第 2 部では、情報幾何の基本的な概念を整理する。まず第 1 節と第 2 節では具体例を扱う。第 1 節では有限集合上の測度の空間を考える。測度の空間を多様体とみなし、さらに統計的に自然な要請に基づいて計量や接続、ダイバージェンスなどの概念を導入する。第 2 節では行列多様体を考える。

第 3 節と第 4 節では、パラメータ空間 [TODO] パラメータ空間は第 1 節と第 2 節ではまだ現れていないのでは？の内在的な幾何構造として双対構造と双対平坦性を整理する。

第 5 節では無限集合上の測度の空間を考える。

第 6 節では統計的推定について考える。

第 1 章　有限集合上の測度の空間

この章では有限集合上の確率分布について考える。有限集合上の確率分布を扱う目的は、これが情報幾何学の最も基本的な具体例だからというだけでなく、パラメータ空間の内在的な幾何や無限集合上の確率分布を扱う際に、幾何学的直感の基礎になると期待できるからである。

この章では全体を通して $n \in \mathbb{Z}_{\geq 0}$, $\mathcal{X} \coloneqq \{1, \ldots, n+1\}$, $\mathcal{A} \coloneqq 2^{\mathcal{X}}$ とし、可測空間 $(\mathcal{X}, \mathcal{A})$ を考える。イメージとしては、$(\mathcal{X}, \mathcal{A})$ は何らかの確率変数の (定義域というよりも) 値域を表している。いまはそれが有限集合ということだから、たとえば「コインの裏表」「サイコロの出目」などといった、値が有限通りしかない確率変数を考えているという雰囲気である。

1.1　有限集合上の測度の空間

まず $\mathcal{X}$ 上の実数値関数全体の集合を考える。

定義 1.1.1. $\mathbb{R}$-代数 $\mathcal{F}(\mathcal{X})$ を

$$\mathcal{F}(\mathcal{X}) \coloneqq \{f \colon \mathcal{X} \to \mathbb{R}\} \tag{1.1.1}$$

と定め、$\mathcal{F}(\mathcal{X})$ の $\mathbb{R}$-ベクトル空間としての (代数的) 双対空間 $\mathcal{F}(\mathcal{X})^*$ を $\mathcal{S}(\mathcal{X})$ とおく。

先に注意しておくと、この章で主役となるのは $\mathcal{F}(\mathcal{X})$ よりもむしろ $\mathcal{S}(\mathcal{X})$ の方である。さて、$\mathcal{F}(\mathcal{X})$ は次の標準的な基底を持つ。

命題 1.1.2 ($\mathcal{F}(\mathcal{X})$ の標準基底)**.** 各 $i \in \mathcal{X}$ に対して

$$e_i(j) \coloneqq \begin{cases} 1 & (j = i) \\ 0 & (j \neq i) \end{cases} \quad (j \in \mathcal{X}) \tag{1.1.2}$$

と定めると、$(e_1, \ldots, e_{n+1})$ は $\mathcal{F}(\mathcal{X})$ の $\mathbb{R}$ 上の基底となる。

証明　$\mathbb{R}$-線型写像 $\mathbb{R}^{n+1} \to \mathcal{F}(\mathcal{X})$, $(x^1, \ldots, x^{n+1}) \mapsto \sum_{i \in \mathcal{X}} x^i e_i$ が写像 $\mathcal{F}(\mathcal{X}) \to \mathbb{R}^{n+1}$, $f \mapsto (f(1), \ldots, f(n+1))$ を逆写像として同型を与える。 □

そこで $(e_1, \ldots, e_{n+1})$ の双対基底を $(\delta^1, \ldots, \delta^{n+1})$ とおく。すると $\mathcal{S}(\mathcal{X})$ について次が成り立つ。

