

第2章：合成関数の微分法

7月4日 清野和彦

前書き（言い訳）

やっぱり時間が足りませんでした。すみません。このプリントに入れる予定だった線積分の説明は後日別プリントで配布します。また、第2節までと第3節以降で変数などの記号の使い方が違ってきます。統一する時間がありませんでしたので、注意してよいでいただけたらと思います。

目次

1	この章の意味と目標	2
1.1	合成関数の微分だけで一つの章を立てるのはなぜか	2
1.2	合成関数の微分公式	5
1.2.1	1変数関数に1変数関数を入れた場合	5
1.2.2	1変数関数に2変数関数を入れた場合	7
1.2.3	2変数関数に1変数関数を2つ入れた場合	9
1.2.4	2変数関数に2変数関数を2つ入れた場合	11
1.2.5	一般の場合	14
1.3	行列を使った記法と写像の微分	14
1.4	「連鎖律」という用語について	19
2	合成関数の微分公式の証明	20
2.1	偏微分の定義に基づいた証明	20
2.1.1	1変数関数に1変数を入れた場合	20
2.1.2	1変数関数に2変数関数を入れた場合	21
2.1.3	2変数関数に1変数関数を2つ入れた場合	22
2.1.4	2変数関数に2変数関数を2つ入れた場合	26
2.1.5	一般の場合	27
2.2	1次近似という視点からの証明	29
2.2.1	1変数関数に1変数関数を入れた場合	29
2.2.2	1変数関数に2変数関数を入れた場合	32
2.2.3	2変数関数に1変数関数を2つ入れた場合	36
2.2.4	2変数関数に2変数関数を2つ入れた場合	39

3	逆写像の微分と変数変換	43
3.1	逆写像の微分公式	43
3.2	変数変換	45
4	ヤコビ行列の意味：多変数写像の視覚化	48
5	座標を使わずに微分を理解する	53
5.1	座標を使わないとはどういうことか	53
5.2	座標を使わずに微分を定義するには	53
5.3	空間とは	54
5.4	写像の微分の定義	57
5.5	合成関数の微分	58
5.6	方向微分	58
5.7	偏微分	59

1 この章の意味と目標

1.1 合成関数の微分だけで一つの章を立てるのはなぜか

高校で 1 変数関数の微分を勉強したとき、まず微分や導関数を定義し、次に多項式や三角関数などのよく出会う関数の導関数を具体的に求め、最後に積、商、合成関数、逆関数の微分法を調べました。これらの「公式」が手に入ったことによって、具体的に微分できる関数がいくつかあるとき、それらを足し引きしてできる関数はもちろん、掛けたり割ったり合成したり逆関数を作ったりしても微分できることになり、「式一本」で書ける関数は何でも微分できるようになったわけです。このように、積や合成関数の微分公式があると、微分できる関数の世界が一気に広がります。

事情は多変数でも同じなはずですが。第 1 章で多変数関数の微分を定義したので、次にすべきことは、基本的な多変数関数の偏導関数を求めることと、積や合成関数の微分公式を得ることでしょう。

ところが、偏微分は注目した変数以外は定数と見なすことによる 1 変数関数としての微分ですし、基本的な多変数関数（つまり、式で書ける多変数関数）は式で書ける 1 変数関数をいくつかそれぞれ別の変数を持つ関数として用意し、それらを足したり引いたり掛けたり割ったり合成したりしてできるものなので、基本的な多変数関数が具体的に与えられたら、それを偏微分することは 1 変数関数の微分の知識の範囲内で完全にできてしまうわけです。

例えば、 $f(x, y) = xy \sin x \cos y$ とすると、 x で偏微分するときは $y \cos y$ は定数

扱い、 y で偏微分するときは $x \sin x$ は定数扱いなので、

$$\begin{aligned}\frac{\partial f}{\partial x}(x, y) &= y \cos y \frac{d}{dx}(x \sin x) = y \cos y (\sin x + x \cos x) \\ \frac{\partial f}{\partial y}(x, y) &= x \sin x \frac{d}{dy}(y \cos y) = x \sin x (\cos y - y \sin y)\end{aligned}$$

というふうに1変数関数の積の微分法で計算できてしまいます。商の微分も全く同様です。また、例えば $\sin z$ に $z = x^2 y^3$ を合成した関数 $g(x, y) = \sin x^2 y^3$ を考えても、やはり x で偏微分するときは y^3 は定数扱い(つまり x^2 の単なる係数)、 y で偏微分するときは x^2 は定数扱い(つまり y^3 の単なる係数)なので、

$$\frac{\partial g}{\partial x}(x, y) = 2xy^3 \cos x^2 y^3, \quad \frac{\partial g}{\partial y}(x, y) = 3x^2 y^2 \cos x^2 y^3$$

と計算できます。なお、多変数関数には逆関数は存在しないので¹、「多変数関数そのものについての逆関数の微分法」というものは存在しません。(多変数で逆関数に当たる概念はずっと後の方でできます。)

多変数関数の微分には偏微分以外にも方向微分や1次近似(全微分可能ということ)があるのではないかと思うかも知れませんが、基本的な関数はすべて C^1 -級、つまり偏導関数が連続関数なので、これらの微分の概念はすべて同じものの別な姿だと思ってよいわけですから(第1章第6節を参照してください)、やはり1変数関数の微分の範囲内に収まってしまいます。

しかし、上の計算は1変数関数のときとはちょっとだけ違っている面があります。というのは、1変数関数のときには、基本的な関数の導関数を求めるのとは別に積、商、合成関数の微分法を「公式」として用意したのです。だから、例えば大学に来て初めてであった逆三角関数のようなものでも、高校のときに用意しておいた公式だけで微分を計算することができたわけです。ところが、上でお見せした偏微分の計算は、偏微分しようとする関数の中に現れている1変数関数が導関数を既に知っている関数だったからできたように見えます。つまり、多変数関数の偏微分そのものに対する「公式」にはなっていないわけです。具体的な関数の姿を使わない理論として偏微分の間の関係を考えるには、いわば抽象的な「公式」を求めておかねばなりません。

とは言え、積や商の微分公式は実はほとんど1変数のときのままです。実際、例えば $f(x, y)$ と $g(x, y)$ の y に定数 b を代入してできる x の1変数関数を $\varphi(x) = f(x, b)$, $\psi(x) = g(x, b)$ と書くことにすれば、 $f(x, y)$ と $g(x, y)$ の積の (a, b) にお

¹ $z = f(x, y)$ の逆関数とは、 $(x, y) = (g_1(z), g_2(z))$ であって $f(g_1(z), g_2(z)) = z$ と $g_1(f(x, y)) = x$, $g_2(f(x, y)) = y$ を満たすものということになりますが、 $(x, y) = (g_1(z), g_2(z))$ というのは z をパラメタとするパラメタ曲線ですので、 $f(x, y)$ の定義域(平面内で広がりを持つ部分)全体を値域とすることはできません。よって、多変数関数には逆関数は存在しません。

ける x による偏微分は、偏微分の定義から

$$\begin{aligned}\frac{\partial(fg)}{\partial x}(a, b) &= \frac{d(\varphi\psi)}{dx}(a) \\ &= \frac{d\varphi}{dx}(a)\psi(a) + \varphi(a)\frac{d\psi}{dx}(a) \\ &= \frac{\partial f}{\partial x}(a, b)g(a, b) + f(a, b)\frac{\partial g}{\partial x}(a, b)\end{aligned}$$

となります。(二つ目の等号で1変数関数の積の微分法を使っています。) x での偏微分を下付添え字で f_x, g_x などと表し、「どこでの偏微分か」を表す (a, b) を省略して書けば、1変数関数の積の微分法の式にもっと似た形の

$$(fg)_x = f_x g + f g_x$$

で表すことができます。商の微分も全く同様に、 $1/f(x, y)$ の (a, b) における x による偏微分は、上の計算と同じ記号を流用すると、

$$\begin{aligned}\frac{\partial}{\partial x} \left(\frac{1}{f} \right) (a, b) &= \frac{d}{dx} \left(\frac{1}{\varphi} \right) (a) \\ &= -\frac{1}{(\varphi(a))^2} \frac{d\varphi}{dx}(a) \\ &= -\frac{1}{(f(a, b))^2} \frac{\partial f}{\partial x}(a, b)\end{aligned}$$

となります。(二つ目の等号で1変数関数の商の微分法を使っています。) つまり、

$$\left(\frac{1}{f} \right)_x = -\frac{f_x}{f^2}$$

と、1変数関数のときと全く同じ式になっているわけです。

というわけで、合成関数の微分公式を求めることだけが残るわけです。もちろん、これも積や商のように簡単に求まるならそれだけのための章を立てたりはしません。章を別立てにした理由は

- 1変数関数の場合より合成関数の概念そのものが捉えにくい。
- 証明が難しく、何をしているのかわかりにくい。
- 合成関数の微分公式をよく見ることで、多変数写像の微分のあるべき姿が見えてくる。

と言った、どれも重要なものばかりです。

1.2 合成関数の微分公式

まず、合成関数の微分公式を証明抜きで紹介します。多変数の合成関数には多くの変数と多くの関数が出てくるので、ここでこのプリントでの記号の使い方も合わせて決めてしまいましょう。読んでいて記号で混乱したらここに返ってみてください。

1変数関数が二つ、例えば $f(x), g(x)$ があったとき、もしも $g(x)$ の値域が $f(x)$ の定義域に含まれているなら、 $g(x)$ を $f(x)$ の x のところに入れることができます。ただ、このように書くと x が二重の意味を持つてしまう (g の独立変数としての意味と、 g の従属変数としての意味) ので、紛れがないように文字を決めましょう。普通「 $y = g(x)$ を $f(y)$ に代入する」というふうに記号を決めると思いますが、この後2変数関数を考えるときに文字が足りなくなってしまうので、1変数のときからギリシャ文字を使わせてください。

あとで2変数関数を考えるときに変数は2つ一組で必要になります。よく使う組は $(x, y), (u, v), (s, t)$ などでしょう。このゼミでは (u, v) の組はベクトルの成分として使ってきたので、 (s, t) を使うことにしましょう。変数が一つしかないときは普通 s より t を使うと思いますが、2変数になったときの混乱を避けるために、1変数のときは s を使うことにします。 x に対応するギリシャ文字は ξ 、 y に対応するギリシャ文字は η なので、 x に代入される関数を $\xi(s)$ や $\xi(s, t)$ 、 y に代入される関数を $\eta(s)$ や $\eta(s, t)$ としましょう。 ξ や η という文字になじみの薄い人も多いと思いますが、是非ここで慣れてください。「 x に対応 x に対応」とか「 y に対応 y に対応」と心の中で唱えながらつき合ってもらえればわりとすぐ慣れると思います。

1.2.1 1変数関数に1変数関数を入れた場合

1変数関数 $f(x)$ に1変数関数 $\xi(s)$ を合成するということは、イメージ的には図1のようになるでしょう。これは「 s に $f(\xi(s))$ を対応させる」という歴とした

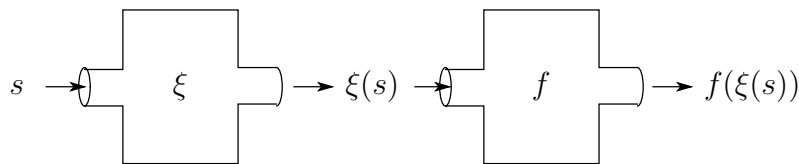


図 1: 「1変数関数を二つ合成する」ということのイメージ

関数ですので、関数らしい記号を決めておくのがよいでしょう。 $f(\xi(s))$ でよいのでは？と思うかも知れませんが、 $f(\xi(s))$ という記号には「関数そのものを表す部分」がないのが問題です。関数 $f(x)$ は (x) を省いて f と書くことができる、つまり f が関数そのものを表しているように、 $f(x)$ に $x = \xi(s)$ を代入してできる関

数にも、このようにして定義された関数だということが一目で分かるような「関数そのものを表す記号」を決めたいわけです。 $f(\xi(s))$ がダメな理由は大きな括弧があるからなので、それを省いてやったらどうか、と思うかも知れませんが、そうすると $f\xi(s)$ となって積と紛らわしくなってしまいます。そこで、 f と ξ の間に「積ではなくて合成だよ」ということを表す印を入れてやればよいのではないかと考えてくるでしょう。積は「 $\cdot$ 」で表すことが多いので、もう少し偉そうに「 $\circ$ 」にするのが一般的です。つまり、

関数 $f \circ \xi$ を

$$f \circ \xi(s) = f(\xi(s))$$

で定義する

というわけです (図2)。

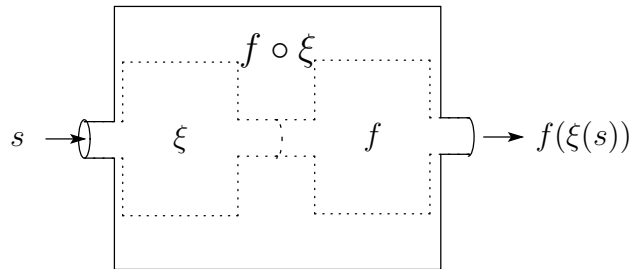


図2: 出力口と入力口を貼り付けて全体を大きな箱に入れる

f の独立変数と ξ の独立変数を別な文字で表したように、物理などの具体的な状況を考えるときには f の独立変数と ξ の独立変数は別な意味を持つ量を考えることが多いと思います。このような場合には、定義域 (を含む $\mathbb{R}$) や値域 (を含む $\mathbb{R}$) をはっきり書き表す「写像の記法」を使うと状況がよりはっきりするでしょう。

$$f \circ \xi: \mathbb{R} \xrightarrow{\xi} \mathbb{R} \xrightarrow{f} \mathbb{R}$$

となります。この記法やイメージ図1と $f \circ \xi$ や $f(\xi(s))$ という記号では f と ξ の順序が逆になっていることに注意してください。

この場合の合成関数の微分公式は高校のときに学んだとおり、

$$\frac{d(f \circ \xi)}{ds}(s) = \frac{df}{dx}(\xi(s)) \frac{d\xi}{ds}(s) \quad \text{すなわち} \quad (f \circ \xi)'(s) = f'(\xi(s)) \xi'(s)$$

です。変数 s を省いて「関数そのもの」の等式として書くと

$$\frac{d(f \circ \xi)}{ds} = \left(\frac{df}{dx} \circ \xi \right) \cdot \frac{d\xi}{ds} \quad \text{すなわち} \quad (f \circ \xi)' = (f' \circ \xi) \cdot \xi'$$

となります²。

注意. 上の書き表し方はかなり丁寧というかくどい書き方です。多くの場合 $x = \xi(s)$ を代入して

$$\frac{d(f \circ \xi)}{ds}(s) = \frac{df}{dx}(x) \frac{d\xi}{ds}(s) \quad \text{あるいは} \quad (f \circ \xi)'(s) = f'(x)\xi'(s)$$

と書くでしょう。さらに、 s と x の対応は分かり切っているなので、それを省いて

$$\frac{d(f \circ \xi)}{ds} = \frac{df}{dx} \frac{d\xi}{ds} \quad \text{あるいは} \quad (f \circ \xi)' = f' \cdot \xi'$$

と書くことの方が多いかも知れません。右側の書き方では、 f は x の関数なのだから f' は x での微分、 ξ は s の関数なのだから ξ' は s による微分しかあり得ないので誤解の余地はないわけです。そのことを逆手に取ると、左側の書き方ではどの変数による微分かをはっきり書いているのだから、 $f \circ \xi$ のことを f と書いてしまっても誤解は起こりえないとも言えます。つまり f そのものは x でしか微分できないのだから、 df/ds と書くことによって f と ξ の合成関数を微分していることまで表せていると考えるわけです。さらに関数 ξ は x に代入されるわけですから、 ξ と書かずに x と書いてしまってもこれまた誤解の余地はないはずです。というわけで、

$$\frac{df}{ds} = \frac{df}{dx} \frac{dx}{ds}$$

という、物理の本などでよく見かける書き方が得られるわけです。

この書き方は、 df や dx をただの数だと思って普通に分数の掛け算をすると両辺が一致するので印象に残りやすいのですが、これから見るように、多変数関数の合成関数の微分公式ではそうはなっていないので、かえって間違いの元になる可能性もあります。便利な記憶法にはあまり頼らず、次節以降で説明する証明を理解するようにしましょう。

1.2.2 1変数関数に2変数関数を入れた場合

次に2変数関数を合成することを考えたいのですが、「入れ物」の方の関数 f と中に入る方の関数 ξ (と η) の両方を一遍に2変数関数にすると混乱しやすいと思うので、徐々に変数の数を増やして行きましょう。

まず $f(x)$ に $x = \xi(s, t)$ を合成することを考えます。この場合は1変数関数に1変数関数を合成する場合と大差ありません。実際、記号は

$$f \circ \xi(s, t) = f(\xi(s, t))$$

と全く同じで済むし、イメージ図も図3を真ん中でくっつけて図4になるだけで、1変数関数に1変数関数を合成するときとの違いは、一番左側のインプット口が二つあるということだけです。

写像の記法では、二つの変数 s, t を平面上の点の座標のように考えて

$$f \circ \xi: \mathbb{R}^2 \xrightarrow{\xi} \mathbb{R} \xrightarrow{f} \mathbb{R}$$

². は掛け算です。(s) がある場合のように $\cdot$ なしでももちろんよいのですが、わかりにくいようなので (s) なしの方では付けることにしました。

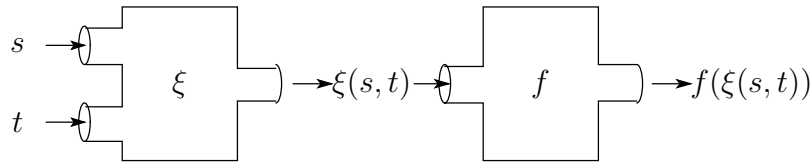


図 3: 2 変数関数の出力を 1 変数関数に入力する

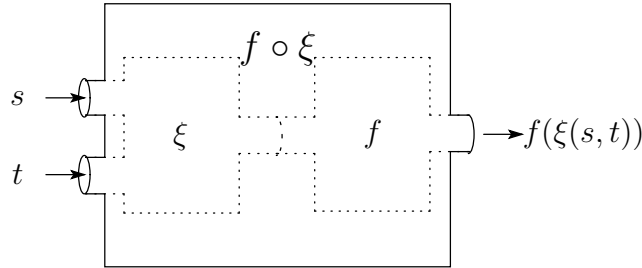


図 4: 出力口と入力口を貼り付けて全体を大きな箱に入れる

と書きます。

この場合の合成関数の微分法は次のようになります。

$$\frac{\partial(f \circ \xi)}{\partial s}(s, t) = \frac{df}{dx}(\xi(s, t)) \frac{\partial \xi}{\partial s}(s, t), \quad \frac{\partial(f \circ \xi)}{\partial t}(s, t) = \frac{df}{dx}(\xi(s, t)) \frac{\partial \xi}{\partial t}(s, t)$$

すなわち

$$(f \circ \xi)_s(s, t) = f'(\xi(s, t)) \xi_s(s, t), \quad (f \circ \xi)_t(s, t) = f'(\xi(s, t)) \xi_t(s, t)$$

です。関数そのものの等式として書くなら、

$$\frac{\partial(f \circ \xi)}{\partial s} = \frac{df}{dx} \circ \xi \cdot \frac{\partial \xi}{\partial s}, \quad \frac{\partial(f \circ \xi)}{\partial t} = \frac{df}{dx} \circ \xi \cdot \frac{\partial \xi}{\partial t}$$

すなわち

$$(f \circ \xi)_s = (f' \circ \xi) \cdot \xi_s, \quad (f \circ \xi)_t = (f' \circ \xi) \cdot \xi_t$$

となります。 f は 1 変数関数なので、その微分には df/dx や f' など 1 変数関数の微分の記号を使っています。

合成関数が 2 変数なので s による偏微分と t による偏微分の二つあって 1 変数関数同士のときより複雑に見えるかも知れませんが、一つ一つを見れば 1 変数関数同士のときの公式とほとんど同じであることがわかるでしょう。その理由は偏微分の定義を思い出してみればすぐわかるので、ここで (1 変数関数の合成関数の微分公式を認めることにして) ここで証明をしてしまってもよいくらいなのですが、話が前後するのを避けるために、証明はすべて次節以降にまわすことにします。

注意. 1変数関数同士のときの注意の最後に書いた記法を使ってみましょう。つまり、 $f \circ \xi$ のことも f と書いてしまい、 ξ のことは x と書いてしまうわけです。すると、例えば s での偏微分は

$$\frac{\partial f}{\partial s} = \frac{df}{dx} \frac{\partial x}{\partial s}$$

となります。これは d と ∂ の違いを無視すれば普通の分数の間の等式になっています。

1.2.3 2変数関数に1変数関数を2つ入れた場合

f が2変数関数 $f(x, y)$ で、これに二つの1変数関数 $\xi(s), \eta(s)$ を合成する場合を考えましょう。この場合は今までとは違って少し複雑になります。イメージ図5を見て下さい。

