

情報幾何学の基礎

大上由人

2024 年 9 月 27 日

目次

1	はじめに	3
2	多様体のアフィン接続	3
2.1	ベクトル場の共変微分	3
2.2	アフィン接続	5
2.3	曲率	7
2.4	捩率	9
2.5	平坦な多様体とアフィン座標系	10
2.6	Riemann 接続	13
2.7	部分多様体	16
3	双対アフィン接続の幾何	20
3.1	双対アフィン接続	20
3.2	双対平坦な多様体の幾何	22
4	確率分布空間の幾何構造	30
4.1	Chentsov の定理	30
4.2	Fisher 計量と α -接続	33
4.3	指数型接続と混合型接続	35
4.4	\mathcal{S}_{n-1} の双対微分幾何	39
4.5	自己平行部分多様体	42
4.6	双対葉層構造	44
5	統計物理学への応用	46
5.1	統計力学	46
5.2	最大エントロピー原理	47

5.3	磁性	50
5.4	平均場近似	51

1 はじめに

Thm: 逆写像定理

点 $\mathbf{a} = (a_1, \dots, a_n)$ を含む領域 $D \subset \mathbb{R}^n$ から \mathbb{R}^n への C^1 級写像

$$\mathbf{f} : D \rightarrow \mathbb{R}^n \quad \mathbf{x} = (x_1, \dots, x_n) \mapsto \mathbf{y} = (f_1(\mathbf{x}), \dots, f_n(\mathbf{x})) \quad (1.1)$$

が、点 $\mathbf{f}(\mathbf{a})$ の近傍で C^1 級の逆写像を持つための必要十分条件は、 \mathbf{f} のヤコビ行列 $\mathbf{J}(\mathbf{x})$ が点 \mathbf{a} で正則であることである。

2 多様体のアフィン接続

2.1 ベクトル場の共変微分

一般の多様体上では、共変微分を用いて平行移動を定義する。以下では初めに共変微分の定義を述べる。

Def: 共変微分

M を多様体とする。以下の 4 条件を満たす写像

$$\nabla : \mathfrak{X}(M) \times \mathfrak{X}(M) \rightarrow \mathfrak{X}(M) \quad (X, Y) \mapsto \nabla_X Y \quad (2.1)$$

を M 上の共変微分という。

1.

$$\nabla_X (Y + Z) = \nabla_X Y + \nabla_X Z \quad (2.2)$$

2.

$$\nabla_X (fY) = (Xf)Y + f\nabla_X Y \quad (2.3)$$

3.

$$\nabla_{X+Y} Z = \nabla_X Z + \nabla_Y Z \quad (2.4)$$

4.

$$\nabla_{fX} Y = f\nabla_X Y \quad (2.5)$$

ex: Euclid 空間

このとき、共変微分 $\nabla_X Y$ は、テンソル性を満たさないが、これらの差はテンソル性を持つ。

∴

$$S(X, Y) = \nabla_X Y - \nabla'_X Y \quad (2.6)$$

$$(2.7)$$

に対して、任意の $f, g \in C^\infty(M)$ に対して、

$$S(fX, gY) = \nabla_{fX}(gY) - \nabla'_{fX}(gY) \quad (2.8)$$

$$= f\nabla_X(gY) - f\nabla'_X(gY) \quad (2.9)$$

$$= f((Xg)Y + g\nabla_X Y) - f((Xg)Y + g\nabla'_X Y) \quad (2.10)$$

$$= fg\nabla_X Y - fg\nabla'_X Y \quad (2.11)$$

$$= fgS(X, Y) \quad (2.12)$$

が成り立つ。 \square

したがって、 S は $(1, 2)$ 型のテンソル場である。したがって、 M 上に何か一つ共変微分を固定して、それとの差を考えると、 M 上の共変微分全体と、 $(1, 2)$ 型のテンソル場全体は、同型である。すなわち、

$$\{M \text{ 上の共変微分全体} \} \simeq \nabla^0 + \{ (1, 2) \text{ 型のテンソル場全体} \} \quad (2.13)$$

となる。

次に、共変微分の局所座標表示を考える。

Def: 接続係数

多様体 M の座標近傍 $(U; x^1, \dots, x^n)$ において、

$$\nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} = \Gamma_{ij}^k \frac{\partial}{\partial x^k} \quad (2.14)$$

によって定義される、 n^3 個の関数 Γ_{ij}^k を接続係数という。

これは、局所座標の取り方に依存することに注意する。別の座標近傍でも同様に、

$$\nabla_{\frac{\partial}{\partial \xi^a}} \frac{\partial}{\partial \xi^b} = \Gamma_{ab}^c \frac{\partial}{\partial \xi^c} \quad (2.15)$$

により、接続係数の組 $\{\Gamma_{ab}^c\}$ が定義される。このとき、座標変換則を考える。

Prop: 接続係数の座標変換

接続係数 Γ_{ij}^k と Γ_{ab}^c の間には、次の関係が成り立つ。

$$\Gamma_{ij}^k = \frac{\partial \xi^a}{\partial x^i} \frac{\partial \xi^b}{\partial x^j} \frac{\partial x^k}{\partial \xi^c} \Gamma_{ab}^c + \frac{\partial^2 \xi^c}{\partial x^i \partial x^j} \frac{\partial x^k}{\partial \xi^c} \quad (2.16)$$

Prf

$$\nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} = \Gamma_{ij}^k \frac{\partial}{\partial x^k} \quad (2.17)$$

$$= \Gamma_{ij}^k \frac{\partial \xi^c}{\partial x^k} \frac{\partial}{\partial \xi^c} \quad (2.18)$$

$$(2.19)$$

である、一方、

$$\nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} = \nabla_{\frac{\partial}{\partial \xi^b}} \frac{\partial}{\partial \xi^b} \quad (2.20)$$

$$= \frac{\partial^2 \xi^b}{\partial x^i \partial x^j} \frac{\partial}{\partial \xi^b} + \frac{\partial \xi^b}{\partial x^j} \nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial \xi^b} \quad (2.21)$$

$$= \frac{\partial^2 \xi^b}{\partial x^i \partial x^j} \frac{\partial}{\partial \xi^b} + \frac{\partial \xi^b}{\partial x^j} \nabla_{(\frac{\partial \xi^a}{\partial x^i} \frac{\partial}{\partial \xi^a})} \frac{\partial}{\partial \xi^b} \quad (2.22)$$

$$= \frac{\partial^2 \xi^b}{\partial x^i \partial x^j} \frac{\partial}{\partial \xi^b} + \frac{\partial \xi^b}{\partial x^j} \frac{\partial \xi^a}{\partial x^i} \nabla_{\frac{\partial}{\partial \xi^a}} \frac{\partial}{\partial \xi^b} \quad (2.23)$$

$$= \frac{\partial^2 \xi^b}{\partial x^i \partial x^j} \frac{\partial}{\partial \xi^b} + \frac{\partial \xi^b}{\partial x^j} \frac{\partial \xi^a}{\partial x^i} \Gamma_{ab}^c \frac{\partial}{\partial \xi^c} \quad (2.24)$$

$$= \frac{\partial^2 \xi^c}{\partial x^i \partial x^j} \frac{\partial}{\partial \xi^c} + \frac{\partial \xi^b}{\partial x^j} \frac{\partial \xi^a}{\partial x^i} \Gamma_{ab}^c \frac{\partial}{\partial \xi^c} \quad (2.25)$$

$$= \left(\frac{\partial^2 \xi^c}{\partial x^i \partial x^j} + \frac{\partial \xi^a}{\partial x^i} \frac{\partial \xi^b}{\partial x^j} \Gamma_{ab}^c \right) \frac{\partial}{\partial \xi^c} \quad (2.26)$$

であるから、

$$\Gamma_{ij}^k \frac{\partial \xi^c}{\partial x^k} = \frac{\partial^2 \xi^c}{\partial x^i \partial x^j} + \frac{\partial \xi^a}{\partial x^i} \frac{\partial \xi^b}{\partial x^j} \Gamma_{ab}^c \quad (2.27)$$

が成り立つ。したがって、

$$\Gamma_{ij}^k = \frac{\partial \xi^a}{\partial x^i} \frac{\partial \xi^b}{\partial x^j} \frac{\partial x^k}{\partial \xi^c} \Gamma_{ab}^c + \frac{\partial^2 \xi^c}{\partial x^i \partial x^j} \frac{\partial x^k}{\partial \xi^c} \quad (2.28)$$

が成り立つ。 \square

ところで、共変微分がテンソル性を満たさないことは上の座標変換の第二項が現れることによって分かる。

2.2 アフィン接続

Def: アフィン接続

多様体 M の各座標近傍に、上の座標変換則を満たすような $C^\infty(M)$ 上の関数の組 $\{\Gamma_{ij}^k\}$ を与えることを、 M にアフィン接続を与えるといい、 $\{\Gamma_{ij}^k\}$ を接続係数という。

以下、アフィン接続を用いて、平行および平行移動を定義する。

Def: 平行

多様体 M にアフィン接続が与えられているとする。 M 上のなめらかな曲線

$$C = \{p(t); t \in [a, b]\} \quad (2.29)$$

に沿って定義されたベクトル場 $Z = \{Z_{p(t)}\}$ が、 C に沿って平行であるとは、

$$\nabla_{\dot{p}(t)} Z_{p(t)} = 0 \quad (2.30)$$

が成り立つことをいう。ここで、

$$\dot{p}(t) = \frac{dx^i}{dt} \frac{\partial}{\partial x^i} \quad (2.31)$$

である。

Prf

上の平行条件を局所座標表示する。

$$Z_{p(t)} = Z^i \frac{\partial}{\partial x^i} \quad (2.32)$$

と成分表示すると、

$$\nabla_{\dot{p}(t)} Z = \nabla_{\frac{dx^i}{dt} \frac{\partial}{\partial x^i}} \left(Z^j \frac{\partial}{\partial x^j} \right) \quad (2.33)$$

$$= \frac{dx^i}{dt} \left(\frac{\partial Z^j}{\partial x^i} \frac{\partial}{\partial x^j} + \frac{dx^j}{dt} Z^j \nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} \right) \quad (2.34)$$

$$= \left(\dot{Z}^k + \dot{x}^i Z^j \Gamma_{ij}^k \right) \frac{\partial}{\partial x^k} \quad (2.35)$$

となる。したがって、

$$\dot{Z}^k + \dot{x}^i Z^j \Gamma_{ij}^k = 0 \quad (2.36)$$

が成り立つ。

ex: 測地線**Def: 平行移動**

多様体 M 上のなめらかな曲線 $C = \{p(t); t \in [a, b]\}$ と、 $\mathbf{p} \in T_{p(a)}M$ が任意に与えられたとき、2.36 の解として定まるベクトル場 Z の $t = b$ における値 $Z_{p(b)}$ を、 \mathbf{p} を $t = a$ から $t = b$ に沿って平行移動して得られた接ベクトルという。

また、この対応は全単射線形写像となり、^a、

$$\Pi_C : T_{p(a)}M \rightarrow T_{p(b)}M \quad (2.37)$$

を定める。これを、曲線 C に沿った平行移動という。

^a 微分方程式の解の一意性から従う。

このとき、平行条件は、

$$\mathbf{v} \in T_{p(a)}M \parallel \mathbf{w} \in T_{p(b)}M \Leftrightarrow \Pi_C \mathbf{v} = \mathbf{w} \quad (2.38)$$

と書ける。

Thm: 共変微分の幾何的意味

多様体 M にアフィン接続が与えられているとし、点 $p(a) \in M$ を通るような M のなめらかな曲線

$$C = \{p(t); t \in [a - \epsilon, a + \epsilon]\} \quad (2.39)$$

を考える。任意のベクトル場 $Y \in \mathfrak{X}(M)$ に対して、

$$(\nabla_{\dot{p}(t)} Y)_{p(a)} = \lim_{t \rightarrow a} \frac{1}{t - a} (\Pi_{p(a)}^{p(t)} Y_{p(t)} - Y_{p(a)}) \quad (2.40)$$

が成り立つ。ここで、 $\Pi_{p(a)}^{p(t)}$ は、 $p(t)$ から $p(a)$ に沿って平行移動する写像である。

Prf

2.3 曲率

平行移動が、経路に依存することを示すために、曲率を導入する。以降、 $\frac{\partial}{\partial x^i}$ を ∂_i と略記する。

Def: 曲率テンソル場

多様体 M にアフィン接続が与えられているとする。 M 上で定義された 3 重 $C^\infty(M)$ -テンソル場 R を、

$$R : \mathfrak{X}(M) \times \mathfrak{X}(M) \times \mathfrak{X}(M) \rightarrow \mathfrak{X}(M) \quad (2.41)$$

$$(X, Y, Z) \mapsto \nabla_X(\nabla_Y Z) - \nabla_Y(\nabla_X Z) - \nabla_{[X, Y]} Z \quad (2.42)$$

によって定義する。このテンソル場を、接続 ∇ の曲率テンソル場という。

局所座標表示すると、

$$R(\partial_i, \partial_j, \partial_k) = R_{ijk}^l \partial_l \quad (2.43)$$

$$(2.44)$$

となる。右辺を具体的に求めると、よって、

$$R_{ijk}^l = \partial_i \Gamma_{jk}^l - \partial_j \Gamma_{ik}^l + \Gamma_{im}^l \Gamma_{jk}^m - \Gamma_{jm}^l \Gamma_{ik}^m \quad (2.45)$$

となる。

Prop: 曲率と平行移動

曲率が 0 であることと、平行移動が経路に依存しないことは同値である。

Prf

2.4 振率

Def: 振率

アフィン接続 ∇ をもつ多様体 M 上で定義された二重 $C^\infty(M)$ -テンソル場 T を、

$$T : \mathfrak{X}(M) \times \mathfrak{X}(M) \rightarrow C^\infty(M) \quad (2.46)$$

$$(X, Y) \mapsto \nabla_X Y - \nabla_Y X - [X, Y] \quad (2.47)$$

によって定義する。このテンソル場を、接続 ∇ の振率テンソル場という。

このとき、 T は $(1, 2)$ 型テンソル場である。

\therefore 共変微分の定義を繰り返し用いる。

$$T(fX, gY) = \nabla_{fX}(gY) - \nabla_{gY}(fX) - [fX, gY] \quad (2.48)$$

$$= f\nabla_X(gY) - g\nabla_Y(fX) \quad (2.49)$$

$$- f(Xg)Y - fg(XY) + g(Yf)X + fg(YX) \quad (2.50)$$

$$= f(Xg)Y + fg\nabla_X Y - f(Yf)X - fg\nabla_Y X \quad (2.51)$$

$$- f(Xg)Y - fg(XY) + g(Yf)X + fg(YX) \quad (2.52)$$

$$= fg(\nabla_X Y - \nabla_Y X - [X, Y]) \quad (2.53)$$

$$= fgT(X, Y) \quad (2.54)$$

となり、 $(1, 2)$ 型テンソル場であることがわかる。 \square

局所座標表示すると、

$$T(\partial_i, \partial_j) = T_{ij}^k \partial_k \quad (2.55)$$

$$(2.56)$$

と定義される。 T_{ij}^k を具体的に求める。左辺について、

$$T(\partial_i, \partial_j) = \nabla_{\partial_i} \partial_j - \nabla_{\partial_j} \partial_i - [\partial_i, \partial_j] \quad (2.57)$$

$$= \nabla_{\partial_i} \partial_j - \nabla_{\partial_j} \partial_i - 0 \quad (2.58)$$

$$= \Gamma_{ij}^k \partial_k - \Gamma_{ji}^k \partial_k \quad (2.59)$$

$$= (\Gamma_{ij}^k - \Gamma_{ji}^k) \partial_k \quad (2.60)$$

となるから、

$$T_{ij}^k = \Gamma_{ij}^k - \Gamma_{ji}^k \quad (2.61)$$

となる。

振率の幾何的な意味を見るために、曲率のときと同様に、微小四辺形をに沿って接ベクトルを平行移動してみる。

$$\{\epsilon_i \partial_i(x) + \epsilon_j \Pi_x^{x+\delta_i} \partial_j(x + \delta_i)\} - \{\epsilon_j \partial_j(x) + \epsilon_i \Pi_x^{x+\delta_j} \partial_i(x + \delta_j)\} \quad (2.62)$$

について考える。これは、それぞれ、 i 方向への微小変化の後に j 方向に微小変化させることと、 j 方向への微小変化の後に i 方向に微小変化させることを表している。まず、

$$\Pi_x^{x+\delta_i} \partial_j(x + \delta_i) = \partial_j(x) + \epsilon_i \Gamma_{ij}^k \partial_k + \mathcal{O}(\epsilon_i^2) \quad (2.63)$$

