情報幾何班 1-2

大上由人

2024 年 9 月 27 日

1　双対アファイン接続の幾何

狭い意味での情報幾何学として、双対アファイン接続の幾何を考える。

1.1　双対アファイン接続

Def: 双対アファイン接続

アファイン接続を持つ Riemann 多様体 (M, g) に対して、双対アファイン接続 ∇^* を、

$$Xg(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla^*_X Z) \quad (X, Y, Z \in \mathfrak{X}(M)) \tag{1.1}$$

により定める。

例えば、Riemann 接続の双対アファイン接続は、Riemann 接続が計量的であることから、

$$Xg(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z) \quad (X, Y, Z \in \mathfrak{X}(M)) \tag{1.2}$$

となり、$\nabla^* = \nabla$ である。(自己双対)
このとき、双対アファイン接続は、一意に定まることを示すことができる。また、共変微分の公理を満たすことも示される。

Def: 双対構造

Riemann 多様体 (M, g) に対して、計量 g に対する双対性を満たすアファイン接続のペア (∇, ∇^*) が与えられたとき、(g, ∇, ∇^*) を M の双対構造という。

計量的であるような接続以外にも、双対構造をもつ接続は存在する。(ここでは例を挙げないが、後に統計多様体の例が挙げられる) このとき、平行移動に対して内積が保存されることはないが、双対接続の定義が計量的であることに似ている。したがって、内積の保存性を以下のように拡張することができる。

Thm: 双対平行移動に対する内積の不変性

なめらかな曲線 $C = \{p(t); t \in [a,b]\}$ に沿った ∇ および ∇^* の平行移動をそれぞれ Π_C, Π_C^* とする。このとき、任意の $\mathbf{v}, \mathbf{w} \in T_{p(a)}M$ に対して、

$$g_{p(b)}(\Pi_C\mathbf{v}, \Pi_C^*\mathbf{w}) = g_{p(a)}(\mathbf{v}, \mathbf{w}) \tag{1.3}$$

が成り立つ。すなわち、二つの接ベクトルのうち、片方を ∇ で平行移動し、もう片方を ∇^* で平行移動したとき、内積は不変である。

Prf

双対アフィン接続の定義式に対して、ベクトル場 X を $\frac{\mathrm{d}}{\mathrm{d}t}$ として代入すると、

$$\frac{\mathrm{d}}{\mathrm{d}t}g(Y,Z) = g(\nabla_{\frac{\mathrm{d}}{\mathrm{d}t}}Y, Z) + g(Y, \nabla^*_{\frac{\mathrm{d}}{\mathrm{d}t}}Z) = 0 \tag{1.4}$$

を得る。ただし、平衡の条件 $\nabla_{\frac{\mathrm{d}}{\mathrm{d}t}}Y = 0, \nabla^*_{\frac{\mathrm{d}}{\mathrm{d}t}}Z = 0$ を用いた。よって、$g(Y,Z)$ は C に沿って平行移動しても不変である。 □

Thm: 曲率

∇-曲率が 0 であることと、∇^*-曲率が 0 であることは同値である。すなわち

$$R = 0 \Leftrightarrow R^* = 0 \tag{1.5}$$

が成り立つ。

Prf.(厳密でない版)

$(\Rightarrow)$ を示す。C の始点と終点が一致していて閉曲線である時を考える。このとき、$R=0$ であることは、

$$\Pi_C\mathbf{v} = \mathbf{v} \tag{1.6}$$

を意味する。このとき、内積の保存性から、

$$g_{p(a)}(\mathbf{v}, \mathbf{w}) = g_{p(a)}(\Pi_C\mathbf{v}, \Pi_C^*\mathbf{w}) = g_{p(a)}(\mathbf{v}, \Pi_C^*\mathbf{w}) \tag{1.7}$$

