Geometry of affine immersions and construction of geometric divergences

Hiroshi Matsuzoe
Nagoya Institute of Technology

Information Geometry And Its Applications IV
In honor of Professor Amari's 80th birthday

1. Affine immersions
2. Statistical manifolds and generalized conformal structures
3. Deformed exponential families
4. Geometric divergences and α-divergences

Appendixes

5. Generalization of Legendre transformation
6. Quantum analogue of affine differential geometry

1 Affine immersions

1.1 Affine immersions

$f : M \to R^{n+1}$: an immersion

ξ: a local vector field along f

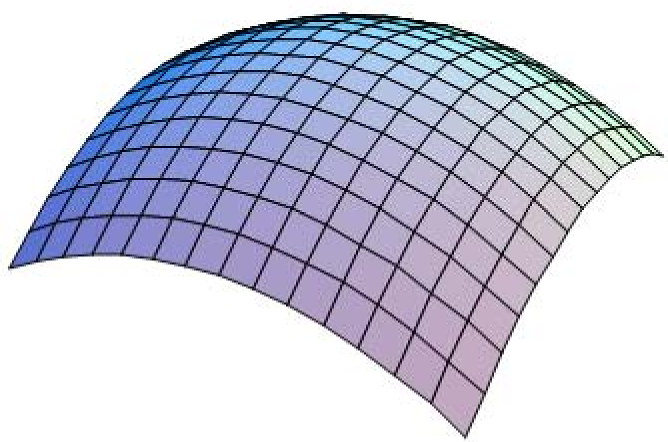
Definition 1.1

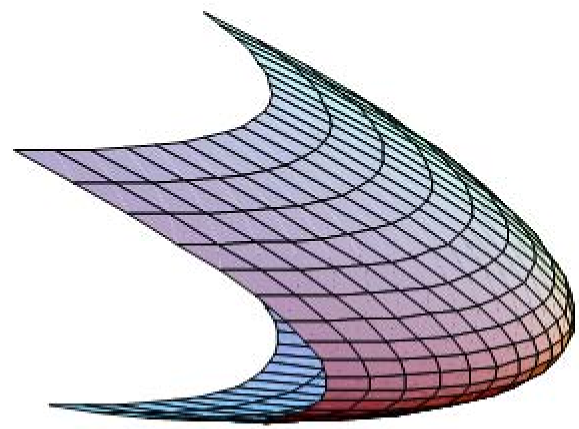
$\{f, \xi\} : M \to R^{n+1}$ is an affine immersion

$\overset{\text{def}}{\Longleftrightarrow}$ For an arbitrary point $p \in M$,

$$T_{f(p)}R^{n+1} = f_*(T_pM) \oplus R\{\xi_p\}$$

ξ: a transversal vector field





$\tilde{f} = Af + b$
$\tilde{\xi} = A\xi$

1 Affine immersions

1.1 Affine immersions

$f : M \to R^{n+1}$: an immersion

ξ: a local vector field along f

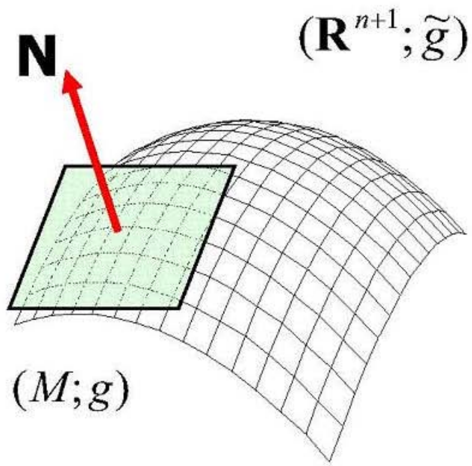
Definition 1.1

$\{f, \xi\} : M \to R^{n+1}$ is an affine immersion

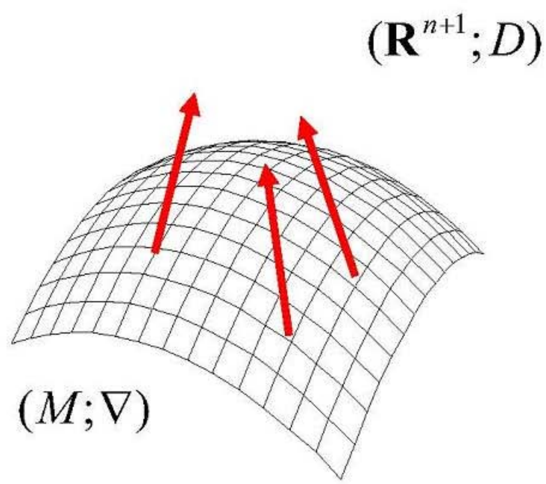
$\overset{\text{def}}{\iff}$ For an arbitrary point $p \in M$,

$$T_{f(p)}R^{n+1} = f_*(T_pM) \oplus R\{\xi_p\}$$

ξ: a transversal vector field



Riemannian geometry



Affine geometry

1 Affine immersions

1.1 Affine immersions

$f : M \to R^{n+1}$: an immersion

ξ: a local vector field along f

Definition 1.1

$\{f, \xi\} : M \to R^{n+1}$ is an affine immersion

$\overset{\text{def}}{\iff}$ For an arbitrary point $p \in M$,

$$T_{f(p)}R^{n+1} = f_*(T_pM) \oplus R\{\xi_p\}$$

ξ: a transversal vector field

D: the standard flat affine connection on R^{n+1}

$$\begin{aligned} D_X f_* Y &= f_*(\nabla_X Y) + h(X, Y)\xi, \\ D_X \xi &= -f_*(SX) + \tau(X)\xi. \end{aligned}$$

$\{f, \xi\}$, $\{\tilde{f}, \tilde{\xi}\}$: affine immersions

$\nabla = \tilde{\nabla}$, $h = \tilde{h}$, $S = \tilde{S}$, $\tau = \tilde{\tau}$

$\iff$ $\{f, \xi\}$, $\{\tilde{f}, \tilde{\xi}\}$ are affinely congruent.

1 Affine immersions

1.1 Affine immersions

$f : M \to R^{n+1}$: an immersion

ξ: a local vector field along f

Definition 1.1

$\{f, \xi\} : M \to R^{n+1}$ is an affine immersion

$\overset{\text{def}}{\Longleftrightarrow}$ For an arbitrary point $p \in M$,

$$T_{f(p)}R^{n+1} = f_*(T_pM) \oplus R\{\xi_p\}$$

ξ: a transversal vector field

D: the standard flat affine connection on R^{n+1}

$$\begin{aligned} D_X f_* Y &= f_*(\nabla_X Y) + h(X, Y)\xi, \\ D_X \xi &= -f_*(SX) + \tau(X)\xi. \end{aligned}$$

∇ : the induced connection
h : the affine fundamental form
S : the affine shape operator
τ : the transversal connection form

1.2 Equiaffine structures and statistical manifolds

D: the standard flat affine connection on R^{n+1}

$$D_X f_* Y = f_*(\nabla_X Y) + h(X,Y)\xi,$$
$$D_X \xi = -f_*(SX) + \tau(X)\xi.$$

f : non-degenerate $\overset{\text{def}}{\Longleftrightarrow}$ h : non-degenerate

$\{f,\xi\}$: equiaffine $\overset{\text{def}}{\Longleftrightarrow}$ $\tau = 0$

ω : the induced volume element (n-form) with respect to $\{f,\xi\}$

$$\overset{\text{def}}{\Longleftrightarrow} \quad \omega(X_1,\dots,X_n) := \det(f_*X_1,\dots,f_*X_n,\xi),$$

where "det" is the standard volume element on R^{n+1}.

∇, τ, ω : induced objects from $\{f,\xi\}$

$$\Longrightarrow \quad (\nabla_Y \omega)(X_1,\dots,X_n) = \tau(Y)\omega(X_1,\dots,X_n)$$

$\tau = 0 \iff \omega$ is parallel with respect to ∇.

(ω: a uniform distribution)

Proposition 1.2

$\{f, \xi\}$: non-degenerate, equiaffine $\Longrightarrow$ (M, ∇, h) is a statistical manifold, 1–conformally flat.

Proposition 1.3

(M, ∇, h) : a simply connected statistical manifold 1-conformally flat

$\Longrightarrow$ There exists $\{f, \xi\}$ which realizes (M, ∇, h) in R^{n+1}.

Fundamental structural equations for affine immersions

Gauss: $$R(X,Y)Z = h(Y,Z)SX - h(X,Z)SY$$

Codazzi: $$(\nabla_X h)(Y,Z) + \tau(X)h(Y,Z) = (\nabla_Y h)(X,Z) + \tau(Y)h(X,Z)$$

$$(\nabla_X S)(Y) - \tau(X)SY = (\nabla_Y S)(X) - \tau(Y)SX$$

Ricci: $$h(X,SY) - h(Y,SX) = (\nabla_X \tau)(Y) - (\nabla_Y \tau)(X)$$

f : non-degenerate $\overset{\text{def}}{\Longleftrightarrow}$ h : non-degenerate

$\{f, \xi\}$: equiaffine $\overset{\text{def}}{\Longleftrightarrow}$ $\tau = 0$

2 Statistical manifolds

M : a manifold (an open domain in R^n)
h : a (semi-) Riemannian metric on M
∇ : an affine connection on M

Definition 2.1 (Kurose)
We say that the triplet (M, ∇, h) is a statistical manifold

$$\overset{\text{def}}{\Longleftrightarrow} \quad (\nabla_X h)(Y, Z) = (\nabla_Y h)(X, Z).$$

$C(X, Y, Z) := (\nabla_X h)(Y, Z)$, the cubic form,
the Amari-Chentsov tensor field

Definition 2.2
∇^*: the dual connection of ∇ with respect to h

$$\overset{\text{def}}{\Longleftrightarrow} \quad Xh(Y, Z) = h(\nabla^*_X Y, Z) + h(Y, \nabla_X Z).$$

(M, ∇^*, h): the dual statistical manifold of (M, ∇, h).

Remark 2.3 (Original definition by S.L. Lauritzen)
(M, g) : a Riemannian manifold
C : a totally symmetric $(0, 3)$-tensor field
We call the triplet (M, g, C) a statistical manifold.