であるから、

$$\epsilon_i \partial_i(x) + \epsilon_j \Pi_x^{x+\delta_i} \partial_j(x + \delta_i) = \epsilon_i \partial_i(x) + \epsilon_j (\partial_j(x) + \epsilon_i \Gamma_{ij}^k \partial_k) + \mathcal{O}(\epsilon_i^2) \quad (2.64)$$

$$= \epsilon_i \partial_i(x) + \epsilon_j \partial_j(x) + \epsilon_j \epsilon_i \Gamma_{ij}^k \partial_k + \mathcal{O}(\epsilon_i^2) \quad (2.65)$$

となる。同様に、

$$\epsilon_j \partial_j(x) + \epsilon_i \Pi_x^{x+\delta_j} \partial_i(x + \delta_j) = \epsilon_j \partial_j(x) + \epsilon_i (\partial_i(x) + \epsilon_j \Gamma_{ji}^k \partial_k) + \mathcal{O}(\epsilon_j^2) \quad (2.66)$$

$$= \epsilon_j \partial_j(x) + \epsilon_i \partial_i(x) + \epsilon_i \epsilon_j \Gamma_{ji}^k \partial_k + \mathcal{O}(\epsilon_j^2) \quad (2.67)$$

となる。したがって、(2.62) は、

$$\epsilon_j \epsilon_i \Gamma_{ij}^k \partial_k - \epsilon_i \epsilon_j \Gamma_{ji}^k \partial_k = \epsilon_j \epsilon_i T_{ij}^k \partial_k \quad (2.68)$$

となる。すなわち、捩率は、ベクトルの和の経路依存性の、面積要素の比例する主要部分を表している。

2.5 平坦な多様体とアフィン座標系

Def: ∇ -平坦

多様体 M にアフィン接続 ∇ が与えられているとする。このとき、 ∇ が平坦であるとは、 R と T が恒等的に 0 であることをいう。

Def: ∇ -アフィン座標系

多様体 M にアフィン接続 ∇ が与えられているとする。 M の座標近傍 $(U; x^1, \dots, x^n)$ において、接続係数 $\{\Gamma_{ij}^k\}$ がすべて恒等的に 0 であるとき、この座標近傍を ∇ -アフィン座標近傍といい、局所座標系 (x^i) を ∇ -アフィン座標系という。

このとき、(2.36) は、

$$\dot{Z}^k = 0 \quad (2.69)$$

と書くことができる。したがって、 ∇ -アフィン座標系における平行移動とは、基底ベクトルが一定となるような移動のことであり、これは斜向座標における平行移動に対応していると考えられる。

Rem

平坦であるという性質は (曲率や捩率がテンソルで定義されているのだから) 座標の取り方に依らないが、アフィン座標系かどうかは座標の取り方に依存する。

Thm: 局所アフィン座標系の存在

多様体 M にアフィン接続 ∇ が与えられているとする。このとき、以下の二つは同値である。

1. M は、 ∇ -平坦である。
2. M の任意の点 $p \in M$ において、 ∇ -アフィン座標近傍が存在する。

Prf

(2) \Rightarrow (1)

接続係数が恒等的に 0 となる局所座標系があったならば、曲率と捩率の定義より、その座標系での曲率と捩率は 0 である。また、曲率と捩率はテンソル場であるから、それは座標系に依らない。したがって、 ∇ は平坦である。

(1) \Rightarrow (2)

任意の座標近傍 $(U; x^1, \dots, x^n)$ に対して、ある座標近傍 $(V; \xi^1, \dots, \xi^n)$ が存在して、 $U \cap V$ で、

$$0 = \Gamma_{ab}^c = \frac{\partial x^i}{\partial \xi^a} \frac{\partial x^j}{\partial \xi^b} \frac{\partial \xi^c}{\partial x^k} \Gamma_{ij}^k + \frac{\partial^2 x^l}{\partial \xi^a \partial \xi^b} \frac{\partial \xi^c}{\partial x^l} \quad \forall a, b, c \quad (2.70)$$

が成り立つことを示す。

$$\frac{\partial x^\ell}{\partial \xi^b} \frac{\partial \xi^b}{\partial x^j} = \delta_j^\ell \quad (2.71)$$

の両辺を x^i で偏微分して

$$\frac{\partial^2 x^\ell}{\partial \xi^a \partial \xi^b} \frac{\partial \xi^a}{\partial x^i} \frac{\partial \xi^b}{\partial x^j} + \frac{\partial x^\ell}{\partial \xi^b} \frac{\partial^2 \xi^b}{\partial x^i \partial x^j} = 0 \quad (2.72)$$

だから、両辺に Jacobi 行列 $\frac{\partial \xi^a}{\partial x^i}$ と $\frac{\partial \xi^b}{\partial x^j}$ の逆行列を掛けて

$$\frac{\partial^2 x^\ell}{\partial \xi^a \partial \xi^b} = - \frac{\partial x^i}{\partial \xi^a} \frac{\partial x^j}{\partial \xi^b} \frac{\partial x^\ell}{\partial \xi^d} \frac{\partial^2 \xi^d}{\partial x^i \partial x^j} \quad (2.73)$$

この式を (2.70) に代入して

$$0 = \frac{\partial x^i}{\partial \xi^a} \frac{\partial x^j}{\partial \xi^b} \frac{\partial \xi^c}{\partial x^k} \Gamma_{ij}^k - \left(\frac{\partial x^i}{\partial \xi^a} \frac{\partial x^j}{\partial \xi^b} \frac{\partial x^\ell}{\partial \xi^d} \frac{\partial^2 \xi^d}{\partial x^i \partial x^j} \right) \frac{\partial \xi^c}{\partial x^\ell} \quad (2.74)$$

$$= \frac{\partial x^i}{\partial \xi^a} \frac{\partial x^j}{\partial \xi^b} \frac{\partial \xi^c}{\partial x^k} \Gamma_{ij}^k - \frac{\partial x^i}{\partial \xi^a} \frac{\partial x^j}{\partial \xi^b} \delta^d \frac{\partial^2 \xi^d}{\partial x^i \partial x^j} \quad (2.75)$$

$$= \frac{\partial x^i}{\partial \xi^a} \frac{\partial x^j}{\partial \xi^b} \left(\frac{\partial \xi^c}{\partial x^k} \Gamma_{ij}^k - \frac{\partial^2 \xi^c}{\partial x^i \partial x^j} \right) \quad (2.76)$$

従って、(2.70) は次式と同値である。

$$\frac{\partial^2 \xi^c}{\partial x^i \partial x^j} = \frac{\partial \xi^c}{\partial x^k} \Gamma_{ij}^k \quad (2.77)$$

あるいはこれを連立方程式の形で書けば

$$\frac{\partial \xi^c}{\partial x^k} = \theta_k^c \quad (2.78)$$

$$\frac{\partial^2 \xi^c}{\partial x^i \partial x^j} = \theta_k^c \Gamma_{ij}^k \quad (2.79)$$

この連立偏微分方程式の可積分条件は

$$\frac{\partial^2 \xi^c}{\partial x^i \partial x^j} = \frac{\partial^2 \xi^c}{\partial x^j \partial x^i} \quad (2.80)$$

$$\frac{\partial^2 \theta_i^c}{\partial x^i \partial x^j} = \frac{\partial^2 \theta_k^c}{\partial x^j \partial x^i} \quad (2.81)$$

となる。これの第 1 式を同値変形していくと、

$$\frac{\partial^2 \xi^c}{\partial x^i \partial x^j} = \frac{\partial^2 \xi^c}{\partial x^j \partial x^i} \quad (2.82)$$

$$\Leftrightarrow \frac{\partial}{\partial x^i} \theta_j^c = \frac{\partial}{\partial x^j} \theta_i^c \quad (2.83)$$

$$\Leftrightarrow \theta_k^c \Gamma_{ij}^k = \theta_k^c \Gamma_{ji}^k \quad (2.84)$$

$$\Leftrightarrow \theta_k^c T_{ij}^k = 0 \quad (2.85)$$

同様に、第 2 式を同値変形していくと、

$$\frac{\partial^2 \theta_k^c}{\partial x^i \partial x^j} = \frac{\partial^2 \theta_k^c}{\partial x^j \partial x^i} \Leftrightarrow \frac{\partial}{\partial x^i} (\theta_l^c \Gamma_{jk}^l) = \frac{\partial}{\partial x^j} (\theta_l^c \Gamma_{ik}^l) \quad (2.86)$$

$$\Leftrightarrow (\theta_m^c \Gamma_{il}^m) \Gamma_{jk}^l + \theta_l^c (\partial_i \Gamma_{jk}^l) = (\theta_m^c \Gamma_{jl}^m) \Gamma_{ik}^l + \theta_l^c (\partial_j \Gamma_{ik}^l) \quad (2.87)$$

$$\Leftrightarrow \theta_m^c \{ \Gamma_{il}^m \Gamma_{jk}^l + \partial_i \Gamma_{jk}^m - \Gamma_{jl}^m \Gamma_{ik}^l - \partial_j \Gamma_{ik}^m \} = 0 \quad (2.88)$$

$$\Leftrightarrow \theta_m^c R_{ijk}^m = 0 \quad (2.89)$$

さて、仮定より $T = 0$ かつ $R = 0$ であるから、連立微分方程式 (3.22) は可積分であることが分かる。まとめると、任意に固定した $p \in U$ における初期条件 $\psi(p) = (\xi^c)_p, (Jf)_p = (\partial_i \xi^c)_p$ を二に与えると、これを満たすような偏微分方程式の解 ξ^c が局所的に存在する。

したがって、特に初期条件を $\det(Jf)_p \neq 0$ とすると、 ξ^c は p の近傍で局所座標系を張る。したがって、 ∇ -アフィン座標系が存在する。 \square

Thm: アフィン座標系の自由度

多様体 M に座標近傍 $(U; x^i)$ と $(V; \xi^a)$ が与えられていて、 (x^i) が局所 ∇ -アフィン座標系であるとする。このとき、

$$(V; \xi^a) \text{ が局所 } \nabla\text{-アフィン座標系である} \Leftrightarrow \exists A \in GL(n, \mathbb{R}) \text{ s.t. } \xi^i = A_j^i x^j + b_i \quad (2.90)$$

が成り立つ。ここで、 $GL(n, \mathbb{R})$ は n 次正則行列全体の集合であり、 b_i は定数である。

Prf

接続係数の座標変換則:

$$\Gamma_{ij}^k = \frac{\partial \xi^a}{\partial x^i} \frac{\partial \xi^b}{\partial x^j} \frac{\partial x^k}{\partial \xi^c} \Gamma_{ab}^c + \frac{\partial^2 \xi^c}{\partial x^i \partial x^j} \frac{\partial x^k}{\partial \xi^c} \quad (2.91)$$

において、 $\Gamma_{ij}^k = 0$ であるという条件のもとで $\Gamma_{ab}^c = 0$ となるための必要十分条件は、第二項が消えることである。すなわち、

$$\frac{\partial^2 \xi^c}{\partial x^i \partial x^j} = 0 \quad (2.92)$$

であり、これは示すべき式と同値である。またこのとき、 A は座標変換の Jacobi 行列なので、特に正則となる。 \square

2.6 Riemann 接続

Def: Riemann 多様体

M 上で定義された $(0,2)$ 型テンソル場 g であって、各点 $p \in M$ で、 g_p が内積を定めるとき、 g を M 上の Riemann 計量という。このとき、 (M, g) を Riemann 多様体という。ここで、

$$g_{ij} = g(\partial_i, \partial_j) \quad (2.93)$$

は正定値対称行列となる。

Def: Riemann 接続/Levi-Civita 接続/Levi-Civita 平行移動

Riemann 多様体 (M, g) の各座標近傍 $(U; x^1, \dots, x^n)$ において、

$$\Gamma_{ij,k} = \Gamma_{ij}^l g_{lk} = \frac{1}{2}(\partial_i g_{jk} + \partial_j g_{ki} - \partial_k g_{ij}) \quad (2.94)$$

により定まる M のアフィン接続を、Levi-Civita 接続という。また、この接続が定める M 上の平行移動を Levi-Civita 平行移動という。

このとき、Levi-Civita 接続は affine 接続を定める。

\therefore

計量

$$g_{ij} = g(\partial_i, \partial_j) \quad (2.95)$$

の座標変換は、

$$\begin{aligned} g_{ik} &= \frac{\partial x^b}{\partial x^i} \frac{\partial x^c}{\partial x^k} g_{bc}, \\ g_{ki} &= \frac{\partial x^c}{\partial x^k} \frac{\partial x^a}{\partial x^i} g_{ca}, \\ g_{ij} &= \frac{\partial x^a}{\partial x^i} \frac{\partial x^b}{\partial x^j} g_{ab}, \end{aligned}$$

で与えられる。そして、

$$\begin{aligned} \partial_i g_{ik} &= \frac{\partial \xi^b}{\partial x^j} \frac{\partial \xi^c}{\partial x^k} \frac{\partial \xi^l}{\partial x^i} \frac{\partial g_{bc}}{\partial \xi^l} + g_{bc} \partial_i \left(\frac{\partial \xi^b}{\partial x^j} \frac{\partial \xi^c}{\partial x^k} \right), \\ \partial_j g_{ki} &= \frac{\partial \xi^c}{\partial x^k} \frac{\partial \xi^a}{\partial x^i} \frac{\partial \xi^l}{\partial x^j} \frac{\partial g_{ca}}{\partial \xi^l} + g_{ca} \partial_j \left(\frac{\partial \xi^c}{\partial x^k} \frac{\partial \xi^a}{\partial x^i} \right), \\ \partial_k g_{ij} &= \frac{\partial \xi^a}{\partial x^i} \frac{\partial \xi^b}{\partial x^j} \frac{\partial \xi^l}{\partial x^k} \frac{\partial g_{ab}}{\partial \xi^l} + g_{ab} \partial_k \left(\frac{\partial \xi^a}{\partial x^i} \frac{\partial \xi^b}{\partial x^j} \right), \end{aligned} \quad (2.96)$$

であるから、

$$\Gamma_{ij,k} = \frac{1}{2} (\partial_i g_{jk} + \partial_j g_{ki} - \partial_k g_{ij}) \quad (2.97)$$

$$= \frac{1}{2} \left[\frac{\partial \xi^a}{\partial x^i} \frac{\partial \xi^b}{\partial x^j} \frac{\partial \xi^c}{\partial x^k} \left(\frac{\partial g_{bc}}{\partial \xi^a} + \frac{\partial g_{ca}}{\partial \xi^b} - \frac{\partial g_{ab}}{\partial \xi^c} \right) \right. \quad (2.98)$$

$$+ g_{bc} \left(\frac{\partial^2 \xi^b}{\partial x^i \partial x^j} \frac{\partial \xi^c}{\partial x^k} + \frac{\partial \xi^b}{\partial x^j} \frac{\partial^2 \xi^c}{\partial x^k \partial x^i} \right) \quad (2.99)$$

$$+ g_{ca} \left(\frac{\partial^2 \xi^c}{\partial x^j \partial x^k} \frac{\partial \xi^a}{\partial x^i} + \frac{\partial \xi^c}{\partial x^k} \frac{\partial^2 \xi^a}{\partial x^i \partial x^j} \right) \quad (2.100)$$

$$\left. - g_{ab} \left(\frac{\partial^2 \xi^a}{\partial x^k \partial x^i} \frac{\partial \xi^b}{\partial x^j} + \frac{\partial \xi^a}{\partial x^i} \frac{\partial^2 \xi^b}{\partial x^j \partial x^k} \right) \right] \quad (2.101)$$

$$= \frac{\partial \xi^a}{\partial x^i} \frac{\partial \xi^b}{\partial x^j} \frac{\partial \xi^c}{\partial x^k} \Gamma_{ab,c} + \frac{\partial \xi^c}{\partial x^k} \frac{\partial^2 \xi^b}{\partial x^i \partial x^j} g_{bc} \quad (2.102)$$

