NOTE ON THE METRIC AND AFFINE CONNECTIONS ON
THE SPACE OF FINITE MEASURES

HUAIYU ZHU

Abstract. Explicit formulas for the metric and affine connections on
the space of finite measures are derived in different coordinate systems.
They are transformed and compared with previous formulas of Čencov,
Amari and others. These formulas are given for finite sample spaces but
in a form that is easily generalizable to infinite sample spaces. An error
in Čencov’s book is corrected.

1. Introduction

1.1 Let $X$ be a sample space with a $\sigma$-field of measurable sets. Let $\mathcal{P}(X)$
and $\mathcal{M}_+(X)$ be the spaces of all probability measures and
finite measures on $X$, respectively. That is,

$$\mathcal{P}(X) = \left\{ p \in \mathcal{M}_+(X) : \int p = 1 \right\}.$$

1.2 Čencov (1982) proved that for $X = \mathbb{N}_m := \{1, 2, \ldots \}$ being a finite set,
there is only one (modulo a constant multiplication factor) invariant metric
$g$ on $\mathcal{P}(X)$ in the category of congruent (sufficient) Markov morphisms,
and there is only one family of equivariant affine connections $\tilde{\nabla}$ indexed by
a real parameter $\gamma$. The metric is defined by the Fisher information matrix.

1.3 Amari (1985) developed the dual affine geometry which associates
these concepts beautifully with the information divergence $D_\gamma$ commonly
used in information theory and statistics. He also extended these concepts
to $\mathcal{M}_+(X)$ and made clear that the $\gamma$-affine structure is simply the affine

Date: August 11, 1998.

1991 Mathematics Subject Classification. Primary 62E10; Secondary 53B20, 94A17.

I am indebted to discussions with S. Amari. This work was supported by TXN, Inc.
structure induced by the $\gamma$-coordinate $l_\gamma(p) = p^\gamma / \gamma$. The parameter $\gamma$ here corresponds to $\alpha$ of Amari with $\alpha = 1 - 2\gamma$. Conversely, Eguchi (1983) showed that the metric $g$ and affine connections $\nabla^\gamma$ are uniquely determined by the second and third order derivatives of the information deviation $D_\gamma$.

1.4 Zhu and Rohwer (1997, 1995) applied these results to Bayesian decision theory and revealed the importance of the dual affine structure on $\mathcal{M}_+(X)$. Additional formulas for

\[(1.2) \quad g_{ij} := g(e_i, e_j), \quad \Gamma^\gamma_{ijk} = g(e_i \nabla^\gamma e_j, e_k),\]

where $e_i$ is the $i$th unit coordinate vector, are also given in different $\tau$-coordinates \(^1\).

1.5 Unfortunately, the representations used in these works are somewhat different and details are often omitted, making it difficult to associate these results. In this note we explicitly spell out all these formulas concerning $g$, $\Gamma^\gamma$ and $D_\gamma$ in any $\tau$-coordinates for both $\mathcal{M}_+(X)$ and $\mathcal{P}(X)$, where $X = \mathbb{N}_m$.

1.6 We do not derive formulas for arbitrary (ie. infinite) $X$, but in the final section we discuss what the corresponding formulas would look like. The algebraic forms are similar, but there will be additional topological issues involved. Some additional related topics are also discussed here.

2. METRIC AND AFFINE CONNECTIONS ON $\mathcal{M}_+(X)$

2.1 Notations Except when explicitly noted, we shall assume $X = \mathbb{N}_m$ and omit explicit reference to it. For differential geometry equations, we shall use the index notation but without the summation convention. Generally, $\gamma, \tau \in [0, 1]$, but the boundary cases $\gamma, \tau \in \{0, 1\}$ often require special considerations.