Example 2.4 (Normal distributions) $\quad (\; l(x;\xi) = \log p(x,\xi) \;)$

$$M = \{ p(x;\xi) \mid \xi = (\xi^1, \xi^2) = (\mu, \sigma), \quad p(x;\xi) = \frac{1}{\sqrt{2\pi(\xi^2)^2}} \exp\left[-\frac{(x-\xi^1)^2}{2(\xi^2)^2} \right] = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2} \right] \}$$

We regard that M is a manifold with local coordinates (μ, σ).

$$g_{ij} = \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial \xi^i} \log p(x,\xi) \right) \left(\frac{\partial}{\partial \xi^j} \log p(x,\xi) \right) p(x,\xi) dx$$

$$= E\left[\frac{\partial l}{\partial \xi^i} \frac{\partial l}{\partial \xi^j} \right] \quad \left(g = -\frac{1}{\sigma^2} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \right)$$ the Fisher information

$$C_{ijk} = E\left[\frac{\partial l}{\partial \xi^i} \frac{\partial l}{\partial \xi^j} \frac{\partial l}{\partial \xi^k} \right]$$ the cubic form or the Amari-Chentsov tensor field

$$\Gamma_{ij,k} = E\left[\frac{\partial^2 l}{\partial \xi^i \partial \xi^j} \frac{\partial l}{\partial \xi^k} \right] = \Gamma^{(0)}_{ij,k} - \frac{1}{2} C_{ijk}$$ ($\nabla^{(0)}$: the Levi-Civita connection w.r.t. g)

$$\Gamma^*_{ij,k} = E\left[\frac{\partial^2 l}{\partial \xi^i \partial \xi^j} \frac{\partial l}{\partial \xi^k} + \frac{\partial l}{\partial \xi^i} \frac{\partial l}{\partial \xi^j} \frac{\partial l}{\partial \xi^k} \right] = \Gamma^{(0)}_{ij,k} + \frac{1}{2} C_{ijk}$$

(M, ∇, g) and (M, ∇^*, g) are statistical manifolds.

2.2 Conformal-Projective structures

Definition 2.5
(M, ∇, h) and $(M, \bar{\nabla}, \bar{h})$ are conformally-projectively equivalent
$\overset{\text{def}}{\Longleftrightarrow}$ There exist two functions ϕ and ψ such that

$$\bar{h}(X, Y) = e^{\phi+\psi} h(X, Y),$$
$$\bar{\nabla}_X Y = \nabla_X Y - h(X, Y)\mathrm{grad}_h \psi + d\phi(Y)\, X + d\phi(X)\, Y$$

(M, ∇, h) and $(M, \bar{\nabla}, \bar{h})$ are α-conformally equivalent
$\overset{\text{def}}{\Longleftrightarrow}$ There exist a function ϕ such that

$$\bar{h}(X, Y) = e^{\phi} h(X, Y),$$
$$\bar{\nabla}_X Y = \nabla_X Y - \frac{1+\alpha}{2} h(X, Y)\mathrm{grad}_h \phi + \frac{1-\alpha}{2}\{d\phi(Y)\, X + d\phi(X)\, Y\}$$

2.2 Conformal-Projective structures

Definition 2.5
(M, ∇, h) and $(M, \bar{\nabla}, \bar{h})$ are conformally-projectively equivalent
$\overset{\text{def}}{\Longleftrightarrow}$ There exist two functions ϕ and ψ such that

$$\begin{aligned} \bar{h}(X, Y) &= e^{\phi+\psi} h(X, Y), \\ \bar{\nabla}_X Y &= \nabla_X Y - h(X, Y)\mathrm{grad}_h \psi + d\phi(Y)\, X + d\phi(X)\, Y \end{aligned}$$

Remark 2.6 (In the case $\phi = \psi$)
(M, g), $(M, \bar{g})$: Riemannian manifolds
∇^0, $\bar{\nabla}^0$: their Levi-Civita connections
If g and $\bar{g}$ are conformally equivalent, i.e. $\bar{g}(X, Y) = e^{2\phi} g(X, Y)$

$$\Longrightarrow \quad \bar{\nabla}^0_X Y = \nabla^0_X Y - h(X, Y)\mathrm{grad}_h \phi + d\phi(Y)\, X + d\phi(X)\, Y$$

(M, ∇^0, g) and $(M, \bar{\nabla}^0, \bar{g})$ are 0-conformally equivalent.

Definition 2.5

(M, ∇, h) and $(M, \bar{\nabla}, \bar{h})$ are conformally-projectively equivalent

$\overset{\text{def}}{\Longleftrightarrow}$ There exist two functions ϕ and ψ such that

$$\begin{aligned} \bar{h}(X, Y) &= e^{\phi+\psi} h(X, Y), \\ \bar{\nabla}_X Y &= \nabla_X Y - h(X, Y)\mathrm{grad}_h \psi + d\phi(Y)\, X + d\phi(X)\, Y \end{aligned}$$

Remark 2.7

ψ is constant: $\Longrightarrow$ $\bar{\nabla}_X Y = \nabla_X Y + d\phi(Y)\, X + d\phi(X)\, Y$

∇ and $\bar{\nabla}$ are projectively equivalent.

(M, ∇, h) and $(M, \bar{\nabla}, \bar{h})$ are (-1)-conformally equivalent

ϕ is constant: $\Longrightarrow$ $\bar{\nabla}_X Y = \nabla_X Y - h(X, Y)\mathrm{grad}_h \psi$

∇ and $\bar{\nabla}$ are dual-projectively equivalent.

(∇^* and $\bar{\nabla}^*$ are projectively equivalent.)

(M, ∇, h) and $(M, \bar{\nabla}, \bar{h})$ are 1-conformally equivalent

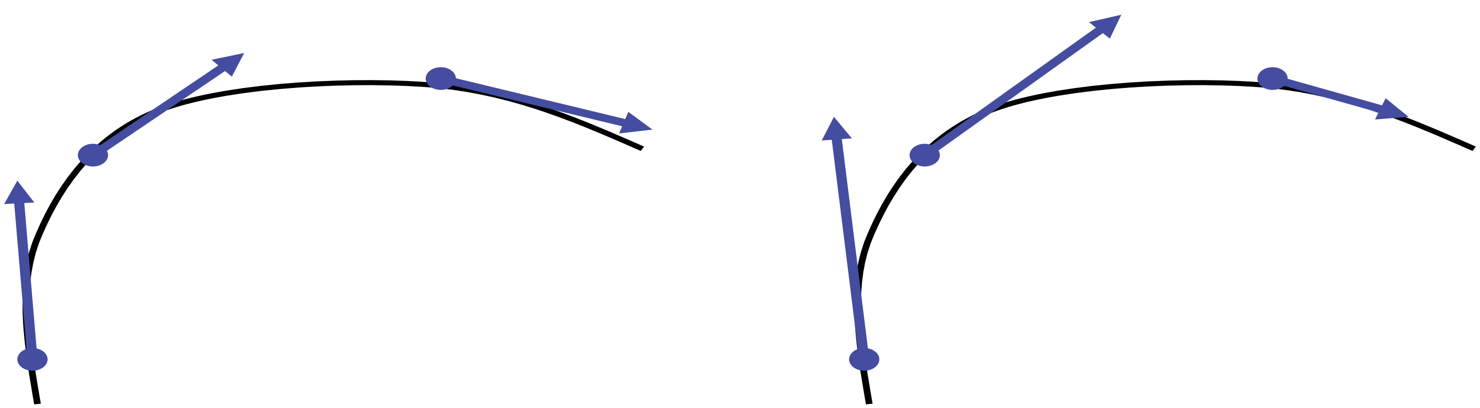
Projective transformation ((−1)-conf. transf.)

$c : I = (-\varepsilon, \varepsilon) \to M :$ a curve on M

$$c \text{ is a geodesic} \iff \nabla_{\frac{d}{dt}} \dot{c} = 0$$
$$c \text{ is a pre-geodesic} \iff \nabla_{\frac{d}{dt}} \dot{c} = \gamma(t)\dot{c}$$

A projective transformation preserves pre-geodesics (unparametrized geodesics).

$$\nabla_{\frac{d}{dt}} \dot{c} = \beta(t)\dot{c} \iff \bar{\nabla}_{\frac{d}{dt}} \dot{c} = \bar{\gamma}(t)\dot{c}$$



Dual-projective transformation (1-conf. transf.)

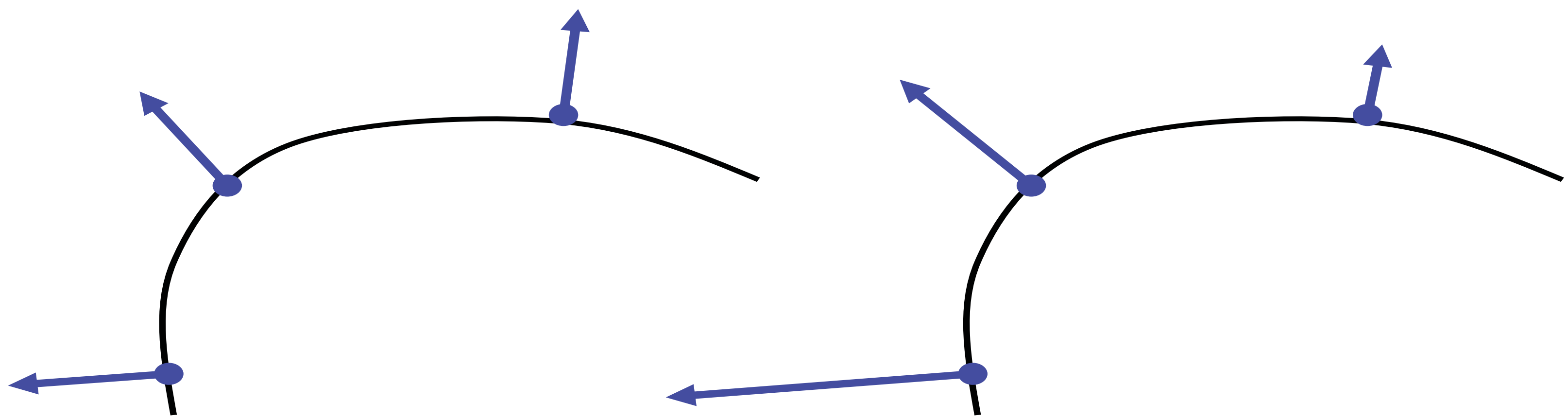
$c : I = (-\varepsilon, \varepsilon) \to M, \qquad \beta(X) = h(X, \dot{c}) :$ tangent 1-form

c is a dual geodesic $\iff \nabla_{\frac{d}{dt}}\beta = 0$

c is a pre-dual geodesic $\iff \nabla_{\frac{d}{dt}}\beta = \gamma(t)\beta$

A projective transformation preserves pre-dual geodesics.

$$\nabla_{\frac{d}{dt}}\beta = \gamma(t)\beta \iff \bar{\nabla}_{\frac{d}{dt}}\beta = \bar{\gamma}(t)\beta$$



2.3 Umbilical points

(M, ∇, h) : a statistical manifold, $n \geq 3$
N : a submanifold of M
ν : the unit normal vector along N
∇' : the induced connection
h' : the induced metric

(N, ∇', h') is a statistical submanifold.

$$\begin{aligned} \nabla_X Y &= \nabla'_X Y + \alpha(X, Y)\nu \\ \nabla_X \nu &= -\beta^\#(X) + \tau(X)\nu \end{aligned}$$

Set $\beta(X, Y) = h'(\beta^\#(X), Y)$.

