

拡張ピタゴラスの定理

2019 年度芝浦祭

令和元年 11 月 1 日

数理科学研究会 3 年
BV17077 横井 健

目次

0	準備	1
1	凸関数の導くダイバージェンスと双対平坦構造	3
1.1	凸関数とダイバージェンス	3
1.2	Legendre 変換と双対性	4
1.3	ダイバージェンスとリーマン幾何: 接空間	6
1.4	凸関数と双対平坦多様体	8
1.5	拡張ピタゴラスの定理	11
1.6	拡張射影定理	12

研究動機

多様体を学習していたとき統計と絡めて何ができるか気になり調べていたとき情報幾何学を見つけ学習したいと思い、学習を始めたところ、導入にダイバージェンスやリーマン計量といった内容があり、そこから取り掛かった。今回は、多様体上に一般化されたピタゴラスの定理について述べる。

0 準備

ここでは必要となる定義、定理の確認が中心である。

定義 0.1 (多様体). 集合 M が以下の条件を満たすとき、 M を n 次元 C^r 級 (可微分) 多様体という。

- i) M は Hausdorff 空間。
- ii) 適当な集合 A を添え字集合とする M の開集合の族 $\{U_\alpha\}_{\alpha \in A}$ と写像 $\psi_\alpha : U_\alpha \rightarrow \mathbb{R}^n$ の族 $\{\psi_\alpha\}_{\alpha \in A}$ があって、以下の 3 条件を満たす。
 - 1) $M = \bigcup_{\alpha \in A} U_\alpha$
 - 2) 各 $\alpha \in A$ に対し、像 $\psi(U_\alpha)$ は \mathbb{R}^n の開集合で、 $\psi_\alpha : U_\alpha \rightarrow \psi_\alpha(U_\alpha)$ は同相写像。
 - 3) $U_\alpha \cap U_\beta \neq \emptyset$ ならば、写像 $\psi_\beta \circ \psi_\alpha^{-1} : \psi_\alpha(U_\alpha \cap U_\beta) \rightarrow \psi_\beta(U_\alpha \cap U_\beta)$ は C^r 級。

上記で定義した多様体に対して計量的な性質として、ダイバージェンスとよぶ 2 点の関数を考える。多様体 M に座標が入るものとして、それをベクトル ξ であらわす。多様体の 2 点 P, Q に対して、次の条件を満たす関数 $D[P : Q]$ を P, Q の間のダイバージェンス (分離度) という。 P, Q の局所座標を ξ_P, ξ_Q 、ダイバージェンスを座標の関数として、 $D[\xi_P : \xi_Q]$ と書く。 D は微分可能。

定義 0.2 (ダイバージェンス). 次の 3 条件を満たす 2 点関数 $D[P : Q]$ をダイバージェンスとよぶ。

1. $D[P : Q] \geq 0$
2. $P = Q$ のとき、このときに限り、 $D[P : Q] = 0$
3. P 点と Q 点が近いとし、それぞれの座標を、 $\xi, \xi + d\xi$ とする。このとき、 $D[\xi : \xi + d\xi]$ をテイラー展開すると、

$$D[\xi : \xi + d\xi] = \frac{1}{2} \sum g_{ij}(\xi) d\xi_i d\xi_j \quad (0.1)$$

と 2 次の項が最初に出るのが、行列

$$G(\xi) = (g_{ij}(\xi))$$

は正定値対象である。

ダイバージェンスは 2 点の分離の度合いを表すが, これは距離ではない. 第一にダイバージェンスは対象でなくともよく. 一般には

$$D[P : Q] \neq D[Q : P]$$

なので, P から Q へのダイバージェンスと呼ばなければならぬ, 三角不等式も成立しない. 式 (0.1) から解るように, ダイバージェンスは座標の二乗の次元をもっている. なのでこれは距離の二乗のようなもの. しかし, 非対象であることに注意.

ここで, $d\xi$ で表される微小な 2 点 $\xi, \xi + d\xi$ の間の距離 ds の二乗を

$$ds^2 = \sum g_{ij}(\xi) d\xi_i d\xi_j \quad (0.2)$$

で定義したとき, ダイバージェンスは微小な距離の二乗を非対称に遠方まで拡張したものとみなすことが出来, 微小線分 $d\xi$ の長さが (0.2) のような 2 次元形式で表される空間をリーマン空間とよぶ.

