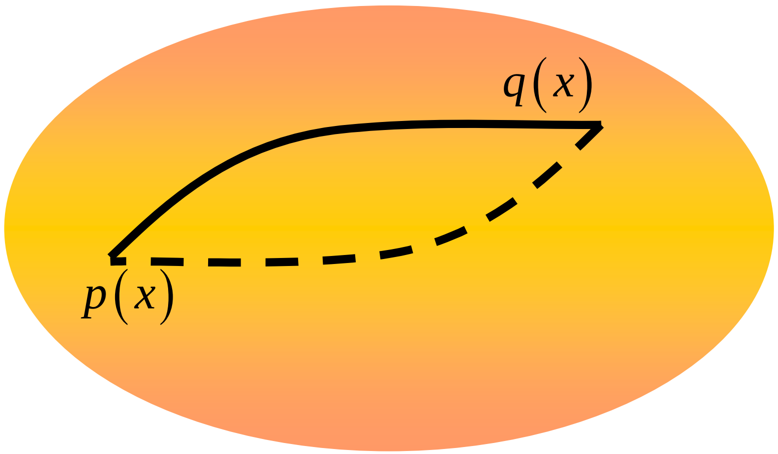
$$(\Pi, \Pi^*)$$

e-geodesic

$$\log r(x,t) = t \log p(x) + (1-t) \log q(x) + c(t)$$

m-geodesic

$$r(x,t) = tp(x) + (1-t)q(x)$$



Mathematical structure of $S = \{ p(x, \xi) \}$

$$g_{ij}(\xi) = E\left[\partial_i l \partial_j l\right]$$

$$T_{ijk}(\xi) = E\left[\partial_i l \partial_j l \partial_k l\right]$$

$$\{\mathbf{M}, \mathbf{g}, \mathbf{T}\}$$

$$l = \log p(x, \xi); \qquad \partial_i = \frac{\partial}{\partial \xi^i}$$

α **-connection**

$$\Gamma^{\alpha}_{ijk} = \{i, j; k\} - \alpha T_{ijk}$$

$\nabla^{\alpha} \leftrightarrow \nabla^{-\alpha}$: **dually coupled**

$$X \langle Y, Z \rangle = \langle \nabla_X Y, Z \rangle + \langle Y, \nabla^*_X Z \rangle$$