となる。両辺に $g^{kl} = \frac{\partial x^k}{\partial \xi^c} \frac{\partial x^l}{\partial \xi^d} g^{cd}$ をかけて縮約すると、

$$\Gamma_{ij}^l = \frac{\partial \xi^a}{\partial x^i} \frac{\partial \xi^b}{\partial x^j} \frac{\partial \xi^c}{\partial x^k} \Gamma_{ab,c} \frac{\partial x^k}{\partial \xi^c} \frac{\partial x^l}{\partial \xi^d} g^{cd} \quad (2.103)$$

$$+ \frac{\partial \xi^c}{\partial x^k} \frac{\partial^2 \xi^b}{\partial x^i \partial x^j} g_{bc} \frac{\partial x^k}{\partial \xi^c} \frac{\partial x^l}{\partial \xi^d} g^{cd} \quad (2.104)$$

$$= \frac{\partial \xi^a}{\partial x^i} \frac{\partial \xi^b}{\partial x^j} \frac{\partial x^l}{\partial \xi^d} \Gamma_{ab}^d + \frac{\partial^2 \xi^c}{\partial x^i \partial x^j} \frac{\partial x^l}{\partial \xi^c} \quad (2.105)$$

となる。したがって、Levi-Civita 接続の接続係数は、計量の座標変換則と同様に変換する。 \square

Thm: 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 Levi-Civita 平行移動によって、内積は不変である。すなわち、

$$g_p(b) (\Pi_b^a \mathbf{v}, \Pi_b^a \mathbf{w}) = g_p(a) (\mathbf{v}, \mathbf{w}) \quad (2.106)$$

が成り立つ。

Prf

$$\frac{d}{dt} g_{p(t)}(Y_{p(t)}, Z_{p(t)}) = 0 \quad (2.107)$$

$$(2.108)$$

であることを示せばよい。

$$\begin{aligned} \frac{d}{dt} g_{p(t)}(Y_{p(t)}, Z_{p(t)}) &= \frac{d}{dt} (Y^i(t)(\partial_i)_{p(t)}, Z^j(t)(\partial_j)_{p(t)}) \\ &= \frac{d}{dt} (Y^i Z^j g_{ij}) \end{aligned} \quad (2.109)$$

$$\begin{aligned} \because \frac{d}{dt} g_{ij} &= \frac{dx^k}{dt} \partial_k g_{ij} \\ &= (\dot{Y}^i Z^j + Y^i \dot{Z}^j) g_{ij} + Y^i Z^j \partial_k g_{ij} \dot{x}^k \end{aligned} \quad (2.110)$$

$$\because \text{平行移動で } \dot{Y}^i = -\Gamma_{kl}^i \dot{x}^k Y^l, \quad \dot{Z}^j = -\Gamma_{kl}^j \dot{x}^k Z^l \quad (2.111)$$

$$\begin{aligned} &= (\partial_k g_{ij} - \Gamma_{ki,j} - \Gamma_{kj,i}) \dot{x}^k Y^i Z^j \\ &= (\partial_k g_{ij} - \frac{1}{2}(\partial_k g_{ij} + \partial_i g_{jk} - \partial_j g_{ki})) \end{aligned} \quad (2.112)$$

$$- \frac{1}{2}(\partial_k g_{ji} + \partial_j g_{ik} - \partial_i g_{kj}) \dot{x}^k Y^i Z^j = 0 \quad (2.113)$$

により、内積は保存される。

□

Prop: 計量的

計量が平行移動で不変であるとき、接続を計量的であるという。ここで、接続 ∇ が計量的であるための必要十分条件は、

$$\partial_k g_{ij} = \Gamma_{ki,j} + \Gamma_{kj,i} \quad \forall i, j, k \quad (2.114)$$

が成り立つことである。座標系に依らない形で書くと、

$$X_g(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z) \quad (X, Y, Z \in \mathfrak{X}(M)) \quad (2.115)$$

と書ける。

上の定理からわかる通り、Levi-Civita 接続は計量的である。

Thm: Riemann 接続の特徴付け

$$\nabla \text{ が Riemann 接続である} \Leftrightarrow \nabla \text{ が計量的であるかつ } T = 0 \quad (2.116)$$

必要性は上の定理で証明した。十分性は、計量係数の条件 (3.27) のインデックスを巡回的に入れ替えて作った 3 つの式

$$\partial_i g_{jk} = \Gamma_{ij,k} + \Gamma_{ik,j} \quad (2.117)$$

$$\partial_j g_{ki} = \Gamma_{jk,i} + \Gamma_{ji,k} \quad (2.118)$$

$$\partial_k g_{ij} = \Gamma_{ki,j} + \Gamma_{kj,i} \quad (2.119)$$

を用い、さらに接続の対称性 $\Gamma_{ij,k} = \Gamma_{ji,k}$ に気をつければ

$$\partial_i g_{jk} + \partial_j g_{ki} - \partial_k g_{ij} = (\Gamma_{ij,k} + \Gamma_{ik,j}) + (\Gamma_{jk,i} + \Gamma_{ji,k}) - (\Gamma_{ki,j} + \Gamma_{kj,i}) \quad (2.120)$$

$$= 2\Gamma_{ij,k} \quad (2.121)$$

となって Riemann 接続の定義式が再現される。 \square

2.7 部分多様体

Def: 部分多様体

$m < n$ とする。 n 次元多様体 N の部分集合 M が N の m 次元多様体であるとは、 M の各点 $p \in M$ において、 p の近傍 U と N の座標近傍 $(V; x^1, \dots, x^n)$ が存在して、

$$M \cap U = \{p \in U; x^{m+1} = \dots = x^n = 0\} \quad (2.122)$$

と表されることをいう。

ex

1 次元部分多様体 $M = \{(x, y) \in \mathbb{R}^2 | y = x^2\}$ $N = \mathbb{R}^2$ であり、 N の標準的座標系 (x, y) に対し

$$\begin{cases} \xi = x \\ \eta = y - x^2 \end{cases} \quad (3.243)$$

で定まる写像 $\varphi : (x, y) \mapsto (\xi, \eta)$ を考えると、

$$\det \left(\frac{\partial \xi^a}{\partial x^i} \right) = \det \begin{pmatrix} 1 & 0 \\ -2x & 1 \end{pmatrix} = 1 \quad (3.244)$$

であるから、逆関数の定理より (ξ, η) は N の局所座標系である。そして、部分集合 M ではたしかに $\eta = 0$ である。

ここで、 N から M に誘導される多様体としての構造を考える。

座標近傍

N の座標近傍 (U, x^1, \dots, x^n) に対して、 M の座標近傍 (U, x^1, \dots, x^m) を、 $U \cap M$ 上での x^i の

制限として定める。

計量

N の Riemann 計量 g に対して、 M 上の計量は、

$$\tilde{g}_p(\mathbf{v}, \mathbf{w}) = g_p(\mathbf{v}, \mathbf{w}) \quad (\mathbf{v}, \mathbf{w} \in T_p M) \quad (2.123)$$

と定める。

接続

接続については厄介である。というのも、

$$\nabla_X Y \in T_p N \quad (2.124)$$

に対して、 $\nabla_X Y$ が M 上のベクトル場であるとは限らないからである。

解決策としては、共変微分したあとのベクトル場を M に射影することが考えられる。すなわち、

$$(\tilde{\nabla}_X Y)_p = \pi((\nabla_X Y)_p) \quad (2.125)$$

と定める。ここで、 π は $T_p N$ から $T_p M$ への射影である。具体的な射影の方法としては、計量が用いられる：

$$\tilde{g}(\tilde{\nabla}_X Y, Z) = g(\nabla_X Y, Z) \quad \forall Z \in \mathfrak{X}(M) \quad (2.126)$$

を満たすように $\tilde{\nabla}$ を定める。

とくに、射影を用いずとも接続を自然に誘導することができる多様体に名前を付けておく。

Def: 自己平行部分多様体

N をアフィン接続 ∇ を持つ多様体とし、 M を N の部分多様体とする。このとき、 M が自己平行であるとは、

$$(\nabla_X Y)_p \in T_p M \quad \forall X, Y \in \mathfrak{X}(M) \quad (2.127)$$

が成り立つことをいう。

また、関連する多様体についても定義しておく。

Def: 全測地的部分多様体

N を Riemann 多様体 (N, g) とし、 M を N の部分多様体とする。このとき、 M が全測地的であるとは、任意の点 $p \in M$ および $p(0) = p$ かつ $\dot{p}(0) \in T_p M$ を満たす任意の ∇ -測地線 $p(t)$ が、十分小さな t に対して M に含まれることをいう。

Thm: 自己平行ならば全測地的

M が自己平行であれば、 M は全測地的である。

Prf

M が自己平行であるということは、

$$(\nabla_X Y)_p \in T_p M \quad \forall X, Y \in \mathfrak{X}(M) \quad (2.128)$$

となることであった。接続係数が、 $\nabla_{\partial_i} \partial_j = \Gamma_{ij}^k \partial_k$ と書けることを思い出すと、これは、接続係数 Γ_{ij}^k が任意の i, j に対して $m+1 \leq k \leq n$ で 0 であることを意味する。したがって、初期条件 $x^k = \dot{x}^k = 0 \quad (m+1 \leq k \leq n)$ での測地線の方程式

$$\ddot{x}^k + \Gamma_{ij}^k \dot{x}^i \dot{x}^j = 0 \quad (2.129)$$

の解は、 $x^k = \dot{x}^k = 0 \quad (m+1 \leq k \leq n)$ である。^{*1}

□

逆は、一般には成り立たない。しかし、次の定理が成り立つ。

Thm: 全測地的ならば自己平行

N の振率が 0 であるとき、 M が全測地的であれば、 M は自己平行である。

Prf

M を局所的に $(x^1, \dots, x^m, 0, \dots, 0)$ で表す N の座標近傍 $(U; x^1, \dots, x^n)$ を考える。 M が全測地的であるとき、 $p \in M$ を通り、 $\dot{p}(0) \in T_p M$ である任意の測地線 $p(t)$ は、 M に含まれる。すなわち、

$$x^k(t) \equiv 0 \quad (m+1 \leq k \leq n) \quad (2.130)$$

が成り立つ。したがって、測地線の方程式は、

$$\Gamma_{ij}^k(p(t)) \dot{x}^i(t) \dot{x}^j(t) = 0 \quad (m+1 \leq k \leq n) \quad (2.131)$$

となる。 i, j がとる範囲は制限していないことに注意されたい。ここで、2.130 より、 $\dot{x}^k(t) = 0 \quad (m+1 \leq k \leq n)$ であるから、

$$\sum_{i,j=1}^m \Gamma_{ij}^k(p(0)) \dot{x}^i(0) \dot{x}^j(0) = 0 \quad (m+1 \leq k \leq n) \quad (2.132)$$

が成り立つ。和の取り方を変えて、

$$\sum_{i,j=1}^m (\Gamma_{ij}^k(p(0)) + \Gamma_{ji}^k(p(0))) \dot{x}^i(0) \dot{x}^j(0) = 0 \quad (m+1 \leq k \leq n) \quad (2.133)$$

これが、任意の $p = p(0)$ および $\{\dot{x}^i(0)\}$ に対して成り立つことから、

$$\Gamma_{ij}^k(p) + \Gamma_{ji}^k(p) = 0 \quad (m+1 \leq k \leq n) \quad (2.134)$$

が成り立つ。いま、振率が 0 であるから、任意の $p \in M$ で、

$$\Gamma_{ij}^k(p) = 0 \quad (2.135)$$

^{*1} 解の一意性からわかる

が成り立つ。したがって、 M は自己平行である。 \square

Thm:

N をアフィン接続 ∇ に関して平坦な n 次元多様体とし、そのアフィン座標系 $(x^i)_{1 \leq i \leq n}$ を任意に固定する。 N の m 次元部分多様体 M が直線近似できるための必要十分条件は、 M が N のアフィン座標系 $(x^j)_{1 \leq j \leq n}$ のアフィン部分空間（の開部分集合）に対応することであり、 M のある局所座標系 $(\xi^a)_{1 \leq a \leq m}$ と、 $\text{rank } A = m$ なる $n \times m$ 行列 A 、および $b \in \mathbb{R}^n$ が存在して次式のようになることである。

$$\begin{bmatrix} x^1 \\ \vdots \\ x^m \\ x^{m+1} \\ \vdots \\ x^n \end{bmatrix} = A \begin{bmatrix} \xi^1 \\ \vdots \\ \xi^m \end{bmatrix} + b \quad (2.136)$$

Prf

必要性を示す

M が自己平行ならば、 M に誘導された曲率、捩率はともにゼロ。ゆえに M 自身 ∇ -平坦 affine 座標系 (ξ^1, \dots, ξ^m) をもつ。 M 上で (x^1, \dots, x^n) は (ξ^1, \dots, ξ^m) の関数

$$0 = \tilde{\nabla}_a \partial_b = \nabla_a \partial_b = \left(\frac{\partial x^i}{\partial \xi^a} \frac{\partial x^j}{\partial \xi^b} \Gamma_{ij}^k + \frac{\partial^2 x^k}{\partial \xi^a \partial \xi^b} \right) \partial_k \quad (2.137)$$

(x^1, \dots, x^n) は N の affine 座標系であったから、 $\Gamma_{ij}^k = 0$.

$$\Rightarrow \frac{\partial^2 x^k}{\partial \xi^a \partial \xi^b} = 0 \quad (1 \leq a, b \leq m, 1 \leq k \leq n) \quad (2.138)$$

これを積分すれば題意を得る。

十分性を示す。 $\Gamma_{ij}^k = 0$ であり、

$$\frac{\partial^2 x^k}{\partial \xi^a \partial \xi^b} = 0 \quad (2.139)$$

が成り立つので、

$$\nabla_a \partial_b = \left(\frac{\partial x^i}{\partial \xi^a} \frac{\partial x^j}{\partial \xi^b} \Gamma_{ij}^k + \frac{\partial^2 x^k}{\partial \xi^a \partial \xi^b} \right) \partial_k = 0 \quad (\forall a, b) \quad (2.140)$$

すべての $p \in M$ で $(\nabla_a \partial_b)_p \in T_p(M)$ となるので、 M は自己平行。 \square

3 双対アフィン接続の幾何

狭い意味での情報幾何学として、双対アフィン接続の幾何を考える。

3.1 双対アフィン接続

Def: 双対アフィン接続

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

$$Xg(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X^* Z) \quad (X, Y, Z \in \mathfrak{X}(M)) \quad (3.1)$$

により定める。

例えば、Riemann 接続の双対アフィン接続は、

$$Xg(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z) \quad (X, Y, Z \in \mathfrak{X}(M)) \quad (3.2)$$

より、 $\nabla^* = \nabla$ である。(自己双対)

このとき、双対アフィン接続は上の定義で一意に定まる。

∴

二つの双対アフィン接続 ∇^* と $\tilde{\nabla}^*$ が定まると仮定すると、

$$Xg(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X^* Z) \quad (3.3)$$

$$= g(\nabla_X Y, Z) + g(Y, \tilde{\nabla}_X^* Z) \quad (3.4)$$

である。右辺を引き算すると、

$$0 = g(Y, (\nabla_X^* - \tilde{\nabla}_X^*)Z) \quad (3.5)$$

となる。 Y は任意なので、

$$\nabla_X^* = \tilde{\nabla}_X^* \quad (3.6)$$

が成り立つ。□

また、 ∇^* は、共変微分の性質を満たす。

∴

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ィン接続のペア (∇, ∇^*) が与えられたとき、 (g, ∇, ∇^*) を M の双対構造という。

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、

$$g_{p(b)}(\Pi_C \mathbf{v}, \Pi_C^* \mathbf{w}) = g_{p(a)}(\mathbf{v}, \mathbf{w}) \quad (3.7)$$

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を、曲線 C に沿って微分することにより、

$$\frac{d}{dt}g(Y, Z) = g(\nabla_{\frac{d}{dt}} Y, Z) + g(Y, \nabla_{\frac{d}{dt}}^* Z) = 0 \quad (3.8)$$

を得る。よって、 $g(Y, Z)$ は C に沿って平行移動しても不変である。 \square

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。

Prf

恒等式

$$[X, Y]g(Z, W) = XYg(Z, W) - YXg(Z, W) \quad (3.9)$$

を考える。左辺について、

$$[X, Y]g(Z, W) = g(\nabla_{[X, Y]} Z, W) + g(Z, \nabla_{[X, Y]}^* W) \quad (3.10)$$

$$(3.11)$$

となる。一方、右辺について、

$$XYg(Z, W) - YXg(Z, W) = g(\nabla_X \nabla_Y Z, W) + g(\nabla_Y Z, \nabla_X^* W) \quad (3.12)$$

$$+ g(\nabla_X Z, \nabla_Y^* W) + g(Z, \nabla_X^* \nabla_Y^* W) \quad (3.13)$$

$$- g(\nabla_Y \nabla_X Z, W) - g(\nabla_X Z, \nabla_Y^* W) \quad (3.14)$$

$$- g(\nabla_Y Z, \nabla_X^* W) - g(Z, \nabla_Y^* \nabla_X^* W) \quad (3.15)$$

$$= g([\nabla_X, \nabla_Y] Z, W) + g(Z, [\nabla_X^*, \nabla_Y^*] W) \quad (3.16)$$