Definition 2.8
$p \in N$
p : a tangentially umbilical point of N in (M, ∇, h)
$\overset{\text{def}}{\Longleftrightarrow}$ ${}^\exists c$: $\alpha_p = c h'_p$
p : a normally umbilical point of N in (M, ∇, h)
$\overset{\text{def}}{\Longleftrightarrow}$ ${}^\exists c$: $\beta_p = c h'_p$

Theorem 2.9 (Kurose '02)
(M, ∇, h) *and* $(M, \bar{\nabla}, \bar{h})$: *simply connected statistical manifolds,* $\dim M = n \geq 3$.

(M, ∇, h) *and* $(M, \bar{\nabla}, \bar{h})$ *are conformally-projectively equivalent*
$\Longleftrightarrow$

(1) $\overline{Ric}(X, Y) - \overline{Ric}(Y, X) = Ric(X, Y) - Ric(Y, X)$

(2) $(\nabla, h) \mapsto (\bar{\nabla}, \bar{h})$ *preserves the tangentially umbilical points and the normally umbilical points of any hypersurface of* M.

Definition 2.8
$p \in N$
p : a tangentially umbilical point of N in (M, ∇, h)
$\overset{\text{def}}{\Longleftrightarrow}$ $^{\exists}c : \quad \alpha_p = ch'_p$
p : a normally umbilical point of N in (M, ∇, h)
$\overset{\text{def}}{\Longleftrightarrow}$ $^{\exists}c : \quad \beta_p = ch'_p$

Definition 2.5
(M, ∇, h) and $(M, \bar{\nabla}, \bar{h})$ are conformally-projectively equivalent
$\overset{\text{def}}{\Longleftrightarrow}$ There exist two functions ϕ and ψ such that

$$\bar{h}(X, Y) = e^{\phi+\psi} h(X, Y),$$
$$\bar{\nabla}_X Y = \nabla_X Y - h(X, Y)\mathrm{grad}_h \psi + d\phi(Y)\, X + d\phi(X)\, Y$$

Proposition 2.10
$D, \tilde{D}$: *contrast functions (divergences) on* M
$(M, \nabla, h), (M, \tilde{\nabla}, \tilde{h})$: *induced statistical manifolds*
ϕ, ψ : *functions on* M.

(1) $\widetilde{D}(p||q) = e^{\phi(p)} D(p||q) \quad \Longrightarrow$
(M, ∇, h) *and* $(M, \tilde{\nabla}, \tilde{h})$ *are (−1)-conformally equivalent.*

(2) $\widetilde{D}(p||q) = e^{\psi(q)} D(p||q) \quad \Longrightarrow$
(M, ∇, h) *and* $(M, \tilde{\nabla}, \tilde{h})$ *are 1-conformally equivalent.*

(3) $\widetilde{D}(p||q) = e^{\psi(p)+\phi(q)} D(p||q) \quad \Longrightarrow$
(M, ∇, h) *and* $(M, \tilde{\nabla}, \tilde{h})$ *are conformally-projectively equivalent.*

Definition 2.5

(M, ∇, h) and $(M, \bar{\nabla}, \bar{h})$ are **conformally-projectively equivalent**

$\overset{\text{def}}{\Longleftrightarrow}$ There exist two functions ϕ and ψ such that

$$\begin{aligned} \bar{h}(X, Y) &= e^{\phi+\psi} h(X, Y), \\ \bar{\nabla}_X Y &= \nabla_X Y - h(X, Y)\mathrm{grad}_h \psi + d\phi(Y)\, X + d\phi(X)\, Y \end{aligned}$$

Definition 2.11

(M, ∇, h) is **conformally-projectively flat**

$\overset{\text{def}}{\Longleftrightarrow}$ (M, ∇, h) is locally conformally-projectively equivalent to some flat statistical manifold.

That is, for each point in M, ${}^{\exists}U \subset M$: a neighborhood,
${}^{\exists}(U, \bar{\nabla}, \bar{h})$: a flat statistical manifold such that
$(U, \nabla|_U, h|_U)$ and $(U, \tilde{\nabla}, \tilde{h})$ are conformally-projectively equivalent.

2.4 Conformal-projective invariants

Definition 2.12
(M, ∇, h) : a statistical manifold
∇^* : the dual connection of ∇ R: the curvature tensor of ∇
$\mathrm{Ric}, \mathrm{Ric}^*$: the Ricci tensors of ∇, ∇^*

$$\begin{aligned} W_{CP}(X,Y)Z = {} & R(X,Y)Z \\ & -\frac{1}{n-2}\{h(Y,Z)\alpha(X) - h(X,Z)\alpha(Y) \\ & \quad +\beta(Y,Z)X - \beta(X,Z)Y\} \\ & +\frac{\mathrm{trace}_h(\mathrm{Ric})}{(n-1)(n-2)}\{h(Y,Z)X - h(X,Z)Y\}, \end{aligned}$$

where
$$\alpha(X) := \frac{1}{n}\{\mathrm{Ric}^{\#}(X) + (n-1)(\mathrm{Ric}^*)^{\#}(X)\}$$
$$\beta(Y,Z) := \frac{1}{n}\{(n-1)\mathrm{Ric}(Y,Z) + \mathrm{Ric}^*(Y,Z)\}.$$

W_{CP} : conformal-projective curvature tensor

Proposition 2.13 (Kurose '02)
Suppose that $n \geq 4$.
(M, ∇, h) is conformally-projectively flat $\iff$ $W_{CP} = 0$.

$$(M, \nabla, h) \quad : \text{a statistical manifold}$$
$$C(X, Y, Z) := (\nabla_X h)(Y, Z), \quad \text{cubic form, Amari-Chentsov tensor}$$
$$K(X, Y) := K_X Y := \nabla_X Y - \nabla^{(0)}_X Y.$$

$$C(X, Y, Z) = -2h(K_X Y, Z),$$
$$K_X Y = \nabla^{(0)}_X Y - \nabla^*_X Y = \frac{1}{2}(\nabla_X Y - \nabla^*_X Y).$$

We may say that K is also a cubic form on (M, ∇, h).

$T^\flat$: the Tchebychev form on (M, ∇, h)
T : the Tchebychev vector field on (M, ∇, h)
$\overset{\text{def}}{\Longleftrightarrow}$

$$T^\flat(X) := \frac{1}{n}\mathrm{trace}\{Y \mapsto K_X Y\} = -\frac{1}{2n}\mathrm{trace}_h\{(Y, Z) \mapsto C(X, Y, Z)\},$$
$$h(X, T) := T^\flat(X), \qquad \text{where } n = \dim M.$$

$\widetilde{K}$: the traceless cubic form on (M, ∇, h)

$\overset{\text{def}}{\Longleftrightarrow}$
$$\widetilde{K}_X Y = K_X Y - \frac{n}{n+2}(h(X, Y)T + T^\flat(Y)X + T^\flat(X)Y)$$

ω, $\omega^{(0)}$: the parallel volume element with respect to ∇, $\nabla^{(0)}$,

$$\Longrightarrow \qquad T^\flat = \frac{1}{2n} d \log \left| \frac{\omega}{\omega^{(0)}} \right| .$$

In particular, if ∇ is flat and $\{\theta^i\}$ is a ∇-affine coordinate system,

$$\Longrightarrow \qquad T^\flat = -\frac{1}{n} d \log \sqrt{\det(g_{ij})}.$$

$T^\flat$: the Tchebychev form on (M, ∇, h)
T : the Tchebychev vector field on (M, ∇, h)
$\overset{\text{def}}{\Longleftrightarrow}$

$$T^\flat(X) := \frac{1}{n}\mathrm{trace}\{Y \mapsto K_X Y\} = -\frac{1}{2n}\mathrm{trace}_h\{(Y, Z) \mapsto C(X, Y, Z)\},$$
$$h(X, T) := T^\flat(X), \qquad \text{where } n = \dim M.$$

$\widetilde{K}$: the traceless cubic form on (M, ∇, h)

$$\overset{\text{def}}{\Longleftrightarrow} \qquad \widetilde{K}_X Y = K_X Y - \frac{n}{n+2}(h(X, Y)T + T^\flat(Y)X + T^\flat(X)Y)$$

Proposition 2.14
(M,∇,h) *and* $(M,\bar{\nabla},\bar{h})$ *are conformally-projectively equivalent*
$\Longrightarrow$ *Their traceless cubic forms coincide:*

$$\widetilde{\bar{K}} = \widetilde{K}$$

Proof:

$$2\bar{K}_X Y = 2K_X Y - h(X,Y)(\mathrm{grad}_h\psi - \mathrm{grad}_h\phi) - (d\psi - d\phi)(Y)X - (d\psi - d\phi)(X)Y.$$

Theorem 2.15
$(M,\nabla,h), (M,\bar{\nabla},\bar{h})$: *statistical manifolds, simply connected*
$Ric, \overline{Ric}$: *symmetric,* $h, \bar{h}$: *conformally equivalent*
$\widetilde{K} = \widetilde{\bar{K}}$
$\Longrightarrow (M,\nabla,h)$ *and* $(M,\tilde{\nabla},\tilde{h})$ *are conformally-projectively equivalent*

A sketch of the proof:

$^{\exists}\phi_1 \Longleftarrow h, \bar{h}$: conformally equivalent
$^{\exists}\psi_1$ s.t. $d\psi_1 = \bar{T}^\flat - T^\flat \Longleftarrow$ Ricci symmetric, simply connected

Set $\psi = \dfrac{1}{2}\phi_1 + \dfrac{n}{n+2}\psi_1, \quad \phi = \dfrac{1}{2}\phi_1 - \dfrac{n}{n+2}\psi_1.$

Then ψ and ϕ give a conformal-projective relation form $\widetilde{K} = \widetilde{\bar{K}}$

Monge-Ampere equations in affine differential geometry

$$T^\flat(X) := \frac{1}{n}\mathrm{trace}\{Y \mapsto K_X Y\} = -\frac{1}{2n}\mathrm{trace}_h\{(Y,Z) \mapsto C(X,Y,Z)\}$$

If ∇ is flat and $\{\theta^i\}$ is a ∇-affine coordinate system,

$$\Longrightarrow \qquad T^\flat = -\frac{1}{n} d\log\sqrt{\det(g_{ij})}.$$

If M is simply connected, $T^\flat$ is integrable, and since ∇ is flat, the metric g is given by a Hessian of the potential function ψ.
$\Longrightarrow$ there exist a function ω on M such that

$$\omega = \det(\partial_i \partial_j \psi)$$

This is nothing but a Monge-Ampere equation.

