

拡張ピタゴラスの定理

BV17077 横井 健

芝浦工業大学 数理科学研究会

November 17, 2019

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.

まず, 今回の発表で使う多様体を定義する.

まず、今回の発表で使う多様体を定義する.

定義 (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.

- ① M は Hausdorff 空間.
- ② 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.

次に多様体の位相的性質と座標系の計量的な性質としてダイバージェンスと呼ぶ2点関数を考える.

定義 (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,

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

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.

注意

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てもよく、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ない、三角不等式も成立しない。式 (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

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

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 (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.

例 (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 (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.).

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

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

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

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

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

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

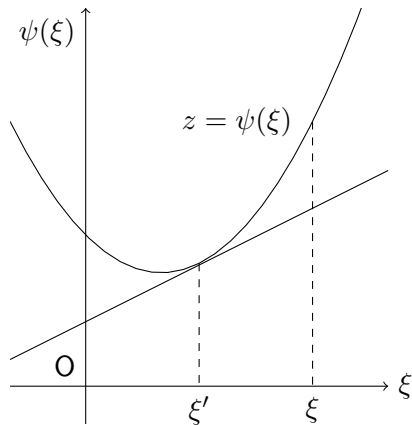
例

$$\psi(\boldsymbol{\xi}) = \frac{1}{2} \sum \xi_i^2 \quad (3)$$

はユークリッド空間の原点からの距離の2乗であり. 2階微分は単位行列である.

凸関数とダイバージェンス

簡単のため1次元における 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.



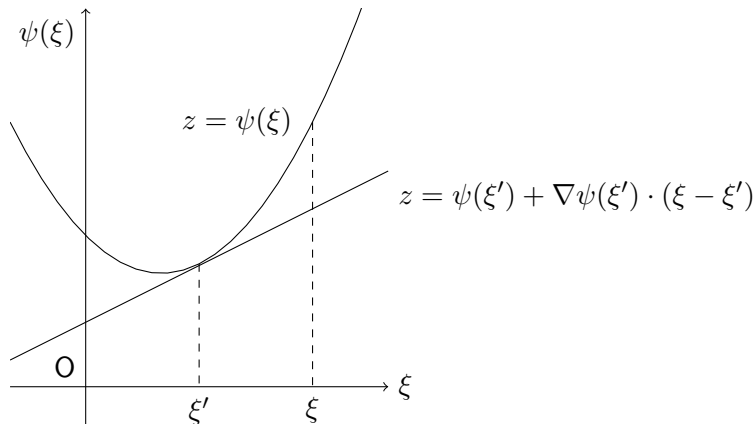
ξ' 点で $z = \psi(\xi')$ に接する超平面の方程式は, z を縦軸として,

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

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

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

凸関数とダイバージェンス



関数 $\psi(\xi)$ が点 ξ で接平面のどれくらい上にあるかは,

$$D[\xi : \xi'] = \psi(\xi) - \psi(\xi') - \nabla\psi(\xi') \cdot (\xi - \xi') \quad (4)$$

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.

式 (3) 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.

凸関数 $\psi(\xi)$ が与えられたとき, この関数の勾配は式 (4) から,

$$\xi^* = \nabla\psi(\xi^*) \quad (5)$$

である.

凸関数 $\psi(\xi)$ が与えられたとき, この関数の勾配は式 (4) から,

$$\xi^* = \nabla\psi(\xi^*) \quad (5)$$

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 (5) 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.

凸関数 $\psi(\xi)$ が与えられたとき, この関数の勾配は式 (4) から,

$$\xi^* = \nabla\psi(\xi^*) \quad (5)$$

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 (5) 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 (5)) を **Legendre 変換** という. この変換は双対である. つまり,

$$\psi^*(\xi^*) = \max_{\xi} \{\xi \cdot \xi^* - \psi(\xi)\} \quad (6)$$

式 (6) によって, ξ の関数 ψ^* を定義する. いま, (6) 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 ξ は最大値を達成するから (6) の右辺を ξ^* で微分すれば, $\xi^* = \nabla \psi(\xi)$ を満たす. (6) を ξ^* で微分すると,

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

式 (6) によって, ξ の関数 ψ^* を定義する. いま, (6) 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 ξ は最大値を達成するから (6) の右辺を ξ^* で微分すれば, $\xi^* = \nabla \psi(\xi)$ を満たす. (6) を ξ^* で微分すると,

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

なので, これが ξ^* から ξ への逆変換となっている.

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.

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

$\psi(\boldsymbol{\theta})$ が凸関数となるようなアフィン座標 $\boldsymbol{\theta}$ を用い、2点 P, Q のアフィン座標を $\boldsymbol{\theta}_P, \boldsymbol{\theta}_Q$ 、双対アフィン座標を $\boldsymbol{\theta}_P^*, \boldsymbol{\theta}_Q^*$ とすると、

$$D[P : Q] = \psi(\boldsymbol{\theta}_P) + \psi^*(\boldsymbol{\theta}_Q^*) - \boldsymbol{\theta}_P \cdot \boldsymbol{\theta}_Q^* \quad (7)$$

が成り立つ。

証明の概要

点 P を固定して, 関数 $Q \mapsto \psi^*(\boldsymbol{\xi}_Q^*) = \boldsymbol{\theta}_Q \cdot \boldsymbol{\theta}_P^* - \psi(\boldsymbol{\theta}_Q)$ とする. これを微分すると,

$$\begin{aligned}d(\psi^*(\boldsymbol{\xi}_Q^*)) &= (\boldsymbol{\theta}_P^* - \nabla\psi(\boldsymbol{\theta}_Q))d\boldsymbol{\theta}_Q \\ &= (\boldsymbol{\theta}_P^* - \boldsymbol{\theta}_Q^*)d\boldsymbol{\theta}_Q\end{aligned}$$

となる.