α-geometry

$$\Gamma^{(\alpha)} = \Gamma^{(0)} - \frac{\alpha}{2} T$$

α-geodesic

α-projection

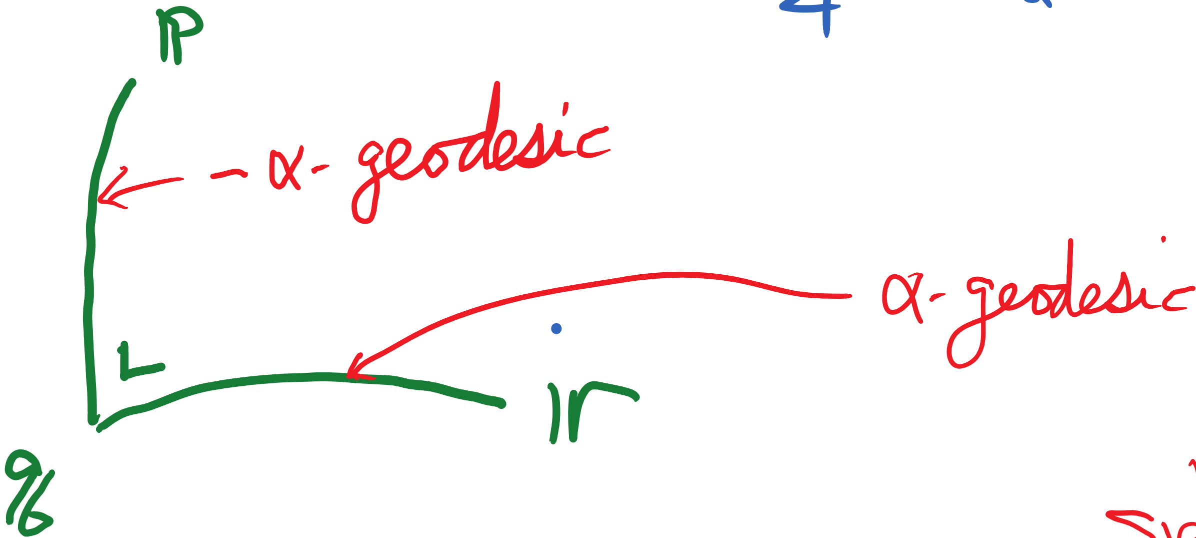
not flat $(\alpha \neq \pm 1)$

extended Pythagorean theorem

α-mean, α-family of prob. distributions

Tsallis q-entropy

$$D_\alpha[p:r] = D_\alpha[p:q] + D_\alpha[q:r] + \frac{1-\alpha^2}{4} D_\alpha[p:q] D_\alpha[q:r]$$



Kurose

spherical geometry
(constant curvature)

Dual Foliations

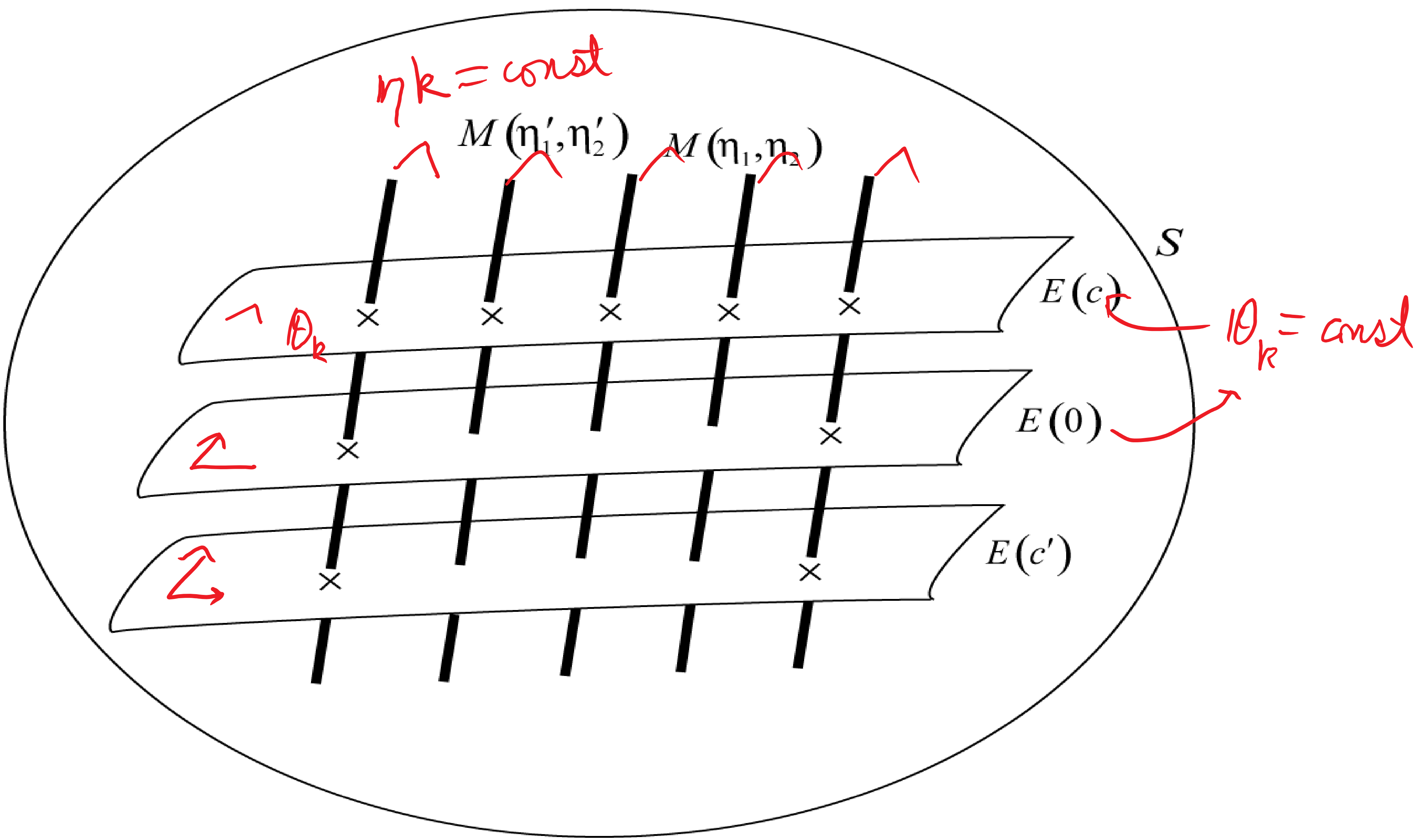
$$\theta = (\theta_1, \cdots, \theta_k ; \theta_{k+1}, \cdots, \theta_n)$$