となる。したがって、

$$g(\nabla_{[X, Y]} Z, W) + g(Z, \nabla_{[X, Y]}^* W) = g([\nabla_X, \nabla_Y] Z, W) + g(Z, [\nabla_X^*, \nabla_Y^*] W) \quad (3.17)$$

となる。曲率の定義が $R(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X, Y]} Z$ であることを用いると、

$$g(R(X, Y, Z), W) = -g(Z, R(X, Y, W)^*) \quad (3.18)$$

となる。したがって、 ∇ -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。 \square

振率については、曲率のような関係が成り立たないことが知られている。

Thm: 双対接続と振率

3.2 双対平坦な多様体の幾何

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、 ∇ -振率と ∇^* -振率がどちらも 0 であることをいう。

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

$$g(\partial_i, \partial_j) = \delta_{ij} \quad (3.19)$$

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。

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、

$$(G_0)_{ij} = g_{p_0} \left(\left(\frac{\partial}{\partial \xi^i} \right)_{p_0}, \left(\frac{\partial}{\partial \eta^j} \right)_{p_0} \right) \quad (3.20)$$

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、

$$x^i = \xi^i, \quad y^i = \sum_j (G_0)_{ij} \eta^j \quad (3.21)$$

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ィン座標系である。また、

$$\frac{\partial}{\partial \xi^i} = \frac{\partial}{\partial x^i} \quad (3.22)$$

$$\frac{\partial}{\partial \eta^i} = \sum_j (G_0)_{kj} \frac{\partial}{\partial y^k} \quad (3.23)$$

であるから、

$$(G_0)_{ij} = g_{p_0}((\frac{\partial}{\partial \xi^i})_{p_0}, (\frac{\partial}{\partial \eta^j})_{p_0}) \quad (3.24)$$

$$= \sum_k g_{p_0}((\frac{\partial}{\partial x^i})_{p_0}, \frac{\partial}{\partial y^k})(G_0)_{kj} \quad (3.25)$$

$$(3.26)$$

であるから、

$$g_{p_0}((\frac{\partial}{\partial x^i})_{p_0}, \frac{\partial}{\partial y^k}) = \delta_{ik} \quad (3.27)$$

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

$$Xg(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}) = g(\nabla_X \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}) + g(\frac{\partial}{\partial x^i}, \nabla_X^* \frac{\partial}{\partial x^j}) = 0 \quad (3.28)$$

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

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。 \square

Def: 双対アフィン座標系

$$g(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial y^j}) = \delta_{ij} \quad (3.29)$$

を満たす ∇ -アフィン座標系 (x^i) と ∇^* -アフィン座標系 (y^i) の組 $\{x^i, y^i\}$ を双対アフィン座標系という。

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

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

と書くことにする。また、直交性は

$$g(\partial_i, \partial^j) = \delta_i^j \quad (3.30)$$

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。

Lem1

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

$$g_{ij} = g(\partial_i, \partial_j) \quad g^{ij} = g(\partial^i, \partial^j) \quad (3.31)$$

とおくと、

$$g_{ij} = \partial_i \eta_j = \partial_j \eta_i \quad g^{ij} = \partial^i \theta^j = \partial^j \theta^i \quad g_{ij} g^{jk} = \delta_i^k \quad (3.32)$$

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、

$$g_{ij} = g(\partial_i, \partial_j) \quad (3.33)$$

$$= g\left(\frac{\partial \eta_k}{\partial \theta^i} \partial^k, \partial_j\right) \quad (3.34)$$

$$= \frac{\partial \eta_k}{\partial \theta^i} g(\partial^k, \partial_j) \quad (3.35)$$

$$= \frac{\partial \eta_k}{\partial \theta^i} \delta_j^k \quad (3.36)$$

$$= \frac{\partial \eta_j}{\partial \theta^i} \quad (3.37)$$

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、

$$g^{ij} = \frac{\partial \theta^i}{\partial \eta^j} \quad (3.38)$$

が成立する。これらを合わせて、

$$g_{ij} g^{jk} = \frac{\partial \eta_i}{\partial \theta^j} \frac{\partial \theta^j}{\partial \eta^k} = \delta_i^k \quad (3.39)$$

もいうことができる。 □

Lem2

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

$$\eta_i = \partial_i \psi \quad \theta^i = \partial^i \varphi \quad \psi(\theta^1, \dots, \theta^n) + \varphi(\eta_1, \dots, \eta_n) - \theta^i \eta_i = 0 \quad (3.40)$$

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、

$$d(\psi + \varphi - \theta^i \eta_i) = d\psi + d\varphi - (d\theta^i) \eta_i - \theta^i (d\eta_i) \quad (3.41)$$

$$= (\partial_i \psi) d\theta^i + (\partial^i \varphi) d\eta_i - \eta_i d\theta^i - \theta^i d\eta_i \quad (3.42)$$

$$= 0 \quad (3.43)$$

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。

□

Lem3

一つ前の補題で定めた関数の組は、計量 g と、

$$g_{ij} = \partial_i \partial_j \psi \quad g^{ij} = \partial^i \partial^j \varphi \quad (3.44)$$

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。

□

Lem4

点 $p \in M$ の θ -座標と、 η -座標をそれぞれ、

$$\theta(p) = (\theta^1(p), \dots, \theta^n(p)) \quad \eta(p) = (\eta_1(p), \dots, \eta_n(p)) \quad (3.45)$$

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、

$$\varphi(\eta(p)) = \max_{q \in M} \{ \theta^i(q) \eta_i(p) - \psi(\theta(q)) \} \quad (3.46)$$

$$\psi(\theta(p)) = \max_{q \in M} \{ \eta_i(q) \theta^i(p) - \varphi(\eta(q)) \} \quad (3.47)$$

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、

$$\begin{aligned} d(\theta^i(q) \eta_i(p) - \psi(\theta(q))) &= (\eta_i(p) - \partial_i \psi(\theta(q))) d\theta^i(q) \\ &= (\eta_i(p) - \eta_i(q)) d\theta^i(q) \end{aligned} \quad (3.48)$$

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

$$\theta^i(p) \eta_i(p) - \psi(\theta(p)) = \varphi(\eta(p)) \quad (3.49)$$

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

$$D(p||q) = \psi(\theta(p)) + \varphi(\eta(q)) - \theta^i(p)\eta_i(q) \quad (3.50)$$

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。

Lem:

$\{(\theta^i), (\eta_i)\}$ と $\{(\tilde{\theta}^i), (\tilde{\eta}_i)\}$ を M 上の双対アフィン座標系であるとし、それぞれのポテンシャルを $\{\psi, \varphi\}$ と $\{\tilde{\psi}, \tilde{\varphi}\}$ とする。このとき、

$$\psi(\theta(p)) + \varphi(\eta(q)) - \theta^i(p)\eta_i(q) = \tilde{\psi}(\tilde{\theta}(p)) + \tilde{\varphi}(\tilde{\eta}(q)) - \tilde{\theta}^i(p)\tilde{\eta}_i(q) \quad (3.51)$$

が成り立つ。

Prf

二組の双対アフィン座標系は、

$$\tilde{\theta}^\lambda = A^\lambda_i \theta^i + a^\lambda \quad (3.52)$$

$$\tilde{\eta}_\lambda = B^i_\lambda \eta_i + b_\lambda \quad (3.53)$$

により関係づけられる。ただし、この時直交条件より、

$$\delta^j_i = g(\partial_i, \partial^j) = A^\lambda_i B^j_\mu g(\partial_\lambda, \partial^\mu) = A^\lambda_i B^j_\lambda \quad (3.54)$$

が成り立つ。したがって、 A, B は互いに逆行列である。したがって、

$$\partial_\lambda = B^i_\lambda \partial_i \quad (3.55)$$

$$\partial^\lambda = A^\lambda_i \partial^i \quad (3.56)$$

が成り立つ。ダイバージェンスの前の補題から、

$$\eta_i = \partial_i \psi \quad (3.57)$$

$$(3.58)$$

であることを用いると、

$$\tilde{\eta}_\lambda = B^i_\lambda \partial_i \psi \quad (3.59)$$

となる。一方、上のアフィン変換を用いると、

$$\tilde{\eta}_\lambda = B^i_\lambda \eta_i + b_\lambda = B^i_\lambda \partial_i \psi + b_\lambda \quad (3.60)$$

となる。したがって、

$$\partial_i \tilde{\psi} = \partial_i \psi + A_j^\lambda b_\lambda \quad (3.61)$$

となる。これを積分して、

$$\tilde{\psi} = \psi + A_j^\lambda b_\lambda \theta^j + C \quad (3.62)$$

となる。次に、Lem2 の三本目の式を用いると、

$$\tilde{\varphi}(\tilde{\eta}) = \tilde{\theta}^\lambda \tilde{\eta}_\lambda - \tilde{\psi}(\tilde{\theta}) \quad (3.63)$$

$$= (A_i^\lambda \theta^i + a^\lambda)(B_\lambda^i \partial_i \psi + b_\lambda) - (\psi + A_j^\lambda b_\lambda \theta^j + C) \quad (3.64)$$

$$= \{\theta^i \eta_i - \psi(\theta)\} + a^\lambda B_\lambda^j \eta_j + a^\lambda b_\lambda - C \quad (3.65)$$

$$= \varphi(\eta) + a^\lambda B_\lambda^j \eta_j + a^\lambda b_\lambda - C \quad (3.66)$$

となる。したがって、

$$\tilde{\psi}(\tilde{\theta}) + \tilde{\varphi}(\tilde{\eta}) - \tilde{\theta}^i \tilde{\eta}_i \quad (3.67)$$

$$= \{\psi(\theta(p)) + A_j^\lambda b_\lambda \theta^j + C\} + \{\varphi(\eta(q)) + a^\lambda B_\lambda^j \eta_j + a^\lambda b_\lambda - C\} \quad (3.68)$$

$$= \psi(\theta(p)) + \varphi(\eta(q)) - \theta^i(p) \eta_i(q) \quad (3.69)$$

となる。したがって、示された。 \square

∇ 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、

$$D(p||q) = D^*(q||p) \quad (3.70)$$

が成り立つ。また、Lem3 の二本目の式と比較することで、

$$D(p||q) \geq 0 \quad (3.71)$$

かつ、

$$D(p||q) = 0 \Leftrightarrow p = q \quad (3.72)$$

が成り立つ。

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、

$$\psi(z) = \varphi(z) = \frac{1}{2} \sum_{i=1}^n (z^i)^2 \quad (3.73)$$

となる。したがって、ダイバージェンスは、

$$D(p||q) = \frac{1}{2} \sum_{i=1}^n (z^i(p))^2 + \frac{1}{2} \sum_{i=1}^n (z^i(q))^2 - \sum_{i=1}^n z^i(p) z^i(q) = \frac{1}{2} \sum_{i=1}^n (z^i(p) - z^i(q))^2 \quad (3.74)$$

となる。

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、

$$D(p||r) = D(p||q) + D(q||r) \quad (3.75)$$

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

$$C_1 = \{p(t); t \in [0, 1] \mid p(0) = q, p(1) = p\} \quad (3.76)$$

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

$$\theta^i(t) = \theta^i(0) + t(\theta^i(1) - \theta^i(0)) \quad (3.77)$$

となる。同様に、 q, r を結ぶ ∇^* -測地線

$$C_2 = \{q(t); t \in [0, 1] \mid q(0) = q, q(1) = r\} \quad (3.78)$$

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

$$\eta_i(t) = \eta_i(0) + t(\eta_i(1) - \eta_i(0)) \quad (3.79)$$

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、

$$\mathbf{v} = \dot{p}(0) = \left(\frac{d\theta^i}{dt}\right)(\partial_i)_q = (\theta^i(1) - \theta^i(0))(\partial_i)_q = (\theta^i(p) - \theta^i(q))(\partial_i)_q \quad (3.80)$$

$$\mathbf{w} = \dot{q}(0) = \left(\frac{d\eta_i}{dt}\right)(\partial^i)_q = (\eta^i(1) - \eta^i(0))(\partial^i)_q = (\eta^i(r) - \eta^i(q))(\partial^i)_q \quad (3.81)$$

となる。これらが直交しているという仮定から、

$$0 = g(\mathbf{v}, \mathbf{w}) \quad (3.82)$$

$$= g((\theta^i(p) - \theta^i(q))(\partial_i)_q, (\eta^j(r) - \eta^j(q))(\partial^j)_q) \quad (3.83)$$

$$= (\theta^i(p) - \theta^i(q))(\eta^j(r) - \eta^j(q))g(\partial_i, \partial^j) \quad (3.84)$$

$$= (\theta^i(p) - \theta^i(q))(\eta^i(r) - \eta^i(q)) \quad (3.85)$$

となる。したがって、

$$D(p||q) + D(q||r) - D(p||r) = \psi(\theta(p)) + \varphi(\eta(q)) - \theta^i(p)\eta_i(q) \quad (3.86)$$

$$+ \psi(\theta(q)) + \varphi(\eta(r)) - \theta^i(q)\eta_i(r) \quad (3.87)$$

$$- \psi(\theta(p)) - \varphi(\eta(r)) + \theta^i(p)\eta_i(r) \quad (3.88)$$

$$= \psi(\theta(q)) + \varphi(\eta(q)) - \theta^i(p)\eta_i(q) - \theta^i(q)\eta_i(r) + \theta^i(p)\eta_i(r) \quad (3.89)$$

$$= \theta^i(q)\eta_i(q) - \theta^i(p)\eta_i(q) - \theta^i(q)\eta_i(r) + \theta^i(p)\eta_i(r) \quad (3.90)$$

$$\because \text{lem2} \quad (3.91)$$

$$= (\theta^i(p) - \theta^i(q))(\eta_i(r) - \eta_i(q)) \quad (3.92)$$

$$= 0 \quad (3.93)$$

となる。したがって、示された。

□

4 確率分布空間の幾何構造

4.1 Chentsov の定理

有限集合 Ω の要素である根元事象を自然数でラベル付けすることにして、サイズ n の根元事象系を

$$\Omega_n = \{1, 2, \dots, n\} \quad (4.1)$$

と表すことにする。また、 Ω_n 上の確率分布全体の集合を、

$$\mathcal{S}_n = \{p : \Omega_n \rightarrow \mathbb{R}_{++}; \sum_{\omega \in \Omega_n} p(\omega) = 1\} \quad (4.2)$$

と表す。ただし、

$$\mathbb{R}_{++} = \{x \in \mathbb{R}; x > 0\} \quad (4.3)$$

である。 \mathcal{S}_n の各確率分布を、

$$p = (p(1), p(2), \dots, p(n)) \quad (4.4)$$

と同一視すると、 \mathcal{S}_n は、 $n - 1$ 次元単体であり、 \mathbb{R}^n 内の超平面

$$H = \{x \in \mathbb{R}^n; \sum_{i=1}^n x_i = 1\} \quad (4.5)$$

と同一視される。

以下、統計多様体を定義する。このとき、統計多様体にかかる条件を考える。

まず、事象のラベル付けは順序付けに依存しない。また、いくつかのラベルを同一視することを考えると、より高次元の統計多様体から低次元の統計多様体が誘導されるはずである。

これらを踏まえて、Markov 埋め込みを考える。

Def:Markov 埋め込み

$2 \leq n \leq \text{lin}\mathbb{N}$ に対して、以下の用の構成される写像

$$f : \mathcal{S}_{n-1} \rightarrow \mathcal{S}_{l-1} \quad (4.6)$$

を Markov 埋め込みという。

1. Ω_l を空でなく互いに交わらない部分集合の族 $\{C_1, C_2, \dots, C_n\}$ に分割する。
2. 各 $j(1 \leq j \leq n)$ に対し、 C_j に台を持つ Ω_l 上の確率分布

$$Q_j = (Q_{(j)}^1, Q_{(j)}^2, \dots, Q_{(j)}^l) \quad (4.7)$$

を付随させる。ただし、 $Q_{(j)}^k$ は、 $k \in C_{(j)}$ ならば正であり、 $k \notin C_{(j)}$ ならば 0 である。

