第 1 章　凸集合と凸関数の基本

1　アファイン接続と凸性

M を多様体、∇ を M 上のアファイン接続とする。

定義 1.1 (平坦アファイン接続). M の開部分集合 $O \overset{\text{open}}{\subset} M$ 上の座標であって、それに関する ∇ の接続係数がすべて 0 となるものを、O 上の **∇-アファイン座標 (∇-affine coordinates)** という。

各 $p \in M$ に対し、p のまわりの ∇-アファイン座標が存在するとき、∇ は M 上**平坦 (flat)** であるという。

命題-定義 1.2 (U 上の標準的な平坦アファイン接続). [TODO] 書き方を修正 $U \overset{\text{open}}{\subset} M$ とする。U 上のアファイン接続 $D\colon \Gamma(TU) \to \Gamma(T^\vee U \otimes TU)$ を、次の規則で well-defined に定めることができる:

- 各 $X \in \Gamma(TU)$ に対し、W の基底が定める U 上の座標 x^i $(i = 1, \ldots, m)$ をひとつ選び、

$$DX := dX^i \otimes \frac{\partial}{\partial x^i} \in \Gamma(T^\vee U \otimes TU) \tag{1.1}$$

  と定める。ただし、X の成分表示を $X = X^i \dfrac{\partial}{\partial x^i}$ とおいた。

さらに、この D は U 上のアファイン接続として平坦である。D を U 上の**標準的な平坦アファイン接続 (standard flat affine connection)** という。

証明　写像として well-defined であることを一旦認め、先に $\mathbb{R}$-線型性、Leibniz 則、平坦性を確かめる。D の $\mathbb{R}$-線型性と Leibniz 則は、外微分 d の $\mathbb{R}$-線型性と Leibniz 則から従う。平坦性は、式 (1.1) で用いた座標 x^i が D-アファイン座標となることから従う。最後に、D が写像として well-defined であることを示す。y^α $(\alpha = 1, \ldots, m)$ を W の基底が定める U 上の座標とすると、

$$\begin{aligned}
dX^i \otimes \frac{\partial}{\partial x^i} &= d\left(X^\alpha \frac{\partial x^i}{\partial y^\alpha}\right) \otimes \frac{\partial y^\alpha}{\partial x^i}\frac{\partial}{\partial y^\alpha} &\text{(1.2)}\\
&= \left(\frac{\partial x^i}{\partial y^\alpha} dX^\alpha + X^\alpha \underbrace{d\left(\frac{\partial x^i}{\partial y^\alpha}\right)}_{=0}\right) \otimes \frac{\partial y^\alpha}{\partial x^i}\frac{\partial}{\partial y^\alpha} &\text{(1.3)}\\
&= dX^\alpha \otimes \frac{\partial}{\partial y^\alpha} &\text{(1.4)}
\end{aligned}$$

となる。ただし「$= 0$」の部分は x^i と y^α の間の座標変換がアファイン変換となることを用いた。これで well-defined 性も示された。 □

定義 1.3 (∇-凸集合). 部分集合 $S \subset M$ が **∇-凸 (∇-convex)** であるとは、任意の $p, q \in S$ に対し、p から q への S 内の ∇-測地線がただひとつ存在することをいう。

定義 1.4 (∇-凸関数). $U \overset{\text{open}}{\subset} M$ を ∇-凸開集合とする。関数 $f \in C^\infty(U)$ が **∇-凸 (∇-convex)** であるとは、U 内の任意の ∇-測地線 $\gamma\colon [0,1] \to U$ に対し、$f \circ \gamma\colon [0,1] \to \mathbb{R}$ が凸関数であることをいう。

2 Hessian

W を m 次元 $\mathbb{R}$-ベクトル空間 $(m \in \mathbb{Z}_{\geq 0})$、$U \overset{\text{open}}{\subset} W$ を開部分集合、D を U 上の標準的な平坦アファイン接続とする。

定義 2.1 (Hessian). C^∞ 関数 $f\colon U \to \mathbb{R}$ に対し、f の **Hessian** を

$$\mathrm{Hess}\, f := Ddf \in \Gamma(T^\vee U \otimes T^\vee U) \tag{2.1}$$

と定義する。

D-アファイン座標を用いると、Hessian の成分表示は簡単な形になる。

命題 2.2 (Hessian の成分表示). x^i $(i = 1, \dots, m)$ を U 上の D-アファイン座標とする。このとき、座標 x^i に関する $\mathrm{Hess}\, f$ の成分表示は

$$\mathrm{Hess}\, f = \frac{\partial^2 f}{\partial x^i \partial x^j}\, dx^i \otimes dx^j \tag{2.2}$$

