Conformal flattening and generalized entropies: an affine differential geometric approach

Atsumi Ohara  University of Fukui

joint work with

Hiroshi Matsuzoe and Shun-ichi Amari

April 11  2014  @Fudan University
1. Introduction

- Purpose:
  A new way to construct the Legendre structure on the family of probability distributions
  - a Legendre pair of conjugate functions,
    generalized entropy, Massieu function
  - divergence
  - generalized exp. distributions (continuous case)
Introduction

- Important dualities in Information geometry
  \((\mathcal{M}, g, \nabla, \nabla^*)\)
- Duality of affine connections
  - a mutually dual pair: \((\nabla, \nabla^*)\) generally nonflat

\[ Zg(X, Y) = g(\nabla_Z X, Y) + g(X, \nabla^*_Z Y) \]

\(X, Y, Z \in \mathcal{X}(\mathcal{M})\)  statistical manifold
Introduction

- Important dualities in Information geometry
  \((\mathcal{M}, g, \nabla, \nabla^*)\)
  - Duality of affine connections
    - a mutually dual pair: \((\nabla, \nabla^*)\) generally nonflat
      \[Zg(X, Y) = g(\nabla_Z X, Y) + g(X, \nabla^*_Z Y)\]
      \(X, Y, Z \in X(\mathcal{M})\) statistical manifold
  - Legendre duality
    - a pair of conjugate functions: \((\psi, \varphi = \psi^*)\)
    - dually flat structure

BGS entropy, Fisher metric, KL divergence, exponential family etc.\(^4\)
Introduction

- Ideas from affine differential geometry
  - Affine immersion
  - 1-conformal flatness [Kurose94]

- Note: Approach using Hessian geometry [Shima07]
  Bregman divergence [Naudts02-04, Eguchi04]
Applications

- Various Legendre structure on generalized exp. family induced from gen. log func.
  - $q$-geometry, Tsallis statistics
  - Kappa log. [Kaniadakis]

- Voronoi partitions on statistical manifolds w.r.t. geometric divergences (non-Bregman type)
  - Efficient algorithm owing to conformality

- Gradient flows on the simplex
Examples: $\alpha$-Voronoi partitions on $S^2$ [OMA12]

$\alpha = 0.6 \ (q=0.2)$ \hspace{1cm} $\alpha = -2 \ (q=1.5)$
2. Preliminaries

- **Def.** (*divergence or contrast function*)
  1) $\rho : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}$ is called a divergence if
     $$\rho(p, r) \geq 0, \ \forall p, r \in \mathcal{M}, \ \rho(p, r) = 0 \iff p = r$$
  2) $\rho$ is said to induce $(g, \nabla, \nabla^*)$ on $\mathcal{M}$ if
     $$g(X, Y) = -\rho[X|Y],$$
     $$g(\nabla_X Y, Z) = -\rho[XY|Z], \ \ g(Y, \nabla^*_X Z) = -\rho[Y|XZ],$$
     where
     $$\rho[X_1 \cdots X_k|Y_1 \cdots Y_l](p) := (X_1)_p \cdots (X_k)_p(Y_1)_q \cdots (Y_l)_q \rho(p, q)|_{p=q}$$
     $p, q \in \mathcal{M}$ and $X_i, Y_j \in \mathcal{X}(\mathcal{M})$
\(\alpha\)-conformal equivalence (1) [Kurose 94]

- **Def.** Two stat. mfds \((M, g, \nabla)\) and \((M, g', \nabla')\) are \(\alpha\)-conf. equiv. if there exists a positive func. \(\phi\) and \(\alpha \in \mathbb{R}\) s.t.

\[
g'(X, Y) = \phi g(X, Y),
\]
\[
g(\nabla'_X Y, Z) = g(\nabla_X Y, Z) - \frac{1+\alpha}{2} d(\ln \phi)(Z)g(X, Y) + \frac{1-\alpha}{2} \{d(\ln \phi)(X)g(Y, Z) + d(\ln \phi)(Y)g(X, Z)\}
\]

- **Def.** A stat. mfd. \((M, g, \nabla)\) is said \(\alpha\)-conformally flat if there exists a flat \((M, g', \nabla')\) that is \(\alpha\)-conf. equiv. with \((M, g, \nabla)\).
$\alpha$-conformal equivalence (2)