3. $(y^1, y^2, \dots, y^l) = f(x^1, x^2, \dots, x^n)$ を、

$$y^k = \sum_{j=1}^n x^j Q_{(j)}^k \quad (4.8)$$

により定義する。行列表記すると、

$$\begin{pmatrix} y^1 \\ y^2 \\ \vdots \\ y^l \end{pmatrix} = \begin{pmatrix} Q_{(1)}^1 & Q_{(2)}^1 & \cdots & Q_{(n)}^1 \\ Q_{(1)}^2 & Q_{(2)}^2 & \cdots & Q_{(n)}^2 \\ \vdots & \vdots & \ddots & \vdots \\ Q_{(1)}^l & Q_{(2)}^l & \cdots & Q_{(n)}^l \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \\ \vdots \\ x^n \end{pmatrix} \quad (4.9)$$

となる。

要請

\mathcal{S}_{n-1} の幾何は、 \mathcal{S}_{l-1} に埋め込まれた部分多様体 $f(\mathcal{S}_{n-1})$ の幾何と同等であるべきである。

Thm:Chentsov の定理

\mathcal{S}_{n-1} 上の $(0, 2)$ テンソル場からなる列 $\{g^{(n)}\}$ が、任意の Markov 埋込 f に対する不変性:

$$g_p^{(n)}(X, Y) = g_{f(p)}^{(n)}(f_*X, f_*Y) \quad (4.10)$$

を満たすものは、定数倍を除いて、

$$g_p^{(n)}(X, Y) = \sum_{\omega=1}^n p(\omega)(X \log p(\omega))(Y \log p(\omega)) \quad (4.11)$$

に限られる。

Prf

4 段階に分けて証明する。

Step1:

\mathcal{S}_{n-1} の重心

$$p_0 = \left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n} \right) \quad (4.12)$$

について考える。^{*2}このとき、事象のラベル付けの不変性から、

$$g_{p_0}^{(n)}\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^i}\right) \quad (4.13)$$

は i に依存せず、また、

$$g_{p_0}^{(n)}\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) \quad (4.14)$$

は i, j に依存しない。したがって、

$$g_{p_0}^{(n)}\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) = \delta_{ij} A^{(n)} + B^{(n)} \quad (4.15)$$

と書くことができる、ところで、点 p_0 における \mathcal{S}_{n-1} の接ベクトルを、

$$X = X^i \frac{\partial}{\partial x^i} \quad (4.16)$$

と表すと、

$$\sum_{i=1}^n X^i = 0 \quad (4.17)$$

が成り立つ。というのも、 \mathcal{S}_{n-1} 上の関数 $h(x) = x^1 + x^2 + \dots + x^n = 1$ に対して、

$$0 = Xh = \sum_{i=1}^n X^i \quad (4.18)$$

が成り立つからである。したがって、

$$g_{p_0}^{(n)}(X, X) = \sum_{i,j=1}^n X^i X^j (\delta_{ij} A^{(n)} + B^{(n)}) = A^{(n)} \sum_{i=1}^n X^i Y^i \quad (4.19)$$

となって、 \mathcal{S}_{n-1} 上では B に依存しない。よって、一般性を失わずに、 $B^{(n)} = 0$ としてよい。

Step2:

ある自然数 K が存在して、 $l = kn$ となっているとき、

$$f(x^1, x^2, \dots, x^n) = \left(\frac{x^1}{k}, \dots, \frac{x^1}{k}, \frac{x^2}{k}, \dots, \frac{x^2}{k}, \dots, \frac{x^n}{k}, \dots, \frac{x^n}{k} \right) \quad (4.20)$$

$$= (y^{1_1}, y^{1_2}, \dots, y^{1_k}, y^{2_1}, y^{2_2}, \dots, y^{2_k}, \dots, y^{n_1}, y^{n_2}, \dots, y^{n_k}) \quad (4.21)$$

^{*2} 事象のラベル付け不変性を見るため?わからんくなってきた

という Markov 埋め込みを考える。このとき、

$$f_* : \frac{\partial}{\partial x^i} \mapsto \sum_{r=1}^k \frac{\partial}{\partial y^{i_r}} \quad (4.22)$$

である。このとき、代入すればわかるように、 \mathcal{S}_{n-1} の重心 p_0 の像 $f(p_0)$ は、 \mathcal{S}_{l-1} の重心である。したがって、Step1 の結果から、

$$A^{(n)} = g_{p_0}^{(n)} \left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^i} \right) \quad (4.23)$$

$$= g_{f(p_0)}^{(l)} \left(f_* \frac{\partial}{\partial x^i}, f_* \frac{\partial}{\partial x^i} \right) \quad (4.24)$$

$$= g_{f(p_0)}^{(l)} \left(\sum_{r=1}^k \frac{\partial}{\partial y^{i_r}}, \sum_{s=1}^k \frac{\partial}{\partial y^{i_s}} \right) \quad (4.25)$$

$$= \frac{1}{k^2} \sum_{r,s=1}^k g_{f(p_0)}^{(l)} \left(\frac{\partial}{\partial y^{i_r}}, \frac{\partial}{\partial y^{i_s}} \right) \quad (4.26)$$

$$= \frac{1}{k^2} \sum_{r,s=1}^k \delta_{rs} A^{(l)} \quad (4.27)$$

$$= \frac{1}{k} A^{(l)} \quad (4.28)$$

となる。したがって、

$$\frac{A^{(n)}}{n} = \frac{A^{(l)}}{l} \quad (4.29)$$

となり、ある定数 λ が存在して、

$$A^{(n)} = \lambda n \quad (4.30)$$

となる。

Step3:

4.2 Fisher 計量と α -接続

Markov 埋め込み $f : \mathcal{S}_{n-1} \rightarrow \mathcal{S}_{l-1}$ のもとでの \mathcal{S}_{n-1} の計量 $g^{(n)}$ に対する不変性の要請は、

$$g_p^{(n)}(X, Y) = g_{f(p)}^{(l)}(f_* X, f_* Y) \quad (4.31)$$

であり、 \mathcal{S}_{n-1} 上のアフィン接続 $\nabla^{(n)}$ に対する不変性の要請は、

$$g_p^{(n)}(\nabla_X^{(n)} Y, Z) = g_{f(p)}^{(l)}(\nabla_{f_* X}^{(l)} f_* Y, f_* Z) \quad (4.32)$$

である。ただし、 f_* は、多様体間の微分写像である。

このとき、4.31 が成り立つならば 4.32 も成り立つ。

∴

また、 \mathcal{S}_{n-1} 上に許容される計量は他にもある。というのも、共変微分の差が $(1, 2)$ テンソル場と一対一に対応すること：

$$S(X, Y) = \nabla_X Y - \tilde{\nabla}_X Y \quad (4.33)$$

が成り立つため、これを計量を用いて射影した、

$$S(X, Y, Z) = g((\nabla_X Y - \tilde{\nabla}_X Y), Z) \quad (4.34)$$

は、 $(0, 3)$ テンソル場に一対一対応する。しかも、Chentsov の定理から、不変性を満たすテンソル場は、定数倍を除いて定まっている。以上をまとめて、以下の定理を得る。

Thm:Chentsov の定理

Markov 埋め込みの下での不変性を満たす Riemann 計量は、定数倍を除いて、

$$g_p(X, Y) = \sum_{\omega=1}^n p(\omega)(X \log p(\omega))(Y \log p(\omega)) \quad (4.35)$$

に限られる。一方、アフィン接続の不変性を満たすようなアフィン接続は、

$$g(\nabla_X^\alpha Y, Z) = g(\tilde{\nabla}_X^\alpha Y, Z) - \frac{\alpha}{2} S_p(X, Y, Z) \quad (4.36)$$

により、実数 α と一対一に対応する。ただし、

$$S_p(X, Y, Z) = \sum_{\omega=1}^n p(\omega)(X \log p(\omega))(Y \log p(\omega))(Z \log p(\omega)) \quad (4.37)$$

である。

Def:Fisher 計量

4.35 を満たす計量 g を Fisher 計量といい、4.36 を満たすアフィン接続 ∇^α を α -接続という。

4.3 指数型接続と混合型接続

Thm:

有限確率分布空間 \mathcal{S}_{n-1} 上の Fisher 計量を g 、 α -接続を ∇^α とする。このとき、以下が成り立つ。

1. 各 $\alpha \in \mathbb{R}$ に対して、 ∇^α の振率は 0 である。
2. 各 $\alpha \in \mathbb{R}$ に対して、 ∇^α と $\nabla^{-\alpha}$ は g に関して互いに双対である。
3. \mathcal{S}_{n-1} は双対構造 $(g, \nabla^{+1}, \nabla^{-1})$ に関して双対平坦である。

Prf

1.

∇^0 は Riemann 接続であるから、振率は 0 である。これを用いると、

$$g(\nabla_X^\alpha Y - \nabla_X^{-\alpha} Y - [X, Y], Z) = g(\nabla_X^0 Y - \nabla_X^0 Y - [X, Y], Z) - \frac{\alpha}{2}(S(X, Y, Z) - S(Y, X, Z)) \quad (4.38)$$

$$= 0 \quad \because S \text{ の対称性} \quad (4.39)$$

となる。したがって、振率は 0 である。

2.

$$Xg(Y, Z) = g(\nabla_X^0 Y, Z) + g(Y, \nabla_X^0 Z) \quad \because \text{Riemann 接続} \quad (4.40)$$

$$= g(\nabla_X^\alpha Y, Z) + g(Y, \nabla_X^{-\alpha} Z) + \frac{\alpha}{2}(S(X, Y, Z) - S(X, Z, Y)) \quad (4.41)$$

$$= g(\nabla_X^\alpha Y, Z) + g(Y, \nabla_X^{-\alpha} Z) \quad \because S \text{ の対称性} \quad (4.42)$$

となる。したがって、 ∇^α と $\nabla^{-\alpha}$ は g に関して互いに双対である。

3.

各 α について振率が 0 であることは述べたから、あとは曲率が 0 であることを示せばよい。以下、 \mathcal{S}_{n-1} が ∇^{-1} -アフィン座標系を持つことを示す。というのも、これが示されれば、 \mathcal{S}_{n-1} が ∇^{-1} -平坦であることが示されるため、この接続について曲率が 0 となり、なおかつそれに双対な接続についても曲率が 0 となるからである。

$$p(\omega) = \sum_{a=1}^{n-1} \xi^a \delta_a(\omega) + \left(1 - \sum_{a=1}^{n-1} \xi^a\right) \delta_n(\omega) \quad (4.43)$$

により定義される座標系が、実は ∇^{-1} -アフィン座標系であることを示す。

計算により、

$$\Gamma_{i,j,k}^\alpha = g(\nabla_{\partial_i}^\alpha \partial_j, \partial_k) \quad (4.44)$$

$$= \frac{1-\alpha}{2} S_{ijk} + \sum_{\omega=1}^n p(\omega) (\partial_i \partial_j \log p(\omega)) (\partial_k \log p(\omega)) \quad (4.45)$$

であることがわかる。ただし、

$$S_{ijk} = \sum_{\omega=1}^n p(\omega) (\partial_i \log p(\omega)) (\partial_j \log p(\omega)) (\partial_k \log p(\omega)) \quad (4.46)$$

である。とくに、 $\alpha = -1$ のとき、

$$\Gamma_{i,j,k}^{-1} = \sum_{\omega=1}^n p(\omega) ((\partial_i \log p(\omega)) (\partial_j \log p(\omega)) + \partial_i \partial_j \log p(\omega)) (\partial_k \log p(\omega)) \quad (4.47)$$

$$= \sum_{\omega=1}^n p(\omega) \left(\frac{p(\omega)^2}{(\partial_i p(\omega)) (\partial_j p(\omega))} + \partial_i \left(\frac{\partial_j p(\omega)}{p(\omega)} \right) \right) \partial_k \log p(\omega) \quad (4.48)$$

$$= \sum_{\omega=1}^n (\partial_i \partial_j p(\omega)) (\partial_k \log p(\omega)) \quad \because \text{積 (商) の微分をばらしている} \quad (4.49)$$

となる。このとき、

$$p(\omega) = \sum_{a=1}^{n-1} \xi^a \delta_a(\omega) + \left(1 - \sum_{a=1}^{n-1} \xi^a \right) \delta_n(\omega) \quad (4.50)$$

を見ると、 $\partial_a \partial_b p(\omega) = 0$ であるから、

$$\Gamma_{a,b,c}^{-1} = 0 \quad (4.51)$$

であることがわかる。 \square

Def: 指数型接続/混合型接続

∇^{-1} を混合型接続といい、 ∇^m と書く。また、 ∇^{+1} を指数型接続といい、 ∇^e と書く。

Def:m-表現/e-表現

各 $X \in T_p \mathcal{S}$ に対して、

$$Xp = (Xp(1), Xp(2), \dots, Xp(n)) \quad (4.52)$$

で定まる確率変数 Xp を、 X の m-表現という。また、その全体を、接空間 $T_p \mathcal{S}$ の m-表現空間といい、 $T_p \mathcal{S}^m$ と書く。

同様に、各 $X \in T_p \mathcal{S}$ に対して、

$$X \log p = (X \log p(1), X \log p(2), \dots, X \log p(n)) \quad (4.53)$$

で定まる確率変数 $X \log p$ を、 X の e-表現という。また、その全体を、接空間 $T_p \mathcal{S}$ の e-表現空間といい、 $T_p \mathcal{S}^e$ と書く。

Thm:m-表現、e-表現の特徴づけ

接空間 $T_p\mathcal{S}$ の m-表現、e-表現空間は、それぞれ、以下のように特徴づけられる。

$$T_p\mathcal{S}^m = \{f : \Omega \rightarrow \mathbb{R}; \sum_{\omega \in \Omega} f(\omega) = 0\} \quad (4.54)$$

$$T_p\mathcal{S}^e = \{f : \Omega \rightarrow \mathbb{R}; \sum_{\omega \in \Omega} f(\omega) = 0, \sum_{\omega \in \Omega} f(\omega)^2 = 1\} \quad (4.55)$$

上の表式を見ると、m-表現においては、 $f(\omega)$ 0 次のモーメントが 0 であることがわかる。また、e-表現においては、 $f(\omega)$ の 1 次のモーメントが 0 であることがわかる。

Prf

m-表現

(左辺) \subset (右辺) は、

$$\sum_{\omega \in \Omega} Xp(\omega)X \left(\sum_{\omega \in \Omega} p(\omega) \right) = X1 = 0 \quad (4.56)$$

であるから、成り立つ。

また、両辺共に、 $n - 1$ 次元の線形空間であるから、線形代数の話を思い出すと、(右辺) = (左辺) である。

e-表現

(左辺) \subset (右辺) は、

$$\sum_{\omega \in \Omega} p(\omega)X \log p(\omega) = \sum_{\omega \in \Omega} p(\omega) \left(\frac{Xp(\omega)}{p(\omega)} \right) = X1 = 0 \quad (4.57)$$

であるから、成り立つ。

また、両辺共に、 $n - 1$ 次元の線形空間であるから、線形代数の話を思い出すと、(右辺) = (左辺) である。

□

Rem

Fisher 計量は、接ベクトルの e-表現と m-表現によって、

$$g(X, Y) = \sum_{\omega \in \Omega} (Xp(\omega))(Y \log p(\omega)) \quad (4.58)$$

と表すことができる。また、接続 ∇^m, ∇^e は、それぞれ、

$$g(\nabla_X^m Y, Z) = \sum_{\omega \in \Omega} (XY \log p(\omega))(Zp(\omega)) \quad (4.59)$$

$$g(\nabla_X^e Y, Z) = \sum_{\omega \in \Omega} (XY p(\omega))(Z \log p(\omega)) \quad (4.60)$$

と表すことができる。

∴

$$g_p(X, Y) = \sum_{\omega \in \Omega} p(\omega) (X \log p(\omega)) (Y \log p(\omega)) \quad (4.61)$$

$$= \sum_{\omega \in \Omega} p(\omega) \frac{X p(\omega)}{p(\omega)} \frac{Y p(\omega)}{p(\omega)} \quad (4.62)$$

$$= \sum_{\omega \in \Omega} (X p(\omega)) (Y p(\omega)) \quad (4.63)$$

であるから、計量について示される。また、接続については、

Def:m-平行移動/e-平行移動

二点 $p, q \in \mathcal{S}$ に対し、

$$\Pi^m : T_p^m \mathcal{S} \rightarrow T_q^m \mathcal{S} : f \mapsto f \quad (4.64)$$

で定まる遠隔平行移動を m-平行移動という。また、

$$\Pi^e : T_p^e \mathcal{S} \rightarrow T_q^e \mathcal{S} : f \mapsto f - E_q[f] \quad (4.65)$$

で定まる遠隔平行移動を e-平行移動という。

平行移動前後で、上で見た、m-表現、e-表現の特徴づけが保たれることがわかる。

実は上で定義した平行移動の微小変化が、共変微分に対応することが以下でわかる。

Thm:

