

誤差逆伝搬法は余接関手である

@phykm

2017 年 3 月 31 日

概要

タイトル通り。誤差逆伝搬法とは、余接関手が反変モノイダルであること（さらに言えば直積多様体の余接バンドルが直和バンドルになること）の言い換えである。

1 余接関手

Definition 1.1. n 次元可微分 (C^p 級) 多様体とは、ハウスドルフ位相空間 M と、 M 上の開被覆 $\{U_i\}_i$ 、この開被覆を定義域とする $\mathbb{R}^n$ の部分集合への同相写像 $\{\phi_i : U_i \rightarrow \phi_i(U_i) (\subset \mathbb{R}^n)\}$ の組 $(M, \{(U_i, \phi_i)\}_i)$ で、次の条件を満たすものとする。すなわち、 $U_i \cap U_j \neq \emptyset$ であるとき、 $\phi_i \circ \phi_j^{-1}$ および $\phi_j \circ \phi_i^{-1}$ の $\phi_j^{-1}(U_i \cap U_j), \phi_i^{-1}(U_i \cap U_j)$ への制限は、この二つの $\mathbb{R}^n$ 部分集合の同相写像かつ C^p 級関数である。

Definition 1.2. $(M, \{(U_i, \phi_i)\}_i), (N, \{(V_j, \psi_j)\}_j)$ を可微分 (C^p 級) 多様体とする。多様体の *morphism* または単に可微分 (C^p 級) 写像とは、連続写像 $f : M \rightarrow N$ であって、次の条件を満たすものとする。すなわち、 $f(U_i) \cap V_j \neq \emptyset$ であるとき、 $\psi_j \circ f \circ \phi_i^{-1}$ を $\phi_i^{-1}(f^{-1}(f(U_i) \cap V_j))$ に制限したものは C^p 級関数である。可逆な *morphism* で移り合うときその二つの多様体は微分同型という。

多様体上の関数または多様体への関数を可微分だというとき、それは各点近傍ごとに、なんらかのチャートで $\mathbb{R}^n$ 間の関数としたときに可微分であることを意味する。

Definition 1.3. $(M, \{(U_i, \phi_i)\}_i), (N, \{(V_j, \psi_j)\}_j)$ を可微分 (C^p 級) 多様体とする。この直積集合は次のように多様体になるので、これを積多様体と呼ぶ。 $(M \times N, \{U_i \times V_j, \phi_i \times \psi_j\}_{i,j})$

なお、 $\{U_i, \phi_i\}$ の組一つをチャートと呼ぶ。多変数解析学の基本的な事実から、チャートの各成分の微分は、余接ベクトルとして線形独立であり、逆に微分が線形独立であるような関数を次元の数だけ得ることで、いつでもチャートを増設できる。そこでチャートの増設によって合流するような多様体のペアを同一視し、以下では単に多様体を M, N のように書く。また p は以下ではある程度高ければこのノートではどうでもいいので、適当に固定するとする。

多様体の圏を **Diff** とする。product($- \times -$) が入っているので、特にモノイダル圏である。すなわち、それぞれの対象は積によって並列でき、写像も並列でき、写像の合成と並列は可換である。具体的には、 $f : M \rightarrow N, g : N \rightarrow L, h : Y \rightarrow Z, k : Z \rightarrow W$ について、 $(g \times k) \circ (f \times h) = (g \circ f) \times (k \circ h)$ 。モノイダル単位対象はゼロ次元多様体 $\{*\}$ (ただの一点) である。

Definition 1.4. M を可微分多様体とする。 $x \in M$ 上の接ベクトルとは、 M 上可微分関数環 $C^\infty(M)$ から

$\mathbb{R}$ への線形写像 d で、次のライプニッツ則を満たすものとする。 $f, g \in C^\infty(M)$

$$d(fg) = d(f)g(x) + f(x)d(g) \quad (1)$$

Fact 1.5. 接ベクトルは各点で M の次元と同じ次元のベクトル空間 $T_x M$ をなす。これを接空間と呼ぶ。 $x \in M$ の近傍で有効なチャート $\{U_i, \phi_i = \{x^j\}_j\}$ をとったとき、 $T_x M$ の基底ベクトルとして

$$\partial_{x^j} = f \mapsto \left(\frac{\partial}{\partial x^j} f \circ \phi^{-1} \right) (\phi_i(x)) \quad (2)$$

をとれる。

$T_x M$ の双対空間 $T_x^* M$ を余接空間と呼ぶ。あるチャートについて $T_x M$ の基底 ∂_{x^j} に対する $T_x^* M$ 双対基底（内部積がクロネッカーデルタになる元のこと）を dx^j と書く。