命題 1.1.3 ($\mathcal{S}(\mathcal{X})$ と符号付き測度の空間)**.** $\mathcal{S}(\mathcal{X})$ は $\mathcal{X}$ 上の符号付き測度全体の空間とみなせる。より詳しく言えば、$\mathcal{X}$ 上の符号付き測度全体の空間は、要素ごとの和とスカラー倍で $\mathbb{R}$-ベクトル空間となり、写像

$$\Phi \colon \mathcal{S}(\mathcal{X}) \to \{\mathcal{X} \text{ 上の符号付き測度}\}, \quad \omega = \sum_{i \in \mathcal{X}} \omega_i \delta^i \mapsto \left(A \mapsto \sum_{i \in A} \omega_i\right) \tag{1.1.3}$$

が $\mathbb{R}$-線型同型を与える。

証明 Φ が $\mathbb{R}$-線型写像であることの証明はルーティンワークである。逆写像は

$$\Psi\colon \{\mathcal{X}\ \text{上の符号付き測度}\} \to \mathcal{S}(\mathcal{X}), \quad \mu \mapsto \sum_{i\in\mathcal{X}} \mu(\{i\})\delta^i \tag{1.1.4}$$

で与えられる。$\Psi\circ\Phi(\omega) = \omega$ の証明に ω の線型性を用いる。$\Phi\circ\Psi(\mu) = \mu$ の証明に μ の加法性を用いる。
[TODO] もう少しちゃんと書く □

上の命題の同型により、$\mathcal{F}(\mathcal{X})$ 上の線型形式 δ^i は、$\mathcal{X}$ 上の Dirac 測度 $\Phi(\delta^i)$ と同一視できる。そこで記号の濫用により、以後 $\Phi(\delta^i)$ も δ^i と書くことにする。

さて、$\mathcal{S}(\mathcal{X})$ の部分集合を定義する。

定義 1.1.4 ($\mathcal{S}(\mathcal{X})$ の部分集合)**.** $\mathcal{S}(\mathcal{X})$ の部分集合を次のように定義する:

$$\mathcal{M}(\mathcal{X}) := \left\{\mu = \sum_{i\in\mathcal{X}} \mu_i\delta^i \in \mathcal{S}(\mathcal{X}) \,\middle|\, \mu_i \geq 0\right\}, \tag{1.1.5}$$

$$\mathcal{P}(\mathcal{X}) := \left\{\mu = \sum_{i\in\mathcal{X}} \mu_i\delta^i \in \mathcal{M}_+(\mathcal{X}) \,\middle|\, \sum_{i\in\mathcal{X}} \mu_i = 1\right\} \tag{1.1.6}$$

すなわち、$\mathcal{M}_+(\mathcal{X})$ は $\mathcal{X}$ 上の有限測度全体の集合であり、$\mathcal{P}_+(\mathcal{X})$ は $\mathcal{X}$ 上の確率測度全体の集合である。また、次のように定義する:

$$\mathcal{M}_+(\mathcal{X}) := \left\{\mu = \sum_{i\in\mathcal{X}} \mu_i\delta^i \in \mathcal{S}(\mathcal{X}) \,\middle|\, \mu_i > 0\right\}, \tag{1.1.7}$$

$$\mathcal{P}_+(\mathcal{X}) := \left\{\mu = \sum_{i\in\mathcal{X}} \mu_i\delta^i \in \mathcal{M}_+(\mathcal{X}) \,\middle|\, \sum_{i\in\mathcal{X}} \mu_i = 1\right\} \tag{1.1.8}$$

すなわち、$\mathcal{M}_+(\mathcal{X})$ は $\mathcal{X}$ 上の正の測度全体の集合であり、$\mathcal{P}_+(\mathcal{X})$ は $\mathcal{X}$ 上の正の確率測度全体の集合である。

$\mathcal{M}_+(\mathcal{X})$ および $\mathcal{P}_+(\mathcal{X})$ は、次の意味で部分多様体になっている。

命題 1.1.5. $\mathcal{M}_+(\mathcal{X})$ は $\mathcal{S}(\mathcal{X})$ の開部分多様体であり、$\mathcal{P}_+(\mathcal{X})$ は $\mathcal{M}_+(\mathcal{X})$ の n 次元部分多様体である。

