IGAIA 4  Bohemia

Information Geometry

---  Historical Episodes and Future
     with Recent Developments

Shun-ichi Amari
RIKEN Brain Science Institute
Prehistory --- Riemannian Geometry

H. Hotteling 1929
Riemannian metric and Fisher information
location-scale model: constant curvature

P. Ch. Maharanobis 1936 Euclidean distance (multivariate-Gaussian)

C. R. Rao 1945 Cramer-Rao Theorem; Riemannian

H. Jeffreys 1946 Bayesian theory and Jeffreys invariant prior
Dual Geometry, Invariance

N. Chentsov 1972 invariance, \{g, T\}, \alpha\text{-connection}

B. Efron 1975 (A. P. Dawid) statistical curvature; higher-order asymptotics

O. Barndorff-Nielsen 1976 exponential family; Legendre transform

S. Amari 1982 duality; curvature and statistics (M. Kumon)

H. Nagaoka and S. Amari 1982 duality, Pythagorean theorem
Amari’s personal history

1958: statistics seminar (master course at U Tokyo)

S. Kullback, “Information and Statistics”
Riemannian metric (suggested by S. Moriguti)

Gaussian $N(\mu, \sigma^2)$: geodesic, constant curvature (Poincare-half plane)

beautiful structure $\rightarrow$ essential meaning?

mathematical engineering

graph and topology of networks: homology
non-Riemannian geometry of materials manifold: dislocations
information systems, learning and neural networks
Statistical curvature and higher-order inference

B. Efron, 1975
Fisher’s idea; exponential connection and mixture connection

A. P. Dawid, 1975  e- and m-connections

S. Amari :  \( \alpha \)-geometry

\[
Error^2 = \frac{1}{n} G^{-1} + \frac{1}{n^2} (H_e^2 + H_m^2 + \Gamma_m^2) + \frac{1}{n^3} K
\]

(Rao, Kano K?  Fisher’s dream)

Amari and M.Kumon  higher-order power of statistical test
Amari paper: Ann. Statist. 1982:

Reviewers (S. Lauritzen and A.P. Dawid)

Chentsov work (handwritten manuscript)


Ann. Probability Theory 7 reviewers
Zeitshrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete
genometry has nothing to do with statistics
IEEE Trans. Inf. Theory Shannon Theory, now well-known
London Workshop: 1984 (D. Cox)

Cox visited Japan in 1983
patron of information geometry

Rao, Efron, Dawid, Barndorff-Nielsen, Lauritzen
Kass, Eguchi many others

Dodson, Critchley, Marriot, Komaki, Zhang, Ay, Pistone, Giblisco, Nielsen, ...
Information Geometry --- lucky naming

Applications area:

- statistics, time-series and systems,
- machine learning, signal processing, optimization
- brain theory, consciousness
- physics, economics, mathematics
  (Banach manifold, affine differential geometry and beyond)
- quantum information, Tsallis entropy
International Conferences

IGAIA series; GSIS series, ...
Many monographs
new journal  (Jun Zhang); where to publish
mailing list and society

still small community; united and cooperative, blessed by all
My recent works

1. Systems complexity and consciousness (IIT)
2. Geometry of score matching (Hyvarinen score)
3. Natural gradient descent and topology of deep learning
4. Canonical divergence
5. Multi-terminal statistical inference
6. Information geometry and Wasserstein distance
Information Integration and Complexity of Systems

-- Stochastic approach

\[ p(x, y) = p(x) p(y \mid x) \]

\( x \): state of the brain \hspace{1cm} \( y \): next state of the brain
Necessary condition; sufficient?
**full model:** \( S_F = \{ p(x, y) \} \)  

**Disconnected model:** 
\[
S_{dis} = \{ q(x, y) \} \quad q(y \mid x) = \prod q(y_i \mid x_i)
\]

measure of interaction : N. Ay  
information integration : Tononi

Barrett and Seth

Many other \( \Phi \)
Measure of information integration, or system complexity $\Phi$

$$\Phi = D_{KL}[\rho : \hat{\rho}]$$
Definition of $\Phi$ : Postulates

1) $\Phi = \min_{q} D[p(x, y) : q(x, y)], \; q \in S_{\text{dis}}$

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

3) Disconnected model: $S_{\text{dis}}$ Markov conditions
Markov Condition

(1 → 2) branch deleted: Markov condition: \( x_1 \rightarrow x_2 \rightarrow y_2 \)

\[
p(x_1, y_2 \mid x_2) = p(x_1 \mid x_2) p(y_2 \mid x_2)
\]

\[
X_1 - X_2 - Y_2
\]

\[
X_2 - X_1 - Y_1
\]

\( S_{dis} : \text{ all } x_i \rightarrow y_j \ (i \neq j) \) deleted
Why KL-divergence?