\(^1\)The covariant differentiation of $v$ with respect to $u$ is usually written as $\nabla_u v$, but the notation $u \nabla v$ makes explicit the multiplicative role of $u$ and is typographically nicer.
2.2 Coordinates  The $\gamma$-coordinate $\xi$ of $p \in M_+$ is an element of a 
space (Amari, 1985, p. 66)

\begin{align*}
(2.1) \quad l_\gamma(p) :&= p^\gamma / \gamma \in M_+^\gamma \cong \mathbb{R}^m, \\
(2.2) \quad l_0(p) :&= \log p \in M_+^0 \cong \mathbb{R}^m.
\end{align*}

The tangent space is a linear space,

\begin{equation}
(2.3) \quad T_p M_+^\gamma = M^\gamma \cong \mathbb{R}^m,
\end{equation}

with norm (except for $\gamma = 0$)

\begin{equation}
(2.4) \quad \|u\|_{1/\gamma} = \left( \sum_i u_i^{1/\gamma} \right)^\gamma = \left( \sum_i p_i \right)^\gamma / \gamma.
\end{equation}

2.3 Tangent spaces and bundles  We shall from now on fix the association

\begin{align*}
(2.5) \quad l_\gamma &: M_+ \to M_+^\gamma, \\
(2.6) \quad T_p l_\gamma &: T_p M_+ \to T_p M_+^\gamma = M^\gamma.
\end{align*}

Although for different $\gamma$ the space $M^\gamma \cong \mathbb{R}^m$ is the same linear space, their 
norms are different. Incidentally, the topologies induced by these norms are the same, even for infinite $X$, but that is beyond the scope of this note.

At each fixed point $p$, the tangent mappings $T_p l_\gamma$ are linear, so the pull-
back of the affine structures by $T_p l_\gamma$ from $T_p M^\gamma$ to $T_p M_+$ are the same 
for all $\gamma$. That is, all the $T_p M^\gamma$ are isomorphic to the same tangent space 
$T_p M_+$. However, because $l_\gamma$ itself is nonlinear, the isomorphism between 
$T_p M^\gamma$ and $T_p M_+$ is different at different $p$. This gives rise to different 
affine connections depending on $\gamma$ as shown below. For most $\gamma$ the whole 
tangent bundle $TM_+$ is not isomorphic to a single linear space.

2.4 Change of coordinates  A change of coordinate from $\xi = p^\gamma / \gamma$ to 
$u = p^\gamma / \tau$ gives

\begin{align*}
(2.7) \quad \frac{\partial \xi_i}{\partial u_j} &= \delta_{ij} p_i^{\gamma-\tau}, \\
(2.8) \quad \frac{\partial \xi_i}{\partial u_j \partial u_k} &= \delta_{ijk} p_i^{\gamma-2\tau} (\gamma - \tau).
\end{align*}

These relations will be used repeatedly in the sequel.
2.5 Metric The metric \( g \) on \( T\mathcal{M}_+ \) is induced by the inner product of \( T_p\mathcal{M}^{1/2}_+ = \mathcal{M}^{1/2} \), which is independent of \( p \). That is
\[
\forall u, v \in \mathcal{M}^{1/2} : \quad g(u, v) = \sum_i u_i v_i.
\]
If \( u = l_{1/2}(p), v = l_{1/2}(q) \), then
\[
g(u, v) = 4 \sum_i \sqrt{p_i q_i}.
\]
Let \( e_i \) denote the \( i \)th standard unit vector in \( \mathcal{M}^{1/2} \), then
\[
g_{ij} := g(e_i, e_j) = \delta_{ij}.
\]
Changing to \( \tau \)-coordinates transforms the tangent vectors
\[
u_i \in \mathcal{M}^{1/2} \rightarrow p_i^{\tau - 1/2} u_i \in \mathcal{M}^\tau,
\]
and the metric is represented as (Zhu and Rohwer, 1995, p. 30)
\[
g_{ij} = p_i^{1 - 2\tau} \delta_{ij}.
\]

2.6 Affine connection The \( \gamma \)-affine connection on \( T\mathcal{M}_+ \) is induced by the affine structure of \( T_p\mathcal{M}^\gamma_+ = \mathcal{M}^\gamma \). That is, for constant vectors,
\[
u, v \in \mathcal{M}^\gamma : \quad \gamma \overline{\nabla} u = \gamma v.
\]
\[
\overline{\Gamma}_{ijk} := g(e_i \overline{\nabla} e_j, e_k) = 0.
\]
Changing to \( \tau \)-coordinates, using both (2.8) and (2.13), gives (Zhu and Rohwer, 1995, p. 30)
\[
\overline{\Gamma}_{ijk} = (\gamma - \tau) p_i^{1 - 3\tau} \delta_{ijk}.
\]
Obviously, it vanishes exactly when \( \gamma = \tau \). That is, only in \( \mathcal{M}^\gamma \) the constant vector fields are \( \gamma \)-parallel transforms of themselves.