$$\eta = (\eta_1, \cdots, \eta_k ; \eta_{k+1}, \cdots, \eta_n)$$

k-cut

$$\xi = (\theta_k : \eta^k)$$

mixed coordinates

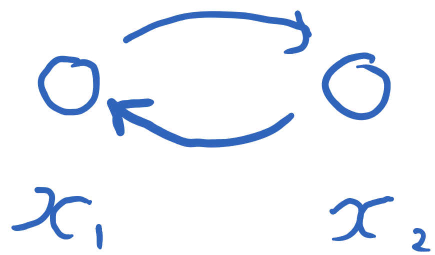


$\eta_k = \text{const}$
$M(\eta_1', \eta_2')$
$M(\eta_1, \eta_2)$
S
$E(c)$
$\theta_k = \text{const}$
θ_k
$E(0)$
$E(c')$

Two neurons: $\{p_{00}, p_{01}, p_{10}, p_{11}\}$

x_1 0011000101101

x_2 0100100110100

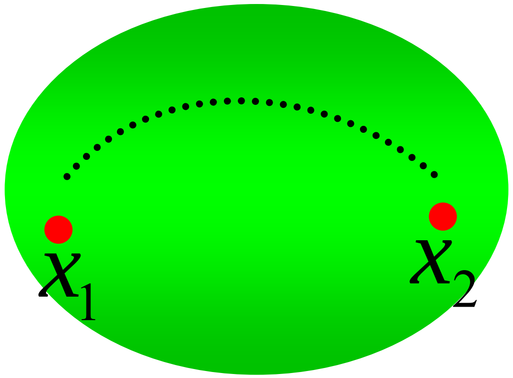


(x_3 0101101001010)

firing rates: $r_1, r_2; r_{12}$

correlation—covariance?