$t = 0$ で $p_0 \in \mathcal{S}$ を通る $X \in \mathcal{X}(\mathcal{S})$ の積分曲線を p_t とする。すなわち、 $\dot{p}_t = X(p_t)$ である。任意のベクトル場 $Y \in \mathcal{X}(\mathcal{S})$ に対して、

$$(\nabla_X^m Y)_{p_0} = \lim_{t \rightarrow 0} \frac{\Pi_0^{(m)t} Y_{p_t} - Y_{p_0}}{t} \quad (4.66)$$

$$(\nabla_X^e Y)_{p_0} = \lim_{t \rightarrow 0} \frac{\Pi_0^{(e)t} Y_{p_t} - Y_{p_0}}{t} \quad (4.67)$$

が成り立つ。

Prf

まず、m-表現について考える。証明の方針としては、平行移動を無限小にしたものが、(4.61) をみたすことを示す。

$$\lim_{t \rightarrow 0} \sum_{\omega \in \Omega} \left(\frac{\Pi_0^{(m)t}(Y_{p_t} p_t(\omega)) - Y_{p_0} p_0(\omega)}{t} \right) (Z_{p_0} \log p_0(\omega)) \quad (4.68)$$

$$= \lim_{t \rightarrow 0} \sum_{\omega \in \Omega} \left(\frac{Y_{p_t} p_t(\omega) - Y_{p_0} p_0(\omega)}{t} \right) (Z_{p_0} \log p_0(\omega)) \quad \because m\text{-平行移動の定義} \quad (4.69)$$

$$= \sum_{\omega \in \Omega} \frac{dY_{p_t} p_t(\omega)}{dt} \Big|_{p=p_0} \quad (4.70)$$

$$= \sum_{\omega \in \Omega} \frac{dx_i}{dt} \frac{\partial Y_{p_t} p_t(\omega)}{\partial x_i} \Big|_{p=p_0} \quad (4.71)$$

$$= \sum_{\omega \in \Omega} v^i \frac{\partial Y_{p_t} p_t(\omega)}{\partial x_i} \Big|_{p=p_0} \quad \because p_t \text{がベクトル場 } X \text{ に沿って動く} \quad (4.72)$$

$$= \sum_{\omega \in \Omega} (XY p(\omega))(Z \log p(\omega)) \Big|_{p=p_0} \quad (4.73)$$

$$= g(\nabla_X^m Y, Z)_{p_0} \because (4.61) \quad (4.74)$$

となる。したがって、示された。

次に、e-表現について考える。

$$\lim_{t \rightarrow 0} \sum_{\omega \in \Omega} \left(\frac{\Pi_0^{(e)t}(Y_{p_t} \log p_t(\omega)) - Y_{p_0} \log p_0(\omega)}{t} \right) (Z_{p_0} p_0(\omega)) \quad (4.75)$$

$$= \lim_{t \rightarrow 0} \sum_{\omega \in \Omega} \left(\frac{Y_{p_t} \log p_t(\omega) - E_{p_0}[Y_{p_t} \log p_t(\omega)] - Y_{p_0} \log p_0(\omega)}{t} \right) (Z_{p_0} p_0(\omega)) \quad (4.76)$$

$$\because e\text{-平行移動の定義} \quad (4.77)$$

$$= \sum_{\omega \in \Omega} \left(\frac{Y_{p_t} \log p_t(\omega) - Y_{p_0} \log p_0(\omega)}{t} \right) (Z_{p_0} p_0(\omega)) \Big|_{p=p_0} \quad (4.78)$$

$$= \sum_{\omega \in \Omega} (XY \log p(\omega))(Z p(\omega)) \Big|_{p=p_0} \quad \because p_t \text{がベクトル場 } X \text{ に沿って動く} \quad (4.79)$$

$$= g(\nabla_X^e Y, Z)_{p_0} \because (4.61) \quad (4.80)$$

となる。したがって、示された。 \square

4.4 \mathcal{S}_{n-1} の双対微分幾何

Def: 統計多様体

4 つ組 $(\mathcal{S}, g, \nabla^m, \nabla^e)$ を統計多様体という。ただし、 \mathcal{S} は有限確率分布空間、 g は Fisher 計量、 ∇^m は m-接続、 ∇^e は e-接続である。

以下、便宜上のため、 ∇^e を ∇ 、 ∇^m を ∇^* と書くことにする。

双対アフィン座標系を考える。前に、

$$p(\omega) = \sum_{i=1}^{n-1} \eta^i \delta_i(\omega) + (1 - \sum_{i=1}^{n-1} \eta^i) \delta_n(\omega) \quad (4.81)$$

により定義される座標系が、実は ∇^m -アフィン座標系であることを示した。この座標系に対しては、 $\partial_i \partial_j p(\omega) = 0$ となるので、 ∇^m の接続係数

$$\Gamma_{ij,k}^m = \sum_{\omega=1}^n (\partial_i \partial_j p(\omega)) (\partial_k \log p(\omega)) \quad (4.82)$$

がすべて 0 である。

つぎに、 ∇ -アフィン座標系を考える。いま、接続係数は、

$$\Gamma_{ij,k}^e = \sum_{\omega=1}^n (\partial_i \partial_j \log p(\omega)) (\partial_k p(\omega)) \quad (4.83)$$

であった。ここで、

$$\log p(\omega) = \sum_{i=1}^{n-1} \left(\log \frac{p(i)}{p(n)} \right) \delta_i(\omega) + \log p(n) \quad (4.84)$$

である。

$$\theta^i = \log \frac{p(i)}{p(n)} \quad (4.85)$$

とおくと、規格化条件より、

$$p(n) = \frac{1}{1 + \sum_{i=1}^{n-1} \exp(\theta^i)} \quad (4.86)$$

であるから、

$$\log p(\omega) = \sum_{i=1}^{n-1} \theta^i \delta_i(\omega) - \log \left(1 + \sum_{i=1}^{n-1} \exp(\theta^i) \right) \quad (4.87)$$

である。さらに、

$$\psi(\theta) = \log \left[1 + \sum_{i=1}^{n-1} \exp(\theta^i) \right] \quad (4.88)$$

とおくと、

$$\log p(\omega) = \sum_{i=1}^{n-1} \theta^i \delta_i(\omega) - \psi(\theta) \quad (4.89)$$

である。これを用いて接続係数を計算すると、

$$\Gamma_{ij,k}^e = \sum_{\omega=1}^n (\partial_i \partial_j \log p(\omega)) (\partial_k p(\omega)) \quad (4.90)$$

$$= -(\partial_i \partial_j \psi(\theta)) \sum_{\omega=1}^n (\partial_k p(\omega)) \quad (4.91)$$

$$= 0 \quad \because \psi \text{ が } \omega \text{ に依存しないかつ、} \sum_{\omega=1}^n p(\omega) = 1 \quad (4.92)$$

となるから、たしかに θ^i は ∇ -アフィン座標系である。

また、このとき、 θ 座標系と η 座標系は、互いに双対の関係にある。というのも、

$$g \left(\frac{\partial}{\partial \eta^i}, \frac{\partial}{\partial \theta^j} \right) = \sum_{\omega=1}^n \left(\frac{\partial}{\partial \eta^i} p(\omega) \right) \left(\frac{\partial}{\partial \theta^j} \log p(\omega) \right) \quad (4.93)$$

$$= \sum_{\omega=1}^n (\delta_i(\omega) - \delta_n(\omega)) \left(\delta_j(\omega) - \frac{\partial \psi}{\partial \theta^j} \right) \quad (4.94)$$

$$= \sum_{\omega=1}^n (\delta_i(\omega) - \delta_n(\omega)) \delta_j(\omega) \quad (4.95)$$

$$= \delta_{ij} \quad (4.96)$$

したがって、たしかに双対座標系である。

次に、対応する双対ポテンシャルを定める。今、

$$\partial_i \psi = \frac{\exp(\theta^i)}{1 + \sum_{j=1}^{n-1} \exp(\theta^j)} \quad (4.97)$$

$$= p(n) e^{\theta^i} \quad \because (4.86) \quad (4.98)$$

$$= p(i) \quad \because \theta \text{ の定義} \quad (4.99)$$

$$= \eta_i \quad (4.100)$$

となることから、 ψ は θ の双対ポテンシャルである。

また、 η の双対ポテンシャルは、ルジャンドル変換により、

$$\varphi = \theta^i \eta_i - \psi(\theta) \quad (4.101)$$

$$= \sum_{i=1}^{n-1} \left(\log \frac{p(i)}{p(n)} \right) p(i) + \log p(n) \quad (4.102)$$

$$= \sum_{i=1}^{n-1} p(i) \log p(i) + \left(1 - \sum_{i=1}^{n-1} p(i) \right) \log p(n) \quad (4.103)$$

$$= \sum_{i=1}^n \eta_i \log \eta_i + \left(1 - \sum_{i=1}^n \eta_i \right) \log \left(1 - \sum_{i=1}^n \eta_i \right) \quad (4.104)$$

である。これより、

$$\partial^i \varphi = \log \eta_i + 1 - \log \left(1 - \sum_{i=1}^n \eta_i \right) - 1 \quad (4.105)$$

$$= \log \frac{\eta_i}{p(n)} \quad (4.106)$$

$$= \theta^i \quad (4.107)$$

となるから、 φ は η の双対ポテンシャルである。

最後にダイバージェンスを求める。

$$D^m(p||q) = \psi(\theta(q)) + \varphi(\eta(p)) - \theta^i(q)\eta_i(p) \quad (4.108)$$

$$= -\log q(n) + \sum_{i=1}^n p(\omega) \log p(\omega) - \sum_{i=1}^{n-1} \left(\log \frac{q(i)}{q(n)} \right) p(i) \quad (4.109)$$

$$= \sum_{\omega=1}^n p(\omega) \log p(\omega) - \sum_{\omega=1}^{n-1} p(\omega) \log p(\omega) - \left(1 - \sum_{\omega=1}^{n-1} p(\omega) \right) \log q(n) \quad (4.110)$$

$$= \sum_{\omega=1}^n p(\omega) \log \frac{p(\omega)}{q(\omega)} \quad (4.111)$$

となる。このダイバージェンスを KL ダイバージェンスという。

4.5 自己平行部分多様体

これまで、統計多様体 (S, g, ∇, ∇^*) について考えてきた。以下では、その自己平行部分多様体について考える。

Def: 指数型分布族

Ω 上の関数 $C(\omega), F_1(\omega), \dots, F_k(\omega)$ 及び \mathbb{R}^k 上の領域 Θ 上を動く k 次元パラメータ $\theta = (\theta^1, \dots, \theta^k) \in \Theta$ を用いて

$$p_\theta(\omega) = \exp \left[C(\omega) + \sum_{i=1}^k \theta^i F_i(\omega) - \psi(\theta) \right] \quad (4.112)$$

と表される確率分属族 $\{p_{\theta; \theta \in \Theta}\}$ を指数型分布族という。ただし、 $\psi(\theta)$ は、

$$\psi(\theta) = \log \left[\sum_{\omega \in \Omega} \exp \left[C(\omega) + \sum_{i=1}^k \theta^i F_i(\omega) \right] \right] \quad (4.113)$$

で定義される。

Thm: —

\mathcal{S} の部分多様体 M に対し、 M が ∇ -自己平行部分多様体であるための必要十分条件は、 M が指数型分布族であることである。

Prf

Prop: M の双対平坦性 —

M が指数型分布族であるとき、 M は双対平坦である。

Prf

Thm: 双対アフィン座標系 —

\mathcal{S} の部分多様体 M が指数型分布族

$$M = \{p_\theta(\omega) \in \mathcal{S}; \log p_\theta(\omega) = C(\omega) + \sum_{i=1}^k \theta^i F_i(\omega) - \psi(\theta)\} \quad (4.114)$$

に対し、

$$\eta = E_{p_\theta}[F_i] = \sum_{\omega \in \Omega} p_\theta(\omega) F_i(\omega) \quad (4.115)$$

とおくと、 η は ∇ -双対アフィン座標系である。

Prf

写像 $\theta \mapsto \eta$ の Jacobi 行列を計算すると、

$$\frac{\partial \eta^i}{\partial \theta^j} = \sum_{\omega \in \Omega} (\partial_j p_\theta(\omega)) F_i(\omega) \quad (4.116)$$

$$= \sum_{\omega \in \Omega} (\partial_j p_\theta(\omega)) (\partial_i \log p_\theta(\omega) + \partial_i \psi(\theta)) \quad (4.117)$$

$$= \sum_{\omega \in \Omega} (\partial_j p_\theta(\omega)) (\partial_i \log p_\theta(\omega)) \quad (4.118)$$

$$= \tilde{g}_{ij} \quad (4.119)$$

となる。ここで、これは正定値行列であるから、逆写像定理によって、 η は局所座標系をなす。そして、

$$\tilde{g}_{ij} = \tilde{g}(\partial_i, \partial_j) \quad (4.120)$$

$$= \tilde{g} \left(\partial_i, \frac{\partial \eta^k}{\partial \theta^j} \partial^k \right) \quad (4.121)$$

$$= \tilde{g}(\partial_i, \partial^k) \tilde{g}_{kj} \quad (4.122)$$

より、

$$\tilde{g}(\partial_i, \partial^k) = \delta_i^k \quad (4.123)$$

となる。この式を、 η_l で偏微分すると、

$$0 = \frac{\partial}{\partial \eta_l} \tilde{g}(\partial_i, \partial^k) \quad (4.124)$$

$$= \tilde{g}(\nabla_{\partial_l}^e \partial_i, \partial^k) + \tilde{g}(\partial_i, \nabla_{\partial_l}^* \partial^k) \quad (4.125)$$

$$= \tilde{g}(\partial_i, \nabla_{\partial_l}^* \partial^k) \quad \because \theta \text{ が } \nabla^e\text{-アフィン座標系であること} \quad (4.126)$$

となる。これが、任意の i について成り立つことから、

$$\nabla_{\partial_l}^* \partial^k = 0 \quad (4.127)$$

となる。したがって、たしかに η は ∇ -双対アフィン座標系である。 \square

∇^m -アフィン座標系についても同様のことができる。

Def: 混合型分布族

Ω 上の確率分布の族 $p_0(\omega), p_1(\omega), \dots, p_k(\omega)$ 領域 H 上を動く k 次元パラメータ $\eta = (\eta_0, \dots, \eta_k) \in H$ を用いて、

$$p_\eta(\omega) = \sum_{i=1}^k \eta_i p_i(\omega) + (1 - \sum_{i=1}^k \eta_i) p_0(\omega) \quad (4.128)$$

と表される確率分布族 $M = \{p_\eta; \eta \in H\}$ を混合型分布族という。

Thm:

\mathcal{S} の部分多様体 M に対し、 M が ∇^m 自己平行部分多様体であるための必要十分条件は、 M が混合型分布族であることである。

4.6 双対葉層構造

統計多様体 $(\mathcal{S}, g, \nabla^e, \nabla^m)$ の ∇^e -自己平行部分多様体である指数型分布族

$$M = \{p_\theta(\omega) \in \mathcal{S}; \log p_\theta(\omega) = C(\omega) + \sum_{i=1}^k \theta^i F_i(\omega) - \psi(\theta)\} \quad (4.129)$$

が与えられたとする。 M の期待値座標系を固定したときに定まる \mathcal{S} 確率分布族

$$\Gamma_\eta = \{q(\omega) \in \mathcal{S}; E_q[F_i] = \eta_i\} \quad (4.130)$$

を考える。任意の $q, r \in \Gamma_\eta$ と、 $\lambda \in [0, 1]$ に対して、

$$E_{\lambda q + (1-\lambda)r}[F_i] = \sum_{\omega \in \Omega} (\lambda q(\omega) + (1-\lambda)r(\omega))p(\omega)F_i(\omega) \quad (4.131)$$

$$= \lambda \sum_{\omega \in \Omega} q(\omega)p(\omega)F_i(\omega) + (1-\lambda) \sum_{\omega \in \Omega} r(\omega)p(\omega)F_i(\omega) \quad (4.132)$$

$$= \lambda E_q[F_i] + (1-\lambda)E_r[F_i] \quad (4.133)$$

$$= \lambda \eta_i + (1-\lambda)\eta_i \quad (4.134)$$

$$= \eta_i \quad (4.135)$$

となるから、 Γ_η は凸集合である。したがって、 Γ_η は、 \mathcal{S} の ∇^m -自己平行部分多様体である。

Thm:

