Draft. `~/research/ig/affine/affine.tex`

Dual Affine Geometry of Finite Measures

Huaiyu Zhu

Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

Email: zhuh@santafe.edu

July 31, 1998

1 Introduction

1.1 This is an attempt to generalize the dual affine geometry to the space of finite measures. It is nonparametric, without assuming a dominating measure. We first describe things in the space of finite measures. Then project them onto space of probability measures. We follow [1, 2, 5].

1.2 In this draft we list what the final representations will look like, but many defintiions can in fact be turned into theorem. Here we rely on concrete representation of tangent, but later we need to expand on the differential definition of tangents. We omit the mixture and exponential geometry first.

1.3 For consistency, we need to check that

- The topologies are equivalent.
- The inner product does not depend on the parameter
- These concepts specialise to the Hilbert bundle of [2].
- The projection to $\mathcal{P}$ coincide with the finite dimensional case.

2 Preliminaries for Finite Measures

2.1 Let Z be a set, $\mathcal{F}$ be a σ-algebra of subsets. $(Z, \mathcal{F})$ is called a measurable space.

2.2 Measures A σ-additive function $\mathcal{F} \to R$, where $R = [0, \infty]$, is called a measure. The signed measures are defined with $R = (-\infty, \infty]$. The finite measures are defined with $R = [0, \infty)$. The signed finite measures are defined with $R = (-\infty, \infty)$. Denote by $\overline{\mathcal{M}}$, $\overline{\mathcal{M}}_+$, $\mathcal{M}$, $\mathcal{M}_+$ and $\mathcal{P}$ the spaces of signed measures, measures, signed finite measures, finite measures, and probability measures.

2.3 Information deviation For $\gamma \in (0, 1)$, the γ-deviation is defined as

$$D_\gamma(p, q) := \begin{cases} \int \dfrac{\gamma p + (1-\gamma)q - p^\gamma q^{1-\gamma}}{\gamma(1-\gamma)} & \gamma \in (0,1), \\ \lim\limits_{\gamma \to 0} D_\gamma(p,q) = \int p - q + q \log(q/p), & \gamma = 0, \\ \lim\limits_{\gamma \to 1} D_\gamma(p,q) = \int q - p + p \log(p/q), & \gamma = 1. \end{cases} \tag{2.1}$$

2.4 Generalized Lebesgue spaces Let $r \in \mathcal{M}_+$. Denote the spaces of finite measures dominated by r,

$$\mathcal{M}_+(r) := \{p : \, p \in \mathcal{M}_+, \, p \prec r\}, \tag{2.2}$$

$$\mathcal{M}(r) := \{p : \, p \in \mathcal{M}, \, p \prec r\}. \tag{2.3}$$

Let $\gamma \in (0, 1)$. Define

$$\mathcal{M}_+^\gamma(r) := \{p^\gamma : \, p \in \mathcal{M}_+(r)\}, \tag{2.4}$$

$$\mathcal{M}^\gamma(r) := \{p_+^\gamma - p_-^\gamma : \, p_i \in \mathcal{M}(r)\}, \tag{2.5}$$

$$\mathcal{M}_+^\gamma := \{p^\gamma : \, p \in \mathcal{M}_+\}, \tag{2.6}$$

$$\mathcal{M}^\gamma := \{p_+^\gamma - p_-^\gamma : \, p_i \in \mathcal{M}\}. \tag{2.7}$$

These spaces have the usual linear structure and norms. The space $\mathcal{M}^\gamma(r)$ was denoted as $L_{1/\gamma, r}$ in [6] and is isomorphic to the usual Lebesgue function space $L_{1/\gamma}(r)$. The space $\mathcal{M}^\gamma$ was denoted $L_{1/\gamma}$ and is the generalized Lebesgue space. The spaces with subscript $+$ consists of non-negative elements. The spaces $\mathcal{M}^\gamma(r)$ and $\mathcal{M}^\gamma$ are Banach spaces, while $\mathcal{M}^{1/2}(r)$ and $\mathcal{M}^{1/2}$ are Hilbert spaces.

3 Tangent

3.1 Notations We generally assume $p, q, r \in \mathcal{M}_+$, $\gamma \in (0,1)$.

3.2 Coordinates Let

$$l_\gamma : \quad \mathcal{M}_+ \to \mathcal{M}_+^\gamma, \qquad l_\gamma(p) := \frac{p^\gamma}{\gamma}. \tag{3.1}$$

$$T_p l_\gamma : \quad \mathcal{M} \to \mathcal{M}^\gamma, \qquad T_p l_\gamma u = p^{\gamma-1} u. \tag{3.2}$$

$$\frac{T_p l_\gamma}{T_p l_\tau} = p^{\gamma-\tau}. \tag{3.3}$$

3.3 Conjugate We often denote

$$\xi := l_\gamma(p) \in \mathcal{M}_+^\gamma, \qquad \eta := l_{1-\gamma}(p) \in \mathcal{M}_+^{1-\gamma}. \tag{3.4}$$