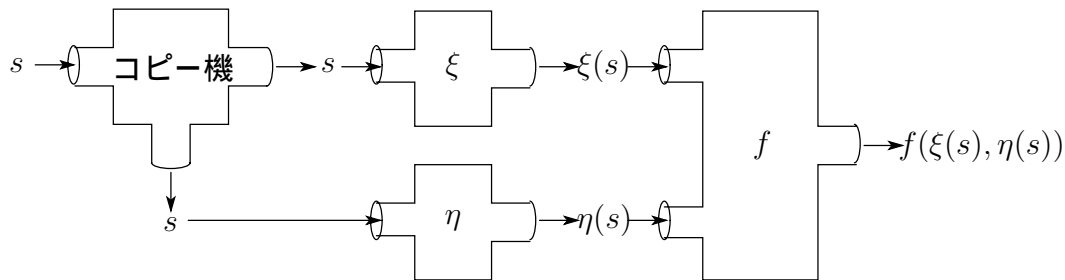


図 5: 2変数関数に1変数関数を二つ合成するには「コピー機」が必要

注意. ここで、中に入れる関数を $\xi(s), \eta(t)$ と別な独立変数を持つ1変数関数にしてはいけません。「関数を合成する」ということの意味は、

元の独立変数(たち)を別な独立変数(たち)の従属変数と見なす

という、いわば「一般化された変数変換」なので、 $f(\xi(s), \eta(t))$ という関数は

元の独立変数 x, y を新しい独立変数 s, t の従属変数と見なす

ということになってしまいます。つまり、この場合 $\xi(s)$ という関数は s の1変数関数ではなく、 s と t を変数に持つ2変数関数だがたまたま t には依存していない、と考えなくてはならないということになるわけです。だから、 $f(\xi(s), \eta(t))$ という合成関数はこのあとで紹介する「2変数関数に2変数関数を合成する」という場合に入ることになります。

さて、これまでと同じようにイメージ図5の出力口と入力口を貼り付けて一つの関数を作ってみましょう。すると、図6のようになります。これまでは大きな箱(つまり新しい関数)の中には古い関数たちが入っているだけでしたが、この場合は「コピー機」まで中に入ってしまったところが違います。これまでの2つの場合には、新しい独立変数 s (や t) の従属変数は x ただ一つだったのに、この場合には s の従属変数が x と y の二つあるところが大きな違いを生むのです。その

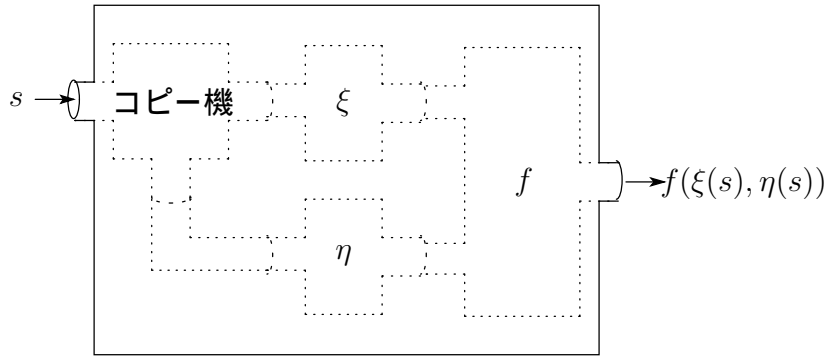


図 6: 大きな箱の中にコピー機まで入ってしまっているところが今までと違う

ことが原因で $\circ$ を使った合成関数の「名付け」もうまくいきません。それを考えるには関数（すなわち行き先が $\mathbb{R}$ ）をはみ出して、 ξ と η を合わせて一つの写像（今の場合は行き先が $\mathbb{R}^2$ ）と見なさなければならぬのです。

しかし、その見方は前小節で「この章の意味」を三つ挙げたうちの最後のものに深く関連しており、合成関数というものを見る視点を変えなければ理解できないので、あとで改めて説明することにし、この小節では合成関数の微分公式を紹介するだけにします。

さて、 $f \circ \xi$ のような記号がないので、

$$\varphi(s) = f(\xi(s), \eta(s))$$

と記号を変えることにします。（ φ は f に当たるギリシャ文字です³。）すると、

$$\frac{d\varphi}{ds}(s) = \frac{\partial f}{\partial x}(\xi(s), \eta(s)) \frac{d\xi}{ds}(s) + \frac{\partial f}{\partial y}(\xi(s), \eta(s)) \frac{d\eta}{ds}(s)$$

すなわち、

$$\varphi'(s) = f_x(\xi(s), \eta(s))\xi'(s) + f_y(\xi(s), \eta(s))\eta'(s)$$

となります。 ξ と η は1変数関数なので $d\xi/ds$ や ξ' という1変数関数の微分の記号を使うことは前の場合と同じです。

既に指摘したように、関数の範疇に留まるならこの場合は $\circ$ を使った合成関数の書き方ができないので、ここから (s) を取り除いて関数そのものの間の等式を作ることにはできませんが、 $x = \xi(s)$, $y = \eta(s)$ はわかっているものとしてしてしまえば、

$$\frac{d\varphi}{ds} = \frac{\partial f}{\partial x} \frac{d\xi}{ds} + \frac{\partial f}{\partial y} \frac{d\eta}{ds}$$

³正確には physics や photograph の ‘ph’ に当たるのだと思います。

あるいは

$$\varphi' = f_x \cdot \xi' + f_y \cdot \eta'$$

と書くことができます。

合成関数の中にコピー機が入っていること、つまり x も y も s に依存していることから、 s で微分するだけなのに x による偏微分の項と y に偏微分の項の両方が出てきてしまうのです。

f が x だけの関数だった場合の1変数関数の合成関数の微分公式と f が y だけの関数だった場合の1変数関数の合成関数の微分公式の両方が足されてしまう

と見れば自然に見えてくるのではないのでしょうか。

注意. このタイプの合成は力学でよく出会います。平面の各点 (x, y) にある「量」 $f(x, y)$ が与えられている一方、その平面上を時刻 t に $(\xi(t), \eta(t))$ という場所にいるように運動している粒子があったとき、時刻 t に粒子の居場所に対応する「量」を対応させることで関数 $\varphi(t)$ ができるわけです。例によって物理では φ という新しい文字を避けて f と書き、 ξ と η のことも x, y と書いてしまうので、

$$\frac{df}{dt} = \frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt}$$

となります。これでも誤解の余地はないというわけです。さらに、時刻による微分は上付きの点で表すことが多いので、

$$\dot{f} = \frac{\partial f}{\partial x} \dot{x} + \frac{\partial f}{\partial y} \dot{y}$$

と書いたりします。

1.2.4 2変数関数に2変数関数を2つ入れた場合

「1変数関数に1変数関数を入れた場合」と「1変数関数に2変数関数を入れた場合」で実質的な違いがなかったのと同じ理由で、「2変数関数に1変数関数を2つ入れた場合」と今から述べる「2変数関数に2変数関数を2つ入れ場合」にも実質的な違いはありません。新しい変数が増えたのもう一つコピー機が必要になり、合成関数の中にはコピー機が2台入っていることになりますが、例えば s という変数で偏微分する場合には、 s の通るコピー機はそのうちの1台だけなので「2変数関数に1変数関数を2つ入れた場合」と同じになるわけです。イメージ図は図7と図8です。

「2変数関数に1変数関数を2つ入れた場合」と同様、写像を使わず関数の範疇だけで考えるなら、 f と ξ と η と $\circ$ という記号で合成関数を表すことはできないので、 φ を使って

$$\varphi(s, t) = f(\xi(s, t), \eta(s, t))$$

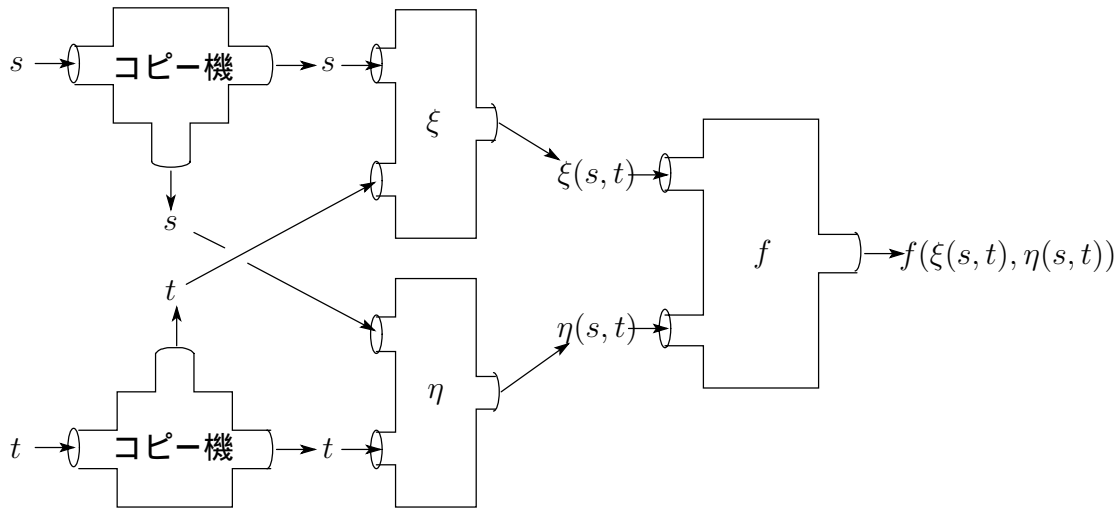


図 7: コピー機が2台必要

と表すことにします。すると、

$$\begin{aligned}\frac{\partial \varphi}{\partial s}(s, t) &= \frac{\partial f}{\partial x}(\xi(s, t), \eta(s, t)) \frac{\partial \xi}{\partial s}(s, t) + \frac{\partial f}{\partial y}(\xi(s, t), \eta(s, t)) \frac{\partial \eta}{\partial s}(s, t) \\ \frac{\partial \varphi}{\partial t}(s, t) &= \frac{\partial f}{\partial x}(\xi(s, t), \eta(s, t)) \frac{\partial \xi}{\partial t}(s, t) + \frac{\partial f}{\partial y}(\xi(s, t), \eta(s, t)) \frac{\partial \eta}{\partial t}(s, t)\end{aligned}$$

すなわち、

$$\begin{aligned}\varphi_s(s, t) &= f_x(\xi(s, t), \eta(s, t))\xi_s(s, t) + f_y(\xi(s, t), \eta(s, t))\eta_s(s, t) \\ \varphi_t(s, t) &= f_x(\xi(s, t), \eta(s, t))\xi_t(s, t) + f_y(\xi(s, t), \eta(s, t))\eta_t(s, t)\end{aligned}$$

となります。 $x = \xi(s), y = \eta(s)$ はわかっているものとしてして省略した書き方をすると、

$$\begin{aligned}\frac{\partial \varphi}{\partial s} &= \frac{\partial f}{\partial x} \frac{\partial \xi}{\partial s} + \frac{\partial f}{\partial y} \frac{\partial \eta}{\partial s} \\ \frac{\partial \varphi}{\partial t} &= \frac{\partial f}{\partial x} \frac{\partial \xi}{\partial t} + \frac{\partial f}{\partial y} \frac{\partial \eta}{\partial t}\end{aligned}$$

あるいは

$$\begin{aligned}\varphi_s &= f_x \cdot \xi_s + f_y \cdot \eta_s \\ \varphi_t &= f_x \cdot \xi_t + f_y \cdot \eta_t\end{aligned}$$

となります。

ゴチャゴチャしてわかりにくく見えるかも知れませんが、考えてみれば偏微分というのは微分しない変数のことは定数と考えるのですから、 s で偏微分

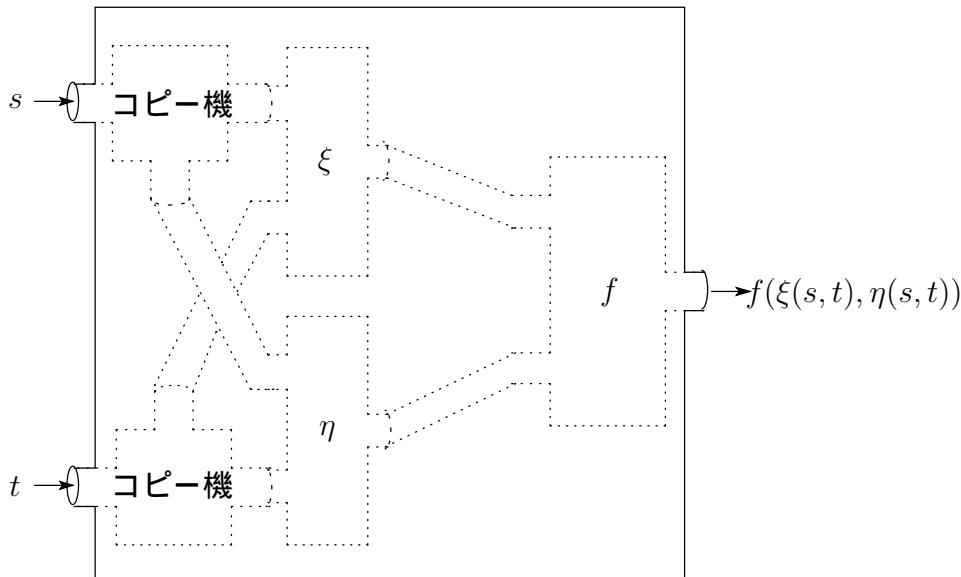


図 8: コピー機が2台入っているが、それぞれの変数の通るコピー機は1台だけ

する場合 t は定数扱いなので公式の中に t による偏微分は出てくるはずがありません。一方、「2変数関数に1変数関数を2つ入れる場合」と同様、 s という変数は x と y の両方を通じて f にかかわっているのですから、 s での偏微分には x による偏微分と y による偏微分の両方が入っていなければおかしいということになります。そうってから公式を見直すと、特にゴチャゴチャしているようには見えなくなるのではないかと思います。

注意. このような f と s, t との関わり方を考えると、次のような方法で公式を書き下せば間違いにくいのではないかと思います。

s で偏微分する場合で考えてみましょう。

まず一つ目の方法として、

1 変数関数の合成関数の微分公式を繰り返す

と考えてみます。まず f が x だけの関数であるかのように考えて

$$\frac{\partial f}{\partial x} \frac{\partial \xi}{\partial s}$$

と1変数関数の微分公式で d を ∂ に変えたものを書きます。次に、 f を y だけの関数であるかのように考えて、このとなりに

$$\frac{\partial f}{\partial y} \frac{\partial \eta}{\partial s}$$

を書きます。そして最後に真ん中に $+$ を書いて

$$\frac{\partial f}{\partial x} \frac{\partial \xi}{\partial s} + \frac{\partial f}{\partial y} \frac{\partial \eta}{\partial s}$$

とするわけです。

もう一つの方法として、

f のすべての変数が s に依存しているのだから、 f の偏微分はすべて出てくるという方に着目してみましょう。この場合にはまず f の偏微分を

$$\frac{\partial f}{\partial x} \quad \frac{\partial f}{\partial y}$$

と間をあけて並べて書いてしまいます。そして、 x による偏微分には x に入っている関数 ξ を偏微分したい変数 s で偏微分したものを掛け、 y による偏微分には y に入っている関数 η を s で偏微分したものを掛けて足し合わせるわけです。

$$\frac{\partial f}{\partial x} \frac{\partial \xi}{\partial s} + \frac{\partial f}{\partial y} \frac{\partial \eta}{\partial s}$$

となります。

まだ証明していないのだから納得はできないかも知れませんが、証明の前にこのように何度も書き下してみても何が起きているのか自分なりの手応えを掴んでおくことが大切だと思い、少ししつこく書いてみました。

1.2.5 一般の場合

このゼミでは原則として2変数の場合しか扱いませんが、ここまでの4つの場合でどのように公式が変化したかをよく見てもらえれば、 f が m 変数関数で、そこに m 個の n 変数関数を入れた場合の微分公式も正しく予想できるだろうと思うので、ここに答を書いておきましょう。

$f(x_1, x_2, \dots, x_m)$ に $x_i = \xi_i(s_1, s_2, \dots, s_n)$ を入れてできる $s_1, s_2, \dots, s_n$ の関数を φ とします。

$$\varphi(s_1, s_2, \dots, s_n) = f(\xi_1(s_1, s_2, \dots, s_n), \xi_2(s_1, s_2, \dots, s_n), \dots, \xi_n(s_1, \dots, s_n))$$

です。これの s_i による偏微分は次のようになります。ただし、長くなるので「どこでの微分か」を省いた書き方で書きます。

$$\frac{\partial \varphi}{\partial s_i} = \frac{\partial f}{\partial x_1} \frac{\partial \xi_1}{\partial s_i} + \frac{\partial f}{\partial x_2} \frac{\partial \xi_2}{\partial s_i} + \dots + \frac{\partial f}{\partial x_m} \frac{\partial \xi_m}{\partial s_i} = \sum_{j=1}^m \frac{\partial f}{\partial x_j} \frac{\partial \xi_j}{\partial s_i}$$

です。予想通りでしたか？

1.3 行列を使った記法と写像の微分

中に入れる方の関数が2変数なら合成関数も2変数になるので、偏微分は2種類あり、その結果公式もほぼ同じものが2つできました。その上、入れ物の方の関数 f も2変数だと、2つの公式の両方に f の二つの偏微分が両方とも顔を出していて、それが複雑そうな見た目の原因であることは上で確認してもらったとおりです。

ところで、第1章で2変数関数 f から df というものを作りました。この df は、抽象的には

点 (a, b) を決めるごとに、 (a, b) からの変位 $\vec{u} = \begin{bmatrix} u \\ v \end{bmatrix}$ に対して、 $f(a, b)$ からの変位を与える線形関数

でしたが、その実体は1行2列の行列

$$\left[\frac{\partial f}{\partial x}(a, b), \frac{\partial f}{\partial y}(a, b) \right]$$

を変位ベクトル $\vec{u}$ に左から掛けるという行列の掛け算によって与えられるものでした。だから、 $df_{(a,b)}$ とこの1行2列の行列は同じものだと思っても何ら差し支えありませんでした。

そこで、 df というもののこのような意味合いをとりあえず忘れて、1行2列の行列を

二つの偏微分を横に並べて一遍に表記する便利な記号

とだけ思ってしまうことにすると、これを使えば「2変数関数に2変数関数を2つ入れた場合」の合成関数の微分公式を $d\varphi$ と df の間の関係式としてスッキリ表せるのではないかと思えてきます。そこで、とりあえずやってみると、式が長くなるので「どこでか」を表す (a, b) や (x, y) を省いて書くことにすれば、

$$\begin{aligned} d\varphi &= \left[\frac{\partial \varphi}{\partial s}, \frac{\partial \varphi}{\partial t} \right] \\ &= \left[\frac{\partial f}{\partial x} \frac{\partial \xi}{\partial s} + \frac{\partial f}{\partial y} \frac{\partial \eta}{\partial s}, \frac{\partial f}{\partial x} \frac{\partial \xi}{\partial t} + \frac{\partial f}{\partial y} \frac{\partial \eta}{\partial t} \right] \\ &= \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right] \begin{bmatrix} \frac{\partial \xi}{\partial s} & \frac{\partial \xi}{\partial t} \\ \frac{\partial \eta}{\partial s} & \frac{\partial \eta}{\partial t} \end{bmatrix} \\ &= df \begin{bmatrix} \frac{\partial \xi}{\partial s} & \frac{\partial \xi}{\partial t} \\ \frac{\partial \eta}{\partial s} & \frac{\partial \eta}{\partial t} \end{bmatrix} \end{aligned}$$

と期待通りまとめられました。「期待通り」というのは「行列の記法の範疇に納まった」という意味です。

さて、ここでまた $d\varphi$ や df の意味を思い出してみましょう。どこまで思い出すのかというと、 $d\varphi$ や df を1行2列の行列だと思うのはよいとして、その行列は「変位ベクトル」という2次元縦ベクトルに左から掛けることによって初めて意味を持つのだ、ということ思い出すのです。すると、上の式変形の出発点である $d\varphi$ は、 (s, t) という独立変数の空間での変位ベクトルに左から掛けることになりませんが、式変形の最後の式を $d\varphi$ と置き換えると、 (s, t) 空間での変位ベクトルにかかっているのは2行2列の行列になってしまいます。しかも、その掛け算をした

結果できあがるベクトルには左から df が掛かってきています。 df を左から掛けることに意味のあるベクトルは (x, y) という変数の空間における変位ベクトルでした。ということは、上の式変形で現れた2行2列の行列

$$\begin{bmatrix} \frac{\partial \xi}{\partial s} & \frac{\partial \xi}{\partial t} \\ \frac{\partial \eta}{\partial s} & \frac{\partial \eta}{\partial t} \end{bmatrix}$$