となる。これが任意の $\mathbf{v}$ について成り立つことから、

$$\Pi_C^*\mathbf{w} = \mathbf{w} \tag{1.8}$$

が成り立つ。これが任意の $\mathbf{w}$ について成り立つことから、∇^*-曲率が 0 であることが示される。逆も同様。 □

捩率については、曲率のような関係が成り立たないことが知られている。

1.2 双対平坦な多様体の幾何

上で定義した双対接続を用いて、双対平坦な多様体の幾何を考える。特に、統計多様体においては、片方の接続が指数型分布族、もう片方の接続が混合型分布族に対応する。

Def: 双対平坦な多様体

双対構造 (g, ∇, ∇^*) を持つ多様体 (M, g) が双対平坦であるとは、∇-曲率と ∇^*-曲率がどちらも 0 かつ、∇-捩率と ∇^*-捩率がどちらも 0 であることをいう。

今回は二つの接続について考えているため、それぞれの接続について局所アフィン座標系をとることができる。さらに、それぞれのアフィン座標系が、アフィン変換に対する任意性を持つことを用いると、以下の定理を示すことができる。

Thm: 局所アフィン座標系の存在

双対構造 (g, ∇, ∇^*) に関して平坦な多様体 M では、各点の周りで

$$g(\partial_i, \partial_j) = \delta_{ij} \tag{1.9}$$

を満たす、局所 ∇-アフィン座標系 (x^i) および局所 ∇^*-アフィン座標系 (y^i) の組 $\{x^i, y^i\}$ をとることができる。このような組 $\{x^i, y^i\}$ を双対アフィン座標系という。

Prf

まず、$p_0 \in M$ を任意にとり、∇ に関するアフィン座標近傍 $(U; \xi^i)$ と、∇^* に関するアフィン座標近傍 $(V; \eta^i)$ を、互いに無関係に取る。そして、

$$(G_0)_{ij} = g_{p_0}\left(\left(\frac{\partial}{\partial \xi^i}\right)_{p_0}, \left(\frac{\partial}{\partial \eta^j}\right)_{p_0}\right) \tag{1.10}$$

とおく。このとき、g_{p_0} は正定値であるから、$\det(G_0) > 0$ である。したがって、G_0 は正則である。そして、

$$x^i = \xi^i, \quad y^i = \sum_j (G_0)_{ij} \eta^j \tag{1.11}$$

とおくと、これが求めるものとなる。実際、前の章の定理より、新たに作った座標系はアファイン座標系である。また、

$$\frac{\partial}{\partial \xi^i} = \frac{\partial}{\partial x^i} \tag{1.12}$$

$$\frac{\partial}{\partial \eta^i} = \sum_j (G_0)_{kj} \frac{\partial}{\partial y^k} \tag{1.13}$$

であるから、

$$(G_0)_{ij} = g_{p_0}\left(\left(\frac{\partial}{\partial \xi^i}\right)_{p_0}, \left(\frac{\partial}{\partial \eta^j}\right)_{p_0}\right) \tag{1.14}$$

$$= \sum_k g_{p_0}\left(\left(\frac{\partial}{\partial x^i}\right)_{p_0}, \left(\frac{\partial}{\partial y^k}\right)_{p_0}\right)(G_0)_{kj} \tag{1.15}$$

$$\tag{1.16}$$

であるから、

$$g_{p_0}\left(\left(\frac{\partial}{\partial x^i}\right)_{p_0}, \left(\frac{\partial}{\partial y^k}\right)_{p_0}\right) = \delta_{ik} \tag{1.17}$$

が成り立つ。また、任意の $X \in \mathfrak{X}(U \cap V)$ に対して、

$$Xg\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) = g\left(\nabla_X \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) + g\left(\frac{\partial}{\partial x^i}, \nabla^*_X \frac{\partial}{\partial x^j}\right) = 0 \tag{1.18}$$

$$\left(\because \text{アフィン座標系の性質から } \nabla_X \frac{\partial}{\partial x^i} = \nabla^*_X \frac{\partial}{\partial x^j} = 0\right)$$

となることから、任意の $X \in \mathfrak{X}(U \cap V)$ に対して、直交性が成り立つ。したがって、求める座標系は存在する。 □

以下では、双対 affine 座標系を用いた局所的な話に限定するため、$U \cap V$ 自身を多様体 M とみなし、直交性をみたす大域的な ∇-affine 座標系を (θ^i)、 ∇^*-affine 座標系を (η_j) で表し、それぞれ θ-座標系、 η-座標系とよぶことにする。また、対応するベクトル場を

$$\partial_i := \frac{\partial}{\partial \theta^i}, \quad \partial^j := \frac{\partial}{\partial \eta_j} \tag{4.40}$$