- **Prop. K** [Kurose 94]
  - A stat. mfd. is of const. curvtr. $\Rightarrow \pm 1$-conf. flat
  - $(\mathcal{M}, g, \nabla)$ and $(\mathcal{M}, g', \nabla')$ are $\alpha$-conf. equiv.
  - $(\mathcal{M}, g, \nabla^*)$ and $(\mathcal{M}, g, \nabla'^*)$ are $-\alpha$-conf. equiv.
  - $(\mathcal{M}, g, \nabla)$ and $(\mathcal{M}, g', \nabla')$ are $\alpha$-conf. equiv., and $\rho$ induces $(g, \nabla, \nabla^*)$
  - $\Rightarrow \rho'(p, q) = \phi(q) \rho(p, q)$ induces $(g', \nabla', \nabla'^*)$

$\rho'(p, q)$ is called the *conformal divergence*. 
3. Conformal flattening of ADG on $S^n$

- Probability simplex

$$S^n := \left\{ \mathbf{p} = (p_i) \middle| p_i > 0, \sum_{i=1}^{n+1} p_i = 1 \right\}$$

- Statistical manifold $(S^n, \nabla, h)$ realized by an affine immersion $(f, \xi)$.
  - $h$: Riemannian metric
  - $\nabla$: affine connection with its dual $\nabla^*$
Affine immersion and a realized geometry

- Immersion $f : S^n \to \mathbb{R}^{n+1}$
- Hypersurface $f(S^n) \subset \mathbb{R}^{n+1}$
- Transversal vector field $\xi$

The affine immersion $(f, \xi)$ realizes $(S^n, \nabla, h)$.

\[ D_X f_*(Y) = f_*(\nabla_X Y) + h(X, Y)\xi \]

$X, Y \in \mathcal{X}(S^n)$
Assumptions

1) \((f, \xi)\): non-degenerate and equiaffine

2) \(f: S^n \ni p = (p_i) \mapsto x = (x^i) \in \mathbb{R}^{n+1}, \quad x^i = L(p_i), \; i = 1, \cdots, n + 1,\)

3) \(L: \mathbb{R}_+ \rightarrow \mathbb{R}, \quad L' > 0 \quad L'' < 0\): generalized log.
   - The inverse is denoted by \(E\), i.e. \(E \circ L = \text{id}\).
Assumptions

1) \((f, \xi)\) : non-degenerate and equiaffine

2) \(f : S^n \ni p = (p_i) \mapsto x = (x^i) \in \mathbb{R}^{n+1}, \quad x^i = L(p_i), \ i = 1, \cdots, n + 1,\)

3) \(L : \mathbb{R}_+ \to \mathbb{R}, \ L' > 0, \ L'' < 0\) : generalized log.
   - The inverse is denoted by \(E\), i.e. \(E \circ L = \text{id}\).

Note

- 1) implies \((S^n, \nabla, h)\) is 1-conf. flat stat. mfd.
  - [Kurose 94] (not necessarily of const. crvtr.)
- It holds that \(E'L' = 1\).
  - \(E' > 0, \ E'' > 0\).
- \(L'' < 0\) \(h\) can be positive definite.
Define a function on $\mathbb{R}^{n+1}$ by

$$\Psi(x) := \sum_{i=1}^{n+1} E(x^i), \quad \text{convex w.r.t. } x$$

- $\frac{\partial \Psi}{\partial x^i} = E'(x^i) = \frac{1}{L'(p_i)}$,

- $f(S^n)$ is a level surface of $\Psi(x) = 1$. 
Conormal vector $\nu$ (in the dual sp. of $\mathbb{R}^{n+1}$)

\[
\nu_i(p) := \frac{1}{\Lambda} \frac{\partial \Psi}{\partial x^i} = \frac{1}{\Lambda(p)} E'(x_i) = \frac{1}{\Lambda(p)} \frac{1}{L'(p_i)},
\]

where

\[
\Lambda(p) := \sum_{i=1}^{n+1} \frac{\partial \Psi}{\partial x^i} \xi^i = \sum_{i=1}^{n+1} \frac{1}{L'(p_i)} \xi^i(p).
\]
Conormal vector $\nu$ (in the dual sp. of $\mathbb{R}^{n+1}$)

\[ \nu_i(p) := \frac{1}{\Lambda} \frac{\partial \Psi}{\partial x^i} = \frac{1}{\Lambda(p)} E'(x_i) = \frac{1}{\Lambda(p)} \frac{1}{L'(p_i)}, \]

where
\[ \Lambda(p) := \sum_{i=1}^{n+1} \frac{\partial \Psi}{\partial x^i} \xi^i = \sum_{i=1}^{n+1} \frac{1}{L'(p_i)} \xi^i(p). \]

