

テーマ：ベクトルと行列、合成関数の微分の公式

・ベクトル

我々が本来やりたい議論は微積分なのだが、ここでいったん微積分を離れて、ベクトルについての議論をする必要がある。ベクトルという用語は数学では非常に多くのものを指す用語なのだが、さしあたり用があるのは数を  $n$  個並べたもの  $((1, 2, 4)$  などのように) である。 $(x_1, x_2, \dots, x_n)$  などというふうに丸括弧の中に  $n$  個の数が入っているものを、 $n$  次元のベクトルと呼ぶことにしよう。高校数学では、 $n = 2$  のときのベクトルを主に扱い、そこではベクトルを矢印の形で図で書いて説明していたと思う。しかし、このような図形的な議論はここではする必要がない。単に数を 2 個並べたものが 2 次元のベクトルである、という考え方をしておいたほうが、後々混乱せずに済むだろうと思う。

このベクトルに対して、足し算と実数倍という概念を定義しておきたい。それは次のようにして行う。いま  $x = (x_1, \dots, x_n)$  と  $y = (y_1, \dots, y_n)$  に対して、新しいベクトル  $x + y$  を、 $(x_1 + y_1, \dots, x_n + y_n)$  というものとして定義する。また、数  $c$  に対して、ベクトル  $cx$  を  $(cx_1, \dots, cx_n)$  として定義する。こうして作られた足し算と実数倍に加えて、もう少し記号を追加しておきたい。すべての座標が 0 であるベクトル  $(0, 0, \dots, 0)$  を単に 0 と書き、ゼロベクトルと呼ぶ。また  $-x = (-x_1, \dots, -x_n)$  とする。すると次の 8 つの性質が成り立つ。

1.  $(x + y) + z = x + (y + z)$ .
2.  $x + y = y + x$ .
3.  $x + 0 = x$ .
4.  $x + (-x) = 0$ .
5.  $(a + b)x = ax + bx$ .
6.  $a(x + y) = ax + ay$ .
7.  $(ab)x = a(bx)$ .
8.  $1x = x$ .

すべて当たり前に見えるかもしれない。しかし実はそうではない。実は、数学では本質的には上の 8 つの性質がすべて成り立つような構造を持っているもののことをすべてベクトルと呼ぶのである。したがって我々が考える  $n$  次元のベクトル以外にも、山のように多

くのベクトルという物が存在する。たとえば、一変数関数  $f$  と  $g$  に対して  $f + g$  という関数を  $(f + g)(x) = f(x) + g(x)$  として、また数  $c$  と関数  $f$  に対して  $cf$  という関数を  $(cf)(x) = cf(x)$  として定義すれば、この足し算と実数倍の概念は上の 8 つの性質をすべて満たす。したがって実は一変数関数もベクトルなのである！ この考え方はこの授業ではこれ以上深入りしないが、多くの応用があるので、念頭に置いておきたい。

### ・内積とノルム

最初に、 $x = (x_1, x_2, \dots, x_n)$  と原点  $0 = (0, 0, \dots, 0)$  の間の距離に当たる概念を考える。 $n = 2$  の場合は、すでに中学で習った三平方の定理から、それは  $\sqrt{x_1^2 + x_2^2}$  となる。これを一般化して、 $n \geq 2$  のときに、 $\sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$  を距離だと思ってよいであろう。この値のことを、ベクトル  $x$  のノルムの値と呼び、通常  $\|x\|$  と書く。つまり、

$$\|x\| = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$$

である。

また、ふたつのベクトル  $x = (x_1, x_2, \dots, x_n)$  と  $y = (y_1, y_2, \dots, y_n)$  の内積  $x \cdot y$  を、 $x_1y_1 + x_2y_2 + \dots + x_ny_n$  で定義する。すぐにわかるように  $x \cdot x = \|x\|^2$  である。それ以外にも内積には多くの性質があるのだが、それについては後で詳しく記す。

ノルムは以下の性質を持つ。

- (1)  $\|x\| \geq 0$  は常に成り立ち、もし  $\|x\| = 0$  ならば  $x = 0$  である。
- (2) 数  $c$  がなんであろうと、 $\|cx\| = |c|\|x\|$  が成り立つ。
- (3)  $\|x + y\| \leq \|x\| + \|y\|$  が成り立つ。(三角不等式)

また内積は以下の性質を持つ。

- (i)  $x \cdot x \geq 0$  であり、もし  $x \cdot x = 0$  ならば  $x = 0$  である。
- (ii)  $x \cdot y = y \cdot x$  が成り立つ。
- (iii)  $(x + y) \cdot (z + w) = x \cdot z + x \cdot w + y \cdot z + y \cdot w$  が成り立つ。
- (iv)  $(cx) \cdot y = c(x \cdot y)