と書くことにする。また、直交性は

$$g(\partial_i, \partial^j) = \delta_i^j \tag{1.19}$$

と表すことにする。

以下、4 つの補題を用いて、ダイバージェンスを定義する。

Lem.1

双対アフィン座標系 $\{(\theta)^i, (\eta)_i\}$ に関する計量 g の成分を、

$$g_{ij} = g(\partial_i, \partial_j) \quad g^{ij} = g(\partial^i, \partial^j) \tag{1.20}$$

とおくと、

$$g_{ij} = \partial_i \eta_j = \partial_j \eta_i \quad g^{ij} = \partial^i \theta^j = \partial^j \theta^i \quad g_{ij} g^{jk} = \delta_i^k \tag{1.21}$$

が成り立つ。

Prf

座標変換則 $\partial_i = \frac{\partial \eta_k}{\partial \theta^i} \partial^k$ を用いると、

$$\begin{aligned}
g_{ij} &= g(\partial_i, \partial_j) && (1.22)\\
&= g\left(\frac{\partial \eta_k}{\partial \theta^i} \partial^k, \partial_j\right) && (1.23)\\
&= \frac{\partial \eta_k}{\partial \theta^i} g(\partial^k, \partial_j) && (1.24)\\
&= \frac{\partial \eta_k}{\partial \theta^i} \delta_j^k && (1.25)\\
&= \frac{\partial \eta_j}{\partial \theta^i} && (1.26)
\end{aligned}$$

となるから、$g_{ij} = \partial_i \eta_j$ が成り立つ。計量の添え字に対する対称性から、$g_{ij} = \partial_j \eta_i$ も成り立つ。また、g^{ij} についても同様にして、

$$g^{ij} = \frac{\partial \theta^i}{\partial \eta^j} \tag{1.27}$$

が成立する。これらを合わせて、

$$g_{ij} g^{jk} = \frac{\partial \eta_i}{\partial \theta^j} \frac{\partial \theta^j}{\partial \eta^k} = \delta_i^k \tag{1.28}$$

もいうことができる。 □

Lem.2

ある C^∞ 関数の組 $\{\psi(\theta^1, \cdots, \theta^n), \varphi(\eta_1, \cdots, \eta_n)\}$ が存在して、

$$\eta_i = \partial_i \psi \quad \theta^i = \partial^i \varphi \quad \psi(\theta^1, \cdots, \theta^n) + \varphi(\eta_1, \cdots, \eta_n) - \theta^i \eta_i = 0 \tag{1.29}$$

が成り立つ。

Prf

一つ前の補題より、$\partial_i \eta_j = \partial_j \eta_i$ であるから、これは、$\eta_i = \partial_i \psi$ と書ける。同様に、$\theta^i = \partial^i \varphi$ と書ける。また、

$$\begin{aligned} \mathrm{d}\big(\psi + \varphi - \theta^i \eta_i\big) &= \mathrm{d}\psi + \mathrm{d}\varphi - (\mathrm{d}\theta^i)\eta_i - \theta^i(\mathrm{d}\eta_i) && (1.30) \\ &= (\partial_i \psi)\mathrm{d}\theta^i + (\partial^i \varphi)\mathrm{d}\eta_i - \eta_i \mathrm{d}\theta^i - \theta^i \mathrm{d}\eta_i && (1.31) \\ &= 0 && (1.32) \end{aligned}$$

となるから、$\psi + \varphi - \theta^i \eta_i$ は定数関数となり、特に積分定数をうまく選ぶと、その値は 0 となる。
□

Lem.3

一つ前の補題で定めた関数の組は、計量 g と、

$$g_{ij} = \partial_i \partial_j \psi \quad g^{ij} = \partial^i \partial^j \varphi \tag{1.33}$$

により関連付けられる。とくに、ψ, φ は狭義凸関数である。

Prf

上の二つの補題と、g_{ij} の正定値性を用いると示される。 □

Lem.4

点 $p \in M$ の θ-座標と、η-座標をそれぞれ、

$$\theta(p) = (\theta^1(p), \cdots, \theta^n(p)) \quad \eta(p) = (\eta_1(p), \cdots, \eta_n(p)) \tag{1.34}$$

