Geometry on Positive Definite Matrices Induced from V-Potentials -Foliated Structure and an Application-

Atsumi Ohara
University of Fukui

&

Shinto Eguchi
Institute of Statistical Mathematics

統計多様体の幾何学とその周辺, 北海道大学 2014 Jan. 11
1. Introduction

- \( PD(n, \mathbb{R}) \) : the set of Positive Definite real symmetric matrices

- logarithmic characteristic func. on \( PD(n, \mathbb{R}) \)
  
  [Vinberg 63], [Faraut & Koranyi 94]

\[
\varphi(P) = -\log \det P, \quad P \in PD(n; \mathbb{R})
\]
\[ \phi(P) = -\log \det P \text{ appears in} \]

- Semidefinite Programming (SDP)
  self-concordant barrier function
- Multivariate Analysis (Gaussian dist.)
  log-likelihood function
  (structured covariance matrix estimation)
- Symmetric cone: log characteristic function
- Information geometry on \( PD(n, \mathbb{R}) \)
  a potential function in standard case
Information geometry on $\mathcal{M}$

Dualistic geometrical structure

$$Xg(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla^*_X Z)$$

$X, Y$ and $Z$ : arbitrary vector fields on $\mathcal{M}$

$g$ : Riemannian metric

$\nabla$, $\nabla^*$ : a pair of dual affine connections
standard IG on $PD(n, \mathbb{R})$

[O, Suda & Amari LAA96]

$\phi(P)$ plays a role of potential function

- $g$: Riemannian metric is the Hesse matrix of $\phi(P)$ (Fisher for Gaussian)

- $\nabla, \nabla^*$: related to the third derivatives of $\phi(P)$

Nice properties: $GL(n, \mathbb{R})$-invariant (unique), KL-divergence, Pythagorean theorem, etc
Purpose of this presentation

- The other convex potentials
  \[ \varphi^{(V)}(P) = V(\det P) \]

- Their different and/or common geometric structures
  - Structure of submanifolds
  - Decomposition properties of divergences
  - Application to non-Gaussian pdf’s
Outline

- Review
  - Dualistic geometry induced by V-potentials
- Foliated structures
  - Submanifold of constant determinant
  - Decomposition of divergence
- Applications to multivariate elliptic pdf’s
  - $\text{GL}(n)$-invariance of geometry of q-Gaussian family induced by beta-divergence
2. Preliminaries and Notation

- $\text{Sym}(n; \mathbb{R})$: the set of $n$ by $n$ real symmetric matrix

  vec. sp. of dimension $N(= n(n + 1)/2)$

- $\{E_i\}_{i=1}^{N}$: arbitrary set of basis matrices

- (primal) affine coordinate system

  $$\text{Sym}(n; \mathbb{R}) \ni X = \sum_{i=1}^{N} x^i E_i$$

- Identification

  $$T_{PD}(n) \ni (\partial/\partial x^i)_P \equiv E_i \in \text{Sym}(n)$$
V-potential function

Def.

\[ \phi^{(V)}(P) = V(\det P), \quad V(s): \mathbb{R}_+ \to \mathbb{R} \]

-The standard case:

\[ V(s) = -\log s \Rightarrow \phi(P) = -\log \det P \]
Def.  
\[ \nu_i(s) = \frac{d\nu_{i-1}(s)}{ds}, \quad i = 1, 2, \cdots, \quad \text{where } \nu_0(s) = V(s) \]

Rem.  The standard case $V = -\log$:
\[ \nu_1(s) = -1, \quad \nu_k(s) = 0, \quad k \geq 2 \]

Prop. 1 (convexity conditions)

The Hessian matrix of the $V$-potential is positive definite on $PD(n, \mathbb{R})$ if and only if

For $\forall s > 0$,

i) $\nu_1(s) < 0$,  \hspace{1cm} ii) $\beta(V)(s) < \frac{1}{n}$, where $\beta(V)(s) = \frac{\nu_2(s)}{\nu_1(s)}$
Assumption: the convexity conditions hold.