- **Note**
\[ \xi_i < 0 \Rightarrow \Lambda(p) < 0, \quad \nu_i(p) < 0 \]

- **Properties**
\[ \langle \nu, \xi \rangle = 1, \quad \langle \nu, f_*(X) \rangle = 0, \quad \forall X \in \mathcal{X}(S^n). \]
Geometric divergence [Kurose 94]

\[ \rho(p, r) = \langle \nu(r), f(p) - f(r) \rangle = \sum_{i=1}^{n+1} \nu_i(r)(L(p_i) - L(r_i)) \]

\[ = \frac{1}{\Lambda(r)} \sum_{i=1}^{n+1} \frac{L(p_i) - L(r_i)}{L'(r_i)}. \]

\((S^n, \nabla, h)\) is also induced from \(\rho\).
Geometric divergence [Kurose 94]

\[ \rho(p, r) = \langle \nu(r), f(p) - f(r) \rangle = \sum_{i=1}^{n+1} \nu_i(r)(L(p_i) - L(r_i)) \]

\[ = \frac{1}{\Lambda(r)} \sum_{i=1}^{n+1} \frac{L(p_i) - L(r_i)}{L'(r_i)}. \]

- (\( S^n, \nabla, h \)) is also induced from \( \rho \).

- **Conformal divergence of** \( \rho \)

\[ \tilde{\rho}(p, r) = \sigma(r)\rho(p, r), \sigma(r) > 0 : \text{conf. factor} \]

- \( \tilde{\rho} \) defines another geometric structure (\( S^n, \tilde{\nabla}, \tilde{h} \))

  \textbf{conformal transformation} of (\( S^n, \nabla, h \))
Conformal flattening

- Conformal flattening = Normalization of the conormal vectors \( \nu(p) \) to the simplex.

- Geometrical explanation
Conformal flattening and a realized geometry

- Set $\sigma(r) = -1/Z(r)$ where $Z(p) := \sum_{i=1}^{n+1} \nu_i(p) = \frac{1}{\Lambda(p)}$
- $Z(p) < 0$ : normalization term
Conformal flattening and a realized geometry

- Set \( \sigma(r) = -1/Z(r) \) where \( Z(p) := \sum_{i=1}^{n+1} \nu_i(p) = \frac{1}{\Lambda(p)} \)

\( Z(p) < 0 \): normalization term

**Prop.** The structure \((S^n, \tilde{\nabla}, \tilde{h})\) induced from the divergence:

\[ \tilde{\rho}(p, r) = -\frac{1}{Z(r)} \rho(p, r), \]

is dually flat.

Further, \( \tilde{\rho} \) is a **canonical** divergence, i.e.,

\[ \tilde{\rho}(p, r) = \psi(\theta) + \varphi(\eta) - \sum_{i=1}^{n} \theta^i \eta_i, \]

\[ \varphi = \psi^* \]
Proposition (ctd.)

where

\[
\begin{align*}
\theta^i(p) &= x^i(p) - x^{n+1}(p) = L(p_i) - L(p_{n+1}), & i = 1, \ldots, n \\
\eta_i(p) &= P_i(p) := \frac{\nu_i(p)}{Z(p)}, & i = 1, \ldots, n \\
\psi(p) &= -x_{n+1}(p) = -L(p_{n+1}) \\
\varphi(p) &= \frac{1}{Z(p)} \sum_{i=1}^{n+1} \nu_i(p)x^i(p) = \sum_{i=1}^{n+1} P_i(p)L(p_i).
\end{align*}
\]
Proposition (ctd.)