They are dual coordinates in the Banach space sense. They are related by

$$(\gamma\xi)^{1/\gamma} = ((1-\gamma)\eta)^{1/(1-\gamma)}. \tag{3.5}$$

implies that

$$\partial_\xi \eta = p^{1-2\gamma}, \qquad \partial_\eta \xi = p^{2\gamma-1}. \tag{3.6}$$

3.4 Manifold The topology of $\mathcal{M}_+$ induced from $\mathcal{M}^\gamma$ by l_γ does not depend on γ. Therefore $\mathcal{M}_+$ is the same Banach manifold modeled on $\mathcal{M}^\gamma$ for each γ. In particular, it is a Hilbert manifold modeled on $\mathcal{M}^{1/2}$. This topology is called the canonical topology. For $\gamma \in \{0,1\}$ the topology will be different, but we don't consider them here.

3.5 Tangent The tangent space $T_p\mathcal{M}_+$ is well defined, and does not depend on γ. (Details) But since we have the tangent space $T_p\mathcal{M}_+^\gamma = \mathcal{M}^\gamma$, we can also have different representations of $T_p\mathcal{M}_+$ by these mappings. That is

$$T_p l_\gamma : \quad T_p\mathcal{M}_+ \to T_p\mathcal{M}_+^\gamma = \mathcal{M}^\gamma. \tag{3.7}$$

3.6 In particular,

$$T_p l_\gamma : \quad T_p\mathcal{P} \to T_p\mathcal{P}^\gamma = \mathcal{M}^\gamma(\textstyle\int p^{1-\gamma} u = 0). \tag{3.8}$$

3.7 The norms of $\mathcal{M}^\gamma$ introduces the Matusita distance on $\mathcal{M}_+$.

4 Metric and Affine connection

4.1 Metric Let $u, v \in \mathcal{M}^{1/2}$. The metric is defined by the natural inner product

$$g(u, v) = \int uv. \tag{4.1}$$

This induces the metric on $\mathcal{M}_+$, which is the Fisher information metric.

4.2 Affine connection Let $u, v \in \mathcal{M}^{1/2}$. The γ-affine connection is defined by the natural affine connection

$$u \overset{\gamma}{\nabla} v = 0. \tag{4.2}$$

This induces the γ-affine connection on $\mathcal{M}_+$, which coincide with Amari's $\alpha = (1 - \gamma)/2$ affine connection.

4.3 Change of coordinates If we change to $u, v, w \in \mathcal{M}^\tau$, then

$$g(u, v) = \int p^{1-2\tau} uv. \tag{4.3}$$

$$u \overset{\gamma}{\nabla} v = (\gamma - \tau) p^{1-3\tau} uv. \tag{4.4}$$

$$g(u \overset{\gamma}{\nabla} v, w) = (\gamma - \tau) \int p^{1-3\tau} uvw. \tag{4.5}$$

4.4 Hilbert bundle This appears to generalize Amari's Hilbert bundle [2, p67] from $\gamma = 0$ to all γ, but in fact, for $\gamma = 0$ this only works for exponential families because of the different topology. Note that although the tangent space at each point p is isomorphic to a Hilbert space, it is a different Hilbert space for each p. So the whole thing is a bundle instead of a single linear space. These spaces coincide with each other only when $\gamma = 1/2$.

4.5 Duality Let $u \in \mathcal{M}^\gamma$, $v \in \mathcal{M}^{1-\gamma}$. Then the duality coupling

$$\int uv \tag{4.6}$$

is the same as the inner product between vectors

$$u \in T_p\mathcal{M}_+^\gamma, \qquad v \in T_p\mathcal{M}_+^{1-\gamma}. \tag{4.7}$$

The anchor point $p \in \mathcal{M}_+$ is irrelevant. This is also called tensor contraction.

5 Dual affine

5.1 Potentials There are potentials

$$\psi_\gamma := \frac{\int p}{1-\gamma} = \gamma \int \xi\eta, \qquad \psi_{1-\gamma} := \frac{\int p}{\gamma} = (1-\gamma)\int \xi\eta. \tag{5.1}$$

$$\partial_\xi \psi_\gamma = \eta, \qquad \partial_\eta \psi_{1-\gamma} = \xi. \tag{5.2}$$

When differentiating, note that ξ and η are not independent.

5.2 The metric tensor in the dual coordinates is trivial.

$$g_{\xi\eta} = \partial_\xi \partial_\eta \psi_{1-\gamma} = I, \qquad g_{\eta\xi} = \partial_\eta \partial_\xi \psi_\gamma = I \tag{5.3}$$

They look the same, but because the identifies are in different spaces, they *are* different by the factor $\psi_{1-\gamma}/\psi_\gamma$. A better notation is to write η on superscripts.