1) \( D[p : q] \geq 0, \quad = 0, \text{ when and only when } p = q \)

2) \( D[p : q] \) invariant under transformations of \( X \)

3) \( D[p : q] = \sum_i d[p(x_i), q(x_i)] \)

4) \( D[p : q] \) induces flat structure dually
Geometric degree of information integration

$$\Phi_{geo} = \min_q D_{KL}[p(x, y) : q(x, y)], \quad q \in S_{dis}$$
Gaussian case

\[ y = A'x + e' \quad \Sigma' = \text{E}[e'e'^T] \]

\[ y = Ax + e \quad \Sigma = \text{E}[ee^T] \]

\[ A': \text{diagonal} \]

\[ \Phi_{geo} = \log \frac{|\Sigma'|}{|\Sigma|} \]
Gaussian case

\[ y = Ax + e \quad \Sigma = \text{E}[ee^T] \]
\[ y = A'x + e' \quad \Sigma' = \text{E}[e'e'^T] \quad A': \text{diagonal} \]

\[ \Phi_{\text{geo}} = \log \left| \frac{\Sigma'}{|\Sigma|} \right| \]
Many other definitions of $\Phi$

Full model
Disconnected models
Full model $S_F$ : graphical model

$$p(x, y) = \exp \left\{ \sum \theta_i^x x_i + \sum \theta_i^y y_i + \theta_{12}^x x_1 x_2 + \theta_{12}^y y_1 y_2 + \sum \theta_{ij}^{xy} x_i y_j \right\} + \text{higher-order terms } -\psi(\theta)$$

exponential family

$\theta$-coordinates $\theta$

$\eta$-coordinates $\eta_i^x = E[x_i], \eta_{ij}^{xy} = E[x_i y_j]$,

There are many disconnected models!!
Split Model $S_H$ : Ay, Barrett & Seth

$$q(x, y) = q(x) \prod q(y_i | x_i)$$

$$\theta_{12}^{XY} = \theta_{21}^{XY} = \theta_Y^{\phi} = 0$$

$$\Phi_H = D_{KL}[p : S_H] = \min_{q \in S_H} D_{KL}[p : q]$$

$$\hat{q} = \prod_{M_S} p : \hat{q}(y | x) = \prod p(y_i | x_i)$$

$$\Phi_H = \sum H[Y_i | X_i] - H[Y | X]$$
$S_H : \quad \theta_{12}^{XY} = \theta_{21}^{XY} = \theta_{12}^Y = 0; \quad \hat{\eta}_q = \eta_q$

**Mixed Coordinates :**

\[ \xi = (\eta_1^X, \eta_1^X, \eta_1^Y, \eta_{11}^{XY}, \eta_{22}^{XY}; \theta_{12}^{XY}, \theta_{21}^{XY}, \theta_{12}^Y) \]

\[ \hat{q} = \prod p = q(x, y; \xi) \]

**problem** \[ 0 \leq \Phi[p] \leq I[X : Y] \]

\[ p_{\text{ind}} = p_X(x) p_Y(y) \Rightarrow I = 0, \quad \Phi > 0 \]