Where

\[ \theta^i(p) = x^i(p) - x^{n+1}(p) = L(p_i) - L(p_{n+1}), \quad i = 1, \ldots, n \]

\[ \eta_i(p) = P_i(p) := \frac{\nu_i(p)}{Z(p)}, \quad i = 1, \ldots, n \]

\[ \psi(p) = -x_{n+1}(p) = -L(p_{n+1}) \]

\[ \varphi(p) = \frac{1}{Z(p)} \sum_{i=1}^{n+1} \nu_i(p)x^i(p) = \sum_{i=1}^{n+1} P_i(p)L(p_i). \]

- (gen. entropy)

- (gen. escort prob.)

- (gen. Massieu function)

- (gen. entropy)

\[ P_i(p) = \frac{E'(x_i)}{\sum_{k=1}^{n+1} E'(x_k)} \]
Remark 1

- When $L = \log$, it recovers the standard IG, i.e.,
  $$(\mathcal{S}^n, \tilde{\nabla}, \tilde{h}) = (\mathcal{S}^n, \nabla^{(e)}, g^F)$$
  and $\tilde{\rho}$ is KL divergence.

- Voronoi partition on $\mathcal{S}^n$ w.r.t. $\rho$
  - Easy to compute via the standard algorithm using the potential $\psi$.  

Summary of our approach

1. Start with a generalized logarithmic function $L$ and the probability simplex $S^n$.

2. Affine immersion realizes a 1-conformally flat geometry $(S^n, \nabla, \bar{h})$. (not dually flat, generally)

3. Conformal flattening of $(S^n, \nabla, \bar{h})$ realizes a dually flat geometry $(S^n, \tilde{\nabla}, \tilde{h})$.
   - The Legendre duality is automatically established.
Example: $q$-geometry

\[
L(t) := \frac{1}{1-q} t^{1-q}, \quad x^i(p) = \frac{1}{1-q} (p_i)^{1-q}
\]

\[
\xi^i(p) = -q(1-q)x^i(p)
\]

realizes $(S^n, \nabla^{(\alpha)}, g^F)$, $q = (1+\alpha)/2$. $\alpha$-geometry

Fisher metric, constant curvature

\[
R^{(\alpha)}(X,Y)Z = \kappa \{ g(Y,Z)X - g(X,Z)Y \}
\]

\[
\kappa = (1 - \alpha^2)/4 = q(1 - q)
\]
Example: $q$-geometry

\[
L(t) := \frac{1}{1-q} t^{1-q}, \quad x^i(p) = \frac{1}{1-q} (p_i)^{1-q} \\
\xi^i(p) = -q(1 - q)x^i(p)
\]

affine immersion

realizes \((S^n, \nabla^{(\alpha)}, g^F), q = (1 + \alpha)/2\).

\(\alpha\)-geometry

\[
\Psi(x) = \sum_{i=1}^{n+1} \left( (1 - q)x^i \right)^{1/(1-q)}, \quad \Lambda(p) = -q, \quad (constant)
\]

conformal flattening

\[
\nu_i(p) = -\frac{1}{q} (p_i)^q, \quad \frac{1}{Z(p)} = \frac{q}{\sum_{k=1}^{n+1} (p_i)^q}
\]

conormal vec. & conf. factor

\[
exp_q(x) := (1 + (1 - q)x)^{1/(1-q)}
\]
Example: $q$-geometry

\[
L(t) := \frac{1}{1 - q} t^{1-q}, \quad x^i(p) = \frac{1}{1 - q} (p_i)^{1-q} \]

affine immersion

\[
\xi^i(p) = -q(1 - q)x^i(p) \]

realizes $(S^n, \nabla^{(\alpha)}, g^F)$, $q = (1 + \alpha)/2$. $\alpha$-geometry

\[
\Psi(x) = \sum_{i=1}^{n+1} ((1 - q)x^i)^{1/1-q}, \quad \Lambda(p) = -q, \quad \text{(constant)}
\]

conformal flattening

\[
\nu_i(p) = -\frac{1}{q} (p_i)^q, \quad -\frac{1}{Z(p)} = \frac{q}{\sum_{k=1}^{n+1} (p_i)^q}
\]

conormal vec. & conf. factor

\[
\eta_i = \frac{(p_i)^q}{\sum_{k=1}^{n+1} (p_k)^q}, \quad \theta^i = \frac{1}{1 - q} (p_i)^{1-q} - \frac{1}{1 - q} (p_{n+1})^{1-q} = \ln_q(p_i) - \psi(p)
\]

\[
\psi(p) = -\ln_q(p_{n+1}), \quad \varphi(p) = \ln_q\left(\frac{1}{\exp_q(S_q(p))}\right) = -S^N_q
\]
Remark 2

- The choice of $\xi$ does not affect on the obtained flattened structure $(S^n, \tilde{\nabla}, \tilde{h})$.