ここでダイバージェンスの例をいくつか挙げる.

例 0.1 (ユークリッド空間). 距離の二乗がダイバージェンスを与える.

$$D[[\xi : \xi']] = \frac{1}{2} \sum (\xi_i - \xi'_i)^2$$

これは対象である.

例 0.2 (確率分布族空間における Kullback-Leibler ダイバージェンス). 2 つの確率分布 $p(x)$ と $q(x)$ のダイバージェンスを

$$D[p(x) : q(x)] = \int p(x) \log \frac{p(x)}{q(x)} dx$$

と定義したものが, Kullback-Leibler ダイバージェンスである. (以下 KL ダイバージェンス, $KL[p : q]$ と略記する.). これは非対称である. 離散分布のときは積分を和分する. これを拡張した α ダイバージェンス^{*1}は, α を実数のパラメータとして,

$$D_\alpha[p(x) : q(x)] = \frac{4}{1 - \alpha^2} \left\{ 1 - \int p(x)^{\frac{1-\alpha}{2}} q(x)^{\frac{1+\alpha}{2}} dx \right\}$$

^{*1} α が 1 または -1 に近づく極限で定義すると 1 ダイバージェンスは $KL[q : p]$, -1 ダイバージェンスは $KL[p : q]$ になる.

1 凸関数の導くダイバージェンスと双対平坦構造

1.1 凸関数とダイバージェンス

多様体 M 上に微分可能な凸関数 $\psi(\xi)$ が与えられたとする。凸関数はその 2 階微分

$$H(\xi) = \left(\frac{\partial^2}{\partial \xi_i \partial \xi_j} \psi(\xi) \right)$$

が正則値行列である。ここで、座標系を $\xi = (\xi_1, \dots, \xi_n)$ を固定して考える。

例 1.1.

$$\psi(\xi) = \frac{1}{2} \sum \xi_i^2 \quad (1.1)$$

はユークリッド空間の原点からの距離の 2 乗であり。ヘッセ行列は単位行列である。

例 1.2. 確率変数を x とする指數型分布族

$$p(x, \theta) = \exp \left\{ \sum \theta_i x_i - \psi(\theta) \right\}$$

は $\theta = (\theta_1, \dots, \theta_n)$ を座標系とする多様体であるが、企画化定数に対応する $\psi(\theta)$ は

$$\psi(\theta) = \log \int \exp \left\{ \sum \theta_i x_i \right\} dx$$

のようにかける。これは凸関数であることはあとで確認する。この他、確率分布族の空間 M では、エントロピーは凹関数、従って、エントロピーの符号を変えた

$$\phi(p) = \sum p_i \log p_i$$

は p の凸関数。

ξ の凸関数 $z = \psi(\xi)$ の 1 点 ξ において、 $z = \psi(\xi)$ に接する超平面を考える。関数が凸であればグラフは必ずこの超平面より上にある。そこで ξ がどれくらい上にあるかを計る。 ξ' 点で $z = \psi(\xi')$ に接する超平面の方程式は、 z を縦軸として、

$$z = \psi(\xi') + \nabla \psi(\xi') \cdot (\xi - \xi')$$

である。ここで $\nabla \psi(\xi)$ は関数 ψ の勾配で、 $\frac{\partial \psi}{\partial \xi_i}$ を成分とするベクトルである。よって、接平面の ξ 点での値は

$$z(\xi) = \psi(\xi') + \nabla \psi(\xi') \cdot (\xi - \xi')$$

である。関数 $\psi(\xi)$ が点 ξ で接平面のどれくらい上にあるかは、

$$D[\xi : \xi'] = \psi(\xi) - \psi(\xi') - \nabla \psi(\xi') \cdot (\xi - \xi') \quad (1.2)$$