M と Γ_η が共有点を持つならば、その点において、 M と Γ_η は直交する。

Prf

5 統計物理学への応用

5.1 統計力学

以下では、物理の話が出てくるため、ちょっとした事項たちを確認しておく。

5.1.1 エントロピー

統計力学における系のエントロピーは、状態のランダムネスを表す量である。また、熱力学においては、状態の遷移可能性を特徴づける量である。

5.1.2 エントロピー最大原理

要請: エントロピー最大原理

マクロ変数の期待値についての制約条件のもとで、エントロピーが最大となる確率分布が平衡状態にある系の確率分布である。

5.1.3 カノニカル分布

エネルギー期待値を一定に保った時の、シャノンエントロピー最大化問題を考える。シャノンエントロピーは確率分布関数の汎関数であるから、変分によって停留点を探す。

$$\tilde{S} = -k_B \sum_i p_i \log p_i - \lambda \left(\sum_i p_i - 1 \right) - \rho \left(\sum_i p_i E_i - U \right) \quad (5.1)$$

である。ここで、 λ と ρ は未定乗数である。微小な確率変分を考えると、

$$\delta \tilde{S} = -k_B \sum_i (p_i + \delta p_i) \log(p_i + \delta p_i) - \lambda \left(\sum_i (p_i + \delta p_i) - 1 \right) - \rho \left(\sum_i (p_i + \delta p_i) E_i - U \right) - \tilde{S} \quad (5.2)$$

$$= -k_B \sum_i \left(\delta p_i \log p_i + (p_i + \delta p_i) \log \left(1 + \frac{\delta p_i}{p_i} \right) \right) - \lambda \left(\sum_i \delta p_i \right) - \rho \left(\sum_i \delta p_i E_i \right) \quad (5.3)$$

となる。ここで、 $\log(1+x) = x + O(x^2)$ であることを用いて、

$$\delta \tilde{S} = -k_B \sum_i (\delta p_i \log p_i + \delta p_i) - \lambda \left(\sum_i \delta p_i \right) - \rho \left(\sum_i \delta p_i E_i \right) \quad (5.4)$$

$$= \sum_i \delta p_i (-k_B \log p_i - k_B - \lambda - \rho E_i) + O(\delta p_i^2) \quad (5.5)$$

となる。したがって、

$$-k_B \log p_i - k_B - \lambda - \rho E_i = 0 \quad (5.6)$$

である。これを変形して、

$$p_i = \exp\left(-1 - \frac{\lambda}{k_B} - \frac{\rho E_i}{k_B}\right) \quad (5.7)$$

したがって、

$$p_i \propto \exp(-\beta E_i) \quad (5.8)$$

となる。ここで、 $\beta = \frac{1}{k_B T}$ である。 $(\rho = \frac{1}{T})$

あとは、規格化条件から、

$$p_i = \frac{1}{Z} \exp(-\beta E_i) \quad (5.9)$$

$$Z = \sum_i \exp(-\beta E_i) \quad (5.10)$$

であることがわかる。これがカノニカル分布である。

5.2 最大エントロピー原理

情報幾何の視点から最大エントロピー原理について考える。統計物理でよく用いられる分布は、カノニカル分布などの指数型分布族であった。

Recall: 指数型分布族

Ω 上の関数 $C(\omega), F_1(\omega), \dots, F_k(\omega)$ 及び \mathbb{R}^k 上の領域 Θ 上を動く k 次元パラメータ $\theta = (\theta^1, \dots, \theta^k) \in \Theta$ を用いて

$$p_\theta(\omega) = \exp \left[C(\omega) + \sum_{i=1}^k \theta^i F_i(\omega) - \psi(\theta) \right] \quad (5.11)$$

と表される確率分属族 $\{p_\theta; \theta \in \Theta\}$ を指数型分布族という。ただし、 $\psi(\theta)$ は、

$$\psi(\theta) = \log \left[\sum_{\omega \in \Omega} \exp \left[C(\omega) + \sum_{i=1}^k \theta^i F_i(\omega) \right] \right] \quad (5.12)$$

で定義される。

カノニカル分布の形を思い出しつつ、指数型分布族において、 $k = 1$ とし、 $F_1(\omega) = -H(\omega)$ とする。そして、 $\theta = 0$ で一様分布

$$u = \left(\frac{1}{n}, \dots, \frac{1}{n} \right) \quad (5.13)$$

を通る 1 次元指数型分布族 (∇^e -測地線) を考える。このとき、この測地線は、

$$p_\theta(\omega) = \exp [-\theta H(\omega) - \psi(\theta)] \quad (5.14)$$

となる。^{*3}このとき、この測地線と、

$$\Gamma_\eta = \{q \in \mathcal{S}; E_q[-H] = \eta\} \quad (5.15)$$

は直交する。

Recall

M と Γ_η が共有点を持つならば、その点において、 M と Γ_η は直交する。また、このときの M と Γ_η の共有点はただ一つに定まる。

したがって、一般化 Pythagoras の定理より、

$$p_{\theta_*} = \operatorname{argmin}_{q \in \Gamma_\eta} \{D^e(u||q)\} \quad (5.16)$$

$$= \operatorname{argmin}_{q \in \Gamma_\eta} \{D^m(q||u)\} \quad (5.17)$$

$$= \operatorname{argmin}_{q \in \Gamma_\eta} \sum_{\omega \in \Omega} q(\omega) \log \frac{q(\omega)}{u(\omega)} \quad (5.18)$$

$$= \operatorname{argmin}_{q \in \Gamma_\eta} \{\log n - S(q)\} \quad (5.19)$$

$$= \operatorname{argmax}_{q \in \Gamma_\eta} \{S(q)\} \quad (5.20)$$

となる。これは、確率変数 $F_1(\omega) = -H$ の期待値が一定であるもとで、Shanon エントロピー $S(q)$ を最大にする確率分布は、 Γ_η から測地線に下した垂線の足 p_{θ_*} に一致することを示している。

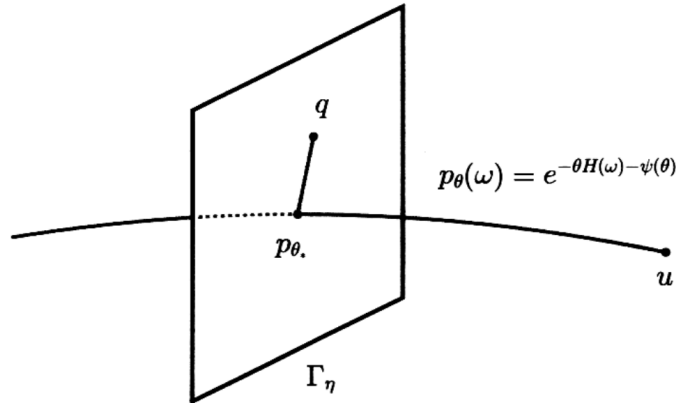


図 6.1 最大エントロピー原理

図 1 エネルギー期待値が一定の面の中で一様分布に最も近い分布が p_{θ_*} である。

^{*3} $\theta = 0$ で一様分布になってほしいので、 $C(\omega) = 0$ とする。

ところで、

$$\log Z(\theta) = \Psi(\theta) \quad (5.21)$$

$$\beta = \theta_* \quad (5.22)$$

とおくと、

$$\operatorname{argmax}_{\langle H \rangle = \text{const}} \{S(q)\} = p_\beta(\omega) \quad (5.23)$$

$$= \exp[-\beta H(\omega) - \psi(\beta)] \quad (5.24)$$

$$= \frac{1}{Z(\beta)} \exp[-\beta H(\omega)] \quad (5.25)$$

となり、見慣れたカノニカル分布の形になる。

以上をまとめると、最大エントロピー原理は、情報幾何の立場では、拘束条件のもとで、一様分布分布との距離のようなもの (KL ダイバージェンス) を最小しようという話になる。

また、測地線の式の両辺の対数をとって、 p_θ に関する期待値をとると、

$$\Psi(\theta) = S(p_\theta) - \theta E_{p_\theta}[H] \quad (5.26)$$

となる。このとき、 S の任意の確率分布 r と、カノニカル分布 p_β との KL ダイバージェンスは、

$$D(r||p_\beta) = \sum_{\omega \in \Omega} r(\omega) \log \frac{r(\omega)}{p_\beta(\omega)} \quad (5.27)$$

$$= -S(r) + \beta E_r[H] + \Psi(\beta) \quad (5.28)$$

$$= -(S(r) - \beta E_r[H]) + (S(p_\beta) - \beta E_{p_\beta}[H]) \quad (5.29)$$

となる。したがって、ダイバージェンスの正值性から、

$$S(p_\beta) - \beta E_{p_\beta}[H] \geq S(r) - \beta E_r[H] \quad (5.30)$$

となる。ここで、

$$F(r) = E_r[H] - \frac{1}{\beta} S(r) \quad (5.31)$$

とおくと、

$$p_\beta = \operatorname{argmin}_r F(r) \quad (5.32)$$

となる。これは、自由エネルギーの最小化原理である。注意されたいこととして、自由エネルギーを最小化するという立場では、拘束条件は特でない。

こうして、拘束条件の下での最小化問題を、拘束条件なしでの最小化問題に言い換えられたわけだが、この議論で鍵になったのは KL ダイバージェンスの正值性であった。ところで、ダイバージェンスの正值性はルジャンドル変換と結びついていた。したがって、エントロピーと自由エネルギー

との関係がルジャンドル変換によって表されるのではないかと考えられる。^{*4}

今、ダイバージェンスの正值性と同等なルジャンドル変換の関係式は、

$$\psi(\theta_\beta) = \max_{r \in \mathcal{S}} \{\theta^i(p_\beta) \eta_i(r) - \varphi(\eta(r))\} \quad (5.33)$$

であった。いま、

$$\theta^1(p_\beta) = \beta \quad (5.34)$$

$$\theta^n(p_\beta) = 0 \quad (n \geq 2) \quad (5.35)$$

$$\eta_1(r) = -E_r[H] \quad (5.36)$$

$$\phi(\eta(r)) = -S(r) \quad (5.37)$$

であるから、

$$\log Z(\theta(p_\beta)) = \max_{r \in \mathcal{S}} \{-\beta E_r[H] + S(r)\} \quad (5.38)$$

となる。とくに $\beta > 0$ のとき、

$$-\frac{1}{\beta} \log Z(\theta(p_\beta)) = \min_{r \in \mathcal{S}} \{E_r[H] - \frac{1}{\beta} S(r)\} \quad (5.39)$$

となる。この右辺は自由エネルギーであり、とくに $r = p_\beta$ のとき最小値をとる。したがって、

$$-\frac{1}{\beta} \log Z(\theta(p_\beta)) = F(p_\beta) \quad (5.40)$$

となる。これは統計物理においてよく知られた関係式である。

5.3 磁性

統計物理の応用として、磁性について考える。磁性は、スピンの持つ磁気モーメントによって引き起こされる現象である。簡単なモデルとして、一次元 Ising リングを考える。

^{*4} 物理の人はルジャンドル変換を最初から用いることが多い。

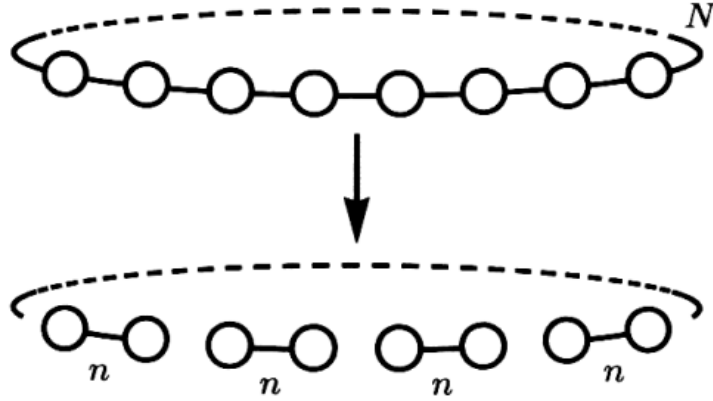


図2 一次元 Ising リング

リングの各点で、スピンと呼ばれる物理量が定義されており、スピンは $\{-1, 1\}$ の値をとる。このスピンの磁場を感じることで、エネルギーが変化する。このとき、ハミルトニアン (エネルギー) は、

$$H = H(\{S_\xi\}) = -h \sum_{\xi=1}^N S_\xi - J \sum_{\xi=1}^N S_\xi S_{\xi+1} \quad J > 0 \quad (5.41)$$

となると仮定する (Ising 模型)。これによると、スピンの値 1 をとるとき、エネルギーが低くなる。また、隣り合うスピンの値がそろったとき、エネルギーが低くなる。

5.4 平均場近似

ハミルトニアン

$$H = H(\{S_\xi\}) = -h \sum_{\xi=1}^N S_\xi - J \sum_{\xi=1}^N S_\xi S_{\xi+1} \quad J > 0 \quad (5.42)$$

をもつ Ising リングを考える。ここで、確率変数 $S_\xi = \{-1, 1\}$ であり、インデックス $\xi + 1$ は N を法として計算する。このとき、対応するカノニカル分布を、

$$q = q(\{S_\xi\}) = \exp(-\beta H - \tilde{\psi}(h, J)) \quad (5.43)$$

とする。ここで、 β は逆温度であり、 $\tilde{\psi}(h, J)$ は規格化定数である。この確率分布は、幾何的には、 $2^N - 1$ 次元統計多様体上の点となる。^{*5}

N 個のスピン全体を隣り合う n 個のスピンごとに分解してみる。このとき、クラスターの数 m とすると、

$$N = m \times n \quad (5.44)$$

^{*5} 次元が 1 つ減っているのは確率の規格化のためである。

となる。もともとのリングにける第 ξ 番目のスピンの、 λ 番目の第 i 番目のスピンであるとする、

$$S_i^\lambda = S_\xi \quad (5.45)$$

$$\xi = (\lambda - 1)n + i \quad (5.46)$$

以下、各クラスターが統計的に独立であると仮定したうえで、カノニカル分布を最もよく近似する n スピン状態を探し出す。ここで、 n スピン系の確率分布は、以下の形で書けると仮定する：

$$\exp \left(\sum_{l=1}^n \sum_{1 \leq i_1 < i_2 < \dots < i_l \leq n} \theta_{i_1 i_2 \dots i_l} S_{i_1} S_{i_2} \dots S_{i_l} - \psi(\theta) \right) \quad (5.47)$$

ここで、 θ は実数パラメータであり、 $\psi(\theta)$ は規格化定数である。要するに、この形は各スピン同士の考え得る相互作用をすべて考慮したような形になっている。省略した記法

$$S_{\{i_1 i_2 \dots i_l\}} := S_{i_1} S_{i_2} \dots S_{i_l}$$

を用いれば、 n スピン系の確率分布は

$$\exp [\theta^a S_a - \psi(\theta)] \quad (5.48)$$

と書くことができる。ここに添字 a は多重インデックス全体の集合

$$I := \bigcup_{l=1}^n I_l, \quad I_l := \{(i_1 i_2 \dots i_l); 1 \leq i_1 < i_2 < \dots < i_l \leq n\}$$

を走るものとする。ここで、確率変数 $\{S_a\}_{a \in I}$ 全体に定数関数 1 を加えた 2^n 個の関数全体は 1 次独立であることに注意。したがって i.i.d. 拡張された m 個のクラスターの確率分布

$$p_\theta := p_\theta(\{\xi_\lambda\}) := \prod_{\lambda=1}^m \exp [\theta^a S_a^{(\lambda)} - \psi(\theta)] \quad (5.49)$$

全体の集合 \mathcal{M}_n は、 $\theta = (\theta^a)_{a \in I}$ を座標系とする P_N の $(2^n - 1)$ -次元部分多様体と見なせる。以下では、各確率分布 p_θ をクラスター状態と呼ぶことにする。

あるクラスター状態 $p_\theta \in \mathcal{M}_n$ でカノニカル分布 q を近似する問題を数学的に定式化するため、次の KL ダイバージェンスを考える：

$$D(p_\theta \| q) = E_{p_\theta} \left[\log \frac{p_\theta}{q} \right] \quad (5.50)$$

前節で見たように、

$$D(p_\theta \| q) = \beta [F(p_\theta) - F(q)] \quad (5.51)$$

と書き直すことができる。したがって、自由エネルギーを最小化する問題を、KL ダイバージェンスを最小化する問題に言い換えることができる。また、一般化 Pythagoras の定理より、これは q