証明 多様体としての同一視 $\mathcal{S}(\mathcal{X}) = \mathbb{R}^{n+1}$ のもとで $\mathcal{M}_+(\mathcal{X}) = \mathbb{R}^{n+1}_{>0} \overset{\text{open}}{\subset} \mathbb{R}^{n+1} = \mathcal{S}(\mathcal{X})$ が成り立つから、$\mathcal{M}_+(\mathcal{X})$ は $\mathcal{S}(\mathcal{X})$ の開部分多様体である。また、C^∞ 写像

$$F\colon \mathcal{M}_+(\mathcal{X}) = \mathbb{R}^{n+1}_{>0} \to \mathbb{R}, \quad (\mu_1, \ldots, \mu_{n+1}) \mapsto \mu_1 + \cdots + \mu_{n+1} \tag{1.1.9}$$

の Jacobi 行列は $JF_{(\mu_1,\ldots,\mu_{n+1})} = [1\ldots1]$ だから、微分 $dF_{(\mu_1,\ldots,\mu_{n+1})}$ は各点 $(\mu_1, \ldots, \mu_{n+1}) \in \mathbb{R}^{n+1}_{>0}$ で全射であり、したがってとくに F は 1 を正則値にもつ。よって $\mathcal{P}_+(\mathcal{X}) = F^{-1}(0)$ は $\mathcal{M}_+(\mathcal{X})$ の $\dim\mathcal{M}_+(\mathcal{X}) - \dim\mathbb{R} = n$ 次元部分多様体である。 □

それぞれの多様体の接空間は次のように表される。

命題 1.1.6 ($\mathcal{S}(\mathcal{X}), \mathcal{M}_+(\mathcal{X}), \mathcal{P}_+(\mathcal{X})$ の接空間).

$$T_\mu \mathcal{S}(\mathcal{X}) = \mathcal{S}(\mathcal{X}), \tag{1.1.10}$$

$$T_\mu \mathcal{M}_+(\mathcal{X}) = \mathcal{S}(\mathcal{X}), \tag{1.1.11}$$

$$T_\mu \mathcal{P}_+(\mathcal{X}) = \mathcal{S}_0(\mathcal{X}) \tag{1.1.12}$$

ただし

$$\mathcal{S}_0 := \left\{ a = \sum_{i \in \mathcal{X}} a_i \delta^i \in \mathcal{S}(\mathcal{X}) \,\middle|\, \sum_{i \in \mathcal{X}} a_i = 0 \right\} \tag{1.1.13}$$

とおいた。

証明 $T_\mu \mathcal{S}(\mathcal{X}) = \mathcal{S}(\mathcal{X})$ および $T_\mu \mathcal{M}_+(\mathcal{X}) = \mathcal{S}(\mathcal{X})$ は、一般にベクトル空間の開部分多様体の接空間がもとのベクトル空間に一致することから従う。また、命題 1.1.5 の証明の写像 F を用いると、部分多様体の接空間の一般論より $T_\mu \mathcal{P}_+(\mathcal{X}) = \operatorname{Ker} dF_\mu$ が成り立つが、ここで

$$dF_{(\mu_1,\dots,\mu_{n+1})} \colon \mathbb{R}^{n+1} \to \mathbb{R}, \quad \begin{pmatrix} a_1 \\ \vdots \\ a_{n+1} \end{pmatrix} \mapsto JF_{(\mu_1,\dots,\mu_{n+1})} \begin{pmatrix} a_1 \\ \vdots \\ a_{n+1} \end{pmatrix} = a_1 + \cdots + a_{n+1} \tag{1.1.14}$$

より $T_\mu \mathcal{P}_+(\mathcal{X}) = \operatorname{Ker} dF_\mu = \mathcal{S}_0(\mathcal{X})$ が従う。 □