は

(s, t) 空間の変位に左から掛けると (x, y) 空間の変位になる

という意味を持つものになっていなければ、上の式変形は単なる形式的なものに過ぎなかったことになってしまいます。

ここで、先ほど保留にした写像という概念を登場させましょう。2変数関数に1変数関数や2変数関数を入れる場合、記号 $\circ$ を使った合成関数の表し方がうまく通用しませんでした。その理由は「コピー機」が邪魔だったからです。この「コピー機」が消えてしまうような設定を作り出してくれるのが写像という概念です。なんだか恐ろしげな言い回しになってしまっていますが、何のことはない、二つの変数 x と y を (x, y) とまとめて一つのもののように扱ったのと同様にして、 $\xi(s, t)$ と $\eta(s, t)$ もまとめてしまおうというだけのことです。つまり、

s と t を決めると $\xi(s, t)$ と $\eta(s, t)$ の値がそれぞれ決まる

と考える代わりに、

(s, t) を決めると $(\xi(s, t), \eta(s, t))$ が決まる

と考えるだけのことです。ただし、微分を考える場合、 $(\xi(s, t), \eta(s, t))$ という値そのものではなく、変位ベクトルが重要であり、変位ベクトルは縦ベクトルであるような記法をここでは採用しているので、はじめから

$$X(s, t) := \begin{bmatrix} \xi(s, t) \\ \eta(s, t) \end{bmatrix}$$

というふうに縦に並べることにします。左辺に書いた $X(s, t)$ は (s, t) に2次元縦ベクトルを対応させる写像

$$X: \mathbb{R}^2 \rightarrow \mathbb{R}^2$$

を一文字で表す⁴ために決めた記号です。値がベクトルなので上に矢印を付けてもよいですが、面倒なのでやめました。

⁴ ξ の大文字の Ξ を使おうかとも思ったのですが、あまりにもなじみがないでしょうから、対応するローマ字の方の大文字にしておきました。

このような写像 X を導入すると、まず、

$$\varphi(s, t) = f \circ X(s, t)$$

という合成関数の記号が復活します。また、そのような形式的なことばかりではなく、 X が

(s, t) 空間の点に (x, y) 空間の点に対応させる

写像であることから、 X の「微分」と呼べるものがあるとすれば、それは

(s, t) 空間の変位ベクトルに (x, y) 空間の変位ベクトルに対応させる

というまさに欲しかったものになるはずです。そう思ってもう一度問題の行列を眺めてみると、第1行は

x 成分を与える関数 ξ の d 、つまり x 成分の変位を与える1行2列の行列 $d\xi$

になっているし、第2行は

y 成分を与える関数 η の d 、つまり y 成分の変位を与える1行2列の行列 $d\eta$

になっていますから、行列全体としては

(x, y) の変位を縦ベクトルとして与える行列

になっているわけです。

以上により、問題の行列は X の微分と呼ぶにふさわしい行列であると言えます。そこで、ちゃんと記号を決めるべきなのですが、 d という記号は関数にしか使わないという習慣が（もちろん意味があって）決まっているので dX と書くわけには行かないのです。そこで、1変数関数のときの記号を流用して $X'(s, t)$ と書く、という選択肢が一つあります。それから、この行列を初めて問題にした（と言われている）ヤコビ (Jacobi) の頭文字をとって $J_X(s, t)$ と書く記法もよく使われます。実際、この行列そのものの名前は「 X の微分」ではなく「 X のヤコビ行列」という方が普通です。以上、記号は

$$X'(s, t) = J_X(s, t) = \begin{bmatrix} \frac{\partial \xi}{\partial s} & \frac{\partial \xi}{\partial t} \\ \frac{\partial \eta}{\partial s} & \frac{\partial \eta}{\partial t} \end{bmatrix}$$

ということになります。これを使うと合成関数の微分公式は

$$d(\varphi \circ X) = df \cdot J_X$$

とまとめることができます。

ここまでは「2変数関数に2変数関数を2つ入れた場合」に限って説明しましたが、全く同じ行列を使った整理が「 n 変数関数に m 変数関数を n 個入れた場合」に成り立つことも実際に計算をしてもらえれば分かります。つまり、

$$f(x_1, x_2, \dots, x_m)$$

に

$$x_i = \xi_i(s_1, s_2, \dots, s_n) \quad i = 1, 2, \dots, m$$

を入れることを、

$$X(s_1, s_2, \dots, s_n) := \begin{bmatrix} \xi_1(s_1, s_2, \dots, s_n) \\ \xi_2(s_1, s_2, \dots, s_n) \\ \vdots \\ \xi_m(s_1, s_2, \dots, s_n) \end{bmatrix}$$

という写像 $X: \mathbb{R}^n \rightarrow \mathbb{R}^m$ との合成だと考えると、 X の微分、すなわちヤコビ行列を

$$X'(s_1, \dots, s_n) = J_X(s_1, \dots, s_n) = \begin{bmatrix} \frac{\partial \xi_1}{\partial s_1} & \frac{\partial \xi_1}{\partial s_2} & \cdots & \frac{\partial \xi_1}{\partial s_n} \\ \frac{\partial \xi_2}{\partial s_1} & \frac{\partial \xi_2}{\partial s_2} & \cdots & \frac{\partial \xi_2}{\partial s_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \xi_m}{\partial s_1} & \frac{\partial \xi_m}{\partial s_2} & \cdots & \frac{\partial \xi_m}{\partial s_n} \end{bmatrix}$$

という m 行 n 列の行列で定義すれば（と言うか、こうであるべきであって）、合成関数の微分公式は

$$d(f \circ X) = df \cdot J_X$$

と表されることになります。

このように一般の m, n で考えてみると、 m や n が1であって悪い理由は何もないことに気づくでしょう。つまり、単なる関数も「 $\mathbb{R}$ への写像」じゃないか、というわけです。そう考えると、

2変数関数 $f(x, y)$ の微分、つまりヤコビ行列は1行2列の行列 df そのもの

であり、

1変数関数 $f(x)$ の微分、つまりヤコビ行列は $f'(x)$ を唯一の成分とする1行1列の行列

だったんだ、というわけです。

ここまで一般的に考えるなら、記号も d とか J とかはやめてシンプルにダッシュで統一してしまうことにすれば、

$$\mathbb{R}^n \xrightarrow{g} \mathbb{R}^m \xrightarrow{f} \mathbb{R}^l$$

に対して（もちろん n, m, l は1でもO.K.です）

$$(f \circ g)' = (f' \circ g) \cdot g'$$

が成り立つことになります。1変数関数のときの公式が完全に再現されるわけです。

1.4 「連鎖律」という用語について

合成関数の微分公式のことを「連鎖律」と呼ぶことが多いのですが、実は、教科書によって（つまり人によって）この言葉で指している公式が違っているのです。そのせいか岩波の数学事典（第3版）には連鎖律という言葉が載っていません。

連鎖律は英語で言うと chain rule となります。（そのまんまですが。）「連鎖」とか「chain」という言葉の意味するところは「連鎖反応」などと同じだと思います。とすると、上で紹介した微分公式を連鎖律と呼ぶ人の心は、

s で偏微分しただけなのに、 f のすべての変数が s とつながっているために、 f のすべての偏微分が出てきてしまう

というところに「連鎖」の気持ちを感じているのだと思います。だから、このように連鎖律という言葉を使う人は1変数関数の合成関数の微分公式のことは連鎖律とは呼ばないようです。

一方、1変数関数の微分公式

$$(f \circ \xi)' = (f' \circ \xi) \cdot \xi'$$

は

全体を微分することが一つ一つの関数の微分を引き起こしている

と見ることができます。二つの関数の合成だとわかりにくいですが、例えば、

$$(f_4 \circ f_3 \circ f_2 \circ f_1)' = (f_4' \circ f_3 \circ f_2 \circ f_1) \cdot (f_3' \circ f_2 \circ f_1) \cdot (f_2' \circ f_1) \cdot f_1'$$

と関数を増やして書いてみると、いかにも「連鎖反応」が起きているふうに見えないでしょうか？これは1変数関数の場合だけの公式ですが、前小節最後の注意で説明したように、関数に留まることをやめて写像とその微分を使うことにすると、この公式は多変数でもそのまま成り立つことが結論されます。この形まで話

を整理してはじめて「連鎖律」と呼ぶのだと考える人もいます。その場合は1変数関数の合成関数の微分公式も連鎖律と呼ぶことになるでしょう。

おそらく前者の使い方の方が普通だと思いますが、後者の使い方をする人も結構いますし、そういう教科書もかなりあります。そういうわけなので、このプリントでは連鎖律という言葉は使わないことにしました。

2 合成関数の微分公式の証明

偏微分は1変数関数の微分の定義から全く安直に定義されるものであるにも拘わらず、多変数関数の微分として十分な情報を持っていること、正確には、問題にしている多変数関数が全微分可能、つまり1次近似をもつものならば、その1次近似式は偏微分によって完全に決定されることを見ました。ということは、合成関数の微分公式も、高校のときに学んだ1変数関数の合成関数の微分公式の証明と同じように証明することも可能なはずだし、また、1次近似の視点から証明することも可能なはずです。証明なんて一つやっておけば十分だという考え方もありますし、とにかく証明して正しいことさえ確認しておけばよいという種類の定理もあるとは思いますが、合成関数の微分公式に関しては両方の証明を見ておくのがよいと思います。と言うのは、微分を1次近似と見る考え方の利点が、高校以来慣れ親しんできた「微分は平均の変化率の極限」という見方と比較することでよりよく理解できると思うからです。

そこで、まず「平均の変化率の極限」という偏微分の定義そのものを使って証明し、そのあと1次近似ということを使って証明することにします。

2.1 偏微分の定義に基づいた証明

前節で紹介した順番の通りに1変数の場合から証明して行きましょう。

2.1.1 1変数関数に1変数を入れた場合

$f(x)$ に $x = \xi(s)$ を合成するわけですが、微分とは本来導関数ではなく「ある1点での微係数」として定義されるものですので、 $s = s_0$ での微分を考えることとし、 $\xi(s_0) = x_0$ とおくことにします。

$f \circ \xi$ の $s = s_0$ での微分の定義式の分子分母に $\xi(s) - \xi(s_0)$ を掛けることで

$$\lim_{s \rightarrow s_0} \frac{f \circ \xi(s) - f \circ \xi(s_0)}{s - s_0} = \lim_{s \rightarrow s_0} \left(\frac{f \circ \xi(s) - f \circ \xi(s_0)}{\xi(s) - \xi(s_0)} \frac{\xi(s) - \xi(s_0)}{s - s_0} \right)$$

と変形します。ここで ξ が s_0 で微分可能なら

$$\lim_{s \rightarrow s_0} \frac{\xi(s) - \xi(s_0)}{s - s_0} = \frac{d\xi}{ds}(s_0)$$

が存在します。さらに、 ξ が s_0 で微分可能なら ξ は s_0 で連続です。つまり、

$$\lim_{s \rightarrow s_0} \xi(s) = \xi(s_0)$$

が成り立ちます。よって、もし f が x_0 で微分可能なら、 $x = \xi(s)$ と置き換えることで、

$$\lim_{s \rightarrow s_0} \frac{f \circ \xi(s) - f \circ \xi(s_0)}{\xi(s) - \xi(s_0)} = \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = \frac{df}{dx}(x_0)$$

となります。以上より、

$$\begin{aligned} \lim_{s \rightarrow s_0} \frac{f \circ \xi(s) - f \circ \xi(s_0)}{s - s_0} &= \lim_{s \rightarrow s_0} \left(\frac{f \circ \xi(s) - f \circ \xi(s_0)}{\xi(s) - \xi(s_0)} \frac{\xi(s) - \xi(s_0)}{s - s_0} \right) \\ &= \left(\lim_{s \rightarrow s_0} \frac{f \circ \xi(s) - f \circ \xi(s_0)}{\xi(s) - \xi(s_0)} \right) \left(\lim_{s \rightarrow s_0} \frac{\xi(s) - \xi(s_0)}{s - s_0} \right) \\ &= \frac{df}{dx}(x_0) \frac{d\xi}{ds}(s_0) \end{aligned}$$

となって極限が存在します。よって、 $f \circ \xi$ は s_0 で微分可能であり、

$$\frac{d(f \circ \xi)}{ds}(s_0) = \frac{df}{dx}(x_0) \frac{d\xi}{ds}(s_0)$$

であることが証明できました。

注意. これでほとんど良いのですが、一つだけ問題があります。それは s_0 のどんなに近くにも $\xi(s) = \xi(s_0)$ となる s が存在する場合、 $\xi(s) - \xi(s_0)$ で割ることができないという問題です。 $\xi(s)$ が s_0 の近くで定数関数であるか、無限に振動して $\xi(s_0)$ と同じ値を無限に取ってしまう場合です。この場合には $\xi'(s_0) = 0$ である一方 $f \circ \xi(s)$ も s_0 の近くで $f \circ \xi(s_0)$ と同じ値を取るなので、やはり $(f \circ \xi)'(s_0) = 0$ となり、結局公式は成り立っています。このようなことを気にすると何をやっているのか大筋が掴みにくくなってしまうので、今後はこのような特別な場合は気にせずに進むことにします。

2.1.2 1変数関数に2変数関数を入れた場合

次に、 $f(x)$ に $x = \xi(s, t)$ を入れた場合を考えましょう。どちらでも同じですので、 s での偏微分を考えます。 $x_0 = \xi(s_0, t_0)$ として (s_0, t_0) における s での偏微分を計算しましょう。

$f \circ \xi$ の (s_0, t_0) における s による偏微分の数値とは、 t に t_0 を代入してできる s の1変数関数

$$\varphi(s) := f \circ \xi(s, t_0)$$

の s_0 における微分の数値です。一方、 $\xi(s, t)$ に $t = t_0$ を代入してできる s の1変数関数を

$$\psi(s) := \xi(s, t_0)$$

と定義すれば、

$$\varphi(s) = f \circ \xi(s, t_0) = f(\xi(s, t_0)) = f(\psi(s)) = f \circ \psi(s)$$

となって、 $\varphi(s)$ は1変数関数 $f(x)$ に1変数関数 $x = \psi(s)$ を入れた合成関数です。よって、 $\varphi(s)$ の微分は先ほど証明した1変数関数の合成関数の微分公式で計算でき、結果は、 f が x_0 で、 ψ が s_0 で微分可能なら φ は s_0 で微分可能で

$$\frac{d\varphi}{ds}(s_0) = \frac{df}{dx}(x_0) \frac{d\psi}{ds}(s_0)$$

です。ここで、 $\psi(s) = \xi(s, t_0)$ なのですから、 $\psi(s)$ の s_0 における微分とは $\xi(s, t)$ の (s_0, t_0) における s による偏微分の値です。

以上をまとめると、 f が x_0 で微分可能で ξ が (s_0, t_0) において s で偏微分可能なら $f \circ \xi$ も (s_0, t_0) において s で偏微分可能で、

$$\frac{\partial(f \circ \xi)}{\partial s}(s_0, t_0) = \frac{df}{dx}(x_0) \frac{\partial \xi}{\partial s}(s_0, t_0)$$

となることが分かりました。

この証明から分かるように、偏微分が注目する変数を定数扱いしての1変数関数の微分に過ぎないことから、偏微分を考える上では中に入れる方の関数の変数がどんなに多くなっても1変数関数を入れる場合と実質上の違いはありません。

2.1.3 2変数関数に1変数関数を2つ入れた場合

以上のことから、この場合、つまり $f(x, y)$ に $x = \xi(s), y = \eta(s)$ を入れた場合が本質的な場合だということになります。気合いを入れていきましょう。

$x_0 = \xi(s_0), y_0 = \eta(s_0)$ とし、

$$\varphi(s) := f(\xi(s), \eta(s))$$

の s_0 での微分を計算します。微分の定義式を書くと、

$$\lim_{s \rightarrow s_0} \frac{\varphi(s) - \varphi(s_0)}{s - s_0} = \lim_{s \rightarrow s_0} \frac{f(\xi(s), \eta(s)) - f(\xi(s_0), \eta(s_0))}{s - s_0} \quad (1)$$

となります。このあとどう式変形をすればよいでしょうか。

しかし、この式(1)をぼんやりと眺めていると、第1章で1変数関数の積の微分を導いたときの式を思い出しませんか？ 実際、 $f(x, y) = xy$ なら $\varphi(s) = \xi(s)\eta(s)$ となって $\varphi(s)$ の微分は単なる積の微分になります。そこで、積の微分法の証明からヒントを得るために、積の微分法の証明のポイントを思い出してみましょう。

$$\begin{aligned} \frac{\xi(s)\eta(s) - \xi(s_0)\eta(s_0)}{s - s_0} &= \frac{\xi(s)\eta(s) - \xi(s_0)\eta(s) + \xi(s_0)\eta(s) - \xi(s_0)\eta(s_0)}{s - s_0} \\ &= \frac{\xi(s) - \xi(s_0)}{s - s_0} \eta(s) + \xi(s_0) \frac{\eta(s) - \eta(s_0)}{s - s_0} \end{aligned}$$

というふうに、分子にわざと $-\xi(s_0)\eta(s) + \xi(s_0)\eta(s)$ という項を水増しするところがミソでした。そこで、式(1)でも同じように、

$$-f(\xi(s_0), \eta(s)) + f(\xi(s_0), \eta(s_0))$$

を間に入れてみましょう。すると、

$$\begin{aligned} & \frac{\varphi(s) - \varphi(s_0)}{s - s_0} \\ &= \frac{f(\xi(s), \eta(s)) - f(\xi(s_0), \eta(s_0))}{s - s_0} \\ &= \frac{f(\xi(s), \eta(s)) - f(\xi(s_0), \eta(s)) + f(\xi(s_0), \eta(s)) - f(\xi(s_0), \eta(s_0))}{s - s_0} \\ &= \frac{f(\xi(s), \eta(s)) - f(\xi(s_0), \eta(s))}{\xi(s) - \xi(s_0)} \frac{\xi(s) - \xi(s_0)}{s - s_0} \end{aligned} \quad (2)$$

$$+ \frac{f(\xi(s_0), \eta(s)) - f(\xi(s_0), \eta(s_0))}{\eta(s) - \eta(s_0)} \frac{\eta(s) - \eta(s_0)}{s - s_0} \quad (3)$$

となります。

最後の式で $s \rightarrow s_0$ としましょう。 ξ と η が s_0 で微分可能なら、第1項(2)の第2因子と第2項(3)の第2因子は、それぞれ

$$\lim_{s \rightarrow s_0} \frac{\xi(s) - \xi(s_0)}{s - s_0} = \frac{d\xi}{ds}(s_0), \quad \lim_{s \rightarrow s_0} \frac{\eta(s) - \eta(s_0)}{s - s_0} = \frac{d\eta}{ds}(s_0)$$

となります。また、 η が s_0 で微分可能なら連続でもあるので、 $\lim_{s \rightarrow s_0} \eta(s) = \eta(s_0)$ ですから、 f が (x_0, y_0) で y によって微分可能なら、式(3)の第1因子で $y = \eta(s)$ とし $s \rightarrow s_0$ とすると

$$\lim_{s \rightarrow s_0} \frac{f(\xi(s_0), \eta(s)) - f(\xi(s_0), \eta(s_0))}{\eta(s) - \eta(s_0)} = \lim_{y \rightarrow y_0} \frac{f(x_0, y) - f(x_0, y_0)}{y - y_0} = \frac{\partial f}{\partial y}(x_0, y_0)$$

となります。残るは(2)の第1因子だけですが、これが問題です。

もしも(2)の第1因子の分子が $f(\xi(s), \eta(s_0)) - f(\xi(s_0), \eta(s_0))$ であれば、つまり η の中身が s でなく s_0 なら、式(3)の第1因子と全く同様の計算で $s \rightarrow s_0$ のとき $f_x(x_0, y_0)$ という目標の値になってくれるのですが、 η の中身が s_0 ではなく s なので安直に $s \rightarrow s_0$ の極限を取ることができません。「そんなことないんじゃない? $s \rightarrow s_0$ のとき $\eta(s) \rightarrow y_0$ なんだから、 $\eta(s)$ を最初から $\eta(s_0) = y_0$ に取り替えておいても問題ないのでは?」と思われるかも知れません。つまり

$$\frac{f(\xi(s), \eta(s)) - f(\xi(s_0), \eta(s))}{\xi(s) - \xi(s_0)} \neq \frac{f(\xi(s), \eta(s_0)) - f(\xi(s_0), \eta(s_0))}{\xi(s) - \xi(s_0)}$$

だけれど、 $s \rightarrow s_0$ の極限を取ってしまえばそんな「小さな違い」はなくなって、めでたく $f_x(\xi(s_0), \eta(s_0))$ に収束するのではないかと。

ところが、これは一般には成り立ちません。例を見ておきましょう。(この関数は第1章で方向微分は可能だが全微分可能でない例として既に調べたことがある関数です。)

例 1. $f(x, y)$ を

$$f(x, y) = \begin{cases} \frac{3x^2y - y^3}{x^2 + y^2} & (x, y) \neq (0, 0) \\ 0 & (x, y) = (0, 0) \end{cases}$$

とし、 $\xi(s) = \eta(s) = s, s_0 = 0$ で考えてみましょう。 $(\xi(s_0), \eta(s_0)) = (0, 0)$ ですので、まず $(0, 0)$ での x による偏微分の値を計算すると、

$$\frac{\partial f}{\partial x}(0, 0) = \lim_{x \rightarrow 0} \frac{f(x, 0) - f(0, 0)}{x} = \lim_{x \rightarrow 0} \frac{0 - 0}{x} = 0$$

となって $f_x(0, 0) = 0$ です。 η の中身が s でなく s_0 だった場合の第1項第1因子は

$$\lim_{s \rightarrow 0} \frac{f(\xi(s), \eta(0)) - f(\xi(0), \eta(0))}{\xi(s) - \xi(0)} = \lim_{s \rightarrow 0} \frac{f(s, 0) - f(0, 0)}{s} = \lim_{s \rightarrow 0} 0 = 0$$