証明の概要

点 P を固定して, 関数 $Q \mapsto \psi^*(\boldsymbol{\xi}_Q^*) = \boldsymbol{\theta}_Q \cdot \boldsymbol{\theta}_P^* - \psi(\boldsymbol{\theta}_Q)$ とする. これを微分すると,

$$\begin{aligned} d(\psi^*(\boldsymbol{\xi}_Q^*)) &= (\boldsymbol{\theta}_P^* - \nabla\psi(\boldsymbol{\theta}_Q))d\boldsymbol{\theta}_Q \\ &= (\boldsymbol{\theta}_P^* - \boldsymbol{\theta}_Q^*)d\boldsymbol{\theta}_Q \end{aligned}$$

となる. よって, 式 (6) の右辺が最大となるのは $\boldsymbol{\theta}_P^* = \boldsymbol{\theta}_Q^*$, 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 $\boldsymbol{\theta}_P = \boldsymbol{\theta}$ とすると,

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

である.

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

$$ds^2 = 2D[\xi : \xi + d\xi] = \sum g_{ij}(\xi) d\xi_i d\xi_j \quad (8)$$

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 (8) 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。その判定は曲率を計算すればできる。

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. 1 点の近傍で, これを線形空間で近似したものを接空間という. 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.

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. 1点の近傍で, これを線形空間で近似したものを接空間という. 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,

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

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ともよい. 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

$$g_{ij}(\xi) = \langle d\xi, d\xi \rangle = \left\langle \sum d\xi_i e_i, \sum d\xi_j e_j \right\rangle \quad (9)$$

とすれば,

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

となるから, 話のつじつまがあう. 基底ベクトル系 $\{e_i\}$ 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、

$$g_{ij}(\xi) = \frac{\partial^2}{\partial \xi_i \partial \xi_j} \psi(\xi) \quad (10)$$

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

$$ds^2 = \sum g_{ij}^*(\boldsymbol{\xi}^*) d\xi_i^* d\xi_j^*,$$

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

である。

ところで、 $\boldsymbol{\xi}$ と $\boldsymbol{\xi}^*$ 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。

アフィ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 (アフィン座標系) $\theta = (\theta_1, \dots, \theta_n)$ 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 $\theta(t)$,

$$\theta(t) = at + b \quad (11)$$

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. これらはアフィン座標系を 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 (11) はどの場所でも接線のベクトルが a で同じである。つまり、方向を変えない。

一般の座標系を用いて平坦性を議論するには、アフィ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、この空間にアフィン座標系が存在する。今回は次のように平坦性を定める。

座標系 θ と凸関数 $\psi(\theta)$ が与えられたときに、この空間にアフィン座標系を θ 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、 θ から $\tilde{\theta}$ への座標系のアフィン変換

$$\tilde{\theta} = A\theta + c \quad (12)$$

を施す。凸関数は、アフィン変換 (10) を施しても凸のままであるから、アフィン不変な概念である。(12) の測地線はこの座標でも同じ形にかける。従って、アフィン座標系はアフィ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 (9) であるから、(10) から分かるように場所 θ に依存する。

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 θ の線形束縛

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

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.

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

$$\theta^*(t) = at + b \quad (13)$$

のようにパラメータ表示で書ける.

凸関数を持つ空間 M は、双対座標 θ^* と双対な凸関数 $\psi^*(\theta^*)$ をもとに、双対平坦性が定義できる。このとき、 θ^* 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

$$\theta^*(t) = at + b \quad (13)$$

のようにパラメータ表示で書ける。

もちろん、双対測地線 (13) は、測地線 (11) とは違う。(13) を座標系 θ で表せば、

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

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.

このことから次の定理が言える.

定理

2つの双対基底系は双対直交系である. すなわち

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

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1). 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.

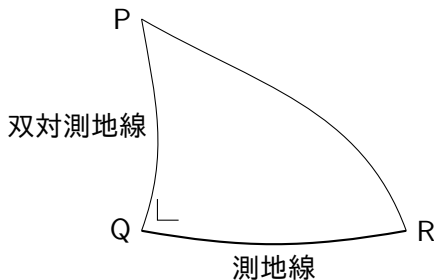


Figure: 拡張ピタゴラスの定理

定理

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.

証明の概要 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 (7) のように書くと、

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

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.

定理

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.

定理

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.

定理

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。

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。実際、凸関数が

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

で与えられたとする。

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。実際、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、アフィン座標系 $\boldsymbol{\theta}$ と双対アフィン座標系 $\boldsymbol{\theta}^*$ 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。

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。実際、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、アフィン座標系 $\boldsymbol{\theta}$ と双対アフィン座標系 $\boldsymbol{\theta}^*$ 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。

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。

- 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.