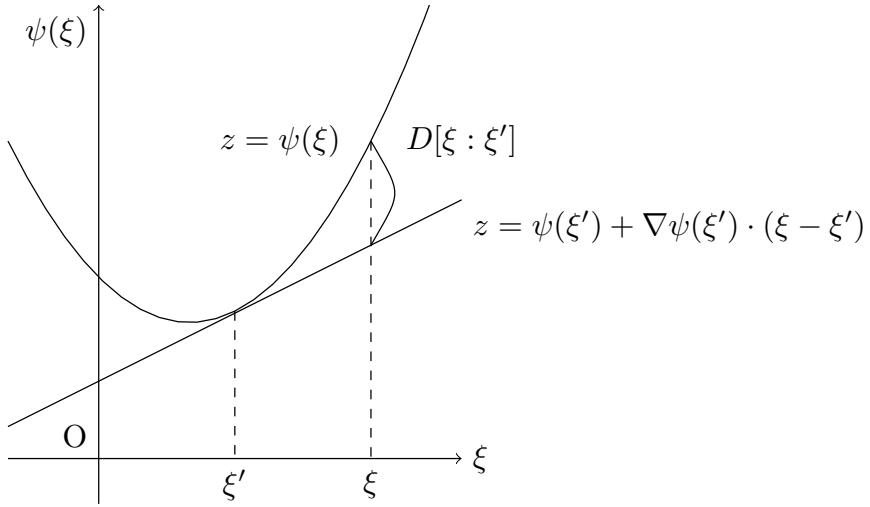


図 1 凸関数によるダイバージェンス

と書ける。これを凸関数 $\psi(\xi)$ から導かれる ξ から ξ' へのダイバージェンスと呼ぶ^{*2}。式 (1.1) からはユークリッド距離の二乗

$$D[\xi : \xi'] = \frac{1}{2} \sum (\xi_i - \xi'_i)^2$$

が導かれる。

離散確率分布空間 S_n では、負のエントロピー

$$\varphi(\mathbf{p}) = \sum p_i \log p_i$$

が凸関数である。 $\sum p_i = 1$ の束縛を無視して計算すると、

$$\frac{\partial}{\partial p_i} \varphi(\mathbf{p}) = \log p_i + 1$$

これを用いると、

$$D[\mathbf{p} : \mathbf{q}] = \sum_{i=0}^n p_i \log \frac{p_i}{q_i}$$

が得られる。これは KL ダイバージェンスである。 $(\sum p_i = 1$ の束縛を無視したが $\sum p_i = 1$ でも答えは同じになる。)

1.2 Legendre 変換と双対性

凸関数 $\psi(\xi)$ が与えられたとき、この関数の勾配は式 (1.2) から、

$$\xi^* = \nabla \psi(\xi^*) \quad (1.3)$$

^{*2} Bregman ダイバージェンス

である. 超平面の傾きは点 ξ によって決まり, 同じものはない. このことから式 (1.3) の変換は 1 対 1 の逆変換も存在し, ξ が多様体 M の座標系であるなら, ξ^* も座標系として使える. これによって, M に 2 つの座標系が与えられて, この変換 (式 (1.3)) を Legendre 変換という. この変換は双対である. つまり,

$$\psi^*(\xi^*) = \max_{\xi} \{\xi \cdot \xi^* - \psi(\xi)\} \quad (1.4)$$

によって, ξ の関数 ψ^* を定義する. いま, (1.4) 式の最大値を満たす ξ を $\bar{\xi} = \xi(\xi^*)$ と書けば, $\psi^*(\xi^*) = \bar{\xi} \cdot \xi^* - \psi(\bar{\xi})$ である. この $\bar{\xi}$ は最大値を達成するから (1.4) の右辺を ξ^* で微分すれば, $\xi^* = \nabla \psi(\xi)$ を満たす. (1.4) を ξ^* で微分すれば,

$$\nabla \psi^*(\xi^*) = \bar{\xi}$$

なので, これが ξ^* から ξ への逆変換となっている.

次節で示すように, ψ^* は凸関数であるから, ψ^* と座標系 ξ^* とで, ダイバージェンスが作られる. これを双対ダイバージェンス呼ぶが,

$$D^*[\xi^* : \xi'^*] = D[\xi' : \xi]$$

になっている. だから座標系によらず実質は 1 つのダイバージェンスで, ただ変数の順序を変えたものになっている.

座標系 ξ^* と凸関数 ψ^* から出発すればその双対として元の ξ と ψ が得られる. これがまさに”お互い様”の関係にある双対性である. ダイバージェンスの非対称性に関する分かりにくいことが双対性で解決できる. ダイバージェンスの順序を変えたものは実は双対ダイバージェンスである.

ダイバージェンスは, 2 つの座標系を用いると, 次の形で双対的に書ける.