2.7 Dual affine The inner product of \( \gamma \) and \( 1 - \gamma \) transforms of two vectors remain constant on any curve. That is (Amari, 1985, p. 68)
\[
g(\gamma \overline{\nabla} v, w) + g(v, 1 - \gamma \nabla w) = (1 - 2\tau) \sum_i p_i^{1 - 3\tau} u_i v_i w_i = (\nabla g)(v, w)
\]
This means $\nabla$ and $\nabla^\gamma$ are dual-affine with respect to $g$ (Amari, 1985).

### 2.8 Potential and deviation

The $\gamma$-potential is defined as (Amari, 1985, p. 83)

$$
\psi_\gamma(p) := \sum_i p_i / (1 - \gamma), \quad \psi_1(p) := \sum_i p_i \log p_i.
$$

The $\gamma$-deviation is defined as (Amari, 1985, p. 85)

$$
D_\gamma(p, q) = \psi_\gamma(p) + \psi_1(q) - l_\gamma(p) \cdot l_1(q).
$$

This expands to (Zhu and Rohwer, 1995, p. 16)

$$
D_\gamma(p, q) = \sum_i \frac{\gamma p_i + (1 - \gamma) q_i - p_i^{\gamma} q_i^{1 - \gamma}}{\gamma (1 - \gamma)}
$$

This expression was also studied in (Vajda, 1989, p. 228) without using the dual affine geometry.

### 2.9 On $P$

Because $\sum p = \sum q = 1$, it simplifies to (Amari, 1985, p. 87)

$$
D_\gamma(p, q) = \frac{1 - \sum_i p_i^{\gamma} q_i^{1 - \gamma}}{\gamma (1 - \gamma)}.
$$

Familiar examples include the Hellinger distance, the Kullback-Leibler deviation and the $\chi^2$-deviation:

$$
D_{1/2}(p, q) = 2 \sum_i (\sqrt{p_i} - \sqrt{q_i})^2,
$$

$$
D_1(p, q) = \sum_i p_i \log \frac{p_i}{q_i},
$$

$$
D_2(p, q) = \sum_i \frac{(p_i - q_i)^2}{2q_i}.
$$

### 2.10 Differentiation

Let $u_i = \frac{p_i}{\tau}, v_i = \frac{q_i}{\tau}$, then

$$
\partial_{u_i} \partial_{v_j} D_\gamma(p, q) = -p_i^{\gamma - \tau} q_i^{1 - \gamma - \tau} \delta_{ij}.
$$

$$
\partial_{u_i} \partial_{u_j} D_\gamma(p, q) = -\gamma q_i^{1 - \gamma} \delta_{ij}.
$$

$$
\partial_{u_i} \partial_{v_j} \partial_{v_k} D_\gamma(p, q) = -\gamma - \tau p_i^{\gamma - 2\tau} q_i^{1 - \gamma} \delta_{ijk}.
$$
Obviously this is indeed consistent with (Eguchi, 1983),

\begin{align}
(2.27) \quad g_{ij} &= -\partial_u \partial_v D_\gamma(p, q) \bigg|_{\gamma = p}, \\
(2.28) \quad \Gamma_{ijk} &= -\partial_u \partial_v \partial_w D_\gamma(p, q) \bigg|_{\gamma = p}.
\end{align}

2.11 Conjugate If the two vectors are represented in mutually dual coordinates, then

\begin{align}
(2.29) \quad u \in \mathcal{M}^\tau, \quad v \in \mathcal{M}^{1-\tau} \implies g(u, v) = \int uv.
\end{align}

2.12 Dual coordinates We have (Zhu and Rohwer, 1995, p. 30)

\begin{align}
(2.30) \quad g^{ij} &= p^{2\tau - 1}\delta^{ij}, \\
(2.31) \quad \Gamma_{ij}^k &= (\gamma - \tau)p_i^{\tau}\delta_{ij}^k.
\end{align}

In particular, for \( \tau = 0 \), i.e. in exponential coordinates,

\begin{align}
(2.32) \quad \Gamma_{ij}^k &= \gamma\delta_{ij}^k.
\end{align}

This will be shown (§3) to be consistent with the equations (12.12) and (12.13) in Čencov, 1982, p. 175).