1.2 Fisher 計量

$\mathcal{S}(\mathcal{X})$ に Riemann 計量を定義する。考えられる Riemann 計量は無数に存在するが、ここでは Fisher 計量と呼ばれる計量を導入する [TODO] なぜ？。$\mathcal{M}_+(\mathcal{X})$ および $\mathcal{P}_+(\mathcal{X})$ には誘導計量を入れる。

定義 1.2.1 (Fisher 計量). 各 $\mu = \mu_i \delta^i \in \mathcal{M}_+(\mathcal{X})$, $a = a_i \delta^i, b = b_i \delta^i \in \mathcal{S}(\mathcal{X})$ に対し

$$\langle a, b \rangle_\mu := \int_{\mathcal{X}} \frac{da}{d\mu} \frac{db}{d\mu} \, d\mu = \sum_{i \in \mathcal{X}} \frac{1}{\mu_i} a_i b_i \tag{1.2.1}$$

と定義する。$\mathcal{M}_+(\mathcal{X})$ 上の Riemann 計量 $\mathfrak{g}$ を

$$\mathfrak{g}_\mu(a, b) := \langle a, b \rangle_\mu \quad (\mu \in \mathcal{M}_+(\mathcal{X}),\ a, b \in T_\mu \mathcal{M}_+(\mathcal{X})) \tag{1.2.2}$$

と定義し、これを **Fisher 計量 (Fisher metric)** と呼ぶ。

$\mathcal{P}_+(\mathcal{X})$ 上に誘導される Fisher 計量は次の形で与えられる。

命題 1.2.2 ($\mathcal{P}_+(\mathcal{X})$ 上の Fisher 計量).

$$g_{ij}(\mu) = \sum_{k=1}^{n} \frac{1}{\mu_k} \delta_{ki} \delta_{kj} + \frac{1}{\mu_{n+1}} \tag{1.2.3}$$

[TODO]

証明 [TODO] □

A. 等長変換群

[TODO]

1.3 Chentsov の定理

前節で導入した Fisher 計量が、統計的に自然な計量であることを示す。本節では $n' \in \mathbb{Z}_{\geq 0}$, $\mathcal{X}' := \{1, \ldots, n'+1\}$, $\mathcal{A}' := 2^{\mathcal{X}'}$ とし、可測空間 $(\mathcal{X}', \mathcal{A}')$ を $(\mathcal{X}, \mathcal{A})$ とあわせて考える。まず、統計量と Markov カーネルの概念を導入する。

定義 1.3.1 (統計量). $\mathcal{X} \to \mathcal{X}'$ なる写像 κ を**統計量 (statistic)** という [1)]。

定義 1.3.2 (Markov カーネル). $\mathcal{X} \to \mathcal{P}(\mathcal{X}')$ なる写像 K を **Markov カーネル (Markov kernel)** といい、ここだけの記法として $K(i)$ $(i \in I)$ の成分表示を

$$K(i) = \sum_{i' \in \mathcal{X}'} K(i)_{i'} \delta^{i'} \tag{1.3.1}$$

と書くことにする。

統計量 $\kappa \colon \mathcal{X} \to \mathcal{X}'$ が与えられると、$K^{\kappa}(i) := \delta^{\kappa(i)}$ $(i \in I)$ とおくことで Markov カーネル K^{κ} が定義される。ただし右辺は $\kappa(i)$ を台とする Dirac 測度である。この意味で、Markov カーネルは統計量の一般化になっている。

次に Markov カーネルの congruent 性を定義する。

定義 1.3.3 (congruent な Markov カーネル). [TODO]

例 1.3.4 (congruent な Markov カーネルの例). $\mathcal{X} := \{0, 1\}$, $\mathcal{X}' := \{1, 2, 3, 4, 5, 6\}$ とし、Markov カーネル $K \colon \mathcal{X} \to \mathcal{P}(\mathcal{X}')$ を [TODO]