定理 1.1. $\psi(\theta)$ が凸関数となるようなアファイン座標 θ を用い, 2 点 P, Q のアファイン座標を θ_P, θ_Q , 双対アファイン座標を θ_P^*, θ_Q^* とすると,

$$D[P : Q] = \psi(\theta_P) + \psi^*(\theta_Q^*) - \theta_P \cdot \theta_Q^* \quad (1.5)$$

が成り立つ.

証明. 点 P を固定して, 関数 $Q \mapsto \psi^*(\xi_Q^*) = \theta_Q \cdot \theta_P^* - \psi(\theta_Q)$ とする. これを微分すると,

$$\begin{aligned} d(\psi^*(\xi_Q^*)) &= (\theta_P^* - \nabla \psi(\theta_Q)) d\theta_Q \\ &= (\theta_P^* - \theta_Q^*) d\theta_Q \end{aligned}$$

となる. よって, 式 (1.4) の右辺が最大となるのは $\theta_P^* = \theta_Q^*$, すなわち $P = Q$ のときであり, このときに限り達成され, その最大値は $\theta_P = \theta$ とすると,

$$\psi(\theta) + \psi^*(\theta^*) - \theta \cdot \theta^* = 0$$

である。

以上のことから,

$$D[P : Q] = \psi(\boldsymbol{\theta}_P) + \psi^*(\boldsymbol{\theta}_Q^*) - \boldsymbol{\theta}_P \cdot \boldsymbol{\theta}_Q^*$$

は明らか。 □

1.3 ダイバージェンスとリーマン幾何: 接空間

ダイバージェンス関数からは、微小な 2 点 ξ と $\xi + d\xi$ の間に二乗距離

$$ds^2 = 2D[\xi : \xi + d\xi] = \sum g_{ij}(\xi) d\xi_i d\xi_j \quad (1.6)$$

が導入される。空間の各点 ξ に、正定値行列 $G = (g_{ij}(\xi))$ が定義され、微小線素 $d\xi$ の長さの二乗が上式 (1.6) で与えられる空間をリーマン空間という。

ユークリッド空間では、正規直行座標をとれば、微小な線素 $d\xi$ の距離の二乗はピタゴラスの定理によって、

$$ds^2 = \sum d\xi^2$$

と書ける。(この場合 $d\xi$ は微小でなくともよい。) これは、行列 G がどの場所 ξ でも単位行列 I に等しい場合である。空間がユークリッド的であっても、極座標 (r, θ) を用いれば微小な線素の長さの二乗は、

$$ds^2 = dr^2 + r^2(d\theta)^2$$

であるから、この場合行列は

$$G(r, \theta) = \begin{bmatrix} 1 & 0 \\ 0 & r^2 \end{bmatrix}$$

となる。しかし、これは座標系を変えることで単位行列に出来る。

どのような座標系をとっても単位行列に出来ない空間は本質的にユークリッド的ではない。その判定は曲率^{*3}を計算すればできる。これは余裕があれば後に述べる。

リーマン空間の接空間について、軽く触れる。空間は一般に曲がっている。例えば、球面は曲がった 2 次元の空間である。1 点の近傍で、これを線形空間で近似したものを受け空間という。各点に別々の接空間があるから、 ξ での接空間を T_ξ と書く。 ξ 点から出る微小ベクトル $d\xi$ は、ベクトル空間 T_ξ に属すると見える。

ベクトル空間は基底ベクトルで張られる。いま、座標軸に沿った接方向のベクトルを基底ベクトルとし、第 i 方向の基底ベクトルを e_i とすると、

$$d\xi = \sum d\xi_i e_i$$

^{*3} 空間の曲がり具合

と表記できる。

線形近似は M の微小な範囲でしか出来ないが、接空間自体は微小でなくてもよい。接空間のベクトル \mathbf{X} を

$$\mathbf{X} = \sum X_i \mathbf{e}_i$$

と表す。

ベクトルの長さはベクトル空間に内積が定義されれば定まり、線素 $d\xi$ の長さ ds の二乗は、 $\langle \cdot, \cdot \rangle$ を用いて、

$$ds^2 = \langle d\xi, d\xi \rangle = \left\langle \sum d\xi_i \mathbf{e}_i, \sum d\xi_i \mathbf{e}_i \right\rangle$$