**Markovian Condition**

\[ Y_1 - X_1 - X_2 - Y_2 \quad \Rightarrow \]

\[ X_1 - X_2 - Y_2; \quad Y_1 - X_1 - X_2 \]
Split Model \( S_{Gr} \)

\[
q(x, y) = q_X(x) \tilde{q}_Y(y) \Pi q(y_i | x_i)
\]

\[
\theta_{12}^{XY} = \theta_{21}^{XY} = 0
\]

\[
q(x_1, y_2 | x_2, y_1) = q(x_1 | x_2, y_1) q(y_2 | x_2, y_1)
\]

\[
0 \leq \Phi \leq I(X : Y)
\]

\[
\hat{q}_X(x) = p_Y(x), \quad \hat{q}_Y(y) = p_Y(y)
\]

\[
\hat{q}(y_i | x_i) = p(y_i | x_i)
\]
Problem: Gaussian channel

\[ p(x, y) : \ y = Ax + \varepsilon \quad \Rightarrow \quad \hat{q} : \ y = \hat{A}x + \varepsilon' \in S_G \]

\[ p(x, y) = \exp \left[ -\frac{1}{2} \left\{ x' \sum_x^{-1} x + (y - Ax)' \sum_\varepsilon^{-1} (y - Ax) - \psi \right\} \right] \]

\( \hat{A} \) is not diagonal \quad \iff \quad \theta_{12}^{XY} = \theta_{21}^{XY} = 0
Mismatched Decoding Model

Best mismatched decoding from y to x

\[ S_M = \{ q_\beta(x, y) = p(x) p_\beta(y) \prod p(y_i | x_i)^\beta \} \]

\[ \Phi^* = D_{KL}[p : S_M] \]
\[ S_H \subset S_{Gr}, S_{Geo} : S_{Gr}, S_H \text{ dually flat} \]
\[ S_M \subset S_{Gr} : S_M, S_{Geo} \text{ not flat} \]

\[ \Phi_{Gr} \leq \Phi_H, \Phi_M ; \Phi_{Geo} \leq \Phi_H \]
Transfer Entropy; Granger causality

\[ TE[x_i \rightarrow y_j] = \min D_{KL}[p : q] \quad q \in (i \rightarrow j) \text{ disconnected} \]
\[ = H[Y_j \mid X - X_i] - H[Y_j \mid X] \]

Non-additive

\[ TE[x_i \rightarrow y_j; x_k \rightarrow y_m] \neq TE[x_i \rightarrow y_j] + TE[x_k \rightarrow y_m] \]
Hierarchy: transfer entropy

Partition of X

\[ \bigcup X_i = X, \quad X_i \cap X_j = \emptyset \]
\[ \bigcup Y_i = Y, \quad Y_i \cap Y_j = \emptyset \]

Partition

subadditivity

cutting branches
split models

\[ X_1 \rightarrow Y_1 \]
\[ X_2 \rightarrow Y_2 \]
Information Geometry of Hyvärinen Game Score

Following Grinwald, Dawid, Parry, Lauritzen, Hyvärinen

\[
L(p, q) = E_{p(x)}[l(x, a)]: a = q(x)
\]

\[
S(x, q) = l(x, q) \quad l(x, q) = \log q(x)
\]

\begin{align*}
S\text{-entropy} & \quad H_S[p : q] = E_p \left[ S(x, q) \right] \\
S\text{-divergence} & \quad D_S[p : q] = H_S[p : q] - H_S[p : p]
\end{align*}
Hyvärinen score

\[
S(x,q) = \bar{I}(x,\xi) + \frac{1}{2}\{I(x,\xi)\}^2
\]

\[
D_s[p:q] = \frac{1}{2}E_p\left[\frac{d}{dx}\{\log p(x) - \log q(x,\theta)\}^2\right]
\]

\[
D[p:cq] = D[p:q]
\]

\[
s(x,\xi) = \partial_\xi \bar{I}(x,\xi) + I(x,\xi) \partial_\xi \hat{I}(x,\xi)
\]
parametric case \[ M = \{ q(\mathbf{x}, \xi) \} : A \]

\[
\min_{\xi} D_s \left[ p : q(\mathbf{x}, \xi) \right]
\]

\[ s(\mathbf{x}, \xi) = \partial_\xi S(\mathbf{x}, \xi) \] : estimating function

\[
\sum_{i=1}^{N} s(\mathbf{x}_i, \xi) = 0
\] : estimating equation

**Information geometry of** \[ D_s[p : q] \]
Asymptotic Analysis of estimator

\[
E\left[\Delta \xi \Delta \xi^T\right] = \frac{1}{N} K^{-1} V K^{-T} \geq G^{-1}
\]

\[K = E\left[ \partial_\xi s(x, \xi) \right]\]