次の定理により、$\mathcal{P}_+(\mathcal{X})$ 上の統計的に自然な計量は Fisher 計量に限られることがわかる。

定理 1.3.5 (Chentsov; Ченцов). すべての空でない有限集合 $\mathcal{X}$ に対し、$\mathcal{P}_+(\mathcal{X})$ 上の Riemann 計量 $h^{\mathcal{X}}$ が与えられているとする。このとき、任意の congruent な Markov カーネル $K \colon \mathcal{X} \to \mathcal{P}(\mathcal{X}')$ ($\mathcal{X}'$ は空でない有限集合) が不変ならば、ある正定数 $\alpha > 0$ が存在して、すべての $\mathcal{X}$ に対し $h^{\mathcal{X}} = \alpha g^{\mathcal{X}}$ が成り立つ。

証明 [TODO] □

1.4 m-接続と e-接続

この節では m-接続と e-接続を導入する。まず $T_\mu \mathcal{M}_+(\mathcal{X})$ や $T_\mu \mathcal{M}_+^*(\mathcal{X})$ の定義から、m, e-平行移動が自ずと現れることをみる。m, e-平行移動から、それぞれに対応する m, e-接続の概念が、ひいては m, e-測地線、m, e-指数写像の概念が導かれる。

[TODO] m-平坦と e-平坦には統計的にどんな意味がある？

1) より一般には κ が可測写像であることも要求されるが、いまは $\mathcal{A} = 2^{\mathcal{X}}$ だから、この要求は自ずから満たされる。

定義 1.4.1 (m, e-平行移動). [TODO]

定義 1.4.2 (m, e-接続). [TODO]

命題 1.4.3 (m, e-測地線と指数写像). [TODO]

証明 [TODO] □

A. アファイン群

[TODO]

1.5 Amari-Chentsov テンソル

[TODO] なんのためにある？ Skewness テンソルとも呼ぶから、3 次のキュムラント (歪度) と関係あり？

定義 1.5.1 (Amari-Chentsov テンソル). [TODO]

定義 1.5.2 (α-接続). [TODO]

A. アファイン群

[TODO]

1.6 ダイバージェンス

前節までに、$\mathcal{M}_+(\mathcal{X})$ 上の幾何学的構造として、Fisher 計量 $\mathfrak{g}$ や m, e-接続 $\nabla^{(m)}, \nabla^{(e)}$ を扱った。本節ではこれらの幾何学的構造がダイバージェンスと呼ばれるひとつの関数により誘導されることをみる。ダイバージェンスとは、空間の 2 点の相違性を表す関数であり、イメージとしては距離のようなものである。最適輸送理論における輸送コストにも似ているが、一般にはダイバージェンスの方が粗いものである [TODO] 本当に?。

定義 1.6.1 (KL ダイバージェンス). [TODO]

命題 1.6.2 (KL ダイバージェンスは Fisher 計量を誘導する). [TODO]

証明 [TODO] □

A. α-ダイバージェンス

Fisher 計量を誘導するダイバージェンスは、実は KL ダイバージェンスだけではない。α-ダイバージェンスと呼ばれる、より広いクラスのダイバージェンスが Fisher 計量を誘導する。

定義 1.6.3 (α-ダイバージェンス)**.**

$$D^{(\alpha)}(\mu\|\nu) := \frac{2}{1-\alpha}\sum_{i\in X}\nu_i + \frac{2}{1+\alpha}\sum_{i\in X}\mu_i - \frac{4}{1-\alpha^2}\sum_{i\in X}\mu_i^{\frac{1+\alpha}{2}}\nu_i^{\frac{1-\alpha}{2}} \tag{1.6.1}$$

[TODO]

命題 1.6.4 (α-ダイバージェンスは Fisher 計量を誘導する)**.** [TODO]