If (M, g, ∇, ∇^*) is doubly projectively flat
$\Longrightarrow$ (M, ∇, g) and (M, ∇^*, g) are spaces of constant curvature.
$\Longrightarrow$ We can choose ξ such that $T^\flat = 0$ (proper affine hypersphere)

3 Geometry of deformed exponential families

q-exponential, q-logarithm ($q > 0$)

$$\exp_q x := (1 + (1-q)x)^{\frac{1}{1-q}} \quad (1 + (1-q)x > 0) \qquad q\text{-exponential}$$

$$\log_q x := \frac{x^{1-q} - 1}{1-q} \quad (x > 0) \qquad q\text{-logarithm}$$

$q \to 1$, these are the standard exponential function, and the standard logarithm function, respectively.

$F_1(x), \ldots, F_n(x)$: random variables on Ω

$\theta = \{\theta^1, \ldots, \theta^n\}$: parameters

$$S = \left\{ p(x,\theta) \;\middle|\; p(x;\theta) > 0, \int_\Omega p(x;\theta)dx = 1 \right\}: \text{ statistical model}$$

Definition 3.1 $S_q = \{p(x;\theta)\}$: q-exponential family

$$\overset{\text{def}}{\iff} S_q := \left\{ p(x;\theta) \;\middle|\; p(x;\theta) = \exp_q \left[\sum_{i=1}^n \theta^i F_i(x) - \psi(\theta) \right], p(x,\theta) \in S \right\}$$

ψ : strictly convex

$\iff \{\partial_1 \log_q p(x;\theta), \ldots, \partial_n \log_q p(x;\theta)\}$ is linearly independent.

Example 3.2 (q-normal distribution (Student's t-distribution))

$$p(x;\mu,\sigma) = \frac{1}{z_q}\left[1 - \frac{1-q}{3-q}\frac{(x-\mu)^2}{\sigma^2}\right]^{\frac{1}{1-q}}$$

Set

$$\theta^1 = \frac{2}{3-q} z_q^{q-1} \cdot \frac{\mu}{\sigma^2}, \qquad \theta^2 = -\frac{1}{3-q} z_q^{q-1} \cdot \frac{1}{\sigma^2}.$$

Then

$$\begin{aligned}
\log_q p_q(x) &= \frac{1}{1-q}(p^{1-q} - 1) \\
&= \frac{1}{1-q}\left\{\frac{1}{z_q^{1-q}}\left(1 - \frac{1-q}{3-q}\frac{(x-\mu)^2}{\sigma^2}\right) - 1\right\} \\
&= \frac{2\mu z_q^{q-1}}{(3-q)\sigma^2}x - \frac{z_q^{q-1}}{(3-q)\sigma^2}x^2 - \frac{z_q^{q-1}}{3-q}\cdot\frac{\mu^2}{\sigma^2} + \frac{z_q^{q-1} - 1}{1-q} \\
&= \theta^1 x + \theta^2 x^2 - \psi(\theta)
\end{aligned}$$

$$\psi(\theta) = -\frac{(\theta^1)^2}{4\theta^2} - \frac{z_q^{q-1} - 1}{1-q}$$

Example 3.3 (discrete distributions)

$\Omega = \{x_0, x_1, \ldots, x_n\}$

$$S = \left\{ p(x;\eta) \;\middle|\; \eta_i > 0,\ \sum_{i=0}^{n} \eta_i = 1,\ p(x;\eta) = \sum_{i=0}^{n} \eta_i \delta_i(x) \right\},$$

$$\eta_0 = 1 - \sum_{i=1}^{n} \eta_i$$

n-dimensional probability simplex

Set $\theta^i = \frac{1}{1-q}\left((\eta_i)^{1-q} - (\eta_0)^{1-q}\right) = \log_q p(x_i) - \log_q p(x_0)$

Then

$$\begin{aligned}\log_q p(x) &= \frac{1}{1-q}(p^{1-q}(x) - 1) = \frac{1}{1-q}\sum_{i=0}^{n} \eta_i^q \delta_i(x) \\ &= \frac{1}{1-q}\left\{ \sum_{i=1}^{n} \left((\eta_i)^{1-q} - (\eta_0)^{1-q}\right)\delta_i(x) + (\eta_0)^{1-q} - 1 \right\}\end{aligned}$$

$$\psi(\theta) = -\log_q \eta_0$$

Remark 3.4 $S = \{p(x;\theta)\}$: (standard) exponential family

$$\begin{aligned} g^F_{ij}(\theta) &= E[(\partial_i \log p(x;\theta))(\partial_j \log p(x;\theta))] \\ &= \partial_i\partial_j\psi(\theta) \quad \text{:the Fisher metric} \\ T^F_{ijk}(\theta) &= E[(\partial_i \log p(x;\theta))(\partial_j \log p(x;\theta))(\partial_k \log p(x;\theta))] \\ &= \partial_i\partial_j\partial_k\psi(\theta) \quad \text{:the cubic form} \end{aligned}$$

Definition 3.5 $S_q = \{p(x;\theta)\}$: a q-exponential family

$$\begin{aligned} g^q_{ij}(\theta) &= \partial_i\partial_j\psi(\theta) : \text{ the } q\text{-Fisher metric} \\ T^q_{ijk}(\theta) &= \partial_i\partial_j\partial_k\psi(\theta) : \text{ the } q\text{-cubic form} \end{aligned}$$

On a deformed exponential family, the Fisher and the Hessian structures are different. (There are two different dually flat structures.)

Set $\Gamma^{q(e)}_{ij,k} := \Gamma^{q(0)}_{ij,k} - \frac{1}{2}T^q_{ijk}, \quad \Gamma^{q(m)}_{ij,k} := \Gamma^{q(0)}_{ij,k} + \frac{1}{2}T^q_{ijk},$

where $\Gamma^{q(0)}_{ij,k}$ is the connection coefficient of the Levi-Civita connection with respect to the q-Fisher metric g^q.

$\nabla^{q(e)}$: the q-exponential connection

$\nabla^{q(m)}$: the q-mixture connection

Proposition 3.6 For S_q, the following hold:

(1) $(S_q, g^q, \nabla^{q(e)}, \nabla^{q(m)})$ is a dually flat space.

(2) $\{\theta^i\}$ is a $\nabla^{q(e)}$-affine coordinate system on S_q.

(3) ψ is the potential of g^q with respect to $\{\theta^i\}$, that is,

$$g^q_{ij}(\theta) = \partial_i \partial_j \psi(\theta).$$

(4) Set the q-expectation of $F_i(x)$ by $\eta_i = E^{esc}_{q,p}[F_i(x)]$.
$\Longrightarrow$ $\{\eta_i\}$ is the dual coordinate system of $\{\theta^i\}$ w.r.t.. g^q.

(5) Set $\phi(\eta) = E^{esc}_{q,p}[\log_q p(x;\theta)]$
$\Longrightarrow$ $\phi(\eta)$ is the potential of g^q with respect to $\{\eta_i\}$.

$P_q(x)$: the escort distribution of $p(x)$ and the q-expectation of $f(x)$

$$\overset{\text{def}}{\Longleftrightarrow} \quad P_q(x) = p(x)^q, \qquad E_{q,p}[f(x)] = \int f(x) P_q(x) dx$$

$E^{esc}_{q,p}[f(x)]$: the normalized q-expectation of $f(x)$

$$\overset{\text{def}}{\Longleftrightarrow} \quad E^{esc}_{q,p}[f(x)] = \int f(x) \frac{P_q(x)}{Z_q(p)} dx, \quad Z_q(p) = \int p(x)^q dx.$$

Proposition 3.6 For S_q, the following hold:

(1) $(S_q, g^q, \nabla^{q(e)}, \nabla^{q(m)})$ is a dually flat space.

(2) $\{\theta^i\}$ is a $\nabla^{q(e)}$-affine coordinate system on S_q.

(3) ψ is the potential of g^q with respect to $\{\theta^i\}$, that is,

$$g^q_{ij}(\theta) = \partial_i \partial_j \psi(\theta).$$

(4) Set the q-expectation of $F_i(x)$ by $\eta_i = E^{esc}_{q,p}[F_i(x)]$.
$\Longrightarrow$ $\{\eta_i\}$ is the dual coordinate system of $\{\theta^i\}$ w.r.t. g^q.

(5) Set $\phi(\eta) = E^{esc}_{q,p}[\log_q p(x;\theta)]$
$\Longrightarrow$ $\phi(\eta)$ is the potential of g^q with respect to $\{\eta_i\}$.

normalized Tsallis relative entropy (q-relative entropy)

$$D^q(p,r) = E^{esc}_{q,p}\left[\log_q p(x) - \log_q r(x)\right] \qquad (\downarrow\ (-\alpha)\text{-divergence})$$

$$= \frac{1 - \int p(x)^q r(x)^{1-q} dx}{(1-q)Z_q(p)} \quad \left(= \frac{q}{Z_q(p)} D^{(1-2q)}(p,r) \right).$$

α-divergence ($\alpha = 1 - 2q$)

$$D^{(1-2q)}(p(x), r(x)) = \frac{1}{q}\int_\Omega p(x)^q \{\log_q p(x) - \log_q r(x)\} dx$$

$D^{(1-2q)}$ induces a non-flat invariant statistical manifold $(S_q, \nabla^{(1-2q)}, g^F)$.

normalized Tsallis relative entropy (q-relative entropy)

$$D^q(p(x), r(x)) = E^{esc}_{q,p}\left[\log_q p(x) - \log_q r(x)\right]$$

$$= \int_\Omega \frac{p(x)^q}{Z_q(p)} \{\log_q p(x) - \log_q r(x)\} dx \quad \left(= \frac{q}{Z_q(p)} D^{(1-2q)}(p, r) \right)$$

D^q induces a Hessian manifold (flat statistical mfd.) $(S_q, \nabla^{q(m)}, g^q)$.

$(S_q, \nabla^{e(m)}, g^q)$ and $(S_q, \nabla^{(2q-1)}, g^F)$ are 1-conformally equivalent, since

$$(D^q(p, q) =)\ D(r, p) = \frac{q}{Z_q(p)} D^{(2q-1)}(r, p).$$

$\nu(x)$ pos. measure $\not\longrightarrow$ $\dfrac{\nu(x)}{Z_q(\nu)}$ prob. measure

Normalization of a positive measure to a probability measure is NOT a trivial problem.