となってちゃんと $f_x(0, 0)$ と一致します。一方、 $\eta(s)$ を使った本物の第1項第1因子を計算すると、

$$\begin{aligned} \lim_{s \rightarrow 0} \frac{f(\xi(s), \eta(s)) - f(\xi(0), \eta(s))}{\xi(s) - \xi(0)} &= \lim_{s \rightarrow 0} \frac{f(s, s) - f(0, s)}{s} \\ &= \lim_{s \rightarrow 0} \frac{1}{s} \left(\frac{3s^3 - s^3}{s^2 + s^2} - \frac{-s^3}{s^2} \right) = \lim_{s \rightarrow 0} 2 = 2 \end{aligned}$$

となって $f_x(0, 0) = 0$ に一致しません。 ■

これで、 $f(x, y)$ は x で偏微分可能というだけではこの場合の合成関数の微分公式が成り立たないことになってしまいました。とは言え、勘のいい人はお気づきだと思いますが、 f が C^1 -級なら大丈夫です。

注意. 実は、 f が全微分可能なら成り立つのですが、ここでの証明方法で全微分可能性だけしか仮定しないと本筋でないとこで難しくなってしまうので、ここでは C^1 -級として話を進めます。(次の小節で全微分可能性だけを仮定して1次近似の視点から証明します。)

なお、 C^1 級まで要求しなくても、微分したい点 (x_0, y_0) で f_y が連続なら十分なのですが⁵、このような細かいことにいちいち拘って文を書くと、ただでさえくどいプリントが無用に長くなってしまいますので C^1 -級としてしまいます。

⁵分子に $-f(\xi(s_0), \eta(s)) + f(\xi(s_0), \eta(s_0))$ を挟み込んだから f_y の方の連続性が効いてくるので、 $-f(\xi(s), \eta(s_0)) + f(\xi(s), \eta(s_0))$ を挟み込めば f_x の方の連続性を使うことが以下の証明を読んでもらえれば分かります。つまり、 (x_0, y_0) で f_x か f_y のどちらか一方だけ連続であれば公式は成り立つのです。しかし、このような関数は(そうなるように無理に作ったものは別として)普通にはありませんので、 C^1 -級としてしまって何の問題もありません。

f が C^1 -級なら、(2) の第1因子は期待通り $f_x(x_0, y_0)$ に収束することを証明しましょう。

証明. まず、 $f(x, y)$ を y だけの関数だと思って1変数関数の平均値の定理を使うと、

$$f(x, y) = f(x, b) + \frac{\partial f}{\partial y}(x, c)(y - b)$$

となる c が b と y の間にいることが分かります。ただし、この c は b や y だけでなく x にも依存していることに注意してください。

ここで x として $\xi(s)$ と $\xi(s_0) = x_0$ を、 y として $\eta(s)$ を、 b として $\eta(s_0) = y_0$ と取ると、

$$\begin{aligned} f(\xi(s), \eta(s)) &= f(\xi(s), \eta(s_0)) + \frac{\partial f}{\partial y}(\xi(s), c)(\eta(s) - \eta(s_0)) \\ f(\xi(s_0), \eta(s)) &= f(\xi(s_0), \eta(s_0)) + \frac{\partial f}{\partial y}(\xi(s_0), c')(\eta(s) - \eta(s_0)) \end{aligned}$$

となる c と c' が $\eta(s)$ と $\eta(s_0)$ の間に存在することがわかります。この2式を式(2)の第1因子に入れると、

$$\begin{aligned} \frac{f(\xi(s), \eta(s)) - f(\xi(s_0), \eta(s))}{\xi(s) - \xi(s_0)} &= \frac{f(\xi(s), \eta(s_0)) - f(\xi(s_0), \eta(s_0))}{\xi(s) - \eta(s_0)} \\ &\quad + \frac{\partial f}{\partial y}(\xi(s), c) \frac{\eta(s) - \eta(s_0)}{\xi(s) - \xi(s_0)} \\ &\quad - \frac{\partial f}{\partial y}(\xi(s_0), c') \frac{\eta(s) - \eta(s_0)}{\xi(s) - \xi(s_0)} \end{aligned}$$

となります。ここで $s \rightarrow s_0$ の極限を取りたいのですが、そのとき

$$\lim_{s \rightarrow s_0} \frac{\eta(s) - \eta(s_0)}{\xi(s) - \xi(s_0)} = \lim_{s \rightarrow s_0} \frac{\eta(s) - \eta(s_0)}{s - s_0} \frac{s - s_0}{\xi(s) - \xi(s_0)} = \frac{\eta'(s_0)}{\xi'(s_0)}$$

であることと、 f_y が連続であること、および c も c' も $\eta(s_0)$ に収束すること (c も c' も $\eta(s)$ と $\eta(s_0)$ の間の数で、 $s \rightarrow s_0$ のとき $\eta(s) \rightarrow \eta(s_0)$ となるからです)

に注意すれば、

$$\begin{aligned}
& \lim_{s \rightarrow s_0} \frac{f(\xi(s), \eta(s)) - f(\xi(s_0), \eta(s))}{\xi(s) - \xi(s_0)} \\
&= \lim_{s \rightarrow s_0} \frac{f(\xi(s), \eta(s_0)) - f(\xi(s_0), \eta(s_0))}{\xi(s) - \xi(s_0)} \\
&\quad + \left(\lim_{s \rightarrow s_0} \frac{\partial f}{\partial y}(\xi(s), c) \right) \left(\lim_{s \rightarrow s_0} \frac{\eta(s) - \eta(s_0)}{\xi(s) - \xi(s_0)} \right) \\
&\quad - \left(\lim_{s \rightarrow s_0} \frac{\partial f}{\partial y}(\xi(s_0), c') \right) \left(\lim_{s \rightarrow s_0} \frac{\eta(s) - \eta(s_0)}{\xi(s) - \xi(s_0)} \right) \\
&= \frac{\partial f}{\partial x}(\xi(s_0), \eta(s_0)) \\
&\quad + \frac{\partial f}{\partial y}(\xi(s_0), \eta(s_0)) \frac{\eta'(s_0)}{\xi'(s_0)} \\
&\quad - \frac{\partial f}{\partial y}(\xi(s_0), \eta(s_0)) \frac{\eta'(s_0)}{\xi'(s_0)} \\
&= \frac{\partial f}{\partial x}(\xi(s_0), \eta(s_0))
\end{aligned}$$

となります。これで示しました。□

以上により、 f が C^1 -級で ξ と η がともに微分可能であるとき、合成関数 $\varphi(s) = f(\xi(s), \eta(s))$ は微分可能で、

$$\frac{d\varphi}{ds}(s) = \frac{\partial f}{\partial x}(\xi(s), \eta(s)) \frac{d\xi}{ds}(s) + \frac{\partial f}{\partial y}(\xi(s), \eta(s)) \frac{d\eta}{ds}(s)$$

となるということが示せました。真に新しい場合だけあってなかなか長かったですね。

2.1.4 2変数関数に2変数関数を2つ入れた場合

これまでの証明から分かるように、合成関数の偏微分は「中身」の関数が多変数であることは全然苦にならず、「入れ物」の関数が多変数であることが面倒のもとです。だから、「1変数関数に2変数関数を入れた場合」の議論と「2変数関数に1変数関数を2つ入れた場合」の議論を単純にくっつけば、

$f(x, y)$ が C^1 -級の関数で $\xi(s, t), \eta(s, t)$ が s について偏微分可能なら、合成関数 $\varphi(s, t) = f(\xi(s, t), \eta(s, t))$ は s で偏微分可能で、

$$\frac{\partial \varphi}{\partial s} = \frac{\partial f}{\partial x}(\xi(s, t), \eta(s, t)) \frac{\partial \xi}{\partial s}(s, t) + \frac{\partial f}{\partial y}(\xi(s, t), \eta(s, t)) \frac{\partial \eta}{\partial s}(s, t)$$

が成り立つ。

という(2変数としては)もっとも一般的な場合の公式が得られます。

とは言え、これだけで終わりにしてしまうとだまされたような感じがするかも知れないので、ちゃんと証明しておきましょう。

証明. (s_0, t_0) における s による偏微分の値を計算してみましょう。 $\varphi(s, t)$ に $t = t_0$ を入れてできる s だけの1変数関数を $\psi(s)$ と書くことにします。

$$\psi(s) := \varphi(s, t_0) = f(\xi(s, t_0), \eta(s, t_0))$$

です。計算したいのは $\psi'(s_0)$ の値です。ここで、 s だけを変数とする1変数関数2つを、

$$\alpha(s) := \xi(s, t_0), \quad \beta(s) := \eta(s, t_0)$$

によって定義すると、 ψ は

$$\psi(s) = f(\alpha(s), \beta(s))$$

というふうに「2変数関数に1変数関数を2つ入れた関数」になっていることがわかります。今、 $f(x, y)$ は C^1 -級、 ξ と η は s で偏微分可能と仮定しているので、特に $\alpha(s), \beta(s)$ は微分可能です。よって、前小節の微分公式が使えて、

$$\frac{d\psi}{ds}(s_0) = \frac{\partial f}{\partial x}(\alpha(s_0), \beta(s_0)) \frac{d\alpha}{ds}(s_0) + \frac{\partial f}{\partial y}(\alpha(s_0), \beta(s_0)) \frac{d\beta}{ds}(s_0)$$

が成り立ちます。ここで、 α, β の定義から

$$\frac{d\alpha}{ds}(s_0) = \frac{\partial \xi}{\partial s}(s_0, t_0), \quad \frac{d\beta}{ds}(s_0) = \frac{\partial \eta}{\partial s}(s_0, t_0), \quad \alpha(s_0) = \xi(s_0, t_0), \quad \beta(s_0) = \eta(s_0, t_0)$$

なので、これらを代入すると、

$$\begin{aligned} \frac{\partial \varphi}{\partial s}(s_0, t_0) &= \frac{d\psi}{ds}(s_0) \\ &= \frac{\partial f}{\partial x}(\xi(s_0, t_0), \eta(s_0, t_0)) \frac{\partial \xi}{\partial s}(s_0, t_0) + \frac{\partial f}{\partial y}(\xi(s_0, t_0), \eta(s_0, t_0)) \frac{\partial \eta}{\partial s}(s_0, t_0) \end{aligned}$$

となります。これは証明したい公式です。

t による偏微分も全く同様です。□

2.1.5 一般の場合

このゼミでは特に証明は原則として2変数の場合しか扱いませんが、一般の場合もほとんど同様に証明できるので、興味のある人のために3変数以上になった場合に注意すべき点だけ書いておきます。

ここまでの議論で「 n 変数関数に n 個の m 変数関数を入れた場合」の公式を証明するには、 $m = 1$ で証明しておけば十分なことが納得してもらえらるでしょう。そして、その場合の証明のポイントが微分の定義式における「分子の水増し」にあることもわかってもらえるものと思います。

「分子の水増し」という方法は1変数関数の積の微分法の証明をまねしたものです。積の微分法の場合、 n 個の関数の積を微分するには2個の関数の場合を何度か使えばできてしまうのに、ここで示したい「 n 変数関数に1変数関数を n 個入れた関数の微分」は $n = 2$ の場合の結果に帰着することはできません。積の微分の場合、例えば、

$$\begin{aligned}(f(x)g(x)h(x))' &= f'(x)(g(x)h(x)) + f(x)(g(x)h(x))' \\ &= f'(x)g(x)h(x) + f(x)(g'(x)h(x) + g(x)h'(x)) \\ &= f'(x)g(x)h(x) + f(x)g'(x)h(x) + f(x)g(x)h'(x)\end{aligned}$$

とできてしまうのに、3変数関数 $f(x, y, z)$ に $x = \xi(s), y = \eta(s), z = \zeta(s)$ を入れてできる s の関数

$$\varphi(s) = f(\xi(s), \eta(s), \zeta(s))$$

の s による微分を2変数の場合の微分公式を使って計算することはできません（試みてみてください）。そこで、証明自体を工夫しなければならなくなるわけですが、その場合のポイントは「分子をどう水増しするか」にあるわけです。

上に書いた3変数の場合でポイントになる式変形を書いてみましょう。式を少しでもスッキリさせるために、 $\xi(s_0) = x_0, \eta(s_0) = y_0, \zeta(s_0) = z_0$ とします。すると、

$$\begin{aligned}\frac{\varphi(s) - \varphi(s_0)}{s - s_0} &= \frac{f(\xi(s), \eta(s), \zeta(s)) - f(x_0, y_0, z_0)}{s - s_0} \\ &= \frac{f(\xi(s), \eta(s), \zeta(s)) - f(x_0, \eta(s), \zeta(s))}{s - s_0} \\ &\quad + \frac{f(\xi(s_0), \eta(s), \zeta(s)) - f(x_0, y_0, \zeta(s))}{s - s_0} \\ &\quad + \frac{f(\xi(s_0), \eta(s_0), \zeta(s)) - f(x_0, y_0, z_0)}{s - s_0} \\ &= \frac{f(\xi(s), \eta(s), \zeta(s)) - f(x_0, \eta(s), \zeta(s))}{\xi(s) - x_0} \frac{\xi(s) - \xi(s_0)}{s - s_0} \\ &\quad + \frac{f(x_0, \eta(s), \zeta(s)) - f(x_0, y_0, \zeta(s))}{\eta(s) - y_0} \frac{\eta(s) - \eta(s_0)}{s - s_0} \\ &\quad + \frac{f(x_0, y_0, \zeta(s)) - f(x_0, y_0, z_0)}{\zeta(s) - z_0} \frac{\zeta(s) - \zeta(s_0)}{s - s_0}\end{aligned}$$

となります。これで $s \rightarrow s_0$ の極限を取ると、 f が C^1 -級なら特に f_y と f_z は (x_0, y_0, z_0) で連続なので、「2変数関数に1変数関数を2つ入れる場合」と全く同様の手続きで示したい微分公式に収束する、というふうになるわけです。

2.2 1次近似という視点からの証明

前節で一応多変数関数における合成関数の微分法を手に入れることができましたが、イマイチ結論の式の意味やそれにたどり着いた筋道がわからないというのが正直なところではないでしょうか。

1変数関数の積の微分法と合成関数の微分法の証明の組み合わせっばい

では、「なぜこのような公式になるのか」は良く分からなくて当然だと思います。

ここでは、関数が全微分可能な場合、偏微分の値は「1次近似（すなわち接平面の式）の係数」であるという視点からすべての関数を全微分可能と仮定して前節の議論を見直してみましょう。つまり、

合成関数の1次近似は、それぞれの関数の1次近似の合成

となっていることを証明することで、合成関数の微分公式を理解しようというわけですね。

なお、1次近似は接線や接平面という図形的な解釈とこそよく馴染む考え方ですが、この小節では1次近似の定義に基づいた式の上での証明のみを行い、幾何学的なイメージについては節を改めて説明することにします。

2.2.1 1変数関数に1変数関数を入れた場合

やることは自体は簡単ですので、すぐに多変数関数の場合を扱ってもよいのですが、概念としては新しいので、焦らずに1変数関数の場合から順番に示して行きましょう。

1変数関数の場合、第1章（プリント29ページ）で説明したように、 a という数が $f'(x_0)$ と一致することは

$$p(x) = a(x - x_0) + f(x_0)$$

という1次式が

$$\lim_{x \rightarrow x_0} \frac{f(x) - p(x)}{x - x_0} = 0 \quad (4)$$

を満たすことと同値でした。つまり、

関数 $f(x)$ と定義域内の x_0 に対し、条件(4)を満たす1次近似式は存在したとしても一つだけであり、そのような $p(x)$ が存在することは f が x_0 で微分可能なことと同値で、そのとき

$$p(x) = f'(x_0)(x - x_0) + f(x_0)$$

である。

この $p(x)$ のことを $f(x)$ の $x = x_0$ における1次近似と呼びました。ここでやりたいことは、

微分可能な関数と微分したい点を決めると1次近似式はただ一つに定まる

ということを使って合成関数の微分公式を証明することです。

さて、 $f(x)$ に $x = \xi(s)$ を入れた関数 $f \circ \xi(s)$ の $s = s_0$ における1次近似式を考えたいわけですが、もちろん、今 $f(x)$ は $x_0 = \xi(s_0)$ で、 $\xi(s)$ は $s = s_0$ で微分可能と仮定するので、 $f(x)$ の x_0 における1次近似

$$p(x) = f'(x_0)(x - x_0) + f(x_0)$$

と $\xi(s)$ の s_0 における1次近似

$$\pi(s) = \xi'(s_0)(s - s_0) + x_0$$

がただ一つだけ存在しています⁶。

われわれが欲しいのは、 $f \circ \xi$ の s_0 における1次近似です。ここで、 $f \circ \xi(s) = f(\xi(s))$ という合成の定義と、1次式に1次式を代入しても1次式のままであることに目を付けて、 $p(x)$ に $x = \pi(s)$ を入れてみることを試みてみましょう。つまり、「 $p \circ \pi(s)$ が $f \circ \xi(s)$ の s_0 における1次近似なのではないか」と考えてみようというわけです。そこで、1次近似の定義式である条件式(4)（で $f(x)$ を $f \circ \xi(s)$ に置き換えたもの）に $p \circ \pi(s)$ を入れて計算してみましょう。すると、まず、

$$\begin{aligned} \lim_{s \rightarrow s_0} \frac{f \circ \xi(s) - p \circ \pi(s)}{s - s_0} &= \lim_{s \rightarrow s_0} \frac{f \circ \xi(s) - p \circ \xi(s) + p \circ \xi(s) - p \circ \pi(s)}{s - s_0} \\ &= \lim_{s \rightarrow s_0} \left(\frac{f(\xi(s)) - p(\xi(s))}{s - s_0} + \frac{p(\xi(s)) - p(\pi(s))}{s - s_0} \right) \end{aligned}$$

と変形できます。ここで、最後の式を第1項と第2項に分けて考えてみましょう。

まず、第1項について。極限を取る前の式を

$$\frac{f(\xi(s)) - p(\xi(s))}{s - s_0} = \frac{f(\xi(s)) - p(\xi(s))}{\xi(s) - \xi(s_0)} \frac{\xi(s) - \xi(s_0)}{s - s_0}$$

と変形しましょう。 ξ は s_0 で微分可能と仮定しているので、

$$\lim_{s \rightarrow s_0} \frac{\xi(s) - \xi(s_0)}{s - s_0} = \frac{d\xi}{ds}(s_0)$$

です。また、微分可能とうことは連続、すなわち $\lim_{s \rightarrow s_0} \xi(s) = \xi(s_0) = x_0$ が成り立っているので、 $x = \xi(s)$ と置き換えることにより、

$$\lim_{s \rightarrow s_0} \frac{f(\xi(s)) - p(\xi(s))}{\xi(s) - \xi(s_0)} = \lim_{x \rightarrow x_0} \frac{f(x) - p(x)}{x - x_0} = 0$$

⁶ $p(x)$ の p は多項式を意味する ‘polynomial’ の頭文字を取りました。また π は円周率みたいでイヤかも知れませんが、 p に当たるギリシャ文字なので使わせてください。

となります⁷。これが0になるのは $p(x)$ が $f(x)$ の $x = x_0$ における1次近似だからです。以上より、第1項については

$$\begin{aligned}\lim_{s \rightarrow s_0} \frac{f(\xi(s)) - p(\xi(s))}{s - s_0} &= \lim_{s \rightarrow s_0} \frac{f(\xi(s)) - p(\xi(s))}{\xi(s) - \xi(s_0)} \frac{\xi(s) - \xi(s_0)}{s - s_0} \\ &= \left(\lim_{s \rightarrow s_0} \frac{f(\xi(s)) - p(\xi(s))}{\xi(s) - \xi(s_0)} \right) \left(\lim_{s \rightarrow s_0} \frac{\xi(s) - \xi(s_0)}{s - s_0} \right) \\ &= 0 \cdot \frac{d\xi}{ds}(s_0) = 0\end{aligned}$$

となることがわかりました。

次に、第2項について。 $p(x)$ は一次式なので、具体的に

$$p(x) = f'(x_0)(x - x_0) + f(x_0)$$

と書いて計算してみましょう。すると、

$$\begin{aligned}\frac{p(\xi(s)) - p(\pi(s))}{s - s_0} &= \frac{f'(x_0)(\xi(s) - x_0) + f(x_0) - f'(x_0)(\pi(s) - x_0) - f(x_0)}{s - s_0} \\ &= f'(x_0) \frac{\xi(s) - \pi(s)}{s - s_0}\end{aligned}$$

となります。そして、 $\pi(s)$ が $\xi(s)$ の s_0 における1次近似であることから、

$$\lim_{s \rightarrow s_0} \frac{p(\xi(s)) - p(\pi(s))}{s - s_0} = f'(x_0) \lim_{s \rightarrow s_0} \frac{\xi(s) - \pi(s)}{s - s_0} = f'(x_0) \cdot 0 = 0$$

となります。

以上より、第1項も第2項も0に収束したので全体として0に収束します。つまり、1次式 $p \circ \xi(s)$ は

$$\lim_{s \rightarrow s_0} \frac{f \circ \xi(s) - p \circ \pi(s)}{s - s_0} = 0$$

を満たすことがわかりました。これは、

$$p \circ \pi(s) \text{ が } f \circ \xi(s) \text{ の } s = s_0 \text{ における1次近似である}$$