証明 [TODO] □

KL ダイバージェンスは α-ダイバージェンスの $\alpha = -1$ の場合である。

命題 1.6.5. KL ダイバージェンスは α-ダイバージェンスの $\alpha = -1$ の場合である。

証明 [TODO] □

B. f-ダイバージェンス

ダイバージェンスを最小化したいという最適化問題への応用を考えると、凸関数を用いて定義されるダイバージェンスのクラスは重要である [TODO] そうなの？。

定義 1.6.6 (f-ダイバージェンス)**.** [TODO]

1.7 指数型分布族

[TODO]

第 2 章　パラメトリックモデル

本章では、有限個のパラメータで特徴付けられるパラメトリックモデルについて述べる。パラメトリックモデルの定式化は **[?**] を参考にした。

2.1　パラメトリックモデル

[TODO]

第 3 章　双対構造

第 1 章では、有限集合上の測度の空間 $\mathcal{M}_+(\mathcal{X})$ について調べた。また $\mathcal{M}_+(\mathcal{X})$ に入る幾何学的構造としては、Fisher 計量 $\mathfrak{g}$ と m, e-接続 $\nabla^{(m)}, \nabla^{(e)}$ を扱ったのであった。本章ではこの構造を抽象化して整理する。すなわち Fisher 計量および m, e-接続という具体的対象から離れて、Riemann 計量 g と g に関し互いに双対的なアファイン接続 ∇, ∇^* の 3 つ組 (g, ∇, ∇^*) について考える。この 3 つ組は双対構造と呼ばれる。本章ではダイバージェンスにより双対構造が誘導されることをみる。

3.1　双対構造

双対構造とは、Riemann 計量 g と g に関し互いに双対的なアファイン接続 ∇, ∇^* の 3 つ組 (g, ∇, ∇) のことである。双対アファイン接続の概念は、計量を保つアファイン接続の概念の一般化である。

定義 3.1.1 (双対アファイン接続). (M, g) を擬 Riemann 多様体、∇, ∇^* を M のアファイン接続とする。∇^* が g に関する ∇ の**双対アファイン接続 (dual affine connection)** であるとは、

$$X(g(Y, Z)) = g(\nabla_X Y, Z) + g(Y, \nabla^*_X Z) \quad (X, Y, Z \in \mathfrak{X}(M)) \tag{3.1.1}$$

が成り立つことをいう。このとき (g, ∇, ∇^*) を M の**双対構造 (dualistic structure)** という。

命題 3.1.2 (双対アファイン接続の存在と一意性). [TODO]

証明　[TODO] 構成は [? , p.96] を参照

一意性を示す。∇', ∇'' を g に関する ∇ の双対アファイン接続とすると、各 $X, Z \in \mathfrak{X}(M)$ に対して

$$0 = g(Y, \nabla'_X Z) - g(Y, \nabla''_X Z) = g(Y, \nabla'_X Z - \nabla''_X Z) \quad (\forall Y \in \mathfrak{X}(M)) \tag{3.1.2}$$

が成り立つから、g の非退化性より $\nabla'_X Z = \nabla''_X Z$ となる。したがって $\nabla' = \nabla''$ である。□

とくに ∇ が計量 g を保つ接続ならば ∇ 自身も ∇ の双対アファイン接続となるから、一意性より $\nabla^* = \nabla$ が成り立つ。

双対アファイン接続は次のような基本性質を持つ。

命題 3.1.3 (双対アファイン接続の基本性質). [TODO]

証明　[TODO] □

命題 3.1.4 (双対アファイン接続と平行移動). [TODO]

証明　[TODO] □

次の定理は情報幾何学の基本定理である [?]。[TODO] なぜ？曲率一定だと測地線の表示がシンプルになるから？

命題 3.1.5 (双対アファイン接続と曲率). [TODO] torsion-free の仮定は不要！

証明 [? , p.226] による。[TODO] □

ところで、双対アファイン接続の概念を Levi-Civita 接続と比較してみたとき、次の定義の意味での「強い」双対性が想起されるかもしれない。

定義 3.1.6 (強い双対アファイン接続). [TODO]