\[V = E\left[ s(x, \xi) s(x, \xi)^T \right] \rightarrow G\]
\[ E[\Delta \xi \Delta \xi^T] = G^{-1} AG^{-1} \]

\[ A = E\left[a(x, \xi) a(x, \xi)^T\right] \]

\[ s(x, \xi) = c\{\partial_\xi \log(x, \xi) + a(x, \xi)\} \]

\[ a \perp \partial_\xi \log q: \text{efficient} \]
Hyvärinen score

\[ S(x, q) = \ddot{i}(x, \xi) + \frac{1}{2} \{ i(x, \xi) \}^2 \]

\[ D_s[p:q] = \frac{1}{2} E_p \left[ \frac{d}{dx} \{ \log p(x) - \log q(x, \theta) \}^2 \right] \]

\[ D[p: cq] = D[p: q] \]

\[ s(x, \xi) = \partial_\xi \ddot{i}(x, \xi) + \dot{i}(x, \xi) \partial_\xi \dot{i}(x, \xi) \]
Hyvärinen estimator

Fisher efficient $\leftrightarrow q$ multivariate Gaussian
Discrete case: \( x \in \text{graph nodes} \)

\[
\Delta f(x) = \frac{1}{N_x} \sum_{x' \in N} \{f(x) - f(x')\}
\]

\( N_x = \text{const} \)
\[
S(x, q) = \left( \frac{\Delta q(x)}{q(x)} \right)^2 - 2\Delta \left\{ \frac{\Delta q(x)}{q(x)} \right\}
\]

\[
D_S[\xi : \xi'] = E_{q(x, \xi)} \left\{ \frac{\Delta q(x, \xi)}{q(x, \xi)} - \frac{\Delta q(x, \xi')}{q(x, \xi')} \right\}^2
\]

\[
\sum_x \Delta f(x) h(x) = \sum_x f(x) \Delta h(x)
\]

\[
\int f'(x) h(x) \, dx = -\int f(x) h'(x) \, dx \quad \text{as} \quad f, h \to 0
\]

\[
x \to \pm\infty
\]
Deep Learning

Self-Organization + Supervised Learning

RBM: Restricted Boltzmann Machine
Auto-Encoder, Recurrent Net

tricks!!

ideas!

Dropout
Contrastive divergence

bi-directional convolution
Mathematical Neurons

\[ y = \varphi\left(\sum w_i x_i - h\right) = \varphi\left(\mathbf{w} \cdot \mathbf{x}\right) \]
Multilayer Perceptrons

\[ y = \sum v_i \varphi(w_i \cdot x) \]

\[ x = (x_1, x_2, \ldots, x_n) \]

\[ f(x, \theta) = \sum v_i \varphi(w_i \cdot x) \]

\[ \theta = (w_1, \ldots, w_m; v_1, \ldots, v_m) \]
Multilayer Perceptron

neuromanifold

space of functions $S$

$y = f(x, \theta)$

$= \sum v_i \varphi (w_i \cdot x)$

$\theta = (v_1, \ldots, v_m ; w_1, \ldots, w_m)$
Backpropagation --- stochastic gradient learning

examples: \((y_1, x_1), \ldots (y_t, x_t)\) --- training set

\[
l(y, x; \theta) = \frac{1}{2} |y - f(x, \theta)|^2
\]

\[
= -\log p(y, x; \theta)
\]

\[
\Delta \theta_t = -\eta_t \frac{\partial l(y_t, x_t; \theta_t)}{\partial \theta}
\]

\[
f(x, \theta) = \sum v_i \varphi(w_i \cdot x)
\]
singularities
Geometry of singular model

\[ y = v\varphi(w \cdot x) + n \quad \text{or} \quad v \mid w \mid = 0 \]
model: 2 hidden neurons

\[ f(x, \theta) = w_1 \varphi(J_1 \cdot x) + w_2 \varphi(J_2 \cdot x) \]

\[ y = f(x, \theta) + \varepsilon \]

\[ \varphi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-\frac{t^2}{2}} dt \]
loss function: \( l(x, y; \theta) = \frac{1}{2} \{y - f(x, \theta)\}^2 \)