4 Geometric divergence and α-divergence

4.1 Review: affine immersions

$f : M \to R^{n+1}$: an immersion
ξ: a local vector field along f

Definition 4.1
$\{f, \xi\} : M \to R^{n+1}$ is an affine immersion
$\overset{\text{def}}{\Longleftrightarrow}$ For an arbitrary point $p \in M$,

$$T_{f(p)}R^{n+1} = f_*(T_pM) \oplus R\{\xi_p\}$$

ξ: a transversal vector field

D: the standard flat affine connection on R^{n+1}

$$\begin{aligned} D_X f_* Y &= f_*(\nabla_X Y) + h(X, Y)\xi, \\ D_X \xi &= -f_*(SX) + \tau(X)\xi. \end{aligned}$$

f : non-degenerate $\overset{\text{def}}{\Longleftrightarrow}$ h : non-degenerate
$\{f, \xi\}$: equiaffine $\overset{\text{def}}{\Longleftrightarrow}$ $\tau = 0$

Proposition 4.2
$\{f, \xi\}$: non-degenerate, equiaffine $\Longrightarrow$ (M, ∇, h) is a statistical manifold, 1–conformally flat.

Proposition 4.3
(M, ∇, h) : a simply connected statistical manifold
1-conformally flat
$\Longrightarrow$ There exists $\{f, \xi\}$ which realizes (M, ∇, h) in R^{n+1}.

Fundamental structural equations for affine immersions

Gauss: $R(X,Y)Z = h(Y,Z)SX - h(X,Z)SY$

Codazzi:
$$(\nabla_X h)(Y,Z) + \tau(X)h(Y,Z) = (\nabla_Y h)(X,Z) + \tau(Y)h(X,Z)$$
$$(\nabla_X S)(Y) - \tau(X)SY = (\nabla_Y S)(X) - \tau(Y)SX$$

Ricci: $h(X,SY) - h(Y,SX) = (\nabla_X \tau)(Y) - (\nabla_Y \tau)(X)$

f : non-degenerate $\overset{\text{def}}{\Longleftrightarrow}$ h : non-degenerate

$\{f, \xi\}$: equiaffine $\overset{\text{def}}{\Longleftrightarrow}$ $\tau = 0$

4.2 Conormal maps and geometric divergences

$\{f, \xi\}$: nondegenerate, equiaffine
R_{n+1} : the dual space of R^{n+1}
$\langle\ ,\ \rangle$: the canonical pairing of R_{n+1} and R^{n+1}.

$v : M \to R_{n+1}$ is the conormal map of $\{f, \xi\}$

$$\overset{\text{def}}{\Longleftrightarrow} \quad \begin{aligned} \langle v(p), \xi_p \rangle &= 1, \\ \langle v(p), f_* X_p \rangle &= 0 \end{aligned}$$

We define a function on $M \times M$ by

$$\rho(p, r) = \langle v(r), f(p) - f(r) \rangle. \tag{1}$$

ρ is called the geometric divergence on M.

The geometric divergence is independent of realization of (M, ∇, h).

cf. affine support function:

$$\begin{aligned} \rho : R^{n+1} \times M &\to R \\ \rho(x, r) &= \langle v(r), x - f(r) \rangle \end{aligned}$$

(M, ∇, h) : a simply connected flat statistical manifold.
($\Longrightarrow$ (M, h, ∇, ∇^*) is a dually flat space.)

$\Longrightarrow$ $\exists\psi$: a function on M (potential function) such that $\dfrac{\partial^2\psi}{\partial\theta^i\partial\theta^j} = g_{ij}$

$\Longrightarrow$ $\{f, \xi\}$: an affine immersion (graph immersion)

$$f : \begin{pmatrix}\theta^1 \\ \vdots \\ \theta^n\end{pmatrix} \mapsto \begin{pmatrix}\theta^1 \\ \vdots \\ \theta^n \\ \psi(\theta)\end{pmatrix}, \quad \xi = \begin{pmatrix}0 \\ \vdots \\ 0 \\ 1\end{pmatrix}$$

v : the conormal map of $\{f, \xi\}$,

$$v = (-\eta_1, \ldots, -\eta_n, 1) \qquad \eta_i = \frac{\partial\psi}{\partial\theta^i}$$

Since $\phi(r) = \sum \eta_i(r)\theta^i(r) - \psi(r)$, we have

$$\begin{aligned}
\rho(p, r) &= \langle v(r), f(p) - f(r)\rangle \\
&= -\sum \eta_i(r)\theta^i(p) + \psi(p) + \sum \eta_i(r)\theta^i(r) - \psi(r) \\
&= \psi(p) + \phi(r) - \sum \eta_i(r)\theta^i(p) \\
&= D(p||r)
\end{aligned}$$

4.3 Realization of q-exponential family and α-divergence

$S_q = \{p(x;\theta) \,|\, p(x;\theta) = \exp_q\left[\sum_{i=1}^{n} \theta^i F_i(x) - \psi(\theta)\right]\}$: q-exponential family

the Hessian manifold $(S_q, \nabla^{(e)q}, g^q)$

$\{f, \xi\}$: an affine immersion (graph immersion)

$$f : \begin{pmatrix} \theta^1 \\ \vdots \\ \theta^n \end{pmatrix} \mapsto \begin{pmatrix} \theta^1 \\ \vdots \\ \theta^n \\ \psi(\theta) \end{pmatrix}, \quad \xi = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

v : the conormal map of $\{f, \xi\}$,

$$v = (-\eta_1, \ldots, -\eta_n, 1) \qquad \eta_i = \frac{\partial \psi}{\partial \theta^i} = E_q^{esc}[F_i(x)]$$

$\rho_q(p(\theta), p(\theta'))$: the geometric divergence of $(S_q, \nabla^{(e)q}, g^q)$

$$\begin{aligned} \rho_q(p(\theta), p(\theta')) &= \langle v(p(\theta')), f(p(\theta)) - f(p(\theta')) \rangle \\ &= E_{q,p(\theta')}^{esc}\left[\log_q p(\theta') - \log_q p(\theta)\right] \\ &= D(p(\theta) || p(\theta')) \end{aligned}$$

the invariant manifold $(S_q, \nabla^{(2q-1)}, g^F)$ $(\alpha = 2q-1)$

$\{f, \bar{\xi}\}$: an affine immersion

$$f : \begin{pmatrix} \theta^1 \\ \vdots \\ \theta^n \end{pmatrix} \mapsto \begin{pmatrix} \theta^1 \\ \vdots \\ \theta^n \\ \psi(\theta) \end{pmatrix}, \qquad \bar{\xi} = \frac{q}{Z_q}\left\{\xi + f_* \mathrm{grad}_h \left(\log \frac{Z_q}{q}\right)\right\}$$

$$Z_q = \int_\Omega p(x;\theta)^q dx$$

v^F : the conormal map of $\{f, \bar{\xi}\}$,

$$v^F = \frac{Z_q}{q}(-\eta_1, \ldots, -\eta_n, 1)$$

$\rho_q^F(p(\theta), p(\theta'))$: the geometric divergence of $(S_q, \nabla^{(2q-1)}, g^F)$

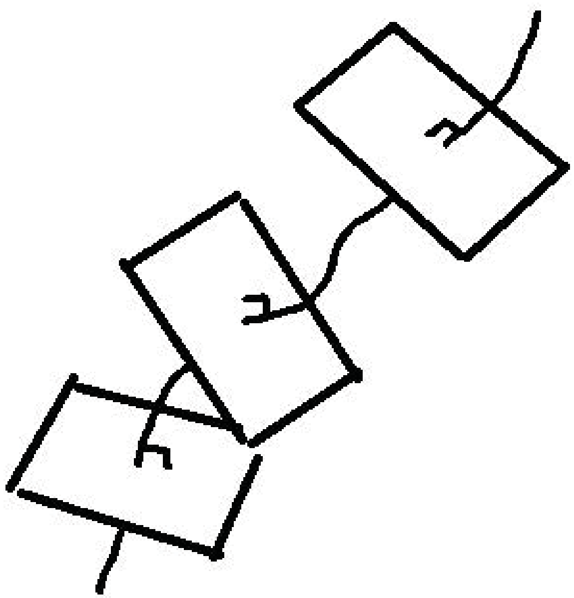
$$\begin{aligned} \rho_q^F(p(\theta), p(\theta')) &= \langle v^F(p(\theta')), f(p(\theta)) - f(p(\theta'))\rangle \\ \left(\begin{array}{c}\text{un-normalized} \\ \text{expectation}\end{array} \Longrightarrow\right) &= \frac{1}{q} E_{q,p(\theta')}\left[\log_q p(\theta') - \log_q p(\theta)\right] \\ &= \frac{1}{q(1-q)}\left\{1 - \int_\Omega p(\theta)^{1-q} p(\theta')^q dx\right\} \\ &= D^{(2q-1)}(p(\theta)||p(\theta')) = D^{(\alpha)}(p(\theta)||p(\theta')) \end{aligned}$$

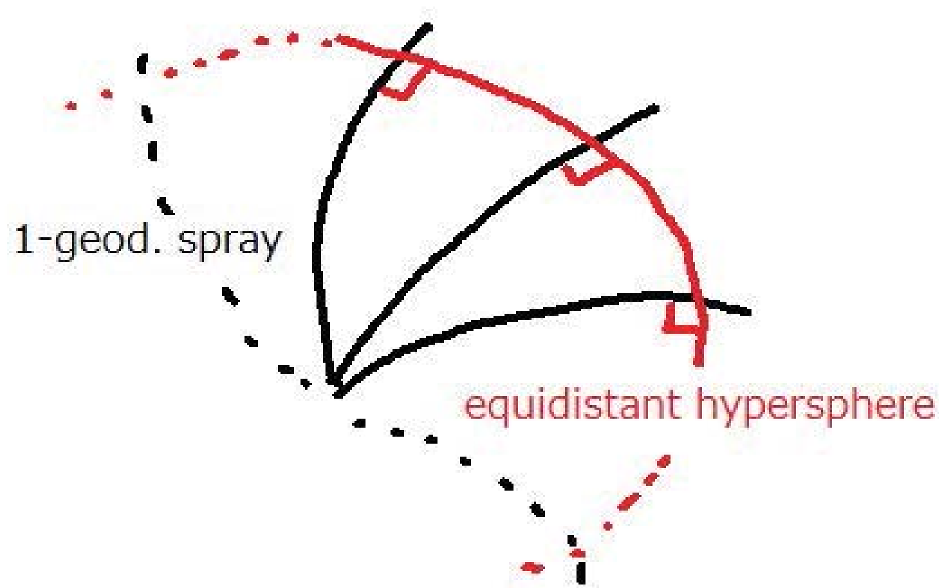
What is the canonical divergence?

(1) If (M, g, ∇, ∇^*) is a dually flat space, the divergence coincides with the well-known canonical divergence.