Correlations of Neural Firing



$\{p(x_1, x_2)\}$

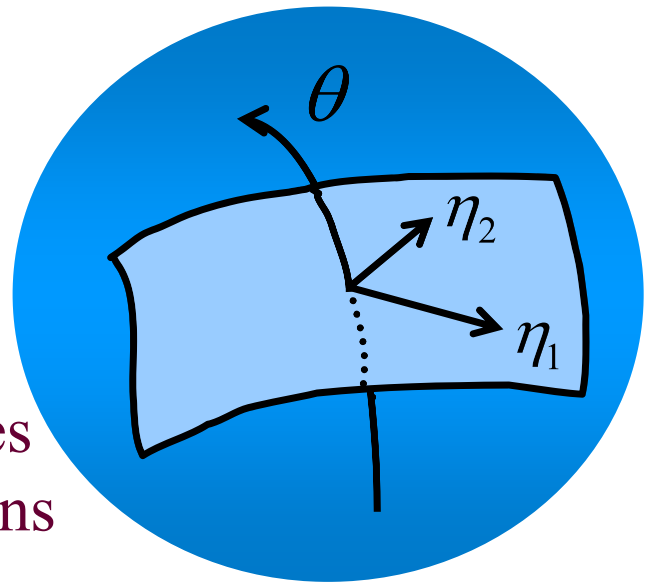
$\{p_{00}, p_{10}, p_{01}, p_{11}\}$

$r_1 = p_{1\cdot} = p_{10} + p_{11}$

$r_2 = p_{\cdot 1} = p_{01} + p_{11}$

$$\theta = \log \frac{p_{11} p_{00}}{p_{10} p_{01}}$$

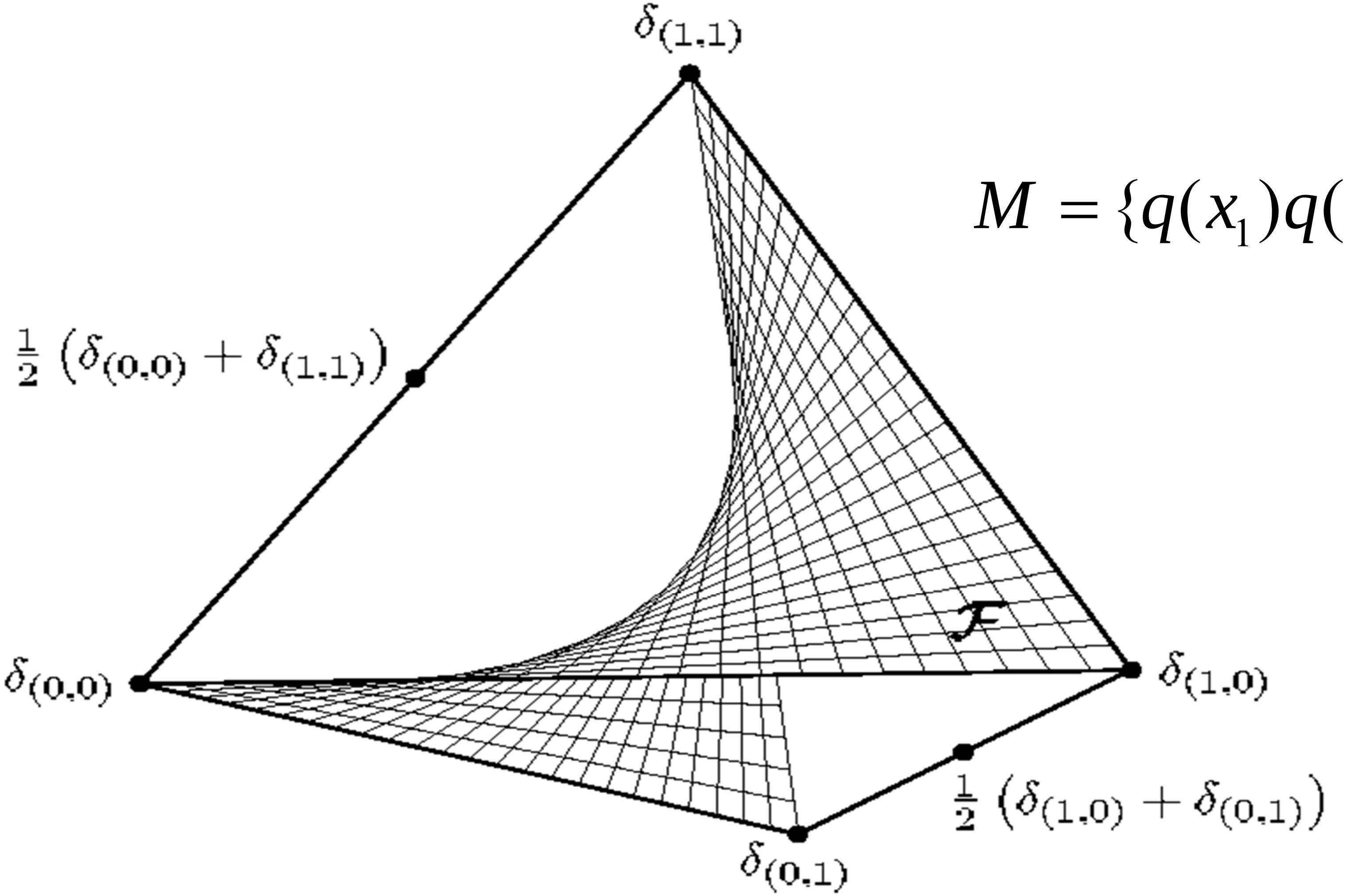
firing rates
correlations



$\{(r_1, r_2), \theta\}$

orthogonal coordinates

Independent Distributions



$$M = \{q(x_1)q(x_2)\}$$

two neuron case

$$r_1, r_2, r_{12}; \quad \theta_1, \theta_2, \theta_{12}$$

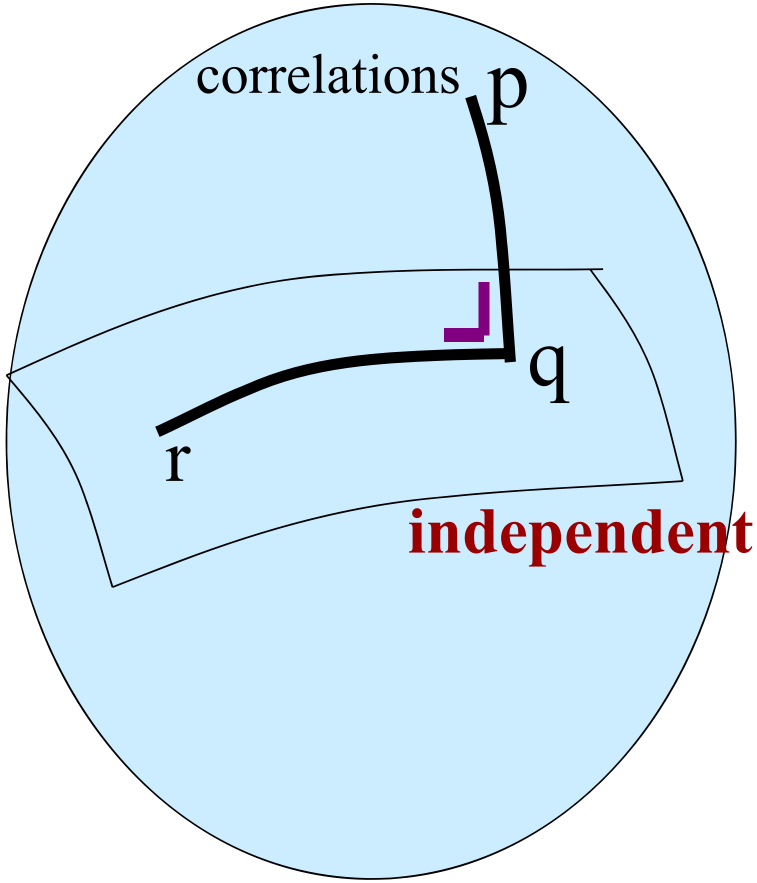
$$\theta_{12} = \log \frac{p_{00} p_{11}}{p_{01} p_{10}} = \log \frac{r_{12}\left(1 + r_{12} - r_1 - r_2\right)}{\left(r_1 - r_{12}\right)\left(r_2 - r_{12}\right)}$$

$$r_{12} = f\left(r_1, r_2, \theta\right)$$

$$r_{12}\left(t\right) = f\left(r_1\left(t\right), r_2\left(t\right), \theta\right)$$

Decomposition of KL-divergence

D[p:r] = D[p:q]+D[q:r]



p,q: same marginals η_1, η_2

r,q: same correlations θ

$$D[p:r] = \sum_x p(x) \log \frac{p(x)}{q(x)}$$

pairwise correlations

covariance: $c_{ij} = r_{ij} - r_i r_j$ **not orthogonal**

independent distributions

$$r_{ij} = r_i r_j, \qquad r_{ijk} = r_i r_j r_k, \cdots$$

How to generate correlated spikes?
(Niebur, Neural Computation [2007])