と表すことにすると、二つ上で定めた関数の組は、互いに Legendre 変換の関係にある。すなわち、

$$\varphi(\eta(p)) = \max_{q \in M} \left\{ \theta^i(q)\eta_i(p) - \psi(\theta(q)) \right\} \tag{1.35}$$

$$\psi(\theta(p)) = \max_{q \in M} \left\{ \eta_i(q)\theta^i(p) - \varphi(\eta(q)) \right\} \tag{1.36}$$

が成り立つ。

Prf

点 p を固定し、関数 $q \mapsto \theta^i(q)\eta_i(p) - \psi(\theta(q))$ を微分してみると、

$$\begin{aligned} d\left(\theta^i(q)\eta_i(p) - \psi(\theta(q))\right) &= (\eta_i(p) - \partial_i\psi(\theta(q)))d\theta^i(q) \\ &= (\eta_i(p) - \eta_i(q))d\theta^i(q) \end{aligned} \tag{1.37}$$

だから上の式の右辺の max は、すべての i で $\eta_i(p) = \eta_i(q)$、すなわち $p = q$ のときかつそのときに限り達成されて、その最大値は

$$\theta^i(p)\eta_i(p) - \psi(\theta(p)) = \varphi(\eta(p)) \tag{1.38}$$

となる。ここで上の補題の最後の等式を用いた。これで上の式が証明された。下の式の証明も全く同様である。 □

以上の準備のもと、ダイバージェンスを定義する。

Def: ダイバージェンス

M を、双対構造 (g, ∇, ∇^*) に関する双対平坦多様体であるとする。このとき、二点 $p, q \in M$ に対して、定まる量

$$D(p||q) = \psi(\theta(p)) + \varphi(\eta(q)) - \theta^i(p)\eta_i(q) \tag{1.39}$$

を ∇-ダイバージェンスという。ただし、$\{(\theta^i), (\eta_i)\}$ は M の大域的な双対アフィン座標系である。

定義にアフィン座標系を用いているが、結局、座標の取り方に依らないことが示される。(ここでは省略)

∇ と ∇^* の役割を入れ替えると、$D(p||q)$ における θ と η、ψ と φ の役割も入れ替わる。したがって、

$$D(p||q) = D^*(q||p) \tag{1.40}$$

が成り立つ。また、Lem3 の二本目の式と比較することで、

$$D(p||q) = \psi(\theta(p)) + \varphi(\eta(q)) - \theta^i(p)\eta_i(q) \tag{1.41}$$

$$= \max_{r \in M} \left\{ \eta_i(r)\theta^i(p) - \varphi(\eta(r)) \right\} + \varphi(\eta(q)) - \theta^i(p)\eta_i(q) \tag{1.42}$$

$$\geq 0 \tag{1.43}$$

かつ、

$$D(p||q) = 0 \Leftrightarrow p = q \quad \because \text{Lem2} \tag{1.44}$$

が成り立つ。したがって、$D(p||q)$ は、M 上の距離のような役割を持つと考えられるが、実際には、対称性 $D(p||q) = D(q||p)$ は成り立たないし、三角不等式も成り立たない。したがって、$D(p||q)$ は、距離の公理を満たさない。

ex:Euclid 空間

Euclid 空間においては、$\nabla = \nabla^*$ である。したがって、双対平坦性はただの平坦性と帰着する。このとき、ポテンシャルは、

$$\psi(z) = \varphi(z) = \frac{1}{2}\sum_{i=1}^{n}(z^i)^2 \tag{1.45}$$

となる。したがって、ダイバージェンスは、

$$D(p||q) = \frac{1}{2}\sum_{i=1}^{n}(z^i(p))^2 + \frac{1}{2}\sum_{i=1}^{n}(z^i(q))^2 - \sum_{i=1}^{n} z^i(p)z^i(q) = \frac{1}{2}\sum_{i=1}^{n}(z^i(p) - z^i(q))^2 \tag{1.46}$$