3. THE METRIC AND AFFINE CONNECTIONS ON \( \mathcal{P} \)

3.1 Equivalent projections The space of probability measures \( \mathcal{P} \) is a submanifold of \( \mathcal{M}_+ \) characterized by \( \sum p_i = 1 \). Therefore not all \( p_i \) are independent. There are at least three ways to “project” the metric and affine connections from \( \mathcal{M}_+ \) to \( \mathcal{P} \):

- Projecting the tangent vectors to \( T\mathcal{P} \) after an infinitesimal parallel transport (Amari, 1985, p. 38–40).
- Using the same formula with all the redundant components as if they are independent (Lauritzen, 1987, p. 187 (3.10)).
- Restricting \( D_\gamma \) to \( \mathcal{P} \), and using (2.27) and (2.28).

Their equivalence is due to the fact that vectors in the tangent space \( T\mathcal{P} \) are constrained by a differential relation implied by \( \sum p_i = 1 \), which in the \( \tau \)
coordinate is

\begin{equation}
\forall u \in T_p \mathcal{P}^\tau \subset \mathcal{M}^\tau : \quad \sum_i p_i^{1-\tau} u_i = 0.
\end{equation}

The expressions for $g$ and $\gamma$ in $T_p \mathcal{P}^\tau$ are therefore

\begin{align*}
(3.2) & \quad g_{ij} = p_i^{1-2\tau} \delta_{ij}, \quad \Gamma_{ijk}^\gamma = (\gamma - \tau) p_i^{1-3\tau} \delta_{ijk}, \\
(3.3) & \quad g^{ij} = p^{2\tau-1} \delta_{ij}, \quad \Gamma_{ij}^k = (\gamma - \tau) p^{\tau-1} \delta_{ij}^k.
\end{align*}

3.2 Čencov's affine connections Čencov (1982, p. 175) showed that the following family of affine connections is unique in the category of Markov morphism on $T_p \mathcal{P}$,

\begin{align*}
(3.4) & \quad X_i \nabla_j X_j = -\gamma(p_j X_i + p_i X_j), \quad i \neq j \\
(3.5) & \quad X_i \nabla_i X_i = \gamma(1 - 2p_i) X_i,
\end{align*}

where $X_i \in T_p \mathcal{P}^0$. This corresponds to

\begin{equation}
\Gamma_{ij}^k = \gamma(\delta_{ij}^k - p_i \delta_j^k - p_j \delta_i^k).
\end{equation}

Since $\forall u, v \in T_p \mathcal{P}^0$,

\begin{equation}
\sum_i p_i u_i = \sum_i p_i v_i = 0,
\end{equation}

this is also equivalent to

\begin{equation}
\sum_{ij} \Gamma_{ij}^k u_i v_j = \gamma u_k v_k, \quad \Gamma_{ij}^k = \gamma \delta_{ij}^k,
\end{equation}

which is identical to (2.32).

The difference between (2.32) and (3.6) for $u \nabla v$ only shows up when $u, v \in T_p \mathcal{M}^0 \setminus T_p \mathcal{P}^0$. In that case (3.6) gives the covariant differentiation of their projections into $T_p \mathcal{P}^0$ while (2.32) gives differentiations in $T_p \mathcal{M}^0$.