(2) The projection property should be hold.

$$\gamma \perp M \text{ at } q \qquad (\text{i.e. } g(\dot\gamma(0), X) = 0, {}^{\forall}X \in T_qM)$$
$$\Longrightarrow \quad D(p,q) = \min_{r \in M} D(p,r)$$





Summary

q-information geometry

$S_q = \{p(x;\theta)\}$: a q-exponential family

- $(S_q, \nabla^{(e)q}, g^q)$: a Hessian manifold (a flat statistical manifold)
- $(S_q, \nabla^{(2q-1)}, g^F)$: an invariant statistical manifold ($\alpha = 2q - 1$)
- $(S_q, \nabla^{e(q)}, g^q)$, $(S_q, \nabla^{(2q-1)}, g^F)$ are 1-conformally equivalent
- $(S_q, \nabla^{(2q-1)}, g^F)$ is 1-conformally flat

affine immersions and geometric divergences

$S_q = \{p(x;\theta)\}$ is realized by affine immersions $S_q \to R^{n+1}$

- $(S_q, \nabla^{e(q)}, g^{(q)})$ is realized by

  $f = (\theta^1, \dots, \theta^n, \psi(\theta))^T, \quad \xi = (0, \dots, 0, 1)^T$

  $\rho_q(p(\theta), p(\theta')) = D(p(\theta)||p(\theta')) \ \left(= E^{esc}_{q,p(\theta')}\left[\log_q p(\theta') - \log_q p(\theta)\right]\right)$

- $(S_q, \nabla^{(2q-1)}, g^F)$ is realized by

  $f = (\theta^1, \dots, \theta^n, \psi(\theta))^T, \quad \bar{\xi} = \frac{q}{Z_q}\left\{\xi + f_* \mathrm{grad}_h \left(\log \frac{Z_q}{q}\right)\right\}$

  $\rho_q^F(p(\theta), p(\theta')) = D^{(2q-1)}(p(\theta)||p(\theta'))$ (α-divergence)

1. Affine immersions
2. Statistical manifolds and generalized conformal structures
3. Deformed exponential families
4. Geometric divergences and α-divergences

Appendixes
5. Generalization of Legendre transformation
6. Quantum analogue of affine differential geometry

5 Generalization of Legendre transformation

5.1 Centroaffine immersions of codimension two

M: an n-dimensional manifold

$f : M \to R^{n+2}$: an immersion

ξ: a local vector field along f

Definition 5.1

$\{f, \xi\} : M \to R^{n+2}$ is a centroaffine immersions of codimension two

$\overset{\text{def}}{\Longleftrightarrow}$

For an arbitrary point $p \in M$,

$$T_{f(x)}R^{n+2} = f_*(T_xM) \oplus R\{\xi_x\} \oplus R\{f(x)\}$$

ξ: a transversal vector field

D: the standard flat affine connection on R^{n+2}

$$\begin{aligned} D_X f_* Y &= f_*(\nabla_X Y) + h(X,Y)\xi + k(X,Y)f, \\ D_X \xi &= -f_*(SX) + \tau(X)\xi + \mu(X)f. \end{aligned}$$

∇ : the induced connection
h : the affine fundamental form
τ : the transversal connection form
S : the affine shape operator

$$\theta(X_1, \cdots, X_n) := \det(f_* X_1, \cdots, f_* X_n, \xi, f)$$

the induced volume element

Proposition 5.2

$$\nabla_X \theta = \tau(X)\theta$$

Definition 5.3

f : non-degenerate $\overset{\text{def}}{\Longleftrightarrow}$ h : non-degenerate

$\{f, \xi\}$: equiaffine $\overset{\text{def}}{\Longleftrightarrow}$ $\tau = 0$

D: the standard flat affine connection on R^{n+2}

$$D_X f_* Y = f_*(\nabla_X Y) + h(X,Y)\xi + k(X,Y)f,$$
$$D_X \xi = -f_*(SX) + \tau(X)\xi + \mu(X)f.$$

∇ : the induced connection
h : the affine fundamental form
τ : the transversal connection form
S : the affine shape operator

$$\theta(X_1, \cdots, X_n) := \det(f_* X_1, \cdots, f_* X_n, \xi, f)$$

the induced volume element

Proposition 5.2

$$\nabla_X \theta = \tau(X)\theta$$

Definition 5.3

f : non-degenerate $\overset{\text{def}}{\Longleftrightarrow}$ h : non-degenerate

$\{f, \xi\}$: equiaffine $\overset{\text{def}}{\Longleftrightarrow}$ $\tau = 0$

Proposition 5.4

$\{f, \xi\} : M \to R^{n+2}$: non-degenerate, equiaffine

$\Longrightarrow$ (M, ∇, h) is a statistical manifold, conformally-projectively flat

5.2 Dual maps and geometric divergences

$\boldsymbol{R}_{n+2}$: the dual vector space of $\boldsymbol{R}^{n+2}$
$\langle\ ,\ \rangle$: the pairing of $\boldsymbol{R}_{n+2}$ and $\boldsymbol{R}^{n+2}$

Definition 5.5
$v, w : M \to \boldsymbol{R}_{n+2}$

$$\overset{\text{def}}{\Longleftrightarrow} \qquad \begin{aligned} \langle v(p), \xi_p\rangle &= 1 & \langle w(p), \xi_p\rangle &= 0,\\ \langle v(p), f(p)\rangle &= 0 & \langle w(p), f(p)\rangle &= 1,\\ \langle v(p), f_*X_p\rangle &= 0 & \langle w(p), f_*X_p\rangle &= 0, \end{aligned}$$

We call v the conormal map of $\{f, \xi\}$

If h is non-degenerate
$\Longrightarrow \{v, w\} : M \to \boldsymbol{R}_{n+2}$ is a centroaffine immersion of codimension two. We call $\{v, w\}$ the dual map of $\{f, \xi\}$.

Proposition 5.6
$\{f, \xi\}$ *induces* $(M, \nabla, h) \quad \Longleftrightarrow \quad \{v, w\}$ *induces* (M, ∇^*, h).

Definition 5.5
$v, w : M \to R_{n+2}$: the dual map of $\{f, \xi\}$.

$$\overset{\text{def}}{\iff} \quad \begin{aligned} \langle v(p), \xi_p \rangle &= 1 \quad \langle w(p), \xi_p \rangle = 0, \\ \langle v(p), f(p) \rangle &= 0 \quad \langle w(p), f(p) \rangle = 1, \\ \langle v(p), f_* X_p \rangle &= 0 \quad \langle w(p), f_* X_p \rangle = 0, \end{aligned}$$

Definition 5.7
$\rho : M \times M \to R$: the geometric divergence

$$\overset{\text{def}}{\iff} \quad \rho(p, q) = \langle v(q), f(p) - f(q) \rangle$$

The geometric divergence ρ is a contrast function, and this is a special form of an affine support function.

Legendre transformation

Proposition 5.8

(M, g, ∇, ∇^*) : *a dually flat space*

$\{\theta^i\}$: *a* ∇*-affine coordinate system*

$\{\eta^i\}$: *a* ∇^**-affine coordinate system*

$$\Longrightarrow \quad \frac{\partial \psi}{\partial \theta^i} = \eta_i, \quad \frac{\partial \phi}{\partial \eta_i} = \theta^i,$$

$$\frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j} = g_{ij}, \quad \frac{\partial^2 \phi}{\partial \eta^i \partial \eta^j} = g^{ij}, \qquad g\left(\frac{\partial}{\partial \theta^i}, \frac{\partial}{\partial \eta_j}\right) = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j), \end{cases}$$

$$\psi(p) + \phi(p) - \sum_{i=1}^{n} \theta^i(p)\eta_i(p) = 0,$$

(M, ∇, g) and (M, ∇^*, g) are flat statistical manifolds.

$$f = \begin{pmatrix} \theta^1 \\ \vdots \\ \theta^n \\ \psi \\ 1 \end{pmatrix}, \qquad \xi = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \end{pmatrix} \implies f_* \frac{\partial}{\partial \theta^n} = \frac{\partial f}{\partial \theta^n} = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \frac{\partial \psi}{\partial \theta^n} \\ 0 \end{pmatrix}$$

$$v = (-\eta_1, \ldots, -\eta_n, 1, \phi), \quad w = (0, \ldots, 0, 0, 1)$$

$$\left\langle v(p), f_* \frac{\partial}{\partial \theta^i} \right\rangle = 0 \iff -\eta_i(p) + \frac{\partial \psi}{\partial \theta^i}(p) = 0$$

$$\langle v(p), f(p) \rangle = 0 \iff \psi(p) + \phi(p) - \sum_{i=1}^{n} \theta^i(p)\eta_i(p) = 0$$

$$\begin{aligned} \rho(p, q) &= \langle v(q), f(p) - f(q) \rangle \\ &= \psi(p) + \phi(q) - \sum_{i=1}^{n} \theta^i(p)\eta_i(q) \end{aligned}$$

6 Quantum analogue of affine differential geometry

6.1 Quasi-statistical manifolds

M : a manifold (an open domain in $\boldsymbol{R}^n$)

h : a non-degenerate $(0,2)$-tensor field on M

∇ : an affine connection on M

$T^{\nabla}(X,Y) = \nabla_X Y - \nabla_Y X - [X,Y]$: the torsion tensor of ∇

Definition 6.1

(M,∇,h): a quasi-statistical manifold

$$\overset{\text{def}}{\iff} \quad (\nabla_X h)(Y,Z) - (\nabla_Y h)(X,Z) = -h(T^{\nabla}(X,Y),Z)$$

In addition, if h is a semi-Riemannian metric, then we say that (M,∇,h) is a statistical manifold admitting torsion (SMAT).

Definition 6.2

∇^*: (quasi-) dual connection of ∇ with respect to h

$$\overset{\text{def}}{\iff} \quad Xh(Y,Z) = h(\nabla^*_X Y, Z) + h(Y, \nabla_X Z).$$

Proposition 6.3

The dual connection ∇^* of ∇ is torsion free.

We remark that $(\nabla^*)^* \neq \nabla$ in general.

Proposition 6.4
If h is symmetric $h(X, Y) = h(Y, X)$
or skew-symmetric $h(X, Y) = -h(Y, X)$
$\Longrightarrow (\nabla^*)^* = \nabla$

Proposition 6.5
(M, ∇^*, h) : ∇^* *is torsion free and dual of* ∇,
h is a non-degenerate $(0, 2)$-tensor field,
$\Longrightarrow$ (M, ∇, h) *is a quasi-statistical manifold.*

Suppose that (M, ∇, h) is a statistical manifold admitting torsion.

(1) (M, ∇, h) is a Hessian manifold
$\iff$ $R^{\nabla} = 0$ and $T^{\nabla} = 0$
$\iff$ (M, h, ∇, ∇^*) is a dually flat space.

(2) (M, ∇, h) is a space of distant parallelism
$\iff$ $R^{\nabla} = 0$ and $T^{\nabla} \neq 0$ $\quad (R^{\nabla^*} = 0, \quad T^{\nabla^*} = 0)$.