ということを意味します。

一方、 $f \circ \xi(s)$ の $s = s_0$ における1次近似は

$$(f \circ \xi)'(s_0)(s - s_0) + f(x_0)$$

であることがわかっています。ということは、

$$p \circ \pi(s) = (f \circ \xi)'(s_0)(s - s_0) + f(x_0)$$

⁷前小節での証明と同様、ここでも s_0 のどんなに近いところにも $\xi(s) = \xi(s_0)$ となる s が存在してしまう場合を考慮しませんでした。以前の注意を参照してください。

でなければならないわけです。 $p \circ \pi(s)$ を具体的に計算すると、

$$\begin{aligned} p \circ \pi(s) &= f'(x_0)(\pi(s) - x_0) + f(x_0) \\ &= f'(x_0)(\xi'(s_0)(s - s_0) + x_0 - x_0) + f(x_0) \\ &= f'(x_0)\xi'(s_0)(s - s_0) + f(x_0) \end{aligned}$$

となります。従って、

$$(f \circ \xi)'(s_0) = f'(x_0)\xi'(s_0)$$

である、ということが結論されます。これで証明できました。

何をしたのかというと、前小節では

$$\lim_{s \rightarrow s_0} \frac{f(\xi(s)) - f(x_0)}{s - s_0} = f'(x_0)\xi'(s_0)$$

であることを証明したのに対し、ここでは、右辺を左辺に移項して分子に載せた

$$\lim_{s \rightarrow s_0} \frac{f(\xi(s)) - f(x_0) - f'(x_0)\xi'(s_0)(s - s_0)}{s - s_0} = 0$$

を証明しただけです。全く同じ内容であり、しかも計算はより追いにくくなってしまっています。しかし、多変数になるとこの方法が途端に威力を発揮するのです。なぜなら、上の極限式は多変数関数にそのまま拡張することができないので前小節のように一つ一つの偏微分を1変数関数の微分と見て適用する以外になかったのが、下の極限式はそのまま多変数関数の場合に適用できて、多変数関数の微分にもっともふさわしい全微分可能性という概念になっているからです。

このような説明を長々とするより実際にその威力を見てもらった方が良いので、説明はこのくらいにして次へ進みましょう。

2.2.2 1変数関数に2変数関数を入れた場合

この場合が1変数関数に1変数関数を入れる場合と本質的に変わらないことは既に見てもらったとおりなのですが、

- 2変数関数でも1次近似は一つしかないことを確認する。
- 二つの偏微分が同時に求まってしまう場面を見る。

ということを簡単な場合で済ませておくのがよいと思うので、一足飛びに進むのはやめて、前小節と同じようにゆっくりと証明して行くことにしましょう。

まず、2変数関数における1次近似の一意性について確認します。このことは第1章で既に示してあります。それは第1章のプリント30ページの定理1

$f(x, y)$ が (x_0, y_0) で全微分可能なとき、 $f(x, y)$ は (x_0, y_0) で x でも y でも偏微分可能であり、1 次近似式 $p(x, y)$ は

$$p(x, y) = f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0)$$

である。

というものです。(ここでの記号に合わせて (a, b) を (x_0, y_0) に変えました。また、元の定理 1 には方向微分に関する部分もあるのですが、今は必要ないので省きました。) つまり、2 変数の 1 次式 $p(x, y)$ が

$$\lim_{(x, y) \rightarrow (x_0, y_0)} \frac{f(x, y) - p(x, y)}{\sqrt{(x - x_0)^2 + (y - y_0)^2}} = 0$$

を満たすことと、

$$p(x, y) = f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0)$$

であることは ($f(x, y)$ が (x_0, y_0) で全微分可能なら) 同値である、というわけです。このことを使って「1 変数関数に 2 変数関数を入れた場合」の合成関数の微分公式を証明してみましょう。

$f(x)$ に $x = \xi(s, t)$ を合成した関数 $f \circ \xi(s, t)$ の (s_0, t_0) における偏微分の値を両方同時に計算します。 $f(x)$ は $x_0 = \xi(s_0, t_0)$ で微分可能、 $\xi(s, t)$ は (s_0, t_0) で全微分可能と仮定します。そして、 $f(x)$ の x_0 における 1 次近似を $p(x)$ 、 $\xi(s, t)$ の (s_0, t_0) における 1 次近似を $\pi(s, t)$ としましょう。具体的には、

$$p(x) = f(x_0) + f'(x_0)(x - x_0)$$

および

$$\pi(s, t) = x_0 + \frac{\partial \xi}{\partial s}(s_0, t_0)(s - s_0) + \frac{\partial \xi}{\partial t}(s_0, t_0)(t - t_0)$$

です。($\xi(s_0, t_0) = x_0$ であることに注意してください。) まず証明することは、

$p(x)$ に $x = \pi(s, t)$ を入れて得られる 1 次式 $p \circ \pi(s, t)$ が $f \circ \xi(s, t)$ の (s_0, t_0) における 1 次近似になっている

ということ、つまり、

$$\lim_{(s, t) \rightarrow (s_0, t_0)} \frac{f \circ \xi(s, t) - p \circ \pi(s, t)}{\sqrt{(s - s_0)^2 + (t - t_0)^2}} = 0$$

が成り立つことです。この式の左辺に前小々節と同様の式変形をすると、

$$\begin{aligned}
 & \lim_{(s,t) \rightarrow (s_0,t_0)} \frac{f \circ \xi(s,t) - p \circ \pi(s,t)}{\sqrt{(s-s_0)^2 + (t-t_0)^2}} \\
 &= \lim_{(s,t) \rightarrow (s_0,t_0)} \frac{f(\xi(s,t)) - p(\xi(s,t)) + p(\xi(s,t)) - p(\pi(s,t))}{\sqrt{(s-s_0)^2 + (t-t_0)^2}} \\
 &= \lim_{(s,t) \rightarrow (s_0,t_0)} \left(\frac{f(\xi(s,t)) - p(\xi(s,t))}{\xi(s,t) - \xi(s_0,t_0)} \frac{\xi(s,t) - \xi(s_0,t_0)}{\sqrt{(s-s_0)^2 + (t-t_0)^2}} \right. \\
 & \quad \left. + \frac{p(\xi(s,t)) - p(\pi(s,t))}{\sqrt{(s-s_0)^2 + (t-t_0)^2}} \right)
 \end{aligned}$$

となります。

最後の式の第1項の第1因子は、 $\xi(s,t)$ が (s_0,t_0) で全微分可能だから連続でもあることと、 $p(x)$ が $f(x)$ の x_0 における1次近似であることから、

$$\lim_{(s,t) \rightarrow (s_0,t_0)} \frac{f(\xi(s,t)) - p(\xi(s,t))}{\xi(s,t) - \xi(s_0,t_0)} = \lim_{x \rightarrow x_0} \frac{f(x) - p(x)}{x - x_0} = 0$$

となります。一方、第2因子は1変数関数同士の場合と違って極限值はありません。どのようにになっているかというと、分子が

$$\xi(s,t) - \xi(s_0,t_0) = \xi(s,t) - \pi(s,t) + \frac{\partial \xi}{\partial s}(s_0,t_0) + \frac{\partial \xi}{\partial t}(s_0,t_0)$$

となっているわけです。

$$\lim_{(s,t) \rightarrow (s_0,t_0)} \frac{\xi(s,t) - \pi(s,t)}{\sqrt{(s-s_0)^2 + (t-t_0)^2}} = 0$$

はわかっているのですが、問題は残りの部分の極限です。記号が長くてわかりにくいので、

$$\frac{\partial \xi}{\partial s}(s_0,t_0) = a, \quad \frac{\partial \xi}{\partial t}(s_0,t_0) = b, \quad s - s_0 = h, \quad t - t_0 = k$$

とおくことにしましょう。調べたいのは、 $(h,k) \rightarrow (0,0)$ のときの

$$\frac{ah + bk}{\sqrt{h^2 + k^2}}$$

の振る舞いです。収束しないことは、例えば $k=0$ として $h \rightarrow +0$ とすると a になるのに、 $h \rightarrow -0$ とすると $-a$ になり、また $h=0$ として $k \rightarrow +0$ とすると b になり、 $k \rightarrow -0$ とすると $-b$ になる、つまり (h,k) の $(0,0)$ への値の近づけ方で極限值が変わってしまうことからわかります。しかし、 $|h| \leq \sqrt{h^2 + k^2}$ 、 $|k| \leq \sqrt{h^2 + k^2}$ であることから、 $(0,0)$ でない任意の (h,k) に対して、

$$\left| \frac{ah + bk}{\sqrt{h^2 + k^2}} \right| \leq |a| \frac{|h|}{\sqrt{h^2 + k^2}} + |b| \frac{|k|}{\sqrt{h^2 + k^2}} \leq |a| + |b|$$

が成り立っています。よって、問題の極限の第1項は、全体としては

$$\begin{aligned} & \lim_{(s,t) \rightarrow (s_0,t_0)} \left| \frac{f(\xi(s,t)) - p(\xi(s,t))}{\xi(s,t) - \xi(s_0,t_0)} \frac{\xi(s,t) - \xi(s_0,t_0)}{\sqrt{(s-s_0)^2 + (t-t_0)^2}} \right| \\ & \leq \lim_{(s,t) \rightarrow (s_0,t_0)} \left| \frac{f(\xi(s,t)) - p(\xi(s,t))}{\xi(s,t) - \xi(s_0,t_0)} (|a| + |b|) \right| = 0 \cdot (|a| + |b|) = 0 \end{aligned}$$

となって、はさみうちの原理により0に収束します。

一方、第2項の方は、

$$\begin{aligned} p(\xi(s,t)) - p(\pi(s,t)) &= f(x_0) + f'(x_0)(\xi(s,t) - x_0) - f(x_0) - f'(x_0)(\pi(s,t) - x_0) \\ &= f'(x_0)(\xi(s,t) - \pi(s,t)) \end{aligned}$$

と具体的に計算でき、 $\pi(s,t)$ が $\xi(s,t)$ の (s_0, t_0) における1次近似であることから、

$$\begin{aligned} \lim_{(s,t) \rightarrow (s_0,t_0)} \frac{p(\xi(s,t)) - p(\pi(s,t))}{\sqrt{(s-s_0)^2 + (t-t_0)^2}} &= f'(x_0) \lim_{(s,t) \rightarrow (s_0,t_0)} \frac{\xi(s,t) - \pi(s,t)}{\sqrt{(s-s_0)^2 + (t-t_0)^2}} \\ &= f'(x_0) \cdot 0 = 0 \end{aligned}$$

となります。

以上により、 $p \circ \pi(s,t)$ が $f \circ \xi(s,t)$ の (s_0, t_0) における1次近似であることが証明できました。

$p \circ \pi(s,t)$ を具体的に計算すると、

$$\begin{aligned} p \circ \pi(s,t) &= f(x_0) + \frac{df}{dx}(x_0)(\pi(s,t) - x_0) \\ &= f(x_0) + \frac{df}{dx}(x_0) \left(x_0 + \frac{\partial \xi}{\partial s}(s_0, t_0)(s - s_0) + \frac{\partial \xi}{\partial t}(s_0, t_0)(t - t_0) - x_0 \right) \\ &= f(x_0) + \frac{df}{dx}(x_0) \frac{\partial \xi}{\partial s}(s_0, t_0)(s - s_0) + \frac{df}{dx}(x_0) \frac{\partial \xi}{\partial t}(s_0, t_0)(t - t_0) \end{aligned}$$

となります。一方、 $f \circ \xi(s,t)$ の (s_0, t_0) における1次近似は

$$f(x_0) + \frac{\partial(f \circ \xi)}{\partial s}(s_0, t_0)(s - s_0) + \frac{\partial(f \circ \xi)}{\partial t}(s_0, t_0)(t - t_0)$$

でなければならないことがわかっています。ということは、この2つの1次式は同じものでなくてはならず、特に s の係数同士、 t の係数同士は一致します。その一致が、

$$\begin{aligned} \frac{\partial(f \circ \xi)}{\partial s}(s_0, t_0) &= \frac{df}{dx}(x_0) \frac{\partial \xi}{\partial s}(s_0, t_0) \\ \frac{\partial(f \circ \xi)}{\partial t}(s_0, t_0) &= \frac{df}{dx}(x_0) \frac{\partial \xi}{\partial t}(s_0, t_0) \end{aligned}$$

という合成関数の微分公式を与えてくれます。2変数の1次式の係数なので2つ一々に得られるというわけです。

2.2.3 2変数関数に1変数関数を2つ入れた場合

次に、前小節で一番苦労した場合を1次近似の方法で考えてみましょう。この場合は、合成関数が1変数なので、前小々節のような「2つの公式が一遍に得られる」という恩恵はあり得ないので、証明が見通しよく簡単になるかどうか、そこがポイントです。

$f(x, y)$ に $x = \xi(s)$ と $y = \eta(s)$ を入れた合成関数を、 $\circ$ を使った記法では書けないので $\varphi(s)$ と書くことにしました。

$$\varphi(s) = f(\xi(s), \eta(s))$$

です。 f は全微分可能、 ξ と η も微分可能、 $x_0 = \xi(s_0)$, $y_0 = \eta(s_0)$ として、 $f(x, y)$ の (x_0, y_0) での1次近似を $p(x, y)$ 、 $\xi(s)$ の s_0 での1次近似を $\pi(s)$ 、 $\eta(s)$ の s_0 での1次近似を $\rho(s)$ とします⁸。具体的には、

$$\begin{aligned} p(x, y) &= f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0) \\ \pi(s) &= x_0 + \frac{d\xi}{ds}(s_0)(s - s_0) \\ \rho(s) &= y_0 + \frac{d\eta}{ds}(s_0)(s - s_0) \end{aligned}$$

です。 $p(x, y)$ に $x = \pi(s)$, $y = \rho(s)$ を入れた合成関数 (s の1次式です) も $\circ$ を使った書き方では書けないので、

$$\psi(s) = p(\pi(s), \rho(s))$$

とおくことにします。示したいことは、例によって $\psi(s)$ が $\varphi(s)$ の s_0 における1次近似であること、つまり、

$$\lim_{s \rightarrow s_0} \frac{\varphi(s) - \psi(s)}{|s - s_0|} = 0$$

が成り立つことです。(あとの式変形を楽にするために分母に絶対値を付けておきました。極限值が0の場合には絶対値がついていてもついていなくても同じなので大丈夫です。むしろ絶対値つきの方が1次近似の本来の定義であるとも言えます。)

⁸ ρ はギリシャ語のアルファベットで π の次の文字で、「ロー」と読みます。ローマ字の 'r' に当たる文字です。

極限を取る前の式を変形して行くと、

$$\begin{aligned}
 \frac{\varphi(s) - \psi(s)}{|s - s_0|} &= \frac{f(\xi(s), \eta(s)) - p(\pi(s), \rho(s))}{|s - s_0|} \\
 &= \frac{f(\xi(s), \eta(s)) - p(\xi(s), \eta(s)) + p(\xi(s), \eta(s)) - p(\pi(s), \rho(s))}{|s - s_0|} \\
 &= \frac{f(\xi(s), \eta(s)) - p(\xi(s), \eta(s))}{\sqrt{(\xi(s) - x_0)^2 + (\eta(s) - y_0)^2}} \frac{\sqrt{(\xi(s) - x_0)^2 + (\eta(s) - y_0)^2}}{|s - s_0|} \\
 &\quad + \frac{p(\xi(s), \eta(s)) - p(\pi(s), \rho(s))}{|s - s_0|}
 \end{aligned}$$

となります。

最後の式の第1項の第2因子で $s \rightarrow s_0$ とすると、 $\xi(s)$ と $\eta(s)$ が微分可能であることから、

$$\begin{aligned}
 \lim_{s \rightarrow s_0} \frac{\sqrt{(\xi(s) - x_0)^2 + (\eta(s) - y_0)^2}}{|s - s_0|} &= \lim_{s \rightarrow s_0} \sqrt{\frac{(\xi(s) - x_0)^2 + (\eta(s) - y_0)^2}{(s - s_0)^2}} \\
 &= \sqrt{\left(\lim_{s \rightarrow s_0} \frac{\xi(s) - x_0}{s - s_0}\right)^2 + \left(\lim_{s \rightarrow s_0} \frac{\eta(s) - y_0}{s - s_0}\right)^2} \\
 &= \sqrt{\left(\frac{d\xi}{ds}(s_0)\right)^2 + \left(\frac{d\eta}{ds}(s_0)\right)^2}
 \end{aligned}$$

となります。

第1項の第1因子について考えてみましょう。 $\xi(s)$ と $\eta(s)$ が微分可能だから連続でもあるので、 $s \rightarrow s_0$ のとき $(\xi(s), \eta(s))$ は (x_0, y_0) に収束します。今、 $p(x, y)$ は $f(x, y)$ の (x_0, y_0) における1次近似なので、 (x, y) がどのように (x_0, y_0) に近づこうとも

$$\frac{f(x, y) - p(x, y)}{\sqrt{(x - x_0)^2 + (y - y_0)^2}} \rightarrow 0$$

が成り立っています。よって、第1項の第1因子は

$$\lim_{s \rightarrow s_0} \frac{f(\xi(s), \eta(s)) - p(\xi(s), \eta(s))}{\sqrt{(\xi(s) - x_0)^2 + (\eta(s) - y_0)^2}} = 0$$

となります。

以上より、 $s \rightarrow s_0$ のとき最後の式の第1項は0に収束することが分かりました。

次に、第2項について考えてみましょう。 $p(x, y)$ は1次式なので、分子を具体的に展開してみましょう。目がチカチカするのを防ぐために、

$$\frac{\partial f}{\partial x}(x_0, y_0) = a, \quad \frac{\partial f}{\partial y}(x_0, y_0) = b$$

とおくことにします。すると、

$$\begin{aligned} p(\xi(s), \eta(s)) - p(\pi(s), \rho(s)) &= f(x_0, y_0) + a(\xi(s) - x_0) + b(\eta(s) - y_0) \\ &\quad - f(x_0, y_0) + a(\pi(s) - x_0) + b(\rho(s) - y_0) \\ &= a(\xi(s) - \pi(s)) + b(\eta(s) - \rho(s)) \end{aligned}$$

となります。 $\pi(s)$ は $\xi(s)$ の、 $\rho(s)$ は $\eta(s)$ の s_0 における1次近似なので、結局、第2項で $s \rightarrow s_0$ とすると、

$$\begin{aligned} \lim_{s \rightarrow s_0} \frac{p(\xi(s), \eta(s)) - p(\pi(s), \rho(s))}{|s - s_0|} &= a \lim_{s \rightarrow s_0} \frac{\xi(s) - \pi(s)}{|s - s_0|} + b \lim_{s \rightarrow s_0} \frac{\eta(s) - \rho(s)}{|s - s_0|} \\ &= a \cdot 0 + b \cdot 0 = 0 \end{aligned}$$

となります。

以上で第1項も第2項も0に収束することが分かったので、

$$\lim_{s \rightarrow s_0} \frac{\varphi(s) - \psi(s)}{|s - s_0|} = 0$$

が成り立ち、 $\psi(s) = p(\pi(s), \rho(s))$ が $\varphi(s) = f(\xi(s), \eta(s))$ の s_0 における1次近似であることが分かりました。 $\psi(s)$ を具体的に計算すると、

$$\begin{aligned} \psi(s) &= p(\pi(s), \rho(s)) \\ &= f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0) \left(x_0 + \frac{d\xi}{ds}(s_0)(s - s_0) - x_0 \right) \\ &\quad + \frac{\partial f}{\partial y}(x_0, y_0) \left(y_0 + \frac{d\eta}{ds}(s_0)(s - s_0) - y_0 \right) \\ &= f(x_0, y_0) + \left(\frac{\partial f}{\partial x}(x_0, y_0) \frac{d\xi}{ds}(s_0) + \frac{\partial f}{\partial y}(x_0, y_0) \frac{d\eta}{ds}(s_0) \right) (s - s_0) \end{aligned}$$

となります。一方、 $\varphi(s)$ の s_0 における1次近似は

$$\varphi(s_0) + \frac{d\varphi}{ds}(s_0)(s - s_0)$$

であることが分かっています。($\varphi(s_0) = f(x_0, y_0)$ です。) よって、この二つの1次式は同じものでなければならず、特に s の係数は一致します。その一致が合成関数の微分公式

$$\frac{d\varphi}{ds}(s_0) = \frac{\partial f}{\partial x}(x_0, y_0) \frac{d\xi}{ds}(s_0) + \frac{\partial f}{\partial y}(x_0, y_0) \frac{d\eta}{ds}(s_0)$$

を与えてくれます。

どうでしたか？やっていることは

合成関数の1次近似は1次近似の合成である

ということであって、前小節の偏微分の定義だけに頼った証明に比べて見通しがよいと感じてもらいたいところなのですが、慣れない議論なので難しく感じたかも知れません。このプリントの残りの部分はすべてこの視点からの議論ですので、読み続けてもらえれば否応なく慣れるだろうと思いますから、心配せずに気楽に読み進んで下さい。

2.2.4 2変数関数に2変数関数を2つ入れた場合

前々小節「1変数関数に2変数関数を入れた場合」の議論と前小節「2変数関数に1変数関数を2つ入れた場合」の議論をくっつけると「2変数関数に2変数関数を2つ入れた場合」の証明になるのですが、全く同じことを繰り返すのも能がないので、ここではもとの関数と1次近似との差である剰余にスポットを当てて証明してみましょう。もちろんやっていることは今までと同じなのですが、剰余という考え方に慣れることは1次近似を2次近似、3次近似と次数を上げた場合、すなわちテイラー近似について考える場合にとっても有用です。