と書ける。

2つの基底ベクトルの内積を

$$g_{ij}(\xi) = \langle d\xi, d\xi \rangle = \left\langle \sum d\xi_i \mathbf{e}_i, \sum d\xi_j \mathbf{e}_j \right\rangle \quad (1.7)$$

とすれば、

$$ds^2 = \sum g_{ij} d\xi_i \xi_j$$

となるから、話のつじつまがあう。基底ベクトル系 $\{\mathbf{e}_i\}$ を、座標系 ξ の座標軸 ξ_i に沿った自然基底という。

接空間のベクトル \mathbf{X} の長さの二乗は

$$\|\mathbf{X}\|^2 = \langle \mathbf{X}, \mathbf{X} \rangle = \sum g_{ij} X_i X_j$$

である。2つのベクトル、 \mathbf{X} と \mathbf{Y} は、その内積が 0 のとき、すなわち

$$\langle \mathbf{X}, \mathbf{Y} \rangle = \sum g_{ij} X_i Y_j$$

のとき直交するという。

多様体上の曲線 $\xi(t)$ を考える。 t はパラメータで、 t に従って、点 ξ が変化する。このとき、曲線の接ベクトルは、

$$\|\dot{\xi}\|^2 = \sum g_{ij} \frac{d\xi_i}{dt} \frac{d\xi_j}{dt}.$$

2つの曲線 $\xi_1(t)$ と $\xi_2(t)$ があって、 $t = 0$ で交わっていたとする。すなわち、

$$\xi_1(0) = \xi_2(0)$$

このとき、もし曲線の接ベクトルがこの点で直交していれば、

$$\langle \dot{\xi}_1(0), \dot{\xi}_2(0) \rangle = 0$$

ならば、2つの曲線は直交するという。

線素ベクトル $d\xi$ の長さは、ダイバージェンスを用いて

$$ds^2 = 2D[\xi : \xi + d\xi]$$

で与えられた。凸関数から導かれるダイバージェンスの場合、これをテイラー展開により計算して、

$$g_{ij}(\xi) = \frac{\partial^2}{\partial \xi_i \partial \xi_j} \psi(\xi) \quad (1.8)$$

であることが確かめられる。一方、双対ダイバージェンスと双対座標系を用いれば、同じ線素の長さは

$$\begin{aligned} ds^2 &= \sum g_{ij}^*(\xi^*) d\xi_i^* d\xi_j^*, \\ g_{ij}^*(\xi) &= \frac{\partial^2}{\partial \xi_i^* \partial \xi_j^*} \psi^*(\xi^*) \end{aligned}$$

である。

ところで、 ξ と ξ^* との座標変換のヤコビ行列は

$$\frac{\partial \xi_i^*}{\partial \xi_j} = g_{ij}, \quad \frac{\partial \xi_i}{\partial \xi_j^*} = g_{ij}^*$$

である。従って逆行列の関係

$$G^* = G^{-1}$$

で結ばれている。

1.4 凸関数と双対平坦多様体

1.4.1 凸関数とアファイン座標

アファイン空間は平坦な空間（曲がっていない空間）である。このとき平坦な座標系（アファイン座標系） $\theta = (\theta_1, \dots, \theta_n)$ が存在して、 t をパラメータとするパラメータ表示で表わした曲線 $\theta(t)$ 、

$$\theta(t) = at + b \quad (1.9)$$

を真っ直ぐな線（平坦な線、測地線）と定義する。ここに、 a, b は定ベクトルである。これらはアファイン座標系を θ で表わし、その成分を上付きの添え字で

$$\theta = (\theta^1, \dots, \theta^n)$$

のように表すことにする。

測地線 (1.9) はどの場所でも接線のベクトルが \mathbf{a} で同じである。つまり、方向を変えない。これはユークリッド空間の直線の持つ性質で、その一般化といえる。ユークリッド空間で成立する、2点間を結ぶ長さが最小であるという直線の性質はここでは無視する。距離の最小性と方向不变性という、ユークリッド空間では一致する2つの性質が、凸関数が導く双対平坦なリーマン空間では分離する^{*4}。

一般の座標系を用いて平坦性を議論するには、アファイン接続が必要であり、そこから導かれるリーマン・クリストッフェル曲率が0になるような空間が平坦である。このとき、この空間にアファイン座標系が存在する。ここではいきなり次のように平坦性を定める。