SMAT with the SLD Fisher metric (Kurose 2007)

$\mathrm{Herm}\,(d)$: the set of all Hermitian matrices of degree d.
$\mathcal{S}$: a space of quantum states

$$\mathcal{S} = \{P \in \mathrm{Herm}\,(d) \mid P > 0,\ \mathrm{trace}P = 1\}$$

$T_P\mathcal{S} \cong \mathcal{A}_0$ $\qquad$ $\mathcal{A}_0 = \{X \in \mathrm{Herm}\,(d) \mid \mathrm{trace}X = 0\}$

We denote by $\widetilde{X}$ the corresponding vector field of X.

For $P \in \mathcal{S},\ X \in \mathcal{A}_0$, define $\omega_P(\widetilde{X})$ $(\in \mathrm{Herm}\,(d))$ by

$$X = \frac{1}{2}(P\omega_P(\widetilde{X}) + \omega_P(\widetilde{X})P)$$

The matrix $\omega(\widetilde{X})$ is the "symmetric logarithmic derivative".

A Riemannian metric and an affine connection are defined as follows:

$$h_P(\widetilde{X}, \widetilde{Y}) = \frac{1}{2}\mathrm{trace}\left(P(\omega_P(\widetilde{X})\omega_P(\widetilde{Y}) + \omega_P(\widetilde{Y})\omega_P(\widetilde{X}))\right),$$

$$\left(\nabla_{\widetilde{X}}\widetilde{Y}\right)_P = h_P(\widetilde{X}, \widetilde{Y})P - \frac{1}{2}(X\omega_P(\widetilde{Y}) + \omega_P(\widetilde{Y})X).$$

The SMAT $(\mathcal{S}, \nabla, h)$ is a space of distant parallelism.
$(R = R^* = 0,\ T^* = 0,\ \text{but } T \neq 0)$

6.2 Affine distributions

$\omega : TM \to R^{n+1}$: a R^{n+1}-valued 1-form
$\xi : M \to R^{n+1}$: a R^{n+1}-valued function

Definition 6.6
$\{\omega, \xi\}$ is an affine distribution
$\overset{\text{def}}{\Longleftrightarrow}$ For an arbitrary point $p \in M$,

$$R^{n+1} = \mathrm{Image}\, \omega_p \oplus R\{\xi_x\}$$

ξ: a transversal vector field

$\{f, \xi\}$: an affine immersion $\Longrightarrow$ $\{df, \xi\}$: an affine distribution

$$\begin{aligned} X\omega(Y) &= \omega(\nabla_X Y) + h(X, Y)\xi, \\ X\xi &= -\omega(SX) + \tau(X)\xi. \end{aligned}$$

∇ : an affine connection $(T^{\nabla}(X, Y) \neq 0$ in general$)$
h : a $(0, 2)$-tensor field $(h(X, Y) \neq h(Y, X)$ in general$)$
S : a $(1, 1)$-tensor field
τ : a 1-form

$$\begin{aligned} X\omega(Y) &= \omega(\nabla_X Y) + h(X,Y)\xi, \\ X\xi &= -\omega(SX) + \tau(X)\xi. \end{aligned}$$

ω : symmetric	$\overset{\text{def}}{\Longleftrightarrow}$	h : symmetric
ω : non-degenerate	$\overset{\text{def}}{\Longleftrightarrow}$	h : non-degenerate
$\{\omega, \xi\}$: equiaffine	$\overset{\text{def}}{\Longleftrightarrow}$	$\tau = 0$

Symmetry and non-degeneracy of ω are independent of ξ.

Proposition 6.7
Set $\tilde{\xi} := \omega(V) + \phi\xi$. Then the induced objects change as follows:

$$\begin{aligned} \nabla_X Y &= \tilde{\nabla}_X Y + \tilde{h}(X,Y)V, \\ h(X,Y) &= \phi\tilde{h}(X,Y), \\ \tilde{S}X - \tilde{\tau}(X)V &= \phi SX - \nabla_X V, \\ \phi\tilde{\tau}(X) &= h(X,V) + d\phi(X) + \phi\tau(X). \end{aligned}$$

Proposition 6.8

Image $(d\omega)_p \subset$ Image ω_p	$\Longleftrightarrow$	*h: symmetric*
Image $(d\xi)_p \subset$ Image ω_p	$\Longleftrightarrow$	$\tau = 0$

Proposition 6.9
$\{\omega, \xi\}$: *non-degenerate, equiaffine*
$\Longrightarrow (M, \nabla, h)$ *is a quasi-statistical manifold.*
$\{\omega, \xi\}$: *symmetric, non-degenerate, equiaffine*
$\Longrightarrow (M, \nabla, h)$ *is a SMAT.*

Fundamental structural equations for affine distributions:

Gauss equation:

$$R(X, Y)Z = h(Y, Z)SX - h(X, Z)SY,$$

Codazzi equations:

$$(\nabla_X h)(Y, Z) + h(Y, Z)\tau(X) - (\nabla_Y h)(X, Z) + h(X, Z)\tau(Y) = -h(T^{\nabla}(X, Y), Z),$$

$$(\nabla_X S)(Y) + \tau(Y)SX - (\nabla_Y S)(X) - \tau(X)SY = -S(T^{\nabla}(X, Y)),$$

Ricci equation:

$$h(X, SY) - (\nabla_X \tau)(Y) - h(Y, SX) + (\nabla_Y \tau)(X) = \tau(T^{\nabla}(X, Y)).$$

SMAT with the SLD Fisher metric (Kurose 2007)

Herm (d) : the set of all Hermitian matrices of degree d.
$\mathcal{S}$: a space of quantum states

$$\mathcal{S} = \{P \in \mathrm{Herm}\,(d) \mid P > 0,\ \mathrm{trace}P = 1\}$$

$T_P\mathcal{S} \cong \mathcal{A}_0 \qquad \mathcal{A}_0 = \{X \in \mathrm{Herm}\,(d) \mid \mathrm{trace}X = 0\}$

We denote by $\widetilde{X}$ the corresponding vector field of X.

For $P \in \mathcal{S},\ X \in \mathcal{A}_0$, define $\omega_P(\widetilde{X})$ $(\in \mathrm{Herm}\,(d))$ and ξ by

$$X = \frac{1}{2}(P\omega_P(\widetilde{X}) + \omega_P(\widetilde{X})P), \qquad \xi = -I_d$$

Then $\{\omega, \xi\}$ is an equiaffine distribution.

The induced quantities are given by

$$h_P(\widetilde{X}, \widetilde{Y}) = \frac{1}{2}\mathrm{trace}\left(P(\omega_P(\widetilde{X})\omega_P(\widetilde{Y}) + \omega_P(\widetilde{Y})\omega_P(\widetilde{X}))\right),$$

$$\left(\nabla_{\widetilde{X}}\widetilde{Y}\right)_p = h_P(\widetilde{X}, \widetilde{Y})P - \frac{1}{2}(X\omega_P(\widetilde{Y}) + \omega_P(\widetilde{Y})X).$$

$$(R = R^* = 0,\ T^* = 0,\ \text{but } T \neq 0)$$

SMAT with the real RLD Fisher metrics (Kurose 2007)

Herm (d) : the set of all Hermitian matrices of degree d.
$\mathcal{S}$: a space of quantum states

$$\mathcal{S} = \{P \in \mathrm{Herm}\,(d) \mid P > 0,\ \mathrm{trace}P = 1\}$$

$T_P\mathcal{S} \cong \mathcal{A}_0 \qquad \mathcal{A}_0 = \{X \in \mathrm{Herm}\,(d) \mid \mathrm{trace}X = 0\}$

For $P \in \mathcal{S},\ X \in \mathcal{A}_0$, set

$$\omega_P(\widetilde{X}) = \frac{1}{2}(P^{-1}X + XP^{-1}), \qquad \xi = -I_d$$

Then $\{\omega, \xi\}$ is an equiaffine distribution.

The induced quantities are given by

$$\begin{aligned} h_P(\widetilde{X}, \widetilde{Y}) &= \frac{1}{2}\mathrm{trace}(P^{-1}(XY + YX)), \\ \omega_P(\nabla_{\widetilde{X}}\widetilde{Y}) &= h_P(\widetilde{X}, \widetilde{Y})I_d - \frac{1}{2}(P^{-1}XP^{-1}Y + YP^{-1}XP^{-1}). \end{aligned}$$

$$(R = R^* = 0,\ T^* = 0,\ \text{but } T \neq 0)$$

6.3 Triviality of quasi-statistical manifolds

(M, ∇, h): a quasi-statistical manifold

∇ is of (weak) constant curvature

$\overset{\text{def}}{\iff}$ There exists a positive function k such that

$$R^{\nabla}(X, Y)Z = k\{h(Y, Z)X - h(X, Z)Y\}$$

Theorem 1

$\{\omega, \xi\}$: *a non-degenerate, equiaffine distribution.*

(M, ∇, h) : *the induced quasi-statistical manifold of* $\{\omega, \xi\}$,

∇ : *weak constant curvature*

$$h^k(X, Y) := kh(X, Y),\ \nabla^k_X Y := \nabla_X Y + d(\log k)(X)Y$$

$\Longrightarrow (M, \nabla^k, h^k)$ *is a statistical manifold of constant curvature* 1.

This theorem implies that a constant curvature quasi-statistical manifold is easily obtained from a standard statistical manifold.

On the other hand, in the case $R = 0$, (i.e., (M, ∇, h) is a space of distant parallelism), we can define non-trivial quasi-statistical manifolds.

Theorem 1

$\{\omega, \xi\}$: *a non-degenerate, equiaffine distribution.*

(M, ∇, h) : *the induced quasi-statistical manifold of* $\{\omega, \xi\}$,

∇ : *weak constant curvature*

$$h^k(X, Y) := kh(X, Y), \ \nabla^k_X Y := \nabla_X Y + d(\log k)(X)Y$$

$\Longrightarrow (M, \nabla^k, h^k)$ *is a statistical manifold of constant curvature* 1.

Fundamental structural equations for affine distributions:

Gauss equation:

$$R(X, Y)Z = h(Y, Z)SX - h(X, Z)SY,$$

Codazzi equations:

$$(\nabla_X h)(Y, Z) + h(Y, Z)\tau(X) - (\nabla_Y h)(X, Z) + h(X, Z)\tau(Y) = -h(T^\nabla(X, Y), Z),$$

$$(\nabla_X S)(Y) + \tau(Y)SX - (\nabla_Y S)(X) - \tau(X)SY = -S(T^\nabla(X, Y)),$$

Ricci equation:

$$h(X, SY) - (\nabla_X \tau)(Y) - h(Y, SX) + (\nabla_Y \tau)(X) = \tau(T^\nabla(X, Y)).$$

6.4 Conormal maps and geometric quasi-divergences

$\{\omega, \xi\}$: nondegenerate, equiaffine
R_{n+1} : the dual space of R^{n+1}
$\langle\ ,\ \rangle$: the canonical pairing of R_{n+1} and R^{n+1}.