なお、偏微分の定義に従った証明と違って、ここでする証明はそのまま一般の場合に適用することができます。興味のある人は確認してみてください。

$f(x, y)$, $\xi(s, t)$, $\eta(s, t)$ をすべて全微分可能な関数とし、 $x = \xi(s, t)$, $y = \eta(s, t)$ として得られる合成関数を $\varphi(s, t)$ としましょう。

$$\varphi(s, t) = f(\xi(s, t), \eta(s, t))$$

です。

$x_0 = \xi(s_0, t_0)$, $y_0 = \eta(s_0, t_0)$ とし、 $f(x, y)$ の (x_0, y_0) における1次近似を $p(x, y)$ 、 $\xi(s, t)$ と $\eta(s, t)$ の (s_0, t_0) における1次近似をそれぞれ $\pi(s, t)$, $\rho(s, t)$ とします。また、 $p(x, y)$, $\pi(s, t)$, $\rho(s, t)$ を具体的な1次式として書いたときの見た目をスッキリさせるために、

$$\begin{aligned} \frac{\partial f}{\partial x}(x_0, y_0) &= a, & \frac{\partial f}{\partial y}(x_0, y_0) &= b, \\ \frac{\partial \xi}{\partial s}(s_0, t_0) &= \alpha_1, & \frac{\partial \xi}{\partial t}(s_0, t_0) &= \alpha_2, \\ \frac{\partial \eta}{\partial s}(s_0, t_0) &= \beta_1, & \frac{\partial \eta}{\partial t}(s_0, t_0) &= \beta_2 \end{aligned}$$

と置くことにします。これを使うと、

$$\begin{aligned} p(x, y) &= f(x_0, y_0) + a(x - x_0) + b(y - y_0) \\ \pi(s, t) &= x_0 + \alpha_1(s - s_0) + \alpha_2(t - t_0) \\ \rho(s, t) &= y_0 + \beta_1(s - s_0) + \beta_2(t - t_0) \end{aligned}$$

となります。さらに、それぞれの剰余を $r(x, y), \sigma(s, t), \tau(s, t)$ とします。つまり、

$$r(x, y) := f(x, y) - p(x, y)$$

$$\sigma(s, t) := \xi(s, t) - \pi(s, t)$$

$$\tau(s, t) := \eta(s, t) - \rho(s, t)$$

と定義するということです。 $p(x, y)$ が $f(x, y)$ の (x_0, y_0) における1次近似であり、 $\pi(s, t)$ が $\xi(s, t)$ の、 $\rho(s, t)$ が $\eta(s, t)$ の (s_0, t_0) における1次近似であるということは、剰余を使って

$$\begin{aligned} \lim_{(x,y) \rightarrow (x_0,y_0)} \frac{r(x,y)}{\sqrt{(x-x_0)^2 + (y-y_0)^2}} &= 0 \\ \lim_{(s,t) \rightarrow (s_0,t_0)} \frac{\sigma(s,t)}{\sqrt{(s-s_0)^2 + (t-t_0)^2}} &= 0 \\ \lim_{(s,t) \rightarrow (s_0,t_0)} \frac{\tau(s,t)}{\sqrt{(s-s_0)^2 + (t-t_0)^2}} &= 0 \end{aligned}$$

が成り立つことと書き表すことができます。

これらの記号を使うと、

$$\begin{aligned} \varphi(s, t) &= f(\xi(s, t), \eta(s, t)) \\ &= p(\xi(s, t), \eta(s, t)) + r(\xi(s, t), \eta(s, t)) \\ &= f(x_0, y_0) + a(\xi(s, t) - x_0) + b(\eta(s, t) - y_0) + r(\xi(s, t), \eta(s, t)) \\ &= f(x_0, y_0) + a(\pi(s, t) + \sigma(s, t) - x_0) + b(\rho(s, t) + \tau(s, t) - y_0) \\ &\quad + r(\xi(s, t), \eta(s, t)) \\ &= f(x_0, y_0) + a(\alpha_1(s - s_0) + \alpha_2(t - t_0) + \sigma(s, t)) \\ &\quad + b(\beta_1(s - s_0) + \beta_2(t - t_0) + \tau(s, t)) + r(\xi(s, t), \eta(s, t)) \\ &= f(x_0, y_0) + (a\alpha_1 + b\beta_1)(s - s_0) + (a\alpha_2 + b\beta_2)(t - t_0) \\ &\quad + a\sigma(s, t) + b\tau(s, t) + r(\xi(s, t), \eta(s, t)) \end{aligned}$$

となります。よって、1次近似式の合成

$$\begin{aligned} \psi(s, t) &:= p(\pi(s, t), \rho(s, t)) \\ &= f(x_0, y_0) + (a\alpha_1 + b\beta_1)(s - s_0) + (a\alpha_2 + b\beta_2)(t - t_0) \end{aligned}$$

を $\varphi(s, t)$ から引いて $\sqrt{(s - s_0)^2 + (t - t_0)^2}$ で割ると

$$\begin{aligned} & \frac{\varphi(s, t) - \psi(s, t)}{\sqrt{(s - s_0)^2 + (t - t_0)^2}} \\ &= \frac{a\sigma(s, t) + b\tau(s, t) + r(\xi(s, t), \eta(s, t))}{\sqrt{(s - s_0)^2 + (t - t_0)^2}} \\ &= a \frac{\sigma(s, t)}{\sqrt{(s - s_0)^2 + (t - t_0)^2}} + b \frac{\tau(s, t)}{\sqrt{(s - s_0)^2 + (t - t_0)^2}} \\ & \quad + \frac{r(\xi(s, t), \eta(s, t))}{\sqrt{(\xi(s, t) - x_0)^2 + (\eta(s, t) - y_0)^2}} \frac{\sqrt{(\xi(s, t) - x_0)^2 + (\eta(s, t) - y_0)^2}}{\sqrt{(s - s_0)^2 + (t - t_0)^2}} \end{aligned}$$

となります。ここで、 $\sigma(s, t)$ と $\tau(s, t)$ は1次近似の剰余なので、既に行ったとおり

$$\lim_{(s, t) \rightarrow (s_0, t_0)} \frac{\sigma(s, t)}{\sqrt{(s - s_0)^2 + (t - t_0)^2}} = \lim_{(s, t) \rightarrow (s_0, t_0)} \frac{\tau(s, t)}{\sqrt{(s - s_0)^2 + (t - t_0)^2}} = 0$$

が成り立っています。また、 ξ も η も全微分可能であることから連続でもあるので、

$$\lim_{(s, t) \rightarrow (s_0, t_0)} \xi(s, t) = x_0, \quad \lim_{(s, t) \rightarrow (s_0, t_0)} \eta(s, t) = y_0$$

です。よって、前小々節と同様に、

$$\begin{aligned} & \lim_{(s, t) \rightarrow (s_0, t_0)} \frac{r(\xi(s, t), \eta(s, t))}{\sqrt{(\xi(s, t) - x_0)^2 + (\eta(s, t) - y_0)^2}} \\ &= \lim_{(x, y) \rightarrow (x_0, y_0)} \frac{r(x, y)}{\sqrt{(x - x_0)^2 + (y - y_0)^2}} = 0 \end{aligned}$$

が成り立ちます。問題は

$$\frac{\sqrt{(\xi(s, t) - x_0)^2 + (\eta(s, t) - y_0)^2}}{\sqrt{(s - s_0)^2 + (t - t_0)^2}}$$

で $(s, t) \rightarrow (s_0, t_0)$ としたときどうなるかです。これは前々小々節でのときと同様に「収束はしないが有界」になります。キチンと証明しましょう。

証明. まず、任意の2実数 u, v に対して

$$\max\{|u|, |v|\} \leq \sqrt{u^2 + v^2} \leq |u| + |v|$$

が成り立つことに注意しましょう。これと三角不等式を使うと、

$$\begin{aligned}
 & \frac{\sqrt{(\xi(s, t) - x_0)^2 + (\eta(s, t) - y_0)^2}}{\sqrt{(s - s_0)^2 + (t - t_0)^2}} \\
 & \leq \frac{|\xi(s, t) - x_0| + |\eta(s, t) - y_0|}{\sqrt{(s - s_0)^2 + (t - t_0)^2}} \\
 & \leq \frac{|a_1||s - s_0| + |a_2||t - t_0| + |\sigma(s, t)| + |b_1||s - s_0| + |b_2||t - t_0| + |\tau(s, t)|}{\sqrt{(s - s_0)^2 + (t - t_0)^2}} \\
 & \leq (|a_1| + |b_1|) \frac{|s - s_0|}{|s - s_0|} + (|a_2| + |b_2|) \frac{|t - t_0|}{|t - t_0|} \\
 & \quad + \frac{|\sigma(s, t)|}{\sqrt{(s - s_0)^2 + (t - t_0)^2}} + \frac{|\tau(s, t)|}{\sqrt{(s - s_0)^2 + (t - t_0)^2}}
 \end{aligned}$$

となります。最後の二つの項は σ と τ が 1 次近似の剰余項であることから $(s, t) \rightarrow (s_0, t_0)$ のとき 0 に収束します。よって、全体としては $(s, t) \rightarrow (s_0, t_0)$ のとき収束はしないとしても絶対値が $|a_1| + |a_2| + |b_1| + |b_2|$ を超えないので有界です。□

以上より、

$$\lim_{(s, t) \rightarrow (s_0, t_0)} \frac{\varphi(s, t) - \psi(s, t)}{\sqrt{(s - s_0)^2 + (t - t_0)^2}} = 0$$

であることが分かりました。これは 1 次式 $\psi(s, t)$ が $\varphi(s, t)$ の (s_0, t_0) における 1 次近似であることを示しています。よって、

$$\psi(s, t) = f(x_0, y_0) + \frac{\partial \varphi}{\partial s}(s_0, t_0)(s - s_0) + \frac{\partial \varphi}{\partial t}(s_0, t_0)(t - t_0)$$

でなければならないことになります。一方、

$$\begin{aligned}
 \psi(s, t) &= p(\pi(s, t), \rho(s, t)) \\
 &= (a\alpha_1 + b\beta_1)(s - s_0) + (a\alpha_2 + b\beta_2)(t - t_0)
 \end{aligned}$$

でした。ということはこの 2 つの 1 次式の対応する係数は一致することになります。 $a, b, \alpha_1, \alpha_2, \beta_1, \beta_2$ を偏微分係数に戻すと、

$$\begin{aligned}
 \frac{\partial \varphi}{\partial s}(s_0, t_0) &= \frac{\partial f}{\partial x}(x_0, y_0) \frac{\partial \xi}{\partial s}(s_0, t_0) + \frac{\partial f}{\partial y}(x_0, y_0) \frac{\partial \eta}{\partial s}(s_0, t_0) \\
 \frac{\partial \varphi}{\partial t}(s_0, t_0) &= \frac{\partial f}{\partial x}(x_0, y_0) \frac{\partial \xi}{\partial t}(s_0, t_0) + \frac{\partial f}{\partial y}(x_0, y_0) \frac{\partial \eta}{\partial t}(s_0, t_0)
 \end{aligned}$$

という合成関数の微分公式になります。

3 逆写像の微分と変数変換

1 変数関数のときには存在する場合があったのに、多変数関数では存在し得ない重要な概念に逆関数というものがありました。しかし、写像まで対象を広げておけば逆写像として復活します。逆写像を持つ写像が自然にでてくるのは直交座標と極座標の兼ね合いを考える場合などの変数変換においてです。物理における変数変換の重要性は言うまでもないでしょう。この節では、逆写像の微分公式を説明した後、それを使って変数変換に対する微分の振る舞いについて考えます。

なお、ここで改めて写像の微分について説明し直す予定だったのですが、第1.3小節で予定していたより詳しく説明してしまったことと、例によって時間の都合により、ここでの説明は割愛します。第1.3小節にもう一度目を通してから以下の小節に進むとよいと思います。

3.1 逆写像の微分公式

n 変数関数が m 個あったとき、その決める写像が逆写像を持つ、つまり、

$$y_1 = f_1(x_1, \dots, x_n), \dots, y_m = f_m(x_1, \dots, x_n) \quad (5)$$

が逆に解けて

$$x_1 = g_1(y_1, \dots, y_m), \dots, x_n = g_n(y_1, \dots, y_m) \quad (6)$$

となったとしましょう。つまり、すべての $i = 1, 2, \dots, m$ について

$$f_i(g_1(y_1, \dots, y_m), \dots, g_n(y_1, \dots, y_m)) = y_i$$

が成り立ち、かつ、すべての $j = 1, 2, \dots, n$ について

$$g_j(f_1(x_1, \dots, x_n), \dots, f_m(x_1, \dots, x_n)) = x_j$$

が成り立っているとします。

m 個の関数(5)たちの決める $\mathbb{R}^n$ から $\mathbb{R}^m$ への写像を

$$F: \mathbb{R}^n \ni \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \rightarrow \begin{bmatrix} f_1(x_1, \dots, x_n) \\ \vdots \\ f_m(x_1, \dots, x_n) \end{bmatrix} \in \mathbb{R}^m$$

とし、 n 個の関数(6)たちの決める $\mathbb{R}^m$ から $\mathbb{R}^n$ への写像を

$$G: \mathbb{R}^m \ni \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \rightarrow \begin{bmatrix} g_1(y_1, \dots, y_m) \\ \vdots \\ g_n(y_1, \dots, y_m) \end{bmatrix} \in \mathbb{R}^n$$

とすると、「逆に解ける」とは

$$F \circ G = \text{「}\mathbb{R}^m \text{の恒等写像」}, \quad G \circ F = \text{「}\mathbb{R}^n \text{の恒等写像」}$$

の両方が成り立つことと同じです。この2式の両辺を微分する、つまりヤコビ行列にすると、合成関数の微分公式によって任意の点で

$$J_F \cdot J_G = E_m, \quad J_G \cdot J_F = E_n$$

が成り立ちます。ただし E_m および E_n はそれぞれ m 次および n 次の単位行列です。（ $\mathbb{R}^n$ の恒等写像とは

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \mapsto \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

と対応させること、つまり第 i 成分が x_i という関数でできている写像のことなので、ヤコビ行列の (i, j) 成分は

$$\frac{\partial x_i}{\partial x_j} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

すなわち単位行列 E_n になります。) つまり、任意の点で J_f と J_g は互いに逆行列だということです。よって、特に J_f は正方行列であり、 $n = m$ となります⁹。

写像の記法では、このような G のことを「 F の逆写像」と呼び F^{-1} と書きますから、それに従って「逆写像の微分公式」は

$$J_{F^{-1}} = J_F^{-1} \quad (7)$$

となります。もちろん右辺は「 J_F という行列の逆行列」という意味です。1変数関数の逆関数の微分法と全く同じ形になるわけです。

2変数のときを成分で書いてみましょう。添え字で書くのはやめて、

$$F = \begin{bmatrix} f(x, y) \\ g(x, y) \end{bmatrix}, \quad F^{-1} = \begin{bmatrix} \varphi(s, t) \\ \psi(s, t) \end{bmatrix}$$

としましょう。すると、

$$J_F = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{bmatrix}, \quad J_{F^{-1}} = \begin{bmatrix} \frac{\partial \varphi}{\partial s} & \frac{\partial \varphi}{\partial t} \\ \frac{\partial \psi}{\partial s} & \frac{\partial \psi}{\partial t} \end{bmatrix}$$

⁹ 「 $AB = E_n$ かつ $BA = E_m$ となるならば $n = m$ 、 A は正則行列で $B = A^{-1}$ 」という事実については数学II（線型代数）の階数のところで学びます。ここでは認めさせてください。

なので、逆写像の微分法の公式は

$$\begin{bmatrix} \frac{\partial \varphi}{\partial s} & \frac{\partial \varphi}{\partial t} \\ \frac{\partial \psi}{\partial s} & \frac{\partial \psi}{\partial t} \end{bmatrix} = \frac{1}{\frac{\partial f}{\partial x} \frac{\partial g}{\partial y} - \frac{\partial f}{\partial y} \frac{\partial g}{\partial x}} \begin{bmatrix} \frac{\partial g}{\partial y} & -\frac{\partial f}{\partial y} \\ -\frac{\partial g}{\partial x} & \frac{\partial f}{\partial x} \end{bmatrix}$$

となります。成分で書けば、例えば

$$\frac{\partial \varphi}{\partial s} = \frac{\frac{\partial g}{\partial y}}{\frac{\partial f}{\partial x} \frac{\partial g}{\partial y} - \frac{\partial f}{\partial y} \frac{\partial g}{\partial x}}$$

などとなって、わけわかりませんね。行列で表示することの便利さが実感できると思います。

3.2 変数変換

逆写像の微分の公式が手に入ったので、「微分の変数変換」が簡単に計算できます。

まず、1変数関数の場合で考えてみましょう。 $f(x)$ を $x = \xi(s)$ により s の関数に変数変換してみましょう。どちらの変数で考えているかハッキリさせるために、 $\xi(s)$ を入れた合成関数を $g(s)$ と書くことにします。つまり、 $g(s) = f(\xi(s))$ です。合成関数の微分法により

$$\frac{dg}{ds}(s) = \frac{df}{dx}(\xi(s)) \frac{d\xi}{ds}(s)$$

となりますが、この式の登場人物のうち dg/ds と $d\xi/ds$ は s の関数で、 df/dx は (今は x に $\xi(s)$ が入っているから s の関数ですが、本来は) x の関数です。だから、この公式は s の関数を x の関数と s の関数で表しているわけで、少々据わりが悪い感じがします。

ここで、もし $\xi(s)$ が逆関数 $s = \sigma(x)$ を持つなら、逆関数の微分法から

$$\frac{d\xi}{ds}(s) = \frac{1}{\frac{d\sigma}{dx}(\xi(s))}$$

ですので、

$$\frac{dg}{ds}(s) = \frac{\frac{df}{dx}(x)}{\frac{d\sigma}{dx}(x)}$$

と、 s での微分をを（本来） x を変数とする関数として表すことができます。（等号の意味は、「右辺の x に $\xi(s)$ を代入するか、左辺の s に $\sigma(s)$ を代入すれば両辺が同じ関数になる」という意味です。）もちろん、同様に、

$$\frac{df}{dx}(x) = \frac{\frac{dg}{ds}(s)}{\frac{d\xi}{ds}(s)}$$

と、 x での微分を（本来） s を変数とする関数として表すこともできます。

ヤコビ行列を使えば、多変数関数の変数変換に対しても全く同じ計算が可能になります。 f を $x_1, \dots, x_n$ を変数とする関数とし、 $\xi_1(s_1, \dots, s_n), \dots, \xi_n(s_1, \dots, s_n)$ という「逆に解ける」関数の組によって f の変数を $x_1, \dots, x_n$ から $s_1, \dots, s_n$ に変換してみましょう。上と同様、どちらの変数で考えているのかわからなくならないために、 $s_1, \dots, s_n$ に変数を変換したあとの関数を $g(s_1, \dots, s_n)$ と書くことにします。変数変換の写像を一文字で $\Phi: \mathbb{R}^n \rightarrow \mathbb{R}^n$ と書くことにする、詳しくは、

$$\Phi(s_1, \dots, s_n) = \begin{bmatrix} \xi_1(s_1, \dots, s_n) \\ \vdots \\ \xi_n(s_1, \dots, s_n) \end{bmatrix}$$

とすると、

$$g = f \circ \Phi$$

なので、合成写像の微分法から

$$J_g(s_1, \dots, s_n) = J_f(x_1, \dots, x_n) J_\Phi(s_1, \dots, s_n)$$

が得られる一方、 Φ の逆写像 Φ^{-1} があって、

$$f = g \circ \Phi^{-1}$$

が成り立っているので、

$$J_f(x_1, \dots, x_n) = J_g(s_1, \dots, s_n) J_{\Phi^{-1}}(x_1, \dots, x_n)$$

という行列の式が得られます。ただし、 J_f と J_g は1行 n 列の行列、つまり df と dg を表す行列です。1変数のときと同様、 J_f と $J_{\Phi^{-1}}$ の本来の変数は $x_1, \dots, x_n$ で、 J_g と J_Φ の本来の変数は $s_1, \dots, s_n$ です。しかし、逆写像の微分公式により、 J_Φ と $J_{\Phi^{-1}}$ は互いに相手の逆行列ですので、上の二つの式は

$$J_g(s_1, \dots, s_n) = J_f(x_1, \dots, x_n) (J_{\Phi^{-1}}(x_1, \dots, x_n))^{-1}$$

及び

$$J_f(x_1, \dots, x_n) = J_g(s_1, \dots, s_n) (J_\Phi(s_1, \dots, s_n))^{-1} \quad (8)$$

となります。

例 2. 極座標 $x = r \cos \theta, y = r \sin \theta$ を考えましょう。これは全単射ではないので無制限では変数変換になっていませんが、例えば $r > 0, 0 \leq \theta < 2\pi$ と制限すれば全単射です。この範囲において、式(8)を具体的に書けば、ヤコビ行列で

$$\begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{bmatrix} = \begin{bmatrix} \frac{\partial g}{\partial r} & \frac{\partial g}{\partial \theta} \end{bmatrix} \begin{bmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{bmatrix}^{-1}$$