座標系 $\boldsymbol{\theta}$ と凸関数 $\psi(\boldsymbol{\theta})$ が与えられたときに、この空間にアファイン座標系を $\boldsymbol{\theta}$ とする平坦性を導入する。 A を $n \times n$ の非特異な行列、 c を定数ベクトルとして、 $\boldsymbol{\theta}$ から $\tilde{\boldsymbol{\theta}}$ への座標系のアファイン変換

$$\tilde{\boldsymbol{\theta}} = A\boldsymbol{\theta} + \mathbf{c} \quad (1.10)$$

を施す。凸関数は、アファイン変換 (1.8) を施しても凸のままであるから、アファイン不变な概念である。(1.10) の測地線はこの座標でも同じ形にかける。従って、アファイン座標系はアファイン変換の中で自由に選べる。 n 本の座標曲線 $\theta^i (i = 1, \dots, n)$ はみな測地線であるから、これは測地座標系といえる。座標曲線 θ^i の接ベクトルが e_i であった。しかし、接ベクトルの内積は (1.7) であるから、(1.8) から分かるように場所 $\boldsymbol{\theta}$ に依存する。これはユークリッド空間とは異なる^{*5}。

M の部分空間 S が、 A を $m \times n$ 行列で \mathbf{b} を m 次元とする $\boldsymbol{\theta}$ の線形束縛

$$A\boldsymbol{\theta} + \mathbf{b} = 0$$

で定義されるとき、これは平坦な部分空間という。平坦な部分空間の2点を結ぶ測地線は S に含まれる。

1.4.2 双対アファイン座標

凸関数を持つ空間 M は、双対座標 $\boldsymbol{\theta}^*$ と双対な凸関数 $\psi^*(\boldsymbol{\theta}^*)$ をもとに、双対平坦性が定義できる。このとき、 $\boldsymbol{\theta}^*$ が M の双対アファイン座標である。双対測地線は双対座標を用いて

$$\boldsymbol{\theta}^*(t) = at + \mathbf{b} \quad (1.11)$$

のようにパラメータ表示で書ける。もちろん、双対測地線 (1.11) は、測地線 (1.9) とは違う。(1.11) を座標系 $\boldsymbol{\theta}$ で表せば、

$$\nabla\psi\{\boldsymbol{\theta}(t)\} = at + \mathbf{b}$$

^{*4} 線形性と計量とを分けて考え、後にこの2つが双対性で結ばれるところに双対微分幾何の極意があるらしい。

ユークリッド空間ではこの2つが一緒になるので、これは自己双対になっている。測地線とは字義どおりなら2点間の距離を定める線であるから、ここで定義する線 (1.9) は平坦線とよぶ方がよい

^{*5} 平坦性と計量とが分離していることがここからも分かる。

となるので $\theta(t)$ は t の 1 次式では書けない。だから双対測地線はもとの座標系の測地線ではなく曲がっている。

部分空間 S^* が θ^* の線形関係

$$A\theta^* + b = 0$$

で定義されるとき、これを双対平坦な部分空間という。双対平坦な部分空間の 2 点を結ぶ双対測地線はこの中に含まれる。

2 つの座標系 θ と θ^* が双対であるということは、計量に関係している。いま、アファイン座標系 θ を用いた自然基底（座標軸 θ^i 方向のベクトル）を e_i とし、双対座標系 θ^* の自然基底を e^{*i} と書く。また、 θ^* の成分は下付き添え字で

$$\theta^* = (\theta_1^*, \dots, \theta_n^*)$$

のように表す。基底 $\{e_i\}$ も $\{e^{*i}\}$ も一般には正規直行系ではない。ユークリッド空間を除けば、 $\langle e_i, e_j \rangle = \delta_{ij}$ は成立しない。つまり基底ベクトルどうしは直交していない。ところが、基底ベクトル e_i と基底ベクトル e^{*j} はいつも直交している。

定理 1.2. 2 つの双対基底系は双対直交系である。すなわち

$$\langle e_i, e_j^* \rangle = \delta_{ij}.$$

証明。まず、基底ベクトル e_i を、基底系 $\{e_j^*\}$ を使って表す。このとき θ の微小な変化 $d\theta$ と対応する θ^* の微小な変化 $d\theta^*$ を考える。両者は同じものを別の座標系で表したものであり、それぞれの基底を使って、

$$d\theta = \sum d\theta^i e_i = \sum d\theta_i^* e_{*i}$$

と書ける。ところが、 $\theta^* = \nabla\psi(\theta)$ から

$$d\theta_i^* = \sum \partial_i \partial_j \psi d\theta^j = \sum g_{ij} d\theta^j$$