$v : M \to R_{n+1}$ is the conormal map of $\{\omega, \xi\}$

$$\overset{\text{def}}{\Longleftrightarrow} \quad \begin{aligned} \langle v(p), \xi_p \rangle &= 1, \\ \langle v(p), \omega(X_p) \rangle &= 0 \end{aligned}$$

We define a function on $TM \times M$ by

$$\rho(X, q) = \langle v(q), \omega(X) \rangle.$$

ρ is called the geometric quasi-divergence on M.

(1) If ω is symmetric, ρ is called the geometric pre-divergence on M.

(2) A SMAT or a quasi statistical manifold can be induced from these divergences.

(3) More generally, a quasi statistical manifold can be induced from a pre-contrast function.

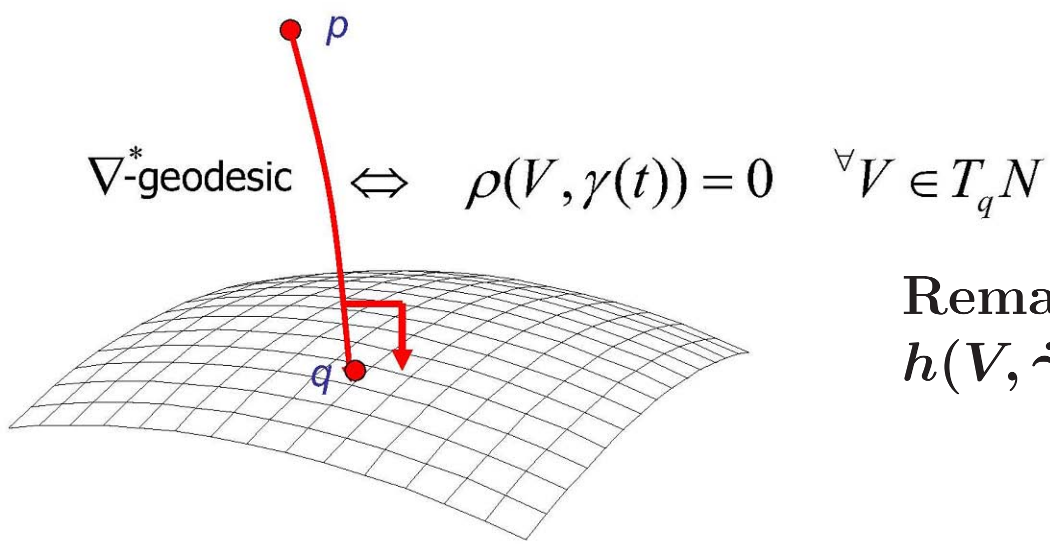
6.5 Generalized projection theorem

Theorem 6.10

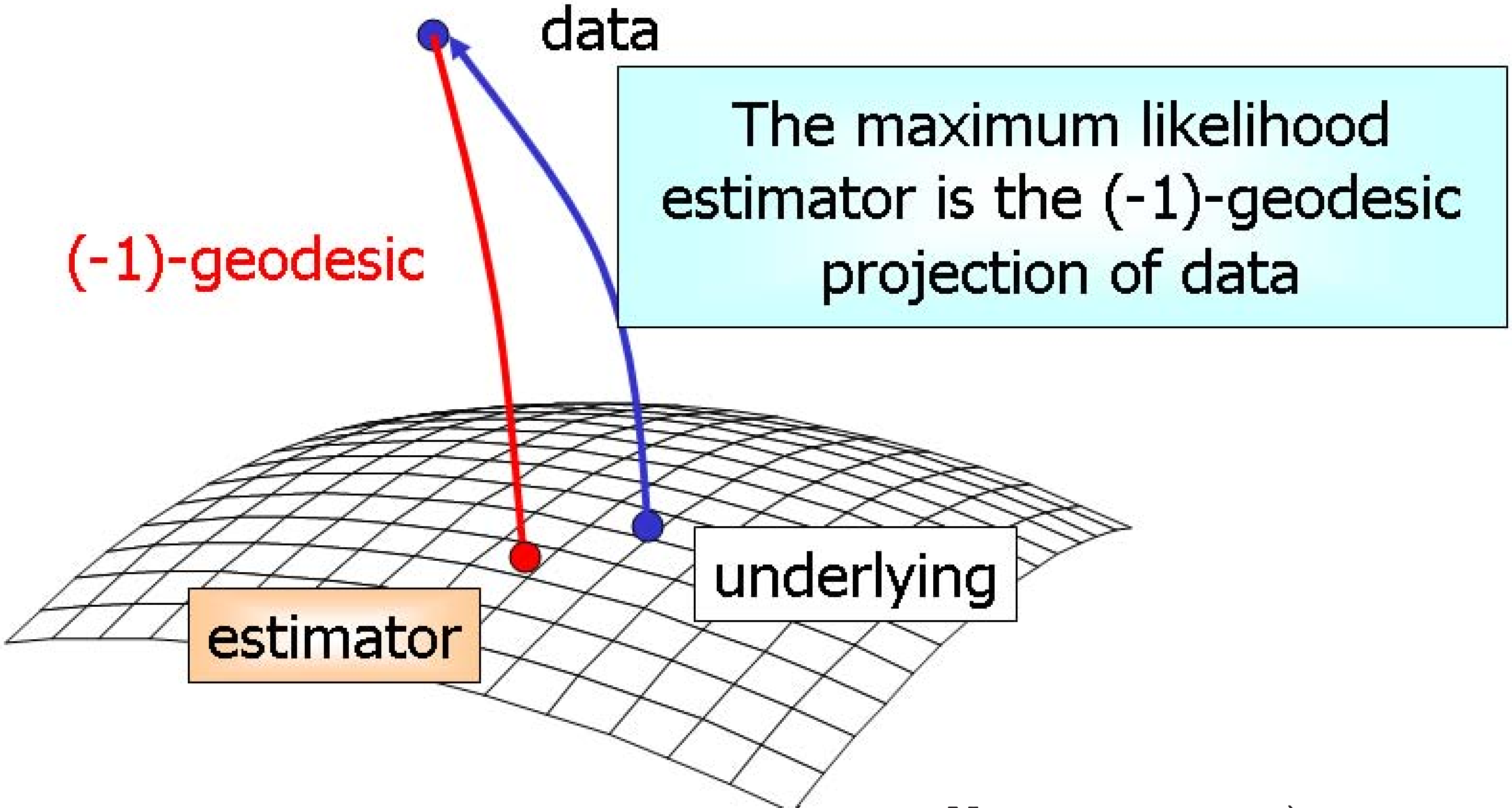
$\{\omega, \xi\}$	:	*an affine distribution to* R^{n+1}
(M, ∇, h)	:	*a quasi statistical manifold induced from* $\{\omega, \xi\}$ *with the quasi-dual connection* ∇^*
ρ	:	*the geometric quasi-divergence on* (M, ∇, h)
$N \subset M$	:	*a submanifold in* M $p \in M \setminus N, \quad q \in N$
γ	:	*the* ∇^* *geodesic connecting* p *and* q

Then

$$\gamma \perp N \text{ at } q \text{ (i.e. } h(\dot{\gamma}(0), V) = 0, {}^{\forall}V \in T_qN) \iff \rho(V, \gamma(t)) = 0$$



Remark 6.11
$h(V, \dot{\gamma}(0)) \neq 0$ *in general*



$$\min_{u \in U} D^{KL}(p(\hat{\eta}), p(u)) \iff \min_{u \in U} \left(-\frac{1}{N} \sum_{i=1}^{N} \log p(x_i; u) \right)$$

$$\implies \sum_{i=1}^{N} s(x_i; u) = 0 \quad \text{estimating equation}$$

$$s^i(x; u) = \frac{\partial}{\partial u^i} \log p(x; u) : \text{score function for } u$$

Statistical manifolds

Affine immersions $\subset$ Dual connections

- divergences, contrast functions
- exponential families, dually flat spaces

Statistical manifolds admitting torsion (SMAT)

Affine distributions $\subset$ Dual connections

- pre-divergence, pre-contrast functions
- quantum IG, non-conservative estimation

Quasi statistical manifolds

Affine distributions $\subset$ Quasi-dual connections

- quasi-divergence, quasi-contrast functions
- symplectic structures, special Kähler manifolds

Summary

q-information geometry

$S_q = \{p(x;\theta)\}$: a q-exponential family

- $(S_q, \nabla^{(e)q}, g^q)$: a Hessian manifold (a flat statistical manifold)
- $(S_q, \nabla^{(2q-1)}, g^F)$: an invariant statistical manifold $(\alpha = 2q-1)$
- $(S_q, \nabla^{e(q)}, g^q)$, $(S_q, \nabla^{(2q-1)}, g^F)$ are 1-conformally equivalent
- $(S_q, \nabla^{(2q-1)}, g^F)$ is 1-conformally flat

affine immersions and geometric divergences

$S_q = \{p(x;\theta)\}$ is realized by affine immersions $S_q \to R^{n+1}$

- $(S_q, \nabla^{e(q)}, g^{(q)})$ is realized by

  $$f = (\theta^1, \ldots, \theta^n, \psi(\theta))^T, \quad \xi = (0, \ldots, 0, 1)^T$$

  $$\rho_q(p(\theta), p(\theta')) = D(p(\theta)||p(\theta')) \ \left(= E^{esc}_{q,p(\theta')}\left[\log_q p(\theta') - \log_q p(\theta)\right]\right)$$

- $(S_q, \nabla^{(2q-1)}, g^F)$ is realized by

  $$f = (\theta^1, \ldots, \theta^n, \psi(\theta))^T, \quad \bar{\xi} = \frac{q}{Z_q}\left\{\xi + f_* \mathrm{grad}_h \left(\log \frac{Z_q}{q}\right)\right\}$$

  $$\rho^F_q(p(\theta), p(\theta')) = D^{(2q-1)}(p(\theta)||p(\theta')) \ \ (\alpha\text{-divergence})$$

Statistical inferences

Dually flat spaces

$(x_1, x_2, \ldots, x_N)$: N-independent observations
$L(\theta) = p(x_1; \theta)p(x_2; \theta) \cdots p(x_N; \theta)$
$\Longrightarrow$ Maximum likelihood estimator, dually flat spaces

Generalized conformal geometry

$(x_1, x_2, \ldots, x_N)$: N-observations, but they are correlated.
$L_q(\theta) = p(x_1; \theta) \otimes_q p(x_2; \theta) \otimes_q \cdots \otimes_q p(x_N; \theta)$
$\Longrightarrow$ anomalous statistical physics, sequential estimations
generalized conformally flat statistical manifolds

Non-integrable geometry

$(x_1, x_2, \ldots, x_N)$: N-independent events, but we cannot observe.
Likelihood functions are complicated
$\Longrightarrow$ non-conservative estimator,
statistical manifolds admitting torsion