- Riemannian metric is
  \[ g_P^{(V)}(X, Y) \]
  \[ = -\nu_1 (\det P) \text{tr}(P^{-1}XP^{-1}Y) + \nu_2 (\det P) \text{tr}(P^{-1}X) \text{tr}(P^{-1}Y) \]

Here,

\[ X, Y \text{ in } \text{sym}(n, \mathbb{R}) \sim \text{tangent vectors at } P \]

Rem. The standard case \( V = -\log \):

\[ g_P^{(V)}(X, Y) = \text{tr}(P^{-1}XP^{-1}Y) \]
Prop. (affine connections)

Let $\nabla$ be the canonical flat connection on $PD(n, \mathbb{R})$. Then the V-potential defines the following dual connection $^*\nabla^{(V)}$ with respect to $g^{(V)}$:

\[
\left( {^*\nabla_{\frac{\partial}{\partial x^i}}^{(V)}} \right)_P = -E_i P^{-1} E_j - E_j P^{-1} E_i - \Phi(E_i, E_j, P) - \Phi^\perp(E_i, E_j, P),
\]

\[
\Phi(X, Y, P) = \frac{\nu_2(s) \text{tr}(P^{-1}X)}{\nu_1(s)} Y + \frac{\nu_2(s) \text{tr}(P^{-1}Y)}{\nu_1(s)} X,
\]

\[
\Phi^\perp(X, Y, P) = \frac{(\nu_3(s) \nu_1(s) - 2\nu_2^2(s)) \text{tr}(P^{-1}X) \text{tr}(P^{-1}Y) + \nu_2(s) \nu_1(s) \text{tr}(P^{-1}XP^{-1}Y)}{\nu_1(s)(\nu_1(s) - n\nu_2(s))}.
\]
(g^(V), ∇, *∇^(V)) : Dually flat structure on PD(n, R) induced by the V-potential

**divergence function**

\[ D^{(V)}(P, Q) = \varphi^{(V)}(P) + \varphi^{(V)*}(Q^*) - \langle Q^*, P \rangle \]
\[ = V(\det P) - V(\det Q) + \langle Q^*, Q - P \rangle. \]

\[ P^* = \text{grad} \varphi^{(V)}(P) = \nu_1(\det P)P^{-1} \]
\[ \varphi^{(V)*}(P^*) = n\nu_1(\det P) - \varphi^{(V)}(P). \]

- a variant of relative entropy,
- Pythagorean type decomposition
3. Group Invariance of the structure \((g^{(V)}, \nabla, \ast \nabla^{(V)})\) on \(PD(n, \mathbb{R})\)

- Linear transformation on \(PD(n, \mathbb{R})\)
  - Congruent transformation: \(\tau_G P = GPG^T, G \in GL(n, \mathbb{R})\),
  - The differential: \((\tau_G)^* X = GXG^T\)
3. Group Invariance of the structure \((g^{(V)}, \nabla, \nabla^{(V)})\) on \(PD(n, \mathbb{R})\)

- **Linear transformation** on \(PD(n, \mathbb{R})\)
  congruent transformation: \(\tau_G P = GPG^T, G \in GL(n, \mathbb{R})\),
  the differential: \((\tau_G)^* X = GXG^T\)

- **Invariance**
  - metric: \(g_{\tilde{P}}(\tilde{X}, \tilde{Y}) = g_P(X, Y)\)
  - connections: \((\tau_G)^*(\nabla_X Y)_P = (\nabla_{\tilde{X}} \tilde{Y})_{\tilde{P}}\)
    and the same for \(*\nabla^{(V)}\)

where
\[
\tilde{P} = \tau_G P, \tilde{X} = (\tau_G)^* X, \tilde{Y} = (\tau_G)^* Y
\]
Prop.

The largest group that preserves the dualistic structure \((g^{(V)}, \nabla, *\nabla^{(V)})\) invariant is

\[
\tau_G \quad \text{with} \quad G \in SL(n, \mathbb{R})
\]

except in the standard case.

Rem. the standard case: \(\tau_G \quad \text{with} \quad G \in GL(n, \mathbb{R})\)

Rem. The power potential of the form:

\[
V(s) = c_1 + c_2 s^\beta
\]

has a special property.
2. Foliated structures

The following foliated structure features the dualistic geometry \((g^{(V)}, \nabla, *\nabla^{(V)})\) induced from every \(V\)-potential.

\[
PD(n, \mathbb{R}) = \bigcup_{s>0} \mathcal{L}_s, \quad \mathcal{L}_s = \{P | P > 0, \det P = s\}.
\]

\[
PD(n, \mathbb{R}) = \bigcup_{P \in \mathcal{L}_s} \mathcal{R}_P. \quad \mathcal{R}_P = \{Q | Q = \lambda P, 0 < \lambda \in \mathbb{R}\}
\]
Prop.
Each leaf $\mathcal{L}_s$ and ray $\mathcal{R}_P$ are orthogonal to each other with respect to $g^{(V)}$.