ですので、成分では

$$\begin{aligned} \frac{\partial f}{\partial x} &= \frac{\partial g}{\partial r} \cos \theta - \frac{1}{r} \frac{\partial g}{\partial \theta} \sin \theta \\ \frac{\partial f}{\partial y} &= \frac{\partial g}{\partial r} \sin \theta + \frac{1}{r} \frac{\partial g}{\partial \theta} \cos \theta \end{aligned}$$

となります。念のためにこの式の意味を言いますと、左辺の f_x や f_y は (x, y) の関数ですが、そこに $x = r \cos \theta, y = r \sin \theta$ を入れてやると右辺の関数になる、というわけです。

この二つの式に対してもう一度同じことをしてやると、例えば、

$$\begin{aligned} f_{xx} &= \frac{\partial f_x}{\partial x} \\ &= \frac{\partial}{\partial r} \left(g_r \cos \theta - \frac{1}{r} g_\theta \sin \theta \right) \cos \theta - \frac{1}{r} \frac{\partial}{\partial \theta} \left(g_r \cos \theta - \frac{1}{r} g_\theta \sin \theta \right) \sin \theta \\ &= \left(g_{rr} \cos \theta + \frac{1}{r^2} g_\theta \sin \theta - \frac{1}{r} g_{r\theta} \sin \theta \right) \cos \theta \\ &\quad - \frac{1}{r} \left(g_{\theta r} \cos \theta - g_r \sin \theta - \frac{1}{r} g_{\theta\theta} \sin \theta - \frac{1}{r} g_\theta \cos \theta \right) \sin \theta \\ &= g_{rr} \cos^2 \theta - \frac{1}{r} g_{r\theta} 2 \sin \theta \cos \theta + \frac{1}{r^2} g_{\theta\theta} \sin^2 \theta + \frac{1}{r} g_r \sin^2 \theta + \frac{1}{r^2} g_\theta 2 \sin \theta \cos \theta \end{aligned}$$

$$\begin{aligned} f_{yy} &= \frac{\partial f_y}{\partial y} \\ &= \frac{\partial}{\partial r} \left(g_r \sin \theta + \frac{1}{r} g_\theta \cos \theta \right) \sin \theta + \frac{1}{r} \frac{\partial}{\partial \theta} \left(g_r \sin \theta + \frac{1}{r} g_\theta \cos \theta \right) \cos \theta \\ &= \left(g_{rr} \sin \theta - \frac{1}{r^2} g_{\theta\theta} \cos \theta + \frac{1}{r} g_{r\theta} \cos \theta \right) \sin \theta \\ &\quad + \frac{1}{r} \left(g_{\theta r} \sin \theta + g_r \cos \theta + \frac{1}{r} g_{\theta\theta} \cos \theta - \frac{1}{r} g_\theta \sin \theta \right) \cos \theta \\ &= g_{rr} \sin^2 \theta + \frac{1}{r} g_{r\theta} 2 \sin \theta \cos \theta + \frac{1}{r^2} g_{\theta\theta} \cos^2 \theta + \frac{1}{r} g_r \cos^2 \theta - \frac{1}{r^2} g_\theta 2 \sin \theta \cos \theta \end{aligned}$$

となりますので、この二つを足しあわせて、

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = \frac{\partial^2 g}{\partial r^2} + \frac{1}{r^2} \frac{\partial^2 g}{\partial \theta^2} + \frac{1}{r} \frac{\partial g}{\partial r}$$

となります。 $f(x, y)$ は (微分さえできれば) 何でも良かったので、

$$\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} = \frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r} + \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2}$$

と書けます。これは、ラプラス作用素 $\partial^2/\partial x^2 + \partial^2/\partial y^2$ を極座標で表す重要な公式です。 ■

4 ヤコビ行列の意味：多変数写像の視覚化

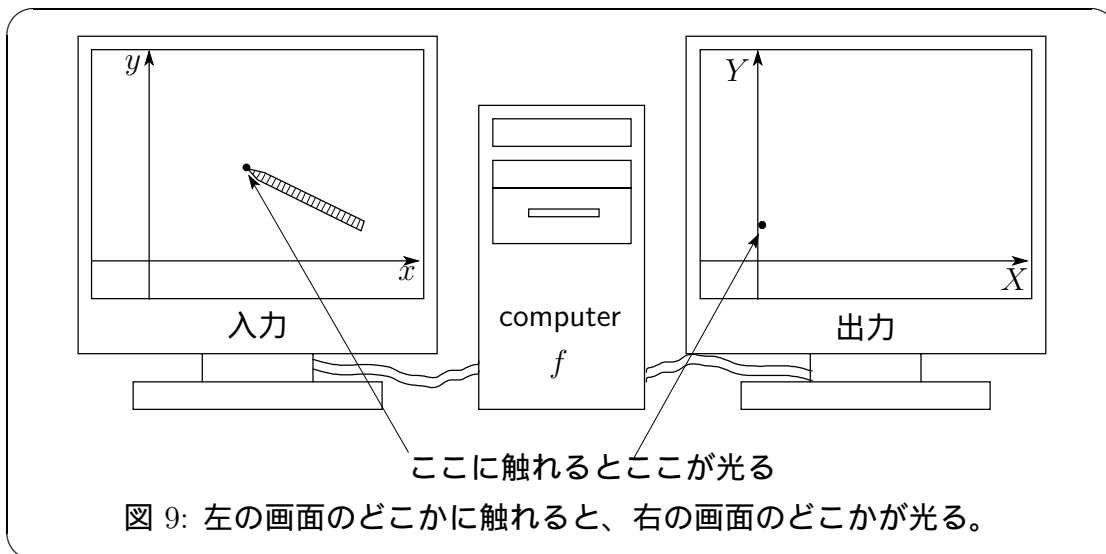
ヤコビ行列というものを持ち出して計算を整理したら、見事に1変数関数の場合と同じ合成関数の微分公式が浮かび上がってきました。何となく「出来過ぎ」な感じがしませんか？ この節では、多変数写像の視覚化を試みることで、このことに「自然な」説明が付けられないか考えてみましょう。

なお、この節までとこの節からで、元になった原稿 (つまり、何年か前に演習のために私が打ったファイル) が違うので、変数などの文字設定がかなり違ってしまっています。統一したかったのですが、とても時間が掛かり結局できませんでしたので、少し注意して読んでください。

多変数の写像 $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ を「視覚化」と言ってもグラフを思い浮かべるのは無理なので、定義域の $\mathbb{R}^n$ と値域の $\mathbb{R}^m$ をそれぞれ別々に視覚化することで満足することにします。実際にイメージを作ろうと思うと、 n, m は3が限度でしょう。その上、図を書こうと思うともはや $n = m = 2$ しか残りません。仕方がないので、以下 $n = m = 2$ で説明します。自分の頭の中で一般の n, m の場合をイメージしようと努力しながら読んでみてください。

今、あなたの前に2つのディスプレイのつながったコンピュータがあるものと想像してください。そして、向かって左側のディスプレイには「入力」、向かって右側のディスプレイには「出力」というラベルが貼ってあります。鉛筆のような形の棒が用意されていて、それで「入力」ディスプレイ上の勝手な1点に触れると「出力」ディスプレイの1点が光ります。「入力」ディスプレイの別な点に触れるとまた「出力」ディスプレイの1点が光りました。(この点はさっきの点とは違うかもしれないし同じかもしれません。) なかなかうまく描けなかったのですが、理解の助けのための図9を見て想像をふくらませてください。これが $\mathbb{R}^2$ から $\mathbb{R}^2$ への写像の「視覚化」(のつもり) です。写像を $f: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ と書くことにすると、「入力」ディスプレイで触れた位置のデータが「computer」に送られ、「computer」内で $f(x, y)$ に当たる計算を施し、その表す位置が「出力」ディスプレイで光るというわけです。

$f(x, y)$ が連続なら、「入力」ディスプレイ上をペンでぐるぐると撫でまわしたとき、「出力」ディスプレイ上の光の点もぐるぐると動き、決して突然消えて別のところに飛び移るようなことはありません。さらに $f(x, y)$ が滑らか (全微分可能とか C^1 -級とか) なときには、「入力」ディスプレイ上の「撫でまわし」が滑らか



なら「出力」ディスプレイ上の光の点の動きも滑らかになることは納得できるでしょう。

この「視覚化」で、ヤコビ行列が何を表す行列かを考えてみましょう。

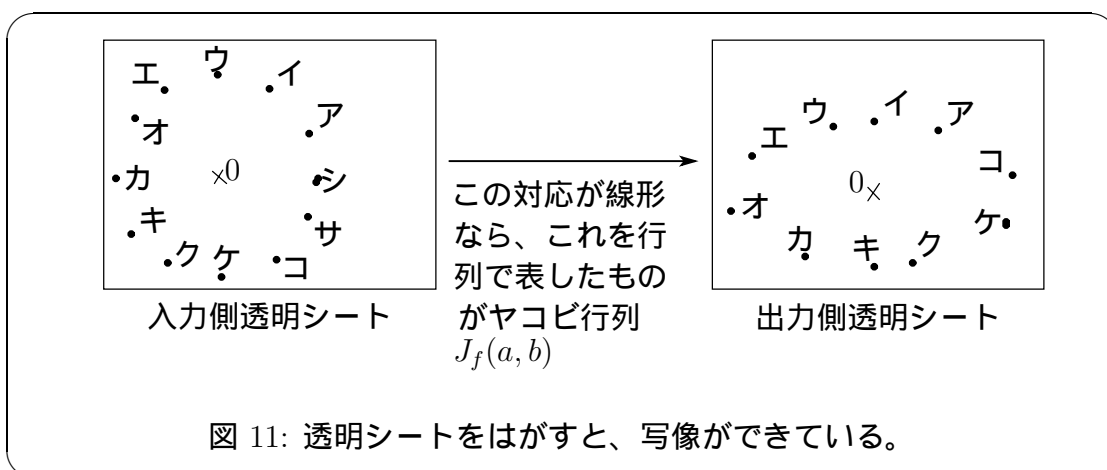
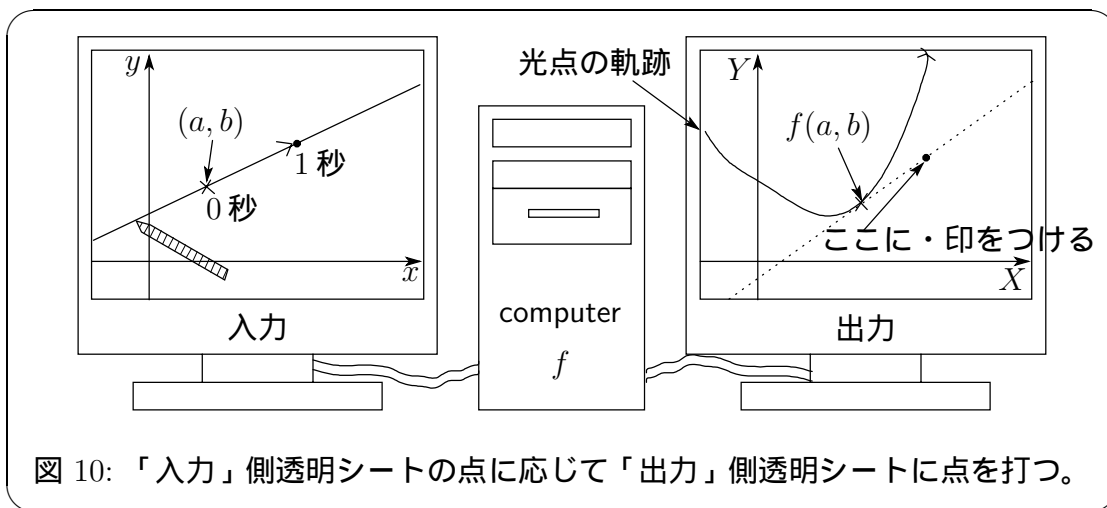
まず、「入力」「出力」の両ディスプレイの表面に OHP シートのような透明なフィルムを貼り付けます。そして、「入力」側の勝手な 1 点 (a, b) を選んで、はっきりわかるようにマジックで×印を書きます。その×印を付けた点にさわると「出力」側のどこか 1 点 $f(a, b)$ が光りますので、そこにもマジックで×印を書いてしまいます。

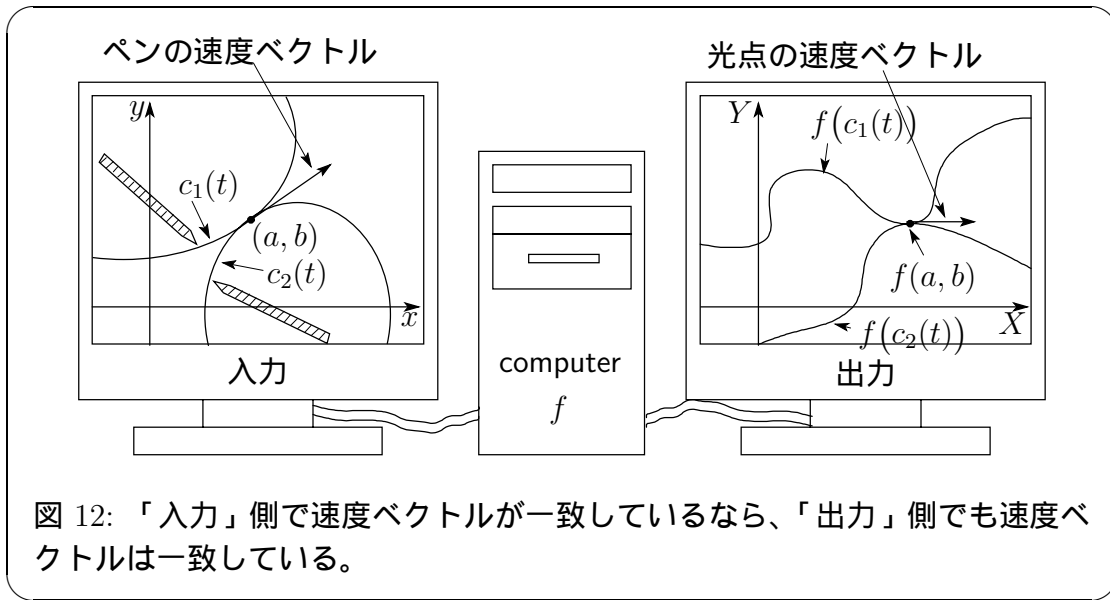
次に、「入力」側で×以外の点の一つを選びそこにも（×以外の）印（例えば・）をつけます。そして、×と・を通るように物差しを当て、ペンを「入力」ディスプレイ上に線を描くように動かします。このとき、ペンを動かす早さは一定で、×から・までちょうど 1 秒かかるように動かしましょう。すると、「出力」ディスプレイ上で光の点の動きが滑らかになります。その光点が×を通ったときの秒速を表す点、つまり、光点の描く曲線の×における接線上の点であって、×を通過したときの速度を保ったまま 1 秒たったときに到達するはずの点に・印を付けます（図 10）。

この作業を「入力」ディスプレイ上の×以外のすべての点に対して行い、両画面の透明シートをはがすと、「入力」透明シートから「出力」透明シートへの写像ができあがっています（図 11）。 f が (a, b) で全微分可能なとき、この新たにできた写像が（両方のシートの×を原点と思って）線型写像という写像になっています¹⁰。になっています。そして、この線型写像を表す行列が $f(x, y)$ の (a, b) におけるヤコビ行列 $J_f(a, b)$ になっています。

この「視覚化」で合成関数の微分が問題の公式に従うことの理由を考えてみましょう。

¹⁰線型写像とは一次変換の別名です。一次変換って知ってるんでしたっけ？





点 (a, b) における $f(x, y)$ のヤコビ行列、すなわち「入力」透明シートから「出力」透明シートへの線型写像を作るとき、「入力」側では点の「等速直線運動」のみを考えました。しかし、「入力」側でも「出力」側と同様「速度ベクトル」を考えて対応を作っても同じ対応ができあがります。つまり、点 (a, b) を通る二つの曲線 $c_1(t)$ と $c_2(t)$ が点 (a, b) を同じ速度で通過するなら、 $f(x, y)$ によって写された曲線 $f(c_1(t))$ と $f(c_2(t))$ の $f(a, b)$ を通過するときの速度ベクトルも一致します(図 12)。これはいかにも当たり前に感じられるでしょう。

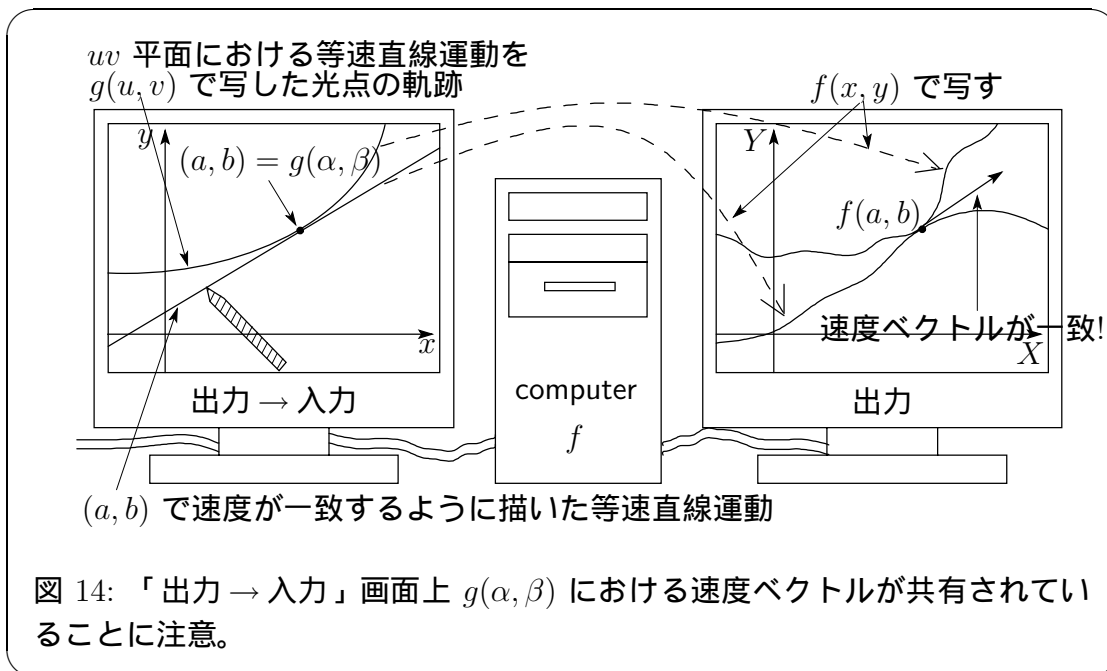
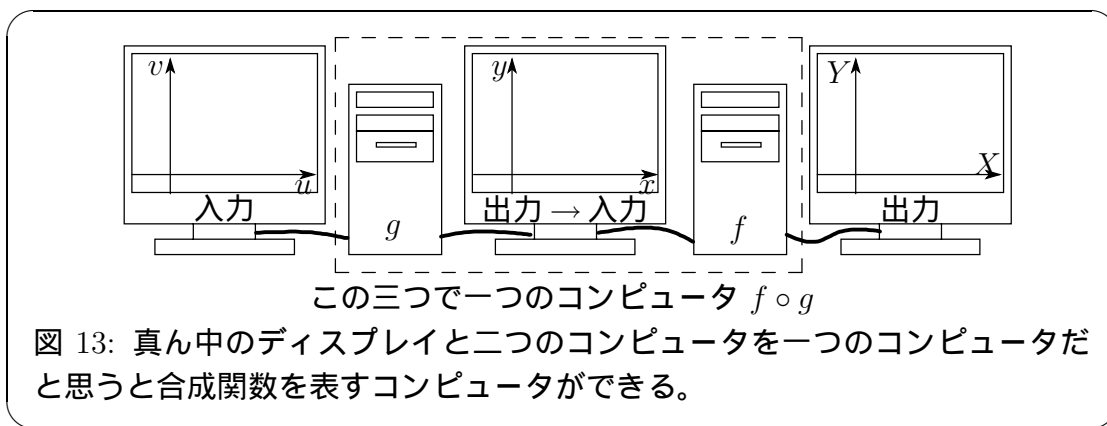
さて、このことを踏まえた上で $f(x, y)$ に (u, v) 平面から (x, y) 平面への写像 $(x, y) = g(u, v)$ を合成してみましょう。 g という名前のコンピュータと f という名前のコンピュータがあり、 g の「出力」用のディスプレイと f の「入力」用のディスプレイが同じで、そのディスプレイは g の出力をそのまま f への入力として送り込んでいるというわけです(図 13)。コンピュータ g と f および真ん中のディスプレイを一つの大きな箱に入れて「コンピュータ $f \circ g$ 」と名前を付ければ合成関数 $f \circ g(u, v)$ を表すコンピュータのできあがりです。

このデカコンピュータ $f \circ g$ に対して前節と同様に透明シート間の線型写像を作ったものが $f \circ g(u, v)$ の (α, β) におけるヤコビ行列 $J_{f \circ g}(\alpha, \beta)$ です。合成関数の微分公式とは、

$$J_{f \circ g}(\alpha, \beta) = J_f(a, b)J_g(\alpha, \beta),$$

つまり、このようにして作った透明シート間の線型写像が、コンピュータ g に対して作った透明シート間の線型写像とコンピュータ f に対して作った透明シート間の線型写像の合成になっているというものでした。