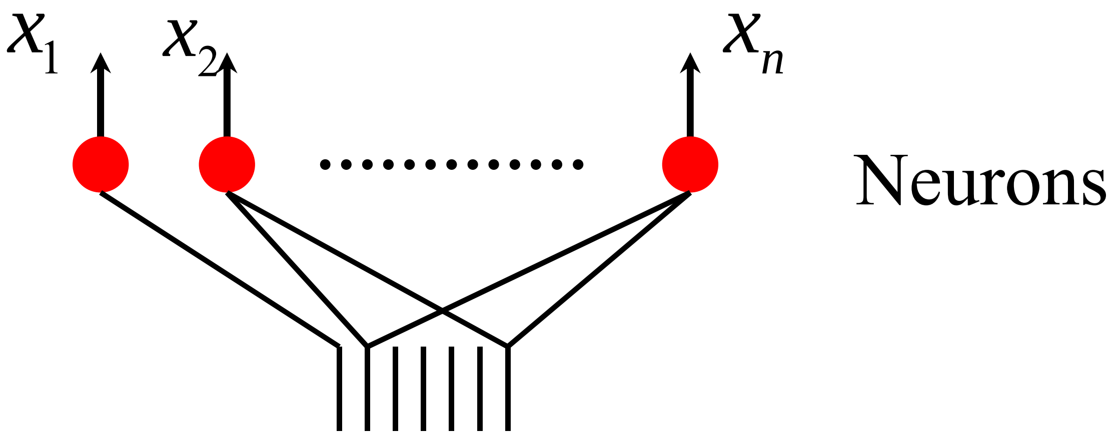
higher-order correlations

Orthogonal higher-order correlations

$$\boldsymbol{\theta} = \left(\theta_i, \theta_{ij}; \quad \cdots, \theta_{1\cdots n}\right)$$

$$\boldsymbol{r} = \left(r_i, r_{ij}; \quad \cdots, r_{1\cdots n}\right)$$

Population and Synfire



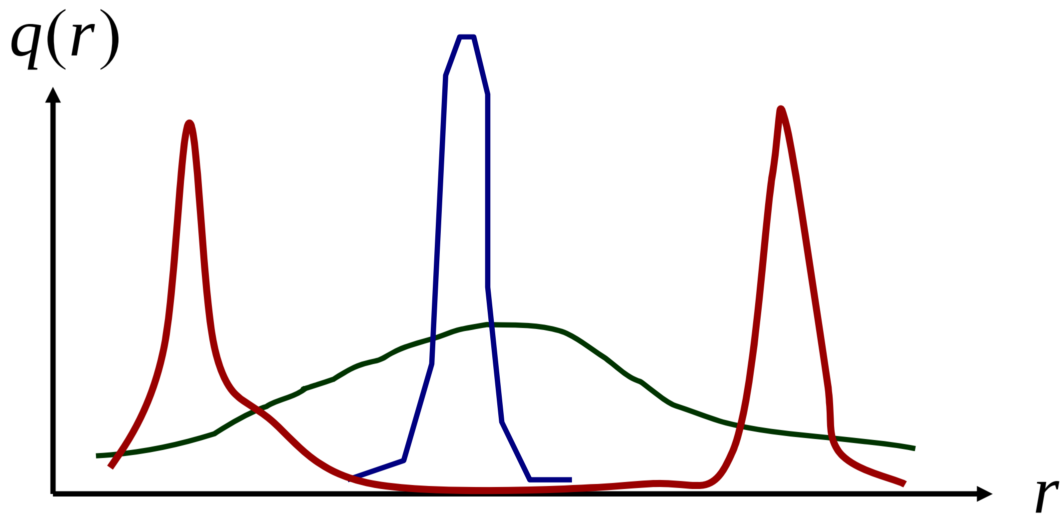
$$x_i = 1(u_i)$$

$$u_i = \text{Gaussian} \qquad E[u_i u_j] = \alpha$$

Synfiring

$$p(\boldsymbol{x}) = p(x_1, \ldots, x_n)$$

$$r = \frac{1}{n}\sum x_i \qquad\qquad q(r)$$



Input-output Analysis

Gross product consumption
Relations among industires
(K. Tsuda and R. Morioka)

n industries

$$(A_{ij}) = \begin{array}{c|cccc} & 1 & \cdots & & n \\ \hline 1 & & & & \\ \vdots & & \left[\begin{array}{cccc} A_{11} & A_{12} & & A_{1n} \\ & \ddots & & \\ & & \ddots & \\ A_{n1} & & & A_{n} \end{array}\right] & & \\ n & & & & \end{array}$$

product

$$n_k \left(\begin{array}{l} A_{i\cdot} = \sum_j A_{ij} \\ A_{\cdot j} = \sum_i A_{ij} \end{array} \right)$$

consumption

$$S_{ij} = \log \frac{A_{ij}\ A_{\cdot\cdot}}{A_{i\cdot}\ A_{\cdot j}}$$

10_k

Mathematical Problems

M: **submanifold of S_n?** n : large yes

Hong van Le

{M, g} **{M, g, T} dually flat: J. Armstrong** no

Affine differential geometry

Hessian manifold

Almost complex structure