Prop.
Every $\mathcal{R}_P$ is simultaneously a $\nabla$- and $\nabla^{(V)}$-geodesic for an arbitrary $V$-potential.
Submanifolds of const det

Induced geometry on $\mathcal{L}_s$ from $(g^{(V)}, \nabla, *\nabla^{(V)})$

**Prop.** For any V, the followings hold:

i) Riemannian metric: $\tilde{g}^{(V)} = -\nu_1(s)\tilde{g}^{(-\log)}$

ii) Divergence: $D^{(V)}(P, Q) = -\nu_1(s)D^{(-\log)}(P, Q)$

iii) Dual connection: $*\tilde{\nabla}^{(V)} = \tilde{\nabla}^{(-\log)}$
Submanifolds of const det

Prop.
Each leaf $(\mathcal{L}_s, \tilde{g}^{(V)})$ is a Riemannian symmetric space

$$\mathcal{L}_s \simeq SL(n, \mathbb{R})/SO(n)$$

$$\iota_s : \mathcal{L}_s \rightarrow \mathcal{L}_s : \text{Involutive isometry of } (\mathcal{L}_s, \tilde{g}^{(V)})$$

$$\iota_s P = - \left( \frac{s}{s^*} \right)^{\frac{1}{n}} \text{grad}\varphi^{(V)}(P) = s^{\frac{2}{n}} P^{-1}, \quad s^* = \frac{(-\nu_1(s))^n}{s}.$$
Submanifolds of const det

- Level surface of both $\varphi^{(V)}$ and $\varphi^{(V)}*$ ADG
- Normal vector field $N$

$$N = -\frac{1}{d\varphi^{(V)}(E)} E.$$  

$$g^{(V)}(X, E) = d\varphi^{(V)}(X), \quad \forall X \in \mathcal{X}(PD(n, \mathbb{R}))$$

- Centro-affine immersion

$$\nabla_X Y = \tilde{\nabla}_X Y + h(X, Y) N,$$

$$\nabla_X N = -A(X) + \tau(X) N.$$  

$$h = \tilde{g}^{(V)}, \quad \tau = 0$$
Submanifolds of const det

Lemma [UOF 00]

- The submfd \( (\mathcal{L}_s, \tilde{\nabla}, \tilde{g}^{(V)}) \) is 1-conformally flat.
- Assume

\[
P, Q \in \mathcal{L}_s, R \in \mathcal{R}_Q, R = \lambda Q, \lambda > 0.
\]

Then

\[
D^{(V)}(P, R) = \mu D^{(V)}(P, Q) + D^{(V)}(Q, R),
\]

where

\[
R^* = \mu Q^*, \text{ i.e., } \mu = \lambda^{-1} \nu_1(\det R)/\nu_1(\det Q) > 0.
\]
Submanifolds of const det

+ Illustration: decomposition of divergence (1)

If $P, Q \in \mathcal{L}_s$ and $R \in \mathcal{R}_Q$ with $R = \lambda Q$, $\lambda > 0$,

then

$$D^{(V)}(P, R) = \mu D^{(V)}(P, Q) + D^{(V)}(Q, R)$$

where $\mu = \lambda^{-1} \nu_1(\det R)/\nu_1(\det Q) > 0$
Prop.
Each leaf $\mathcal{L}_s$ is a homogeneous space of constant negative curvature $k_s = 1/(\nu_1(s)n)$.

$$R(X, Y)Z = k_s\{g(Y, Z)X - g(X, Z)Y\}.$$  

Shown by Gauss eq. and $\mathbf{A} = k_s \mathbf{I}$

Lemma (modified Pythagorean)[Kurose94]

$P, Q, R \in \mathcal{L}_s$ \quad $\tilde{\gamma}_{PQ} \perp \tilde{\gamma}_{QR}^*$ at $Q \Rightarrow$


Decomposition of divergences (2)
Submanifolds of const det

+ Combining the two decomposition results

**Prop.**

If $\tilde{\nabla}$-geodesic $\tilde{\gamma}$ and $\tilde{\nabla}$-geodesic $\tilde{\gamma}^*$ are orthogonal at $R$, then

$$D^{(V)}(P, S) = D^{(V)}(P, R) + \kappa D^{(V)}(R, S)$$

$$\kappa = \lambda \{ 1 - \kappa_s D^{(V)}(Q, R) \} > 0$$

$$Q \in \mathcal{L}_s \cap \mathcal{R}_P$$
3. Application to multivariate statistics

- Non Gaussian distribution (generalized exponential family)
  - Robust statistics
    - beta-divergence,
    - Machine learning, and so on
  - Nonextensive statistical physics
    - Power distribution,
    - generalized (Tsallis) entropy, and so on
U-model and U-divergence

- U-model

**Def.**

Given a convex function $U$ on $\mathbb{R}$ and set $u = U'$, U-model is a family of elliptic pdf's specified by $P$:

$$
\mathcal{M}_U = \left\{ f(x, P) = u \left( -\frac{1}{2} x^T P x - c_U(\det P) \right) : P \in PD(n, \mathbb{R}) \right\}
$$

$c_U(\det P)$ : normalizing const.
Rem. When $U=\exp$, the U-model is the family of Gaussian distributions.