\( y \): teacher signal \( \theta_0 \)  stochastic descent learning

\[ \dot{\theta} = -\eta \left< \frac{\partial l(x_i, y_i, \theta_i)}{\partial \theta} \right> \] backprop: vanilla gradient
Natural Gradient Stochastic Descent

\[ \dot{\theta} = -\eta G^{-1}(\theta_t) \langle \nabla_{\theta} (x_t, y_t, \theta_t) \rangle \]

\[ \nabla_{\theta} = \frac{\partial}{\partial \theta} \]

\[ G(\theta) = \langle \nabla_{\theta} l \nabla_{\theta} l \rangle \quad : \quad \text{Fisher Information Matrix} \]

invariant; steepest descent
Natural gradient is superior

Steepest descent; invariant  Yan Ollivier

Fisher-efficient

Natural gradient is non-vanishing even in multiple layers

Good at singular regions  (avoid plateaus: Milnor attractor)
Adaptive Natural Gradient

\[
G_{t+1}^{-1} = (1 + \varepsilon) G_t^{-1} - \varepsilon \nabla l(\mathbf{x}_i) \nabla^T l(\mathbf{x}_i) G_t^{-1}
\]

\[G^{-1} \to \infty, \quad \nabla l \to 0 \text{ at singularities}\]
Singular Region in Parameter Space

\[ R(w, J) = \{ \theta | J_1 = J_2 = J, w_1 + w_2 = w \} \]

\[ \cup \{ \theta | w_1 = 0, w_2 = w, J_2 = J \} \]

\[ \cup \{ \theta | w_1 = w, w_2 = 0, J_1 = J \} \]

\[ f(x, \theta) = w_1 \varphi(J_1 \cdot x) + w_2 \varphi(J_2 \cdot x) \]
Coordinate transformation

\[ v = \frac{w_1 J_1 + w_2 J_2}{w_1 + w_2}, \]

\[ w = w_1 + w_2, \]

\[ u = J_2 - J_1, \]

\[ z = \frac{w_2 - w_1}{w_1 + w_2} \]

\[ \xi = (v, w, u, z) \]
Singular Region

\[ R(w, J) = \{ u = 0 \} \cup \{ z = \pm 1 \} \]
Milnor attractor

Fig. 5. Critical set with local minima and plateaus.
Fig. 2: trajectories
Saddle and plateau
Topology of singular $\mathbb{R}$

blow-down coordinates : $\mathbf{x} = (\tau, \sigma, e)$

\[
\tau = c_1 \left(1 - z^2\right) u^2, \quad u = |u|
\]

\[
\sigma = c_2 z \left(1 - z^2\right) u^3,
\]

\[
e = \frac{u}{|u|} \in S_n, \quad |e| = 1
\]
Singular Region

\[ R(w, J) = \{ u = 0 \} \cup \{ z = \pm 1 \} \]
\[ e \circ \bigcirc \xrightarrow{\tau} x \times S_n \xrightarrow{P_n} \sigma \]

\[ \tau^3 + \sigma^2 = 0 \]
Sphere $S^n$ and Projective space $P^n$
natural gradient learning near singularity

\[
\frac{d}{dt}\begin{pmatrix} \tau \\ \sigma \end{pmatrix} = -\eta \begin{pmatrix} \tau \\ \sigma \end{pmatrix} \quad : \quad \text{true model} \in R
\]

\[
\frac{d}{dt}\begin{pmatrix} \tau \\ \sigma \end{pmatrix} = O(1) \quad : \quad \text{true model} \notin R
\]

Milnor attractor
Canonical Divergence in Manifold of Dual Affine Connections

Nihat Ay and S. Amari
Divergence and metric

\[ D[p:q] \geq 0 \]

\[ D[\xi: \xi + d\xi] = \frac{1}{2} g_{ij}(\xi) d\xi^i d\xi^j + O\left(|d\xi|^3\right) \]

\( G \) : Riemannian metric, positive-definite
Divergence and dual affine connections

\[ \Gamma_{ijk} \sim \nabla \quad \Gamma^*_{ijk} \sim \nabla^* \]

\[ \Gamma_{ijk} = -\partial_i \partial_j \partial'_k D[\xi : \xi']_{\xi' = \xi} \]

\[ \Gamma^*_{ijk} = -\partial'_i \partial'_j \partial_k D[\xi : \xi']_{\xi' = \xi} \]

\[ \partial_i = \frac{\partial}{\partial \xi^i} \quad \partial'_j = \frac{\partial}{\partial \xi'^j} \]
Dual geometry

\[ \{M, g, \nabla, \nabla^*\} \]