これを用いれば

$$\begin{aligned} e_i &= \sum g_{ij} e^{*j} \\ e^{*j} &= \sum g^{*ji} e_i \end{aligned}$$

が成立する。ここで、内積を

$$\langle e_i, e_j \rangle = g_{ij}, \langle e^{*i}, e^{*j} \rangle = g^{*ij}$$

のように表した。両者は互いに逆行列の関係にある。これは、2 つの基底系が計量の意味で相反的または双対的であることを示している。□

1.5 拡張ピタゴラスの定理

凸関数をもとに、多様体にダイバージェンスと 2 種類の測地線を定義し、双対平坦な空間構造を導くことが出来た、さらに、リーマン計量を定義されて、2 つの直交性も判定できる。これ等のことが揃うと、ユークリッド空間において成立するピタゴラスの定理や射影定理がここでも成立することを示せる。

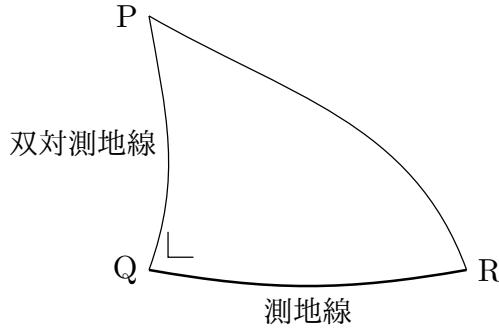


図 2 拡張ピタゴラスの定理

3 点 P, Q, R を考える。 P と Q を結ぶ双対測地線が、 Q と R を結ぶ測地線と直交していたとしよう（図 2）。別に P と Q を結ぶ必要はないが、これを結べば、双対空間における直角三角形が出来上がる。

定理 1.3. 双対平坦空間において直角三角形において、拡張ピタゴラスの定理が成立する。すなわち

$$D[P : R] = D[P : Q] + D[Q : R]$$

となる。

証明。 双対平坦空間において、ダイバージェンスは凸関数を用いて (1.5) のように書ける。これをもとに少し演算すると

$$D[P : Q] + D[Q : R] - D[P : R] = (\theta_P^* - \theta_Q^*) \cdot (\theta_Q - \theta_R)$$

が成立する。一方、 P と Q を結ぶ双対測地線は、

$$\theta^*(t) = (1-t)\theta_P^* + t\theta_Q^*$$

で、その接線方向は $\dot{\theta}^*(t) = \theta_Q^* - \theta_P^*$ である。また、 Q と R を結ぶ測地線は

$$\theta(t) = (1-t)\theta_Q + t\theta_R$$

で、その接線方向は $\dot{\theta}(t) = \theta_R - \theta_Q$ である。この 2 つが直交することから

$$(\theta_P^* - \theta_Q^*) \cdot (\theta_Q - \theta_R) = 0$$

従って、ピタゴラスの定理が成立する。 □

以上のことと用いることで次の双対ピタゴラスの定理が導かれる。

定理 1.4. P と Q を結ぶ測地線が Q と R を結ぶ双対測地線と直交するとき、

$$D^*[P : R] = D^*[P : Q] + D^*[Q : R]$$

が成立する。

もちろん $D^*[P : R] = D[R : P]$ なので上記の定理が成り立つことはすぐわかる。ダイバージェンスの変数の位置を入れ替えるだけですべては双対に出来る。

これ等を拡張ピタゴラスの定理と呼ぶ。これは、ピタゴラスの定理を含み、その拡張になっているからである。すなわち、凸関数が

$$\psi(\boldsymbol{\theta}) = \frac{1}{2} \sum \theta_i^2$$

で与えられたとする。すると、ダイバージェンスはユークリッド距離の二乗であり、アファイン座標系 $\boldsymbol{\theta}$ と双対アファイン座標系 $\boldsymbol{\theta}^*$ は一致するから、測地線も双対測地線も同じでユークリッド空間の直線である。そして、上記の定理はまさしくピタゴラスの定理に他ならない。つまり、この定理はピタゴラスの定理を特別な場合として含んでいる。