しかし、実は強い双対アファイン接続は Levi-Civita 接続に限られるため、これは文字通り強すぎる制約である。

命題 3.1.7 (強い双対アファイン接続は Levi-Civita 接続). [TODO]

証明 [TODO] □

3.2 統計的構造

有限集合上の測度の空間には、Fisher 計量 $\mathfrak{g}$ と Amari-Chentsov テンソル T の組 $(\mathfrak{g}, T)$ によっても幾何学的構造が定まるのであった。本節ではこれを抽象化して統計的構造と呼ばれるものを導入する。後で見るように統計的構造は、捩れのないアファイン接続の対からなる双対構造と等価である。

[TODO] 捩れなしであることはなぜ重要なのか？そもそもねじれとは何か？

定義 3.2.1 (統計的構造). [TODO]

命題 3.2.2 (双対アファイン接続と Levi-Civita 接続).

$$\frac{1}{2}(\nabla + \nabla^*) = \nabla^{\mathrm{LC}} \tag{3.2.1}$$

[TODO]

証明 [TODO] □

3.3 ダイバージェンス

ダイバージェンスの概念を定義する。ダイバージェンスは双対構造を誘導する方法の 1 つであり、またダイバージェンスにより誘導されるアファイン接続は捩れなしである。同じことだが、ダイバージェンスは統計的構造を誘導する。

定義 3.3.1 (ダイバージェンス). M を多様体とする。C^∞ 関数

$$D\colon M \times M \to \mathbb{R}, \quad (p, q) \mapsto D(p \| q) \tag{3.3.1}$$

が**ダイバージェンス (divergence)** であるとは、次が成り立つことをいう:

(1) 任意の $p, q \in M$ に対し $D(p\|q) \geq 0$ である。
(2) $p, q \in M$ に関し $D(p\|q) = 0 \iff p = q$ である。
(3) [TODO]

ただし、以下の記法を用いる:

- [TODO]

命題 3.3.2. (1) $D[\partial_i\|\cdot] = D[\cdot\|\partial_i] = 0$ である。
(2) $D[\partial_i\partial_j\|\cdot] = D[\cdot\|\partial_i\partial_j] = -D[\partial_i\|\partial_j]$ である。

証明は **[?]** によった。

証明 (1) [TODO]
(2) [TODO] □

定義 3.3.3 (ダイバージェンスから誘導される双対構造)**.** [TODO]

ダイバージェンスの分解可能性を定義する。

定義 3.3.4 (ダイバージェンスの分解可能性)**.** [TODO]

第 4 章　双対平坦性

情報幾何学では、双対構造がさらに双対平坦な場合を考えることが多い。[TODO] why?

4.1　平坦性とアファインチャート

[TODO] アファイン座標と正規座標はどう違う？正規座標は局所的には必ず存在するが...

定義 4.1.1 (アファインチャート [2]). M を多様体、∇ を M のアファイン接続とする。M のチャート (U, φ) が **∇-アファインチャート (∇-affine chart)** であるとは、(U, φ) により定まる TM の局所フレームに関する ∇ の接続形式が 0 であることをいう。

命題 4.1.2 (アファインチャートの存在). M を多様体、∇ を M のアファイン接続とする。このとき次は同値である:

(1)　M は ∇-平坦である。
(2)　各 $p \in M$ に対し、p のまわりのアファインチャートが存在する。

証明　[TODO] □

平坦性とアファインチャートのアナロジーで双対平坦性と双対アファインチャートを定義する。

定義 4.1.3 (双対平坦). [TODO]

定義 4.1.4 (双対アファインチャート). [TODO]

命題 4.1.5 (双対アファインチャートの存在). [TODO]

証明　[TODO] □

平行移動と測地線は、双対アファイン座標によって非常に簡単な形に書ける。

命題 4.1.6 (平行移動と測地線の双対アファイン座標表示). [TODO]

証明　[TODO] □

双対平坦な双対構造を誘導するダイバージェンスは平坦であるという。

2)　「アファインチャート」という名前は [?] で使われている「アファイン座標近傍」に由来する。

定義 4.1.7 (ダイバージェンスの平坦性). [TODO]