- $(S^n, \tilde{\nabla}, \tilde{h})$ is 1-conformally equivalent with $(S^n, \nabla, h)$.

\[
\tilde{h} = -\frac{1}{Z} h, \quad \text{conformal}
\]

\[
\nabla^* \sim \tilde{\nabla}^* \quad \text{projectively equivalent}
\]
4. Gradient flow w.r.t. \((\mathcal{S}^n, \nabla, \tilde{h})\)

- One of the fundamental flows
  - Relation with the H–theorem

- The Riemannian metric \(\tilde{h}\) is extended as a diagonal form on \(\mathbb{R}_{++}^{n+1}\).

\[
\tilde{h}_{ii}(p) = -\frac{1}{Z(p)\Lambda(p)} \frac{L''(p_i)}{L'(p_i)}, \quad \tilde{h}_{ij} = 0
\]

- \(\phi(x) := \frac{L'(x)}{L''(x)}\)
\(V(p)\): a function on \(\mathbb{R}_+^{n+1}\)

\(f_i(p)\):
\[
f_i = \frac{\partial V}{\partial p_i}, \quad i = 1, \cdots, n + 1,
\]

gradient flow on \(S^n\) w.r.t. \(\tilde{h}\) that maximizes \(V(p)\)
(Cf. [Harper 11])

\[
\dot{p}_i = -Z(p)\Lambda(p)\phi(p_i)(f_i(p) - \bar{f}_\phi(p)) \quad \text{(GF)}
\]

\[
\bar{f}_\phi(p) := \sum_{i=1}^{n+1} \tilde{\phi}(p_i) f_i(p), \quad \tilde{\phi}(p_i) := \phi(p_i) / \sum_{k=1}^{n+1} \phi(p_k)
\]
Special case: $L(t) = t^{1-q}/(1 - q)$

- $q$-geometry $(S^n, \tilde{\nabla}, \tilde{h})$ (previous example)

$$
\phi(t) = -\frac{t}{q} \quad , \quad \tilde{h}(p) = -\frac{1}{Z(p)}g^F(p)
$$

$g^F$: Fisher (Shahshahani) metric
Special case: 
\[ L(t) = t^{1-q}/(1-q) \]

- \( q \)-geometry \((S^n, \tilde{\nabla}, \tilde{h})\) (previous example)
  \[ \phi(t) = -\frac{t}{q}, \quad \tilde{h}(p) = -\frac{1}{Z(p)}g^F(p) \]

  \( g^F \): Fisher (Shahshahani) metric

- Gradient flow (GF) for \( q \)-geometry
  \[ \dot{p}_i = -Z(p)p_i(f_i(p) - \bar{f}(p)), \quad \bar{f}(p) := \sum_{i=1}^{n+1} p_i f_i(p) \]

- Cf. the replicator equation \((q=1)\)
  \[ \dot{p}_i = p_i(f_i(p) - \bar{f}(p)) \]
Prop. For each $q > 0$, the gradient flow (GF) for the $q$-geometry traces the same trajectory with the replicator equation ($q = 1$), but the different velocity by the conformal factor $-Z(p)$. 
For nonintegrable $f_i$

- **Assumptions:**
  1) $f_i(p) := \frac{L''(p_i)}{(L'(p_i))^2} \sum_{j=1}^{n+1} a_{ij} P_j(p), \quad a_{ij} = -a_{ji}$
  2) The (GF) has an equilibrium $r$ in $S^n$.

- **Prop.** The (GF) conserves $\tilde{\rho}(p, r)$. The first integral

- **Rem.** Cf. [Tokita 04] for the case $L = \log$. 

References

- J. Naudts, Reviews in Mathematical Physics, 16, 6, 809-822 (2004).
Thank you for your attention.