1.6 拡張射影定理

ユークリッド空間において、一点 P とこれを通らない滑らかな曲面 S があったとする。点 P に一番近い点を曲面上に探したい。答えとしては点 P を曲面に射影すればよい。射影した点を P_S とする。射影とは P と P_S を結ぶ直線が、曲面と直交する点のことである。その証明にはピタゴラスの定理が使える。すなわち、 P_S からほんの少し離れた曲面上の点を R とすると、ピタゴラスの定理によって、

$$D[P : R] = D[P : P_S] + D[P_S : R] \quad (1.12)$$

となる。よって、 P に一番近い曲面上の点は、射影で得られる P_S である。

上式 (1.12) は P_S が $D[P : R](R \in S)$ のクリティカルポイントであることを示している。しかし、 S が平坦な部分空間ならこれでよいが、 S の形によってこれは極大点や鞍点であったりする。

同様なことが双対平坦空間でもいえる。双対平坦空間で、一点 P とそれを含まない曲面 S を考える(図 3)。 S の点 P_S が、点 P と P_S を結ぶ測地線が曲面と直交するとき、 P_S^* を P の S への双対射影という。測地線が曲面と直交するとは、測地線の接線が曲面の全ての接ベクトルと直交することである。

定理 1.5 (射影定理). 曲面 S が与えられたとき, P から S へのダイバージェンス

$$D[P : S] = \min_{R \in S} D[P : R]$$

を最小にする S の点 P_S は, P の S への双対射影 P_S^* である. 一方, P の S への双対ダイバージェンス

$$D^*[P : S] = \min_{R \in S} D[P : R]$$

を最小にする点 P_S^* は, P の S への射影である.

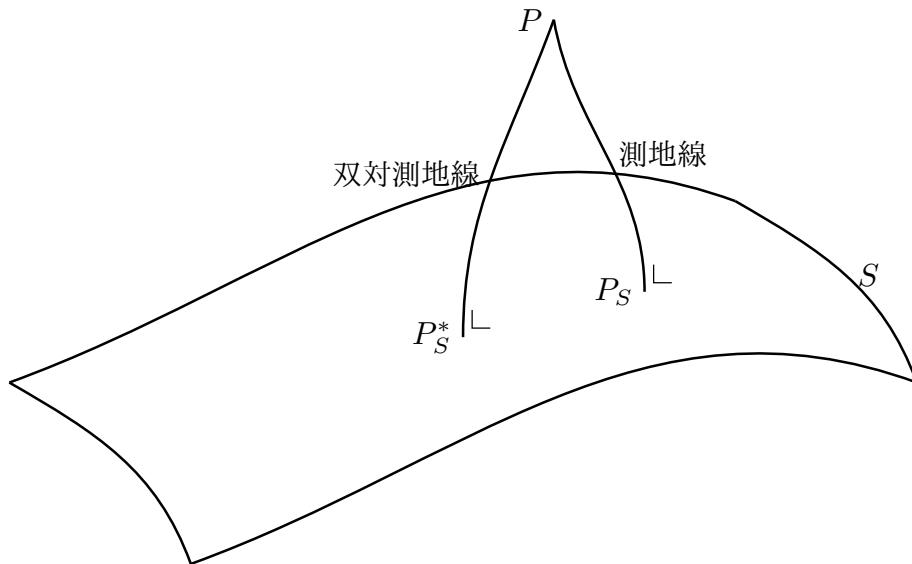


図 3 射影定理

注意 1.1. P から S への最小点を求めるとき, 射影は必要条件であるが十分ではない, 射影 P_S は極大点であったり鞍点であったりすることもある. 次の定理は射影の一意性を保証する.

定理 1.6. S が双対平坦の曲面であるとき, P から S への射影は一意である. また, 平坦な曲面であるとき, S への双対射影は一意である

この定理は, 曲面 S が平坦(双対平坦)のときには, ピタゴラスの定理が大局的に成立することから確かめられる.

今後の課題

今回拡張ピタゴラスの定理とその周辺のことに対して学習したが今後は今回では触れられなかった情報幾何学の応用まで取り掛かりたい.

参考文献

- [1] 甘利 俊一 著, 情報幾何学の新展開, サイエンス社, 2014 年.
- [2] 藤原 彰夫 著, 情報幾何学の基礎, 牧野書店, 2015 年.