Definition 1.6. M を可微分多様体とする。 $TM = \cup_{x \in M} T_x M$ および $T^*M = \cup_{x \in M} T_x^* M$ は次のようなチャートによって再び多様体になる。これを接バンドル、余接バンドルと呼ぶ。 *1M の U_i 上チャートを $\phi_i = \{x^j\}_j$ と書くとする。射影を

$$\pi_{tn} : TM \rightarrow M \quad (3)$$

$$v \in T_x M \mapsto x \quad (4)$$

$$\pi_{ct} : T^*M \rightarrow M \quad (5)$$

$$e \in T_x^* M \mapsto x \quad (6)$$

とし、 $\pi_{tn}^{-1}(U_i), \pi_{ct}^{-1}(U_i)$ を定義域として、チャート $\bar{\phi}_i$ を

$$\bar{\phi}_i : \pi_{tn}^{-1}(U_i) \rightarrow \mathbb{R}^{2n} \quad (7)$$

$$v = v^j \partial_{x^j} \in T_x M \mapsto (\phi(x), \{v^j\}_j) \quad (8)$$

$$\bar{\phi}_i : \pi_{ct}^{-1}(U_i) \rightarrow \mathbb{R}^{2n} \quad (9)$$

$$e = e_j dx^j \in T_x^* M \mapsto (\phi(x), \{e_j\}_j) \quad (10)$$

で定める。位相構造は、 $\bar{\phi}_i$ による開集合の逆像、および $\pi_{ct,tn}^{-1}$ による開集合の逆像を開基とする。

接バンドル、余接バンドルを構成することは関手である。

Definition 1.7. 次の関手を接関手、余接関手と呼ぶ。

$$T : \mathbf{Diff} \rightarrow \mathbf{Diff} \quad (11)$$

$$M \mapsto TM \quad (12)$$

$$f : M \rightarrow N \mapsto df : TM \rightarrow TN \quad (13)$$

$$T^* : \mathbf{Diff} \rightarrow \mathbf{Diff}^{op} \quad (14)$$

$$M \mapsto T^*M \quad (15)$$

$$f : M \rightarrow N \mapsto \partial f : T^*M \rightarrow T^*N \quad (16)$$

*1 本当はベクトルバンドルという概念があり、これはそれを構成することにあたるののだが、筆者の現時点の知識では、余接バンドルをバンドルの圏への関手として書くことができなかった。バンドル概念の一般化が待たれる。

ただし、ここでの $df, \partial f$ は (明示的には) 以下で定義される。 $v \in T_x M, x \in U, e \in T_y N, f(x) \in V, f(x) = y$ で $\phi = \{x^i\}_i, \psi = \{y^j\}_j$ は U, V を定義域にもつ M, N のチャートとする。

$$df(v^i \partial_{x^i}) = v^i \frac{\partial y^j}{\partial x^i} \partial_{y^j} \quad (17)$$

$$\partial f(e_j dy^j) = e_j \frac{\partial y^j}{\partial x^i} dx^i \quad (18)$$

$$\frac{\partial y^j}{\partial x^i} = \left(\frac{\partial y^j(f \circ \phi)}{\partial x^i} \right) (\phi^{-1}(x)) \quad (19)$$

しばし、可微分写像の接関手の像は微分写像、余接関手の像は引き戻しと呼ばれる。それらは、接ベクトル、余接ベクトルの成分にたいしては (互いに逆向きの) 行列演算になっていることが確認できる。

Fact 1.8. この関手は **Diff** のもつ直積モノイダル構造に適い、モノイダル関手でもある。すなわち、自然に以下のような同型がある。

$$T(M \times N) \simeq TM \times TN \quad (20)$$

$$T(\{*\}) \simeq \{*\} \quad (21)$$

$$T^*(M \times N) \simeq T^*M \times T^*N \quad (22)$$

$$T^*(\{*\}) \simeq \{*\} \quad (23)$$

さらにこの同型は、各点ファイバーごとに線形空間の直和同型

$$T_x N \oplus T_y N \simeq T_{(x,y)}(N \times M) \quad (24)$$

$$T_x^* N \oplus T_y^* N \simeq T_{(x,y)}^*(N \times M) \quad (25)$$

でもある。

したがって、直積多様体から、または直積多様体への微分写像や引き戻しも、個々の多様体への微分写像や引き戻しの行列演算の和として実行できる。具体的には、

$$f : M \times M' \rightarrow N \quad (26)$$

$$df(v^i \partial_{x^i} \oplus v'^i \partial_{x'^i}) = \left(x^i \frac{\partial y^j}{\partial x^i} + x'^i \frac{\partial y^j}{\partial x'^i} \right) \partial_{y^j} \quad (27)$$