3.3 Special cases Here we correct a small but confusing error on (Čencov, 1982, p. 177), where he gave the mixture, metric, and exponential connections as corresponding to $\gamma = -1, 3/2, 0$, while in fact they should be
\( \gamma = 1, 1/2, 0. \) Here we follow the convention of regarding covariant vectors as differentials. Let

\[
(3.9) \quad V := \partial p, \quad X := fV, \quad f := p(1 - p).
\]

\[
(3.10) \quad u = 2\sqrt{p}, \quad U = \partial u = \sqrt{fV} = f^{-1/2}X.
\]

Then the mixture connection is given by \( VV = 0, \)

\[
(3.11) \quad X \cdot X = fV(fV) = fVFV = (1 - 2p)VX,
\]

corresponding to \( \gamma = 1. \) The metric connection is given by \( UU = 0, \)

\[
(3.12) \quad X \cdot X = \sqrt{fU}(\sqrt{fU}) = fV\sqrt{fU} = \frac{1}{2}(1 - 2p)VX,
\]

corresponding to \( \gamma = 1/2. \) The exponential connection is of course given by \( X_0X = 0, \) corresponding to \( \gamma = 0. \)

The proof for all \( \gamma \) goes almost identically. Let \( w := L_\gamma(p), \) \( W := \partial w = f^{-\gamma}X_0 = f^{1-\gamma}V. \) Then the \( \gamma \)-connection as defined here is given by \( WW = 0, \)

\[
(3.13) \quad X_0X = f'W(f'W) = f'Wf'W = \gamma X_0f f'W = \gamma (1 - 2p)VX,
\]

which shows the \( \gamma \) is the same parameter as used by Čencov.

4. ELIMINATING REDUNDANT COORDINATES

4.1 Constraint on tangent space The formulas in (Amari, 1985) eliminates redundant coordinate using \( \sum_i p_i = 1. \) Here we use a slightly different approach which is perhaps somewhat simpler. Denote \( \sum' := \sum_{i=1}^{n-1}, \)

and generally use a primed symbol to denote corresponding object on \( \mathcal{P} \)

with index running from 1 through \( n - 1. \) For \( u \in T \mathcal{P}, \)

\[
(4.1) \quad \sum_i p_i^{1-\gamma}u_i + p_n^{1-\gamma}u_n = 0.
\]

\[
(4.2) \quad u_n = -\sum_i \left( \frac{p_i}{p_n} \right)^{1-\gamma}u_i.
\]

This can be used to simply substitute out the last element in each sum to arrive at the desired formula, as the following paragraphs show.
4.2 Metric  We have

\begin{equation}
(4.3)  \quad g(u, v) = \sum_{ij} g_{ij}u_i v_j = \sum_{i} p_i^{1-2\tau} u_i v_i + p_n^{1-2\tau} u_n v_n = \sum_{i} g_\tau^{ij} u_i v_i
\end{equation}

\begin{equation}
(4.4)  \quad g_{ij} := p_i^{1-2\tau} \delta_{ij} + p_n^{-1}(p_i p_j)^{1-\tau}.
\end{equation}

In particular,

\begin{equation}
(4.5)  \quad \tau = 0 \implies g_\tau^{ij} = p_i \delta_{ij} + p_n^{-1} p_i p_j.
\end{equation}

\begin{equation}
(4.6)  \quad \tau = 1 \implies g_\tau^{ij} = p_i^{-1} \delta_{ij} + p_n^{-1},
\end{equation}

the \(\tau = 1\) formula being identical to that on (Amari, 1985, p. 31).

4.3 \(\gamma\)-affine connection  We have

\begin{equation}
(4.7)  \quad g(\nabla^{\gamma} v, w) = \sum_{ijk} \Gamma_{ijk}^\gamma u_i v_j w_k
\end{equation}

\begin{equation}
= (\gamma - \tau) \left( \sum_{i} p_i^{-1+3\tau} u_i v_i w_i + p_n^{-1+3\tau} u_n v_n w_n \right)
\end{equation}

\begin{equation}
= \sum_{ijk} \Gamma_{ijk}^\gamma u_i v_j w_k
\end{equation}

\begin{equation}
(4.8)  \quad \Gamma_{ijk}^\gamma := (\gamma - \tau) \left( p_i^{-1+3\tau} \delta_{ijk} - p_n^{-2}(p_i p_j p_k)^{1-\tau} \right)
\end{equation}

In particular,

\begin{equation}
(4.9)  \quad \tau = 0 \implies \Gamma_{ijk}^\gamma = \gamma \left( p_i \delta_{ijk} - p_n^{-2} p_i p_j p_k \right).
\end{equation}

\begin{equation}
(4.10)  \quad \tau = 1 \implies \Gamma_{ijk}^\gamma = (\gamma - 1) \left( p_i^{-2} \delta_{ijk} - p_n^{-2} \right),
\end{equation}

the \(\tau = 1\) formula being identical to that on (Amari, 1985, p. 43).