から部分多様体 \mathcal{M}_n に下した ∇^e -射影を求める問題となる。

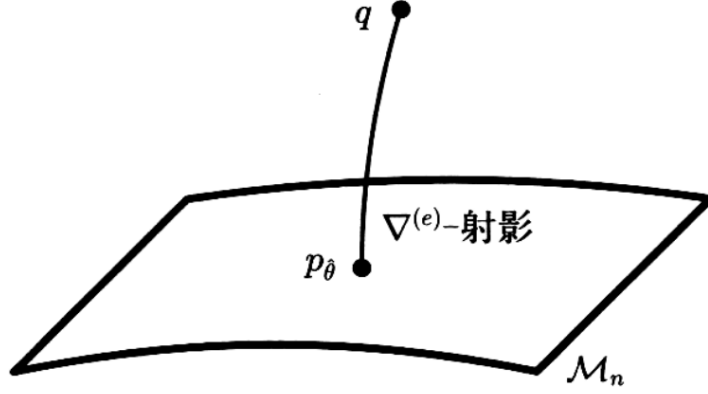


図3 カノニカル分布 q とクラスター状態 p_θ の関係

Thm:

クラスターの大きさは $n \geq 2$ とし, \mathcal{M}_n 上の関数 $\theta \mapsto D(p_\theta \| q)$ を最小化する点を $\hat{\theta}$ とする. このとき

$$\hat{\theta}^{(1)} = \hat{\theta}^{(n)} \quad (5.52)$$

$$\hat{\theta}^{(2)} = \hat{\theta}^{(3)} = \dots = \hat{\theta}^{(n-1)} = \beta h \quad (5.53)$$

$$\hat{\theta}^{(12)} = \hat{\theta}^{(23)} = \dots = \hat{\theta}^{(n-1, n)} = \beta J \quad (5.54)$$

であり, 他のパラメータ値はすべてゼロである. そして, 唯一残っている未知パラメータ $\hat{\theta}^{(1)}$ は次の方程式を満たす.

$$\hat{\theta}^{(1)} = \beta (h + JE_{p_{\hat{\theta}}} [S_1]) \quad (5.55)$$

これをもとのハミルトニアンと見比べてみると, $\theta^{(1)}, \theta^{(n)}$ 以外は, もとのハミルトニアンの形をそのまま遺伝させたものになっている。

Prf

いま、

$$\log p_\theta = \sum_{\lambda=1}^m l^{(\lambda)} \quad (5.56)$$

$$l^{(\lambda)} = \theta^a S_a^{(\lambda)} - \psi(\theta) \quad (5.57)$$

とおく。このとき、

$$E_{p_\theta}[\partial_{\theta^a} l^{(\lambda)}] = E_{p_\theta}[S_a^{(\lambda)}] - \partial_{\theta^a} \psi(\theta) \quad (5.58)$$

$$= E_{p_\theta}[S_a^{(\lambda)}] - E_{p_\theta}[S_a^{(\lambda)}] \quad (5.59)$$

$$= 0 \quad (5.60)$$

となる。Fisher 計量 $g = (g_{ab})$ は、

$$g_{ab} = E_{p_\theta}[\partial_{\theta^a} \log p_\theta \partial_{\theta^b} \log p_\theta] \quad (5.61)$$

$$= E_{p_\theta} \left[\partial_{\theta^a} \left(\sum_{\lambda=1}^m l^{(\lambda)} \right) \partial_{\theta^b} \left(\sum_{\lambda=1}^m l^{(\lambda)} \right) \right] \quad (5.62)$$

$$= \sum_{\lambda=1}^m E_{p_\theta} [\partial_{\theta^a} l^{(\lambda)} \partial_{\theta^b} l^{(\lambda)}] \quad \because \text{異なるクラスターは独立} \quad (5.63)$$

$$= m E_{p_\theta} [\partial_{\theta^a} l^{(1)} \partial_{\theta^b} l^{(1)}] \quad \because \text{クラスターの等価性} \quad (5.64)$$

となる。 $\theta = (\theta^a)$ の双対座標系 $\eta = (\eta_a)$ は

$$\eta_a := \partial_a \psi(\theta) = E_{p_\theta} [S_a^{(\lambda)}]$$

で定義され、これは等式

$$\partial_a \ell^{(\lambda)} = S_a^{(\lambda)} - \eta_a \quad (5.65)$$

を満たす。

さて、

$$\partial_a D(p_\theta \| q) = \partial_a \sum_{\{\xi\}} p_\theta (\log p_\theta - \log q) \quad \because \text{KL ダイバージェンスの定義} \quad (5.66)$$

$$= \sum_{\{\xi\}} \partial_a p_\theta (\log p_\theta - \log q) + p_\theta \frac{\partial_a p_\theta}{p_\theta} \quad (5.67)$$

$$= \sum_{\{\xi\}} p_\theta (\partial_a \log p_\theta) (\log p_\theta - \log q) \quad \because \text{第二項は規格化により消える} \quad (5.68)$$

$$(5.69)$$

であるから、関数 $\theta \mapsto D(p_\theta \| q)$ が最小値をとるのは、

$$E_{p_\theta} [(\log p_\theta - \log q) (\partial_a \log p_\theta)] = 0 \quad (\forall a \in I) \quad (5.70)$$

が成り立つときに限られる。ここで、補助的な変数 $x = (x^a)_{a \in I}$ および $y = (y^a)_{a \in I}$ を以下のよ

うに導入する:

$$x^a = \begin{cases} \theta^{\langle i \rangle} - h & a = \langle i \rangle \\ \theta^{\langle ij \rangle} - J & a = \langle ij \rangle \text{ かつ } j = i + 1 \\ \theta^{\langle ij \rangle} & a = \langle ij \rangle \text{ かつ } j > i + 1 \\ \theta^{\langle i_1 i_2 \dots i_\ell \rangle} & a = \langle i_1 i_2 \dots i_\ell \rangle \text{ かつ } \ell \geq 3 \end{cases} \quad (5.71)$$

$$y^a = \begin{cases} \eta_{\langle n \rangle} & a = \langle 1 \rangle \\ \eta_{\langle 1 \rangle} & a = \langle n \rangle \\ 0 & \text{その他} \end{cases} \quad (5.72)$$

Lem.1

(5.70) 式は、以下の方程式と同値である:

$$x^a = Jy^a \quad (\forall a \in I) \quad (5.73)$$

Prf.(Lem.1)

いま、

$$\log q = h \sum_{\xi=1}^N S_\xi + J \sum_{\xi=1}^N S_\xi S_{\xi+1} - \tilde{\psi}(h, J) \quad (5.74)$$

$$= \sum_{\lambda=1}^m \left\{ h \sum_{i=1}^n S_i^{(\lambda)} + J \sum_{i=1}^{n-1} S_i^{(\lambda)} S_{i+1}^{(\lambda)} + J S_n^{(\lambda)} S_1^{(\lambda+1)} \right\} - \tilde{\psi}(h, J) \quad (5.75)$$

ここに上つき添字 $\lambda + 1$ は m を法として考える。したがって、

$$\log p_\theta - \log q = \sum_{\lambda=1}^m \left\{ \sum_{i=1}^n (\theta^{\langle i \rangle} - h) S_i^{(\lambda)} + \sum_{i=1}^{n-1} (\theta^{\langle i, i+1 \rangle} - J) S_{i, i+1}^{(\lambda)} \right. \quad (5.76)$$

$$\left. - J S_n^{(\lambda)} S_1^{(\lambda+1)} + \sum_{j>i+1} \theta^{\langle ij \rangle} S_{ij}^{(\lambda)} + \sum_{|b| \geq 3} \theta^b S_b^{(\lambda)} \right\} \quad (5.77)$$

$$- m\psi(\theta) + \tilde{\psi}(h, J) \quad (5.78)$$

$$= \sum_{\lambda=1}^m \left\{ L^{(\lambda)} - J S_n^{(\lambda)} S_1^{(\lambda+1)} \right\} - m\psi(\theta) + \tilde{\psi}(h, J) \quad (5.79)$$

を得る。ただし、

$$L^{(\lambda)} := \sum_{i=1}^n (\theta^{\langle i \rangle} - h) S_i^{(\lambda)} + \sum_{i=1}^{n-1} (\theta^{\langle i, i+1 \rangle} - J) S_{i, i+1}^{(\lambda)} \quad (5.80)$$

$$+ \sum_{j>i+1} \theta^{\langle ij \rangle} S_{ij}^{(\lambda)} + \sum_{|b| \geq 3} \theta^b S_b^{(\lambda)} \quad (5.81)$$

である。したがって、

$$E_{p_\theta} \left[\left(\sum_{\lambda=1}^m L^{(\lambda)} \right) \left(\sum_{\mu=1}^m \partial_a \ell^{(\mu)} \right) \right] = E_{p_\theta} \left[\left(\sum_{\lambda=1}^m J S_n^{(\lambda)} S_1^{(\lambda+1)} \right) \left(\sum_{\mu=1}^m \partial_a \ell^{(\mu)} \right) \right] \quad (5.82)$$

さて、異なるクラスターに属する確率変数は独立であることから、クロスタームは消える。したがって、これは、

$$\sum_{\lambda=1}^m E_{p_\theta} \left[L^{(\lambda)} \partial_a \ell^{(\lambda)} \right] = J \sum_{\lambda=1}^m E_{p_\theta} \left[S_n^{(\lambda)} S_1^{(\lambda+1)} \left(\partial_a \ell^{(\lambda)} + \partial_a \ell^{(\lambda+1)} \right) \right] \quad (5.83)$$

と同値である。右辺を計算すると、

$$(\text{右辺}) = J \sum_{\lambda=1}^m E_{p_\theta} \left[\left(\partial_{\langle n \rangle} \ell^{(\lambda)} + \eta_{\langle n \rangle} \right) \left(\partial_{\langle 1 \rangle} \ell^{(\lambda)} + \partial_{\langle n \rangle} \ell^{(\lambda+1)} + \eta_{\langle 1 \rangle} \right) \left(\partial_a \ell^{(\lambda)} + \partial_a \ell^{(\lambda+1)} \right) \right] \quad (5.84)$$

\because (5.57) より

$$\begin{aligned} &= J \sum_{\lambda=1}^m \left\{ E_{p_\theta} \left[\left(\partial_{\langle n \rangle} \ell^{(\lambda)} + \eta_{\langle n \rangle} \right) \left(\partial_a \ell^{(\lambda)} \right) \right] E_{p_\theta} \left[\left(\partial_{\langle 1 \rangle} \ell^{(\lambda+1)} + \eta_{\langle 1 \rangle} \right) \right] \right. \\ &\quad \left. + E_{p_\theta} \left[\partial_{\langle n \rangle} \ell^{(\lambda)} + \eta_{\langle n \rangle} \right] E_{p_\theta} \left[\left(\partial_{\langle 1 \rangle} \ell^{(\lambda+1)} + \eta_{\langle 1 \rangle} \right) \left(\partial_a \ell^{(\lambda+1)} \right) \right] \right\} \quad (5.85) \end{aligned}$$

$$\begin{aligned} &= J \sum_{\lambda=1}^m \left\{ E_{p_\theta} \left[\left(\partial_{\langle n \rangle} \ell^{(\lambda)} \right) \left(\partial_a \ell^{(\lambda)} \right) \right] E_{p_\theta} (\eta_{\langle 1 \rangle}) \right. \\ &\quad \left. + E_{p_\theta} (\eta_{\langle n \rangle}) E_{p_\theta} \left(\left(\partial_{\langle 1 \rangle} \ell^{(\lambda+1)} \right) \left(\partial_a \ell^{(\lambda+1)} \right) \right) \right\} \quad (\because E_{p_\theta} [\partial_a \ell^{(\lambda)}] = 0) \quad (5.86) \end{aligned}$$

$$= J (g_{a\langle n \rangle} \eta_{\langle 1 \rangle} + g_{a\langle 1 \rangle} \eta_{\langle n \rangle}) \quad (5.87)$$

$$= J g_{ab} y^b \quad (y \text{ の定義より}) \quad (5.88)$$

を得る。左辺は次のように計算される。

$$(\text{左辺}) = \sum_{\lambda=1}^m \left(E_{p_\theta} \left(\sum_{i=1}^n (\theta^{(i)} - h) S_i^{(\lambda)} + \sum_{i=1}^{n-1} (\theta^{(i,i+1)} - J) S_{i,i+1}^{(\lambda)} \right. \right. \quad (5.89)$$

$$\left. + \sum_{j>i+1} \theta^{(ij)} S_{ij}^{(\lambda)} + \sum_{|b|\geq 3} \theta^b S_b^{(\lambda)} \right) (\partial_a \ell^{(\lambda)}) \quad (5.90)$$

$$= \sum_{\lambda=1}^m \left(E_{p_\theta} \left(\sum_{i=1}^n (\theta^{(i)} - h) (\partial_a \ell^{(\lambda)} + E_{p_\theta}(S)) + \sum_{i=1}^{n-1} (\theta^{(i,i+1)} - J) (\partial_a \ell^{(\lambda)} + E_{p_\theta}(S)) \right. \right. \quad (5.91)$$

$$\left. + \sum_{j>i+1} \theta^{(ij)} (\partial_a \ell^{(\lambda)} + E_{p_\theta}(S)) + \sum_{|b|\geq 3} \theta^b (\partial_a \ell^{(\lambda)} + E_{p_\theta}(S)) \right) (\partial_a \ell^{(\lambda)}) \quad (5.92)$$

$$= \sum_{i=1}^n g_{a\langle i} (\theta^{(i)} - h) + \sum_{i=1}^{n-1} g_{a\langle i,i+1} (\theta^{(i,i+1)} - J) \quad (5.93)$$

$$+ \sum_{j>i+1} g_{a\langle ij} \theta^{(ij)} + \sum_{|b|\geq 3} g_{ab} \theta^b \quad (5.94)$$

$$= g_{ab} x^b \quad (5.95)$$

以上まとめると、

$$g_{ab} x^b = J g_{ab} y^b \quad (\forall a \in I) \quad (5.96)$$

と同値であることが分かる。さて、計量 g は正定値であるから、上式はさらに

$$x^a = J y^a \quad (\forall a \in I) \quad (5.97)$$

と同値であることが分かり、補題が証明された。 \square

補題より、(5.70) 式を満たすとき、

$$\hat{\theta}^{(1)} = h + J \hat{\eta}_{\langle n} \quad (5.98)$$

$$\hat{\theta}^{(n)} = h + J \hat{\eta}_{\langle 1} \quad (5.99)$$

$$\hat{\theta}^{(2)} = \hat{\theta}^{(3)} = \dots = \hat{\theta}^{(n-1)} = h \quad (5.100)$$

$$\hat{\theta}^{(12)} = \hat{\theta}^{(23)} = \dots = \hat{\theta}^{(n-1,n)} = J \quad (5.101)$$

であって、他のパラメータ値はすべてゼロであることが結論される。

Lem.2

連立方程式

$$\begin{cases} \theta^{\langle 1 \rangle} = h + J\eta_{\langle n \rangle} \\ \theta^{\langle n \rangle} = h + J\eta_{\langle 1 \rangle} \end{cases} \quad (5.102)$$

は、 $J > 0$ のもとで、ただ一つの対称解

$$\theta^{\langle 1 \rangle} = \theta^{\langle n \rangle} \quad (5.103)$$

を持つ。

Prf.(Lem2)

連立方程式より

$$\theta^{\langle 1 \rangle} - \theta^{\langle n \rangle} = -J(\eta_{\langle 1 \rangle} - \eta_{\langle n \rangle}) = -J\left(\frac{\partial}{\partial \theta^{\langle 1 \rangle}} - \frac{\partial}{\partial \theta^{\langle n \rangle}}\right)\psi(\theta) \quad (5.104)$$

となるので、座標変換 $(\theta^{\langle 1 \rangle}, \theta^{\langle n \rangle}) \rightarrow (X, Y) := (\theta^{\langle 1 \rangle} - \theta^{\langle n \rangle}, \theta^{\langle 1 \rangle} + \theta^{\langle n \rangle})$ を施すと、上の式は

$$X = -2J\frac{\partial \psi}{\partial X} \quad (5.105)$$

と同値である。関数 $\psi(\theta)$ は θ に関して狭義凸であるから、 $\partial^2 \psi / \partial X^2 > 0$ であり、従って $X \mapsto \partial \psi / \partial X$ は単調増加関数である。しかも、関数 $\psi(\theta)$ は $\theta^{\langle 1 \rangle}$ と $\theta^{\langle n \rangle}$ について対称であるので、

$$\left. \frac{\partial \psi}{\partial X} \right|_{X=0} = 0 \quad (5.106)$$

を満たす。従って、各 Y に対し、方程式 (5.105) は唯一の解 $X = 0$ を持つ。 □

以上の二つの補題を合わせて、Thm. が証明される。 □