このことは、今までの考察でほとんど明らかでしょう。なぜなら、 $J_{f \circ g}(\alpha, \beta)$ という写像は、「入力」ディスプレイ上 (α, β) を通る等速直線運動を $g(u, v)$ で「出



「出力 → 入力」ディスプレイに写したものをそのまま $f(x, y)$ に入力して「出力」ディスプレイに描かれた光点の動きの $f(g(\alpha, \beta))$ における速度ベクトルに写す写像であり、一方、 $J_f(a, b)J_g(\alpha, \beta)$ という写像は、「入力」ディスプレイ上 (α, β) を通る等速直線運動を $g(u, v)$ で「出力 → 入力」ディスプレイに写した光点の動きの $g(\alpha, \beta)$ における速度ベクトルを速度とする「出力 → 入力」ディスプレイ上の等速直線運動を $f(x, y)$ で「出力」ディスプレイに写した光点の動きの $f(g(\alpha, \beta))$ における速度ベクトルに写す写像だからです。なんだかゴチャゴチャしてわかりにくいように感じるかもしれませんが、図で考えれば一発で理解できます。図 14 を見てください。

5 座標を使わずに微分を理解する

5.1 座標を使わないとはどういうことか

前小々節の説明では、ディスプレイ上に xy 座標や XY 座標が書いてありました。それは、「写像 f は x と y という数を使って X と Y という数を計算する手続き」という気持ちがあるからです。つまり、関数とか写像とかいうものは「式で与えられるもの」という気分があるからです。しかし、理屈の上では平面上の点そのものに別な平面上の点を対応させさえすれば、具体的な計算式があろうとなかろうと「平面から平面への写像」です。つまり、「式 f を具現化するコンピュータ」を考えるのではなく、とにかく「入力ディスプレイの1点をさわると出力ディスプレイの1点が光る」という現象そのものを「写像 f 」と思ってよいわけです。

そう思うと、座標 x, y などはあとからくっつけた恣意的なもので、 f の存在とは無関係なものとなります。つまり f は点そのものに点そのものを対応させているのです。

「点そのもの」などと言っても結局その点を指定するには座標がいるじゃないか、と思われるかも知れません。中学あるいは高校以来、平面や空間の考察には座標を使うのがあまりにも当たり前で、「点そのもの」と言われてもピンとこないでしょう。思い出してほしいのは、小中学校のときの図形の考え方です。たとえば、「三角形の内角の和は 180° である。」とか「二等辺三角形の二つの底角は等しい。」とかというのを証明せよと言われたら、今でも座標など使わずに三角形の3頂点に A, B, C などの「名前」を直接つけて考えるでしょう。そのように「平面や空間の点に直接名前を付ける」考え方が「座標を使わない」考え方です。つまり、私を「清野和彦」と指定するのが座標を使わない名前の付け方、東大の教官で横浜市に住んでいて身長が 186 cm で、という指定の仕方が座標による名前の付け方という感じです。もっとわかりやすい例は地名でしょう。たとえば「駒場東大前」というのが座標を使わない名前、緯度と経度による指定や、どこどこ社の何とかいう地図の何ページの、というのが座標を使った名前です。

さて、「平面や空間」といちいち言うのは面倒ですし直線や「4次元空間」なども考えたいので、以降は空間という言葉で直線や平面や空間など全部を表すことにし「次元」によって言葉をかえることをやめます。もちろん具体的なイメージは、前小々節の図から座標軸を取り除いたもので十分です。

5.2 座標を使わずに微分を定義するには

座標を使わないということは「変数」という数を使わないということなので、微分として偏微分を採用することはできません。というわけで、方向微分が全微分をうまくモディファイすることになります。どちらもできるのですが、この節は「理屈重視」なので、ことあるごとに「多変数関数の微分としてもっともふさわし

い」と言ってきた全微分、つまり1次近似で考えることにしましょう。

f を n 次元空間 X から m 次元空間 Y への写像、 x 及び p を X の点として、

$$\lim_{x \rightarrow p} \frac{f(x) - f(p) - A(x-p)}{\|x-p\|} = 0 \quad (9)$$

を満たす $A(x-p)$ を f の p における微分の定義としようというわけです。が、この式が意味を持つためには、いくつか考えなければならないことがあります。

Q1 X の点 p と q に対し $p-q$ とは何か。

Q2 その上で $\|p-q\|$ とは何か。

Q3 $A(p-q)$ とは何か。

Q4 その上で Y の点 $f(p)$ と $f(q)$ に対し $f(p) - f(q) - A(p-q)$ とは何か。

Q5 それを $\|p-q\|$ で「割る」とは何か。

Q6 最後に、その「極限」とは何か。

これらの事柄は平面や空間についての常識的な直感を満たすように自然に意味をつけなければなりません。その「常識的な直感」を次の小々節で明確に言い表してみましょう。

5.3 空間とは

ここまではまるで当たり前にわかっていることのように「空間」という言葉を使ってきましたが、そもそも空間とは何のことでしょうか。たとえば4次元空間とは何でしょう。「 $\mathbb{R}^4$ のことだ」と言ってしまうのももちろんかまわないのですが、それでは「座標によらない直接の名前」と「座標を使った名前」の区別が付かないので、ここでの目的には適いません。

ここでは、「空間」というものの満たすべき性質を列挙することで「空間とは何か」という質問に答えたことにするしかないでしょう。そのために、まず(抽象的な意味での)ベクトル空間について復習しておきましょう。ベクトル空間というと $\mathbb{R}^n$ を思い出します。というより、ベクトル空間とは $\mathbb{R}^n$ のことであるとさえ言いたくなるかもしれません。しかし、線形代数で勉強したように、足し算と実数倍の定義された集合のことをベクトル空間というのであって¹¹、なにも、数を並べたものだけがベクトルなわけではありませんでした。数列や多項式、関数なども考えようによってはベクトルと見なせることを線形代数の講義で学ばれたでしょうし、前期の微分方程式のところで実際にそういう見方をしました。もちろん、有

¹¹ 正確な定義は線形代数の講義や教科書に譲ります。

有限次元ベクトル空間は $\mathbb{R}^n$ と「実質的に同じ」になってしまうので、 $\mathbb{R}^n$ だけ考えれば有限次元ベクトル空間については十分です。しかし、一般のベクトル空間を考えるのに必ず $\mathbb{R}^n$ との同型を通じて考える、つまり座標を適当にとって考えるというのは、座標によらない性質を見極める上ではかえって不便になってしまいます。そこで、以下では「 n 次元ベクトル空間 V 」という言い方をし、 $\mathbb{R}^n$ と書くことをわざと避けることにします。みなさんも我慢して $\mathbb{R}^n$ と考えないようにして読んでください。

次に、ここでいう「空間」の前段階としてアファイン空間を説明します。まず定義を書いてしまいましょう。わかりにくいかもしれませんが、すぐ後で平面や空間との具体的な関係を書きますので、よくわからなくてもとりあえず読んでみてください。

アファイン空間とは集合 X とベクトル空間 V と $X \times X$ から V への写像 φ の組であって、次の2条件を満たすものである。

1. X の任意の元 x_0 に対して $\varphi(x_0, \cdot): X \rightarrow V$ は全単射である。
2. X の任意の3つの元 x, y, z に対して $\varphi(x, z) = \varphi(x, y) + \varphi(y, z)$ が成り立つ。

この定義からすると、アファイン空間とは (X, V, φ) と書かれるべきものですが、ふつつ面倒なので X だけ書きます。「ベクトル空間 V の付随するアファイン空間 X 」とか「 V をベクトル空間とするアファイン空間 X 」とかとも言います。

実にわかりにくい定義ですね。平面を例にとって説明しましょう。 X を平面（すなわち2次元空間）とし、 V を「方向と大きさ」ベクトルのなすベクトル空間とします。たとえば X は「東京」とか「大阪」とかの「地点」の集合で、 V は「南30m」とか「北西53km」などの集合です。そして、 $\varphi: X \times X \rightarrow V$ は、たとえば（横浜, 東京）に横浜から東京までの方向と距離のベクトル、たとえば「北北東30km」¹²を対応させる写像です。こう考えると、条件1、2の意味も明らかでしょう。条件1は「平面上のあらゆる点は、任意の点からの方向と距離で表せるし、ある点から方向と距離が同じ2点は存在しない」ということです。もっと具体的に言えば、たとえば「大阪」から「稚内」への「方向と距離」ベクトルもあるし「大阪」から「波照間島」への「方向と距離」ベクトルも存在し、また、「大阪」という始点と「方向と距離」ベクトルを指定すれば、それによって定まる場所の一つだけだというごく当たり前のことです。また、条件2は「大阪」から「東京」への「方向と距離」ベクトルが「大阪」から「名古屋」への「方向と距離」ベクトルと「名古屋」から「東京」への「方向と距離」ベクトルを足したものと一致するという、これまたごく当たり前のことです。

$\varphi: X \times X \rightarrow V$ という、いかにも数学っぽいものは、 X の2元 x, y に対し x を始点 y を終点とするベクトルを $\varphi(x, y)$ と書きましようと言っているだけのこ

¹²全然正確ではありません。

とです。前小々節のイメージで言えば、ディスプレイが「平面 X 」で、OHP シートが「付随するベクトル空間 V 」に当たります。

ここで、 X から 1 点 p を特別な点として選んでみましょう。すると、アファイン空間 X はベクトル空間 V と同一視できてしまいます。つまり、たとえば「駒場東大前」という 1 点を選んでしまえば、どんな場所もそこからの「方向と距離」ベクトルだけで指定できるというわけです。(これが高校で学んだ「平面の位置ベクトル表示」というやつです。) つまり、 $X = \{p + \vec{v}; \vec{v} \in V\}$ と書けるというわけです。一方、どんなベクトル空間 V もそれ自身をベクトル空間とするアファイン空間と見なせます。 $\varphi(\vec{v}, \vec{w}) = \vec{w} - \vec{v}$ です。結局アファイン空間とはベクトル空間を解釈し直したものの、特に有限次元アファイン空間とは $\mathbb{R}^n$ のことにすぎないことがわかりました。

ところで、よく「ベクトルとは向きと大きさを持った量のことだ」と説明することがありますが、単なるベクトル空間 V の元 $\vec{v}$ には大きさはありません。 $\vec{v}$ と $\vec{w}$ が 1 次独立でないとき、つまり平行なときに限り $\vec{v}$ と $\vec{w}$ の比があるだけで、 $\vec{v}$ と $\vec{w}$ が 1 次独立なら比さえありません。特に注意しなければならないことは、 V の元の列 $\vec{v}_1, \vec{v}_2, \dots$ があっても、その極限という概念がないことです。それを考えるためには、ベクトルの大きさという概念を別に導入してやらなければならないのです。「ベクトルの」大きさなので、やはりベクトルの性質と相性のよいものでなければならないでしょう。 V の元 $\vec{v}$ の大きさを $\|\vec{v}\|$ と書くことにしたとき、その「相性の良さ」とは次の条件を満たすことです。

1. 任意の $\vec{v}$ に対して $\|\vec{v}\| \geq 0$ であり、 $\|\vec{v}\| = 0$ となるのは $\vec{0}$ だけである。
2. 任意の実数 c に対して $\|c\vec{v}\| = |c|\|\vec{v}\|$ が成り立つ。
3. 任意の 2 つのベクトル $\vec{v}$ と $\vec{w}$ に対して $\|\vec{v} + \vec{w}\| \leq \|\vec{v}\| + \|\vec{w}\|$ が成り立つ。

この性質を満たすことをはっきり宣言するために、 $\|\vec{v}\|$ のことを $\vec{v}$ の大きさと言わずにノルムと言います。条件 1 は「ベクトルの大きさは 0 以上で、特に大きさ 0 のベクトルは $\vec{0}$ のみ」ということ、条件 2 は「 $\vec{v}$ と $\vec{w}$ が平行なとき、ベクトルとしての比と大きさの比とは絶対値が一致する」ということ、条件 3 は「三角形の 2 辺の長さの和は残りの 1 辺の長さ以上」といういわゆる「三角不等式」のことです。

「東京」「大阪」などを使ったアファイン空間の例の場合、「方向と大きさ」ベクトルという言葉からわかるように X に付随するベクトル空間 V にはノルムがあります。微分を考えるには極限が必要なので、我々の必要な「空間」とは、結局

アファイン空間であって付随するベクトル空間にノルムのあるもの

です。 X の 2 点 p, q に対してベクトル $\varphi(p, q)$ のノルム $\|\varphi(p, q)\|$ が「 p と q の距離」だというわけです。

以上、今まで身近だった「空間」という概念をいかにもわざと難しくしてしまったような印象しかないかもしれません。なぜこんな七面倒くさいことをしたかと

いうと、空間から空間への写像の微分はその空間たちに付随するベクトル空間の間の線形写像だと解釈するのが自然だからなのです。

5.4 写像の微分の定義

微分の定義式(9)を、写像 $f: X \rightarrow Y$ 、 X の点 p に対して書き直すと、

$$\lim_{x \rightarrow p} \frac{f(x) - f(p) - A(x - p)}{\|x - p\|} = 0$$

となります。ここで、 X が V をベクトル空間とする空間、 Y が W をベクトル空間とする空間なら、「考えなければならないこと」Q1～Q6に次のように自然に解答することができます。

A1 $x - p$ は p を始点とし x を終点とするベクトル。つまり $x - p = \varphi(p, x) \in V$ 。

A2 $\|x - p\|$ はベクトル $x - p$ のノルム。つまり p と x の「距離」。

A3 A は V から W への線形写像。よって特に $A(x - p) \in W$ 。

A4 $x - p \in V$ であるのと同様に $f(x) - f(p) \in W$ なので、 $f(x) - f(p) - A(x - p) = (f(x) - f(p)) - A(x - p)$ と解釈して $f(x) - f(p) - A(x - p) \in W$ 。

A5 $f(x) - f(p) - A(x - p) \in W$ で $\|x - p\|$ が0でない実数なので、 $f(x) - f(p) - A(x - p)$ というベクトルを $\|x - p\|$ で割ることができる。

A6 極限は V において $\|x - p\| \rightarrow 0$ としたときの W における極限のこと。

この見方がもっとよくわかるように定義式を書き換えると次のようになります。

$$\lim_{\vec{v} \rightarrow \vec{0}} \frac{(f(p + \vec{v}) - f(p)) - A(\vec{v})}{\|\vec{v}\|} = \vec{0} \quad (10)$$

$p \in X$ 、 $\vec{v} \in V$ 、 $A: V \rightarrow W$ で、左辺の $\vec{0}$ は V の0ベクトル、右辺の $\vec{0}$ は W の0ベクトルです。

この定義によると、点 $p \in X$ における空間 (X, V) から空間 (Y, W) への写像の微分とは、 V から W への線形写像 A で式(10)を満たすものだということになります。つまり、「入力」ディスプレイ X から「出力」ディスプレイ Y への写像 f の点 p での微分は、 p を原点として貼り付けたOHPシート V から $f(p)$ を原点として貼り付けたOHPシート W への線形写像だというわけです。

5.5 合成関数の微分

このまま一気に座標を使わずに合成関数の微分法を示してしましましょう。

3つの空間 (X, V) 、 (Y, W) 、 (Z, U) と X から Y への写像 f および Y から Z への写像 g があり、 f は p で微分可能、 g は $q = f(p)$ で微分可能としましょう。つまり、

$$\lim_{\vec{v} \rightarrow \vec{0}} \frac{(f(p + \vec{v}) - f(p)) - A(\vec{v})}{\|\vec{v}\|} = \vec{0}$$

を満たす線形写像 $A: V \rightarrow W$ と、

$$\lim_{\vec{w} \rightarrow \vec{0}} \frac{(g(q + \vec{w}) - g(q)) - B(\vec{w})}{\|\vec{w}\|} = \vec{0}$$

を満たす線形写像 $B: W \rightarrow U$ が存在するとします。すると、

$$\begin{aligned} g \circ f(p + \vec{v}) &= g(f(p + \vec{v})) \\ &= g(f(p) + A(\vec{v}) + o(\vec{v})) \quad (\vec{v} \rightarrow \vec{0}) \\ &= g(f(p)) + B(A(\vec{v}) + o(\vec{v})) + o(A(\vec{v}) + o(\vec{v})) \quad (\vec{v} \rightarrow \vec{0}) \\ &= g \circ f(p) + B \circ A(\vec{v}) + o(\vec{v}) \quad (\vec{v} \rightarrow \vec{0}) \end{aligned}$$

となつて、 $g \circ f$ も p で微分可能であることが証明され、それが f の p における微分 $A: V \rightarrow W$ と g の $f(p)$ における微分 $B: W \rightarrow U$ との合成写像 $B \circ A: V \rightarrow U$ であるという合成写像の微分法がいとも簡単に得られてしまいました。

5.6 方向微分

さて、今まで説明してきたように、点 p を一つにとって固定して考えても微分は写像になります。具体的には V から W への写像です。よって、点 $p \in X$ だけでなく方向と距離 $\vec{v} \in V$ を決めて初めて W の元が決まります。これが p における $\vec{v}$ 方向微分の意味です。つまり、微分とは「 p を始点としたベクトルに $f(p)$ を始点としたベクトルを対応させる線形写像」であり、方向微分とは「 p を始点とするベクトルを具体的に与えたときにそれが微分によって写される $f(p)$ を始点としたベクトル」のことです。

ところで、この説明だと、 $p \in X$ と $\vec{v} \in V$ を決めることによって $f(p) \in Y$ と $A(\vec{v}) \in W$ が決まるのですから、点 $p + \vec{v} \in V$ と点 $f(p) + A(\vec{v}) \in Y$ の間の何らかの関係を決めているような印象を受けるかもしれません。しかし、微分は極限で決めるものですので、極限をとる前の2点の関係を直接語るものにはなり得ないはずです。それでは、方向微分における $\vec{v}$ はどのように解釈すればよいのでしょうか。

今までは空間 (X, V) の V を X の2点間の方向と距離と言っていましたが、微分という極限と相性がよいようにするには V の元を X の何かの極限の集合であるように考え直した方がよいでしょう。 X という「位置の集合」に関連のあるものの極限としてすぐに思いつくのは、 X を動く動点の「速度」ベクトルでしょう。つまり、 X 内を動いている点が速度 $\vec{v}$ で点 p を通過した場合、その運動を写像 f で Y に写すと速度 $A(\vec{v})$ で点 $f(p)$ を通過するということです。

さて、 X は特にアフィン空間ですので、直線という概念が意味を持ちます。だから、 X 内の動点で p を通過するときの速度が $\vec{v}$ であるもので最も簡単なものは、速度がずっと一定値 $\vec{v}$ であるような等速直線運動で p を通るものでしょう。具体的に式で書くなら、時刻 $t = 0$ で p を通過するものとして

$$c(t) = p + t\vec{v}$$

と書けます。これによって、 $\vec{v}$ 方向微分 $A(\vec{v})$ は $f \circ c(t)$ という1変数写像の $t = 0$ における微分と見ることができます。これが「微分の視覚化」のときに使った見方です。

5.7 偏微分

ここで初めて座標を導入してみましょう。 V はベクトル空間ですから基底を一つ決めることにより $\mathbb{R}^n$ と同一視できます。さらに、 X の1点 p を固定します。すると、 X の点は $p + \vec{v}$ と一意的に表せるのですから $\mathbb{R}^n$ と全単射がつきます。これが空間の点の座標表示です。

(Y, W) にも同様に一つ座標表示を導入しておくと、写像 $f: X \rightarrow Y$ の点 p における微分である線形写像 $A: V \rightarrow W$ は、線形代数で学んだように行列で表すことができます。その行列が f の p におけるヤコビ行列だと言うわけです。

また、 V のベクトル $(1, 0, \dots, 0)$ 方向微分が第1変数による偏微分です。実際、前節の最後に説明したように $(1, 0, \dots, 0)$ 方向微分は $\mathbb{R}$ から X への写像 $c(t) = (a_1, a_2, \dots, a_n) + t(1, 0, \dots, 0) = (a_1 + t, a_2, \dots, a_n)$ と f との合成写像 $f(a_1 + t, a_2, \dots, a_n)$ の t による微分です。これは点 $(a_1, a_2, \dots, a_n)$ における第1変数による偏微分の定義そのものです。

微分は線形写像ですから、 V の一組の基底に対する行き先が決まれば、一般の元の行き先もそれで決まります。つまり、 $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ が V の基底のとき、 V の任意の元は $\vec{v} = a_1\vec{v}_1 + \dots + a_n\vec{v}_n$ と書けます。このとき点 p における f の微分が $\vec{v}$ を写す先は $A(\vec{v}) = a_1A(\vec{v}_1) + \dots + a_nA(\vec{v}_n)$ となります。特に、座標を一つ決めてしまえば、 $(1, 0, \dots, 0), \dots, (0, 0, \dots, 1)$ を基底にとれますので、

$$A(\vec{v}) = a_1f_{x_1}(p) + \dots + a_nf_{x_n}(p)$$

と、偏微分の 1 次結合でかけることになります。ヤコビ行列

$$\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

が f の微分と呼ばれるのはこのようなわけなのです。