4.2 凸関数と Legendre 変換

[TODO] geodesically convex とはどう違う？

まず通常の凸関数の定義を復習する。

命題 4.2.1 (凸関数). A を $\mathbb{R}$-ベクトル空間の凸部分集合とする。関数 $f\colon A \to \mathbb{R}$ が**凸 (convex)** であるとは、任意の $x, y \in A$, $x \neq y$ および任意の $t \in (0,1) \subset \mathbb{R}$ に対して

$$f(tx + (1-t)y) < tf(x) + (1-t)f(y) \tag{4.2.1}$$

が成り立つことをいう。

注意 4.2.2. 関数の凸性は座標に依存する。この意味を理解するため、多様体 $M = \mathbb{R}_{>0}$ 上の関数 $f\colon \mathbb{R}_{>0} \to \mathbb{R}$ の座標表示について考えてみよう。実は、f がある座標で凸であったとしても別の座標でも凸とは限らない。たとえば $f\colon \mathbb{R}_{>0} \to \mathbb{R}$, $x \mapsto x$ は座標 $\varphi\colon \mathbb{R}_{>0} \to \mathbb{R}_{>0}$, $x \mapsto \sqrt{x}$ のもとで $f \circ \varphi(t) = f(t^2) = t^2$ だから凸だが、座標 $\varphi'\colon \mathbb{R}_{>0} \to \mathbb{R}_{>0}$, $x \mapsto x$ のもとでは $f \circ \varphi'(t) = f(t) = t$ だから凸ではない。

命題 4.2.3 (凸性はアファイン変換で不変). [TODO]

証明 [TODO] □

定義 4.2.4 (Legendre 変換). [TODO]

4.3 Bregman 幾何

凸関数をポテンシャルとして双対平坦構造が誘導される。

定義 4.3.1 (Bregman ダイバージェンス). [TODO]

4.4 一般化 Pythagoras の定理

定理 4.4.1 (一般化 Pythagoras の定理). [TODO]

証明 [TODO] □

4.5 射影定理

定義 4.5.1 (測地射影). [TODO]

定理 4.5.2 (射影定理)**.** M を多様体、D を M 上のダイバージェンス、(g, ∇, ∇^*) を D により誘導された双対構造、$S \subset M$ を部分多様体とする。このとき次が成り立つ:

(1) 任意の $p \in M$ に対し、$D(p\|\square)$ を最小化する S 上の点 $q \in S$ は p の S への ∇^*-射影である。

(2) 任意の $p \in M$ に対し、$D^*(p\|\square)$ を最小化する S 上の点 $q \in S$ は p の S への ∇-射影である。

証明 [TODO] □

定理 4.5.3 (一意性)**.** S が平坦のとき [TODO]

証明 [TODO] □

第 5 章　無限集合上の測度の空間

5.1　無限集合上の測度の空間

[TODO]

5.2　パラメータ付き測度モデル

定義 5.2.1 (パラメータ付き測度モデル). [TODO]

パラメトリックモデルとは、M が有限次元で $\mathbf{p}$ の値が確率分布の場合である。

定義 5.2.2 (パラメトリックモデル). [TODO]

第 6 章　正定値行列の空間

[TODO] $n \times n$ 行列を n 元集合上の測度とみなす？

[TODO] Stiefel 多様体と最適化についても触れる？

6.1　正定値対称行列の空間

定義 6.1.1 (正定値対称行列の空間)**.** [TODO]

6.2　ダイバージェンス

第 7 章　統計的推定と情報幾何

7.1　漸近理論 (1 次オーダー)

測地射影を用いて漸近理論を幾何学的に考察する。

[TODO]

7.2　漸近理論 (高次オーダー)

高次オーダーの漸近理論には曲率が現れる。

[TODO]

7.3　ノンパラメトリック統計

定義 7.3.1 (Estimating function)**.** [TODO]