$$\partial f(e_j dy^j) = e_j \left(\frac{\partial y^j}{\partial x^i} dx^i + \frac{\partial y^j}{\partial x'^i} dx'^i \right) \quad (28)$$

$$f : M \rightarrow N_1 \times N_2 \quad (29)$$

$$df(v^i \partial_{x^i}) = v^i \left(\frac{\partial y_1^j}{\partial x^i} \partial_{y_1^j} + \frac{\partial y_2^j}{\partial x^i} \partial_{y_2^j} \right) \quad (30)$$

$$\partial f(e_{1j} dy_1^j \oplus e_{2j} dy_2^j) = \left(e_{1j} \frac{\partial y_1^j}{\partial x^i} + e_{2j} \frac{\partial y_2^j}{\partial x^i} \right) dx^i \quad (31)$$

と計算することが許される。

2 誤差逆伝搬法

ニューラルネットワークのエッジを可微分多様体、ノードを可微分写像だとする。すなわち、

Definition 2.1. 入力層、出力層、バイアスとウェイトをあわせたパラメータ全体をいずれも **Diff** の対象 D, N, W とする。

Definition 2.2. (DAG) ニューラルネットワークとは、モノイダル圏 **Diff** の射 $F : D \times W \rightarrow N$ である。

実際には各多様体は $\mathbb{R}^n$ のような単純なものであることがほとんどであるが、ここでは一般化しておく。それぞれの対象は **Diff** のモノイダル構造から、一般に複数の多様体の直積でありえる。ノードの複雑な接続は基本的な可微分写像の合成と直積で構成できる。

Definition 2.3. 教師データを $T : \{*\} \rightarrow D$ 、 T に関する誤差関数を $E_T : N \rightarrow \mathbb{R}$ とする。具体的には N がリーマン多様体で、なんらかの距離が入っており、 T に対応する N 上のラベルに対する距離関数などが例として上がる。しかし経験誤差のように対称でない距離も考えて良い。特定のパラメータを $P : \{*\} \rightarrow W$ とする。

ニューラルネットワーク $F : D \times W \rightarrow N$ であるとする。教師データ $T \times \text{id}_W$ をネットワークに代入し、誤差関数 $E_T : N \rightarrow \mathbb{R}$ を合成することで、パラメータからの誤差関数

$$E_T \circ F \circ (T \times \text{id}_W) : W \rightarrow \mathbb{R} \quad (32)$$

を考えることができる。そして D についての学習ステップは、この関数の勾配によって、最小点へ接近させることである。従って要求されているタスクは、 W 上の微分 1 形式を得ることであるが、その最初のタネは $\mathbb{R}$ 上の自明な 1 形式 $dx \in T_e \mathbb{R}, e = E_T \circ F \circ (T \times P)\{*\} \in \mathbb{R}$ である。この係数は学習係数に相当するので自由に決めて良い。ここから W 上 1 形式を得る方法は明らかだろう。もちろん

$$T^*(E_T \circ F \circ (T \times \text{id}_W))(dx) = \partial(E_T \circ F \circ (T \times \text{id}_W))(dx) \quad (33)$$

がそれである。ところで、これはどのように計算すればいいだろうか？ 前節で見たように、 T^* は直積に関して反変モノイダル関手であり、かつ (余) 接バンドルは各点ファイバーごとに直和になるのだから、ネットワークを構成する $E_T \circ F \circ (T \times \text{id}_W)$ を (30,31) を用いながら自由に分解し（それはネットワークのノード単位になる）、その可微分写像ごとに $T^*(-)$ を計算すればよい。(18) のように、個々のそれは微分行列の計算をノードの数だけ行うことになるが、これは誤差逆伝搬法にほかならない。

ところで、多くの勾配法の説明には、以上の状況に合わせると説明不足な点があり、もし本気でパラメータの空間を多様体と考えるのなら、計量を定義しないといけない。なぜならば勾配は 1 形式であり、多様体の微小な変換をもたらすのはベクトル場であって、両者はまったく別物だからである。逆に言えば、このことに言及がないタイプの勾配法は、暗黙のうちにその座標系のユークリッド計量を使っているのだから、一般の多様体上では通用しない。なんらかの計量を宣言して固定することが求められる。