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

\[ \{M, g, T\}, \quad T_{ij} = \Gamma^*_{ijk} - \Gamma_{ijk} \]

\[ \Gamma_{ijk}^{\pm \alpha} = \Gamma_{ijk}^{\circ} \pm \frac{\alpha}{2} T_{ijk} \quad \Gamma^{\circ}: \text{Levi-Civita connection} \]
Dual geometry $\rightarrow$ canonical divergence

$M : \text{ dually flat} : \exists \psi(\theta), \varphi(\eta)$

$D[\theta : \theta'] = \psi(\theta) + \varphi(\eta') - \theta \cdot \eta'$

Bregman divergence
Exponential map: $\xi(t)$ geodesic

\[ \nabla_{\dot{\xi}} \dot{\xi} = 0 \]
\[ \xi(0) = p \]
\[ \xi(1) = q \]
\[ \dot{\xi}(0) = X = \log_p q \]
Exponential map divergence

\[ D[p:q] = \left\| X(p:q) \right\|^2 \]

\[ D_\alpha[p:q] = \left\| X_\alpha(p:q) \right\|^2 \]

\( \alpha \)-divergence
Theorem 1. Exponential map divergence induces $\alpha = -3$ geometry

Theorem 2. $\alpha = -\frac{1}{3}$ exponential map divergence recovers the original geometry

Standard divergence: $D_{\text{stan}}[p : q] = \|X_{-1/3}(p, q)\|^2$
\[ D[p:q] = \int_0^1 \langle X_t(q,p), \dot{\xi}_{q,p}(t) \rangle \, dt \]

\[ = \int_0^1 t \| \dot{\xi}_{q,p}(t) \|^2 \, dt \]

\[ \int_0^1 w(t) \| \dot{\xi}_{q,p}(t) \|^2 \, dt \]
Divergence and projection

\[ \hat{p} = \arg \min_{q \in S} D[p : q] \]

projection theorem:

\[ X = c \, \text{grad}_q \, D[p : q] \]
Data Compression in Multiterminal Statistical Inference

Shun-ichi Amari
RIKEN Brain Science Institute
A long standing problem
T. Berger; Csiszar, Ahlswede, Burnashev, Han, Amari

correlated sources $X$, $Y$

data compression and statistical inference

$$p(x, y; q) ; \text{iid}$$

$X : x_1, x_2 \cdots x_n$  $k_X$ bits

$Y : y_1, y_2 \cdots y_n$  $k_Y$ bits

$\hat{q}$
\( X : x_1 x_2 \cdots x_n \rightarrow k_X \text{ bits} \)

\( Y : y_1 y_2 \cdots y_n \rightarrow k_Y \text{ bits} \)

\[ p(x, y; q) ; \text{iid} \quad \text{Prob}\{x = y\} = q \]

binary : \( x, y = 0,1 \); \( \text{Prob}\{x = 1\} = \text{Prob}\{y = 1\} = \frac{1}{2} \)
Encoding: data compression

$X^n$  

$|x| = 2^n$

$X^n$  

$|c| = 2^{k_x}$
One-bit helper case

\[ k_X = 1, \quad k_Y = n \quad \quad c = \text{sgn}(a \cdot x) \]

\[ c \quad y \]

\[ X : x_1 x_2 \cdots x_n \rightarrow c \quad 1\text{bit} \]

\[ Y : y_1 y_2 \cdots y_n \quad \rightarrow \quad n\text{ bits} \]
Is single-bit encoding optimal?

It is optimal

\[ q = \frac{1}{2} \quad (x, y \text{ independent}), \]

but not for general \( q \).
Fisher information: \( k_X = 1, \ k_Y = n \)
Kingo Kobayashi: parity encoding

\[ x_1 \oplus x_2 \oplus \cdots \oplus x_s \]

in progress!
Information Geometry and Transportation Problem (Wasserstein distance)

entropic relaxation: \( \min <c, p> - a\{-H(p)\} \) dual

New Paper

S. Pal and T-K L. Wong,
Exponentially concave function and a new information geometry

Portfolio theory, transportation problem and information geometry
(dually projectively flat)