**U-divergence:**

**Natural closeness measure** on the U-model

$$D_U(f, g) = \int [U(\xi(g)) - U(\xi(f)) - \{\xi(g) - \xi(f)\}] f dx,$$

where $\xi$ is the inverse function of $u$.

Rem. When $U=\exp$, the U-divergence is the Kullback-Leibler divergence (relative entropy).
Example: beta-model and beta-divergence (1)

- Beta-model $M_{\beta}$
  - For $\beta \neq 0$ and $\beta \neq -1$

$$U(s) = \begin{cases} \frac{1}{\beta + 1} (1 + \beta s)^{(\beta+1)/\beta}, & s \in I_\beta = \{s \in \mathbb{R} | 1 + \beta s > 0\} \\ 0, & \text{otherwise} \end{cases}$$

$$u(s) = \begin{cases} \frac{dU(s)}{ds} = (1 + \beta s)^{1/\beta}, & s \in I_\beta = \{s \in \mathbb{R} | 1 + \beta s > 0\} \\ 0, & \text{otherwise} \end{cases}$$

$$\xi(t) = \frac{t^\beta - 1}{\beta}, \quad t > 0$$

- q-exponential and q-logarithmic functions
Example: beta-model and beta-divergence (2)

- Beta-divergence

\[ D_\beta(f, g) = \int \frac{g(x)^{\beta+1} - f(x)^{\beta+1}}{\beta + 1} - \frac{f(x)\{g(x)^\beta - f(x)^\beta\}}{\beta} \, dx \]
IG induced from divergences

- Divergence induces stat mfd structure.

\[
g^{(D)}(X, Y) = -D(X|Y),
\]

\[
g^{(D)}(\nabla^{(D)}_X Y, Z) = -D(XY|Z),
\]

\[
g^{(D)}(*\nabla^{(D)}_X Y, Z) = -D(Z|XY),
\]

where

\[
D(X_1 \cdots X_n|Y_1 \cdots Y_m)(p) = (X_1)_p \cdots (X_n)_p (Y_1)_q \cdots (Y_m)_q D(p, q)\bigg|_{p=q}
\]
Relation between the U- and V-geometries

Prop.

IG on $\mathcal{M}_U$ induced from $D_U$ coincides with $(g^{(V)}, \nabla, *\nabla^{(V)})$ derived from the following V-potential function:

$$V(s) = s^{-\frac{1}{2}} \int U \left( -\frac{1}{2} x^T x - c_U(s) \right) dx + c_U(s), \quad s > 0.$$
Group invariance for the power potentials $V(s) = c_1 + c_2 s^\beta$

Prop.

$V$ is of the power form $\iff$

1) Orthogonality is $GL(n)$-invariant.

2) The dual affine connections derived from the power potentials are $GL(n)$-invariant.

Hence,

- Both $\nabla$- and $^*\nabla^{(V)}$-projections are $GL(n)$-invariant.
Thm [O & Eguchi 13]
IG on $\mathcal{M}_\beta$ induced from $D_\beta$ coincides with $(g^{(V)}, \nabla, *g^{(V)})$ on $PD(n, \mathbb{R})$ induced from

$$V(s) = \begin{cases} 
\frac{1}{\beta} + c^+ s^{1/(2n\beta)}, & \beta > 0 \\
\frac{1}{\beta} + c^- s^{1/(2n\beta)}, & -\frac{2}{n+2} < \beta < 0
\end{cases}$$

Implication: statistical inference on $\mathcal{M}_\beta$ using $D_\beta$ is GL(n)-invariant.
Conclusions

- Derived dualistic geometry is invariant under the \( SL(n,R) \)-group actions.
  - For power function, dual connections and orthogonality are \( GL(n,R) \)-invariant.
- Each leaf is a homogeneous manifold with a negative constant curvature.
- Decomposition of the divergence function
- Correspondence between the U- and V-geometries
  - Statistical inference on Beta-model using dual projections are \( GL(n,R) \)-invariant.
Main References


T. Kanamori and A. Ohara,