4.4 Dual coordinate  The corresponding formulas for the metric are

\begin{equation}
(4.11)  \quad g^{ij} = p_i^{2\tau-1} \delta^{ij} - (p_i p_j)^\tau.
\end{equation}

\begin{equation}
(4.12)  \quad \tau = 0 \implies g^{ij} = p_i^{-1} \delta^{ij} - 1.
\end{equation}

\begin{equation}
(4.13)  \quad \tau = 1 \implies g^{ij} = p_i \delta^{ij} - p_i p_j.
\end{equation}
and those for the affine connection are

\[ \Gamma_{ij}^k = (\gamma - \tau) \left( p_i^{-\gamma} \delta_{ij}^k - \delta_{ij} p_i^{1-2\tau} p_k^{1-\tau} - p_n^{-1} (p_i p_j)^{1-\tau} p_k^{1-\tau} \right). \]  

(4.14)

\[ \tau = 0 \implies \Gamma_{ij}^k = \gamma \left( \delta_{ij}^k - \delta_{ij} p_i - p_n^{-1} p_i p_j \right). \]  

(4.15)

\[ \tau = 1 \implies \Gamma_{ij}^k = (\gamma - 1) \left( p_i^{-1} \delta_{ij}^k - \delta_{ij} p_i^{-1} p_k - p_n^{-1} p_k \right). \]  

(4.16)

5. DISCUSSION OF RELATED ISSUES

5.1 It is obvious that the generalization of these equations to arbitrary sample space \( X \) requires calculus on the fractional powers of measures. It can be shown (Zhu, 1998), using ideas in (Neveu, 1965) that the linear spaces \( \mathcal{M}^\gamma(X) \) spanned by these elements are Banach spaces with the desired duality, if we restrict attention to \( \gamma \in (0, 1) \). These spaces generalize the corresponding classical Lebesgue function spaces and act as enveloping tangent spaces. The cases of \( \{0, 1\} \) are very important but need special treatment because of nonequivalence of topologies. In this paper the issue of topology does not arise because \( X \) is assumed to be finite.

5.2 The formulas for the metric and affine connections on arbitrary \( X \) look very similar to those given here. In nonparametric fashion, we have, for \( p \in \mathcal{M}_+, u, v, w \in T_p \mathcal{M}_+ = \mathcal{M}^\gamma,

\[ g(u, v) = \int p^{1-2\gamma} uv, \]  

(5.1)

\[ g(u \nabla v, w) = (\gamma - \tau) \int p^{1-3\gamma} uvw. \]  

(5.2)

\[ u \nabla v = (\gamma - \tau) p^{1-3\gamma} uv \in \mathcal{M}^{1-\gamma}. \]  

(5.3)

These are calculations in the generalized Lebesgue spaces.

5.3 Campbell (1986) showed that the extension of \( g \) from \( \mathcal{P}(X) \) to \( \mathcal{M}_+(X) \) is not unique under classical Markov morphisms. It is likely that the same is true for \( \nabla^\gamma \). The metric and affine connections defined here may be called the canonical versions. We conjecture that if the Markov morphisms are relaxed to allow unnormalized mappings, the canonical versions are again unique equivariant extensions. On the other hand, it is shown elsewhere that
the extensions of the canonical versions from $X = \mathbb{N}_m$ to arbitrary $X$ are unique.

5.4 Amari (1985) uses functions of random variables and their differentials where we use functions of measures and their differentials as models of tangent vectors. It is trivial to verify that they are equivalent, as long as all variables with the same distributions are identified.

REFERENCES


SANTA FE INSTITUTE, 1399 HYDE PARK ROAD, SANTA FE, NM 87501
E-mail address: zuhu@santeafe.edu