となる。これは、Euclid 空間における距離の二乗に一致する。

ところで、Euclid 空間において、距離の二乗と結びつく定理として、Pythagoras の定理がある。これは、双対平坦な多様体へ拡張することができる。

Thm: 拡張 Pythagoras の定理

双対平坦多様体 (M, g, ∇, ∇^*) において、点 p, q を結ぶ ∇-測地線と q, r を結ぶ ∇^*-測地線が計量 g に関して直交しているなら、

$$D(p||r) = D(p||q) + D(q||r) \tag{1.47}$$

が成り立つ。

Prf

測地線は一次元自己平行部分多様体であったから、∇-測地線は、θ 座標系を用いて表現することができ、また、θ 座標系はアフィン座標系であるから接続係数は 0 である。したがって、測地線の方程式を思い出すと、θ 座標系において、測地線は直線で表すことができる。また、∇^*-測地線も同様に直線で表すことができる。

このことから、q, p を結ぶ ∇-測地線

$$C_1 = \{p(t); t \in [0, 1] \quad p(0) = q, p(1) = p\} \tag{1.48}$$

を $p(t) = (\theta^1(t), \cdots, \theta^n(t))$ と表すと、

$$\theta^i(t) = \theta^i(0) + t(\theta^i(1) - \theta^i(0)) \tag{1.49}$$

となる。同様に、q, r を結ぶ ∇^*-測地線

$$C_2 = \{q(t); t \in [0,1] \quad q(0) = q, q(1) = r\} \tag{1.50}$$

を $q(t) = (\eta_1(t), \cdots, \eta_n(t))$ と表すと、

$$\eta_i(t) = \eta_i(0) + t(\eta_i(1) - \eta_i(0)) \tag{1.51}$$

となる。これを用いて、点 q におけるそれぞれの曲線の接ベクトルを計算すると、

$$\mathbf{v} = \dot{p}(0) = \left(\frac{\mathrm{d}\theta^i}{\mathrm{d}t}\right)(\partial_i)_q = (\theta^i(1) - \theta^i(0))(\partial_i)_q = (\theta^i(p) - \theta^i(q))(\partial_i)_q \tag{1.52}$$

$$\mathbf{w} = \dot{q}(0) = \left(\frac{\mathrm{d}\eta_i}{\mathrm{d}t}\right)(\partial^i)_q = (\eta^i(1) - \eta^i(0))(\partial^i)_q = (\eta^i(r) - \eta^i(q))(\partial^i)_q \tag{1.53}$$

となる。これらが直交しているという仮定から、

$$\begin{aligned}
0 &= g(\mathbf{v}, \mathbf{w}) && (1.54)\\
&= g((\theta^i(p) - \theta^i(q))(\partial_i)_q, (\eta^j(r) - \eta^j(q))(\partial^j)_q) && (1.55)\\
&= (\theta^i(p) - \theta^i(q))(\eta^j(r) - \eta^j(q)) g(\partial_i, \partial^j) && (1.56)\\
&= (\theta^i(p) - \theta^i(q))(\eta^i(r) - \eta^i(q)) && (1.57)
\end{aligned}$$

となる。したがって、

$$\begin{aligned}
D(p||q) + D(q||r) - D(p||r) &= \psi(\theta(p)) + \varphi(\eta(q)) - \theta^i(p)\eta_i(q) && (1.58)\\
&\quad + \psi(\theta(q)) + \varphi(\eta(r)) - \theta^i(q)\eta_i(r) && (1.59)\\
&\quad - \psi(\theta(p)) - \varphi(\eta(r)) + \theta^i(p)\eta_i(r) && (1.60)\\
&= \psi(\theta(q)) + \varphi(\eta(q)) - \theta^i(p)\eta_i(q) - \theta^i(q)\eta_i(r) + \theta^i(p)\eta_i(r) && (1.61)\\
&= \theta^i(q)\eta_i(q) - \theta^i(p)\eta_i(q) - \theta^i(q)\eta_i(r) + \theta^i(p)\eta_i(r) && (1.62)\\
&\quad \because \mathrm{lem2} && (1.63)\\
&= (\theta^i(p) - \theta^i(q))(\eta_i(r) - \eta_i(q)) && (1.64)\\
&= 0 && (1.65)
\end{aligned}$$

となる。したがって、示された。 □