Its representation in one particular space

$$g_{\xi\xi} = \partial_\xi \partial_\xi \psi_\gamma = \partial_\xi \eta = p^{1-2\gamma}, \qquad g_{\eta\eta} = \partial_\eta \partial_\eta \psi_\gamma = \partial_\eta \xi = p^{2\gamma-1}. \tag{5.4}$$

5.3 Deviation Amari'd definition through Legendre transform

$$D_\gamma(p,q) = \psi_\gamma(p) + \psi_{1-\gamma}(q) - \int \xi(p)\eta(q) \tag{5.5}$$

$$= \int \frac{\gamma p + (1-\gamma)q - p^\gamma q^{1-\gamma}}{\gamma(1-\gamma)}. \tag{5.6}$$

5.4 Equivalence Given two curves

$$q_s, \qquad r_t, \tag{5.7}$$

$$q_0 = r_0 = p. \tag{5.8}$$

Using the cosine theorem [7], the tangent vectors are

$$u := \partial_s l_\gamma(q_s)|_{s=0} \in \mathcal{M}^\gamma, \tag{5.9}$$

$$v := \partial_t l_{1-\gamma}(r_t)|_{t=0} \in \mathcal{M}^{1-\gamma}. \tag{5.10}$$

The inner product is

$$\int uv \tag{5.11}$$

$$= \lim_{s,t\to 0} \frac{1}{st} \int \frac{(q_s^\gamma - p^\gamma)(r_t^{1-\gamma} - p^{1-\gamma})}{\gamma(1-\gamma)} \tag{5.12}$$

$$= \lim_{s,t\to 0} \frac{1}{st} (D_\gamma(q_s, p) + D_\gamma(p, r_t) - D_\gamma(q_s, r_t)) \tag{5.13}$$

$$= \lim_{s,t\to 0} \frac{1}{st} \left(D_{1/2}(q_s, p) + D_{1/2}(p, r_t) - D_{1/2}(q_s, r_t)\right) \tag{5.14}$$

$$= \lim_{s,t\to 0} \frac{4}{st} \int (\sqrt{q_s} - \sqrt{p})(\sqrt{r_t} - \sqrt{p}). \tag{5.15}$$

This is the same inner product as on $\mathcal{M}^{1/2}$. Therefore it is independent of γ.

5.5 Dual affine Given two vectors

$$u \in \mathcal{M}^\gamma, \qquad v \in \mathcal{M}^{1-\gamma} \tag{5.16}$$

at one point of a curve p_s. Their parallel transport to any other point remain the same. Therefore the inner product

$$\int uv \tag{5.17}$$

is invariant along the curve (in fact, any curve). This shows that these are dual affine coordinates (connections, geometries).

5.6 Converse The metric and affine connections can be recovered from the deviation by differentiation, following [4].

6 Projection to $\mathcal{P}$

6.1 The following two approaches are equivalent [5]:

Restrict D_γ to $\mathcal{P}$, and use the differentiation representation of [4].

Project in the coordinate form following [1].

6.2 The formulas are exactly the same, for $u, v, w \in T_p\mathcal{P}^\tau$,

$$g(u, v) = \textstyle\int p^{1-2\tau} uv. \tag{6.1}$$

$$u\overset{\gamma}{\nabla}v = (\gamma - \tau)p^{1-3\tau} uv. \tag{6.2}$$

$$g(u\overset{\gamma}{\nabla}v, w) = (\gamma - \tau)\int p^{1-3\tau} uvw. \tag{6.3}$$

The only difference is that u, v, w cannot move freely in $\mathcal{M}^\tau$, but this is already taken care of implicitly.

7 Discussions

7.1 It seems that all concepts can be represented neatly. The only thing lacking is a decent notation for differentials, and for index notations. Maybe we'll never need them?

References

[1] S. Amari. *Differential-Geometrical Methods in Statistics*, volume 28 of *Springer Lecture Notes in Statistics*. Springer-Verlag, New York, 1985.

[2] S. Amari. Differential geometrical theory of statistics. In Amari et al. [3], chapter 2, pages 19–94.

[3] S. Amari, O. E. Barndorff-Nielsen, R. E. Kass, S. L. Lauritzen, and C. R Rao, editors. *Differential Geometry in Statistical Inference*, volume 10 of *IMS Lecture Notes Monograph*. Inst. Math. Statist., Hayward, CA, 1987.

[4] S. Eguchi. Second order efficiency of minimum contrast estimators in a curved exponential family. *Ann. Statist.*, 11:793–803, 1983.

[5] S. L. Lauritzen. Statistical manifolds. In Amari et al. [3], chapter 4, pages 163–216.

[6] H. Zhu. Generalized Lebesgue spaces and applications to statistics. Technical Report SFI 98-06-44, Santa Fe Institute, 1998. submitted to J. Functional Analysis.

[7] H. Zhu and R. Rohwer. Information geometric measurements of generalisation. Technical Report NCRG/4350, Aston University, 1995. `ftp://cs.aston.ac.uk/neural/zhuh/generalisation.ps.Z`.