となる。とくに f の C^∞ 性より $\mathrm{Hess}\, f$ は対称テンソルである。

証明 $(\mathrm{Hess}\, f)(\partial_i, \partial_j) = \langle D_{\partial_i} df, \partial_j \rangle = \partial_i \langle df, \partial_j \rangle - \langle df, \underbrace{D_{\partial_i} \partial_j}_{=0} \rangle = \partial_i(\partial_j f) = \dfrac{\partial^2 f}{\partial x^i \partial x^j}$ より従う。 □

命題 2.3 (凸性と Hessian の関係). [TODO] strict, strong など

3 Legendre 変換

定義 3.1 (Legendre 変換). $U \subset W$ を開集合、$f\colon U \to \mathbb{R}$ を C^∞ 関数であって $\nabla f\colon U \to W^\vee$ が単射であるものとする。関数

$$f^\vee\colon U' \to \mathbb{R}, \quad y \mapsto \left\langle (\nabla f)^{-1}(y), y \right\rangle - f((\nabla f)^{-1}(y)) \quad \text{where} \quad U' := \nabla f(U) \tag{3.1}$$

を f の **Legendre 変換 (Legendre transform)** という。

例 3.2 (Legendre 変換の例). 具体的な指数型分布族に対し、対数分配関数の Legendre 変換を計算してみる。

- Bernoulli 分布族 (i.e. 2 元集合上の full support な確率分布の族): 対数分配関数は $\psi\colon \mathbb{R} \to \mathbb{R},\ \theta \mapsto \log(1 + \exp\theta)$ であった。よって $\nabla\psi(\theta) = \dfrac{\exp\theta}{1 + \exp\theta}$ であり、$(\nabla\psi)^{-1}(\eta) = \log\eta - \log(1 - \eta)$ である。したがって $\psi^\vee(\eta) = \eta\log\eta + (1 - \eta)\log(1 - \eta)$ である。
- 正規分布族: 対数分配関数は $\psi\colon \mathbb{R} \times \mathbb{R}_{<0} \to \mathbb{R},\ \theta \mapsto -\dfrac{(\theta^1)^2}{4\theta^2} - \dfrac{1}{2}\log(-\theta^2) + \dfrac{1}{2}\log\pi$ であった。

よって $\nabla\psi(\theta) = \begin{pmatrix} -\dfrac{\theta^1}{2\theta^2} & \dfrac{(\theta^1)^2}{4(\theta^2)^2} - \dfrac{1}{2\theta^2} \end{pmatrix}$ であり、$(\nabla\psi)^{-1}(\eta) = \dfrac{1}{\eta_2 - (\eta_1)^2}\begin{pmatrix} \eta_1 \\ -1/2 \end{pmatrix}$ である。したがって $\psi^\vee(\eta) = -\dfrac{1}{2}\left(1 + \log 2\pi + \log(\eta_2 - (\eta_1)^2)\right)$ である。

本稿では、とくに次の状況を考えることになる。

命題 3.3. [TODO] 単射の証明などは補題に切り出す $U \subset W$ を凸開集合、$f\colon U \to \mathbb{R}$ を C^∞ 関数であって $\operatorname{Hess} f$ が U 上各点で (対称であることも含む意味で) 正定値であるものとする。このとき、次が成り立つ:

(1) ∇f は局所微分同相である。とくに $U' := \nabla f(U)$ は $W^\vee$ の開集合である。
(2) $\nabla f\colon U \to U'$ は微分同相である。とくに ∇f は単射である。

したがって $f^\vee$ が定義でき、$f^\vee$ は次をみたす:

(3) $f^\vee\colon U' \to \mathbb{R}$ は C^∞ 関数である。
(4) $\nabla f^\vee = (\nabla f)^{-1}$ が成り立つ。とくに $\nabla f^\vee$ は単射である。
(5) 各 $y \in U'$ に対し $(\operatorname{Hess} f^\vee)_y = ((\operatorname{Hess} f)_x)^{-1}$ が成り立つ (ただし $x := (\nabla f)^{-1}(y)$)。とくに $(\operatorname{Hess} f^\vee)_y$ は正定値である。

証明 (1) 命題の仮定より $\operatorname{Hess} f$ は U 上各点で正定値だから、∇f の微分は各点で線型同型である。したがって ∇f は局所微分同相であり、とくに開写像である。よって $U' = \nabla f(U)$ は $W^\vee$ の開集合である。

(2) $u, \widetilde{u} \in U$, $u \neq \widetilde{u}$ を固定し、$[0,1]$ を含む $\mathbb{R}$ の開区間 I であって、すべての $t \in I$ に対し $(1-t)u + t\widetilde{u}$ が U に属するようなものをひとつ選ぶ (U は W の凸開集合だからこれは可能)。さらに $\varphi\colon I \to U$, $t \mapsto f((1-t)u + t\widetilde{u})$ と定めると、平均値定理より、ある $\tau \in (0,1)$ が存在して

$$\begin{aligned} \langle \nabla f(\widetilde{u}) - \nabla f(u), \widetilde{u} - u \rangle &= \varphi'(1) - \varphi'(0) && (3.2)\\ &= \varphi''(\tau) \qquad (\text{平均値定理}) && (3.3)\\ &= \left\langle (\operatorname{Hess} f)_{(1-\tau)u+\tau\widetilde{u}}, (\widetilde{u} - u)^2 \right\rangle && (3.4)\\ &> 0 \qquad (\operatorname{Hess} f \text{ は正定値}) && (3.5) \end{aligned}$$

が成り立つ。よって $\nabla f(\widetilde{u}) \neq \nabla f(u)$ である。したがって ∇f は単射である。このことと (1) より $\nabla f\colon U \to U'$ は微分同相である。

(3) $\nabla f\colon U \to U'$ が微分同相ゆえに $(\nabla f)^{-1}\colon U' \to U$ は C^∞ だから、$f^\vee$ は C^∞ 関数である。

(4) $f^\vee$ の定義式を ∇ で微分すると、すべての $y \in U'$ に対し

$$(\nabla f^\vee)(y) = (\nabla f)^{-1}(y) + \left\langle y, \nabla(\nabla f)^{-1}(y) \right\rangle - \left\langle (\nabla f)((\nabla f)^{-1}(y)), \nabla(\nabla f)^{-1}(y) \right\rangle = (\nabla f)^{-1}(y) \tag{3.6}$$

が成り立つ。よって $(\nabla f)^{-1} = \nabla f^\vee$ である。

(5) (4) より

$$\begin{aligned} (\operatorname{Hess} f^\vee)_y &= d(\nabla f^\vee)_y && (3.7)\\ &= d((\nabla f)^{-1})_y && (3.8)\\ &= (d(\nabla f)_x)^{-1} && (3.9) \end{aligned}$$

$$= ((\mathrm{Hess}\, f)_x)^{-1} \tag{3.10}$$

となる。 □

系 3.4 (Legendre 変換の対合性)**.** $f^{\vee\vee} = f$.

証明 Legendre 変換の定義より、すべての $x \in U$ に対し

$$\begin{aligned} f^{\vee\vee}(x) &= \left\langle x, (\nabla f^{\vee})^{-1}(x) \right\rangle - f^{\vee}((\nabla f^{\vee})^{-1}(x)) && (3.11)\\ &= \langle x, \nabla f(x) \rangle - f^{\vee}(\nabla f(x)) \qquad (\nabla f^{\vee} = (\nabla f)^{-1}) && (3.12)\\ &= \langle x, \nabla f(x) \rangle - \left(\left\langle \nabla f(x), (\nabla f)^{-1}(\nabla f(x)) \right\rangle - f((\nabla f)^{-1}(\nabla f(x))) \right) && (3.13)\\ &= f(x) && (3.14) \end{aligned}$$

が成り立つ。よって $f^{\vee\vee} = f$ である。 □

4 Fourier-Laplace 変換

[TODO] ちゃんと書く。cf. [?]

定義 4.1 (Fourier-Laplace 変換)**.** V を有限次元 $\mathbb{R}$-ベクトル空間、μ を V 上の測度とする。

$$L_\mu(\theta) := \int_{v \in V} e^{\langle \theta, v \rangle}\, d\mu(v) \quad (\theta \in V^{\vee} \otimes \mathbb{C}) \tag{4.1}$$

と定め、L_μ を **Fourier-Laplace 変換 (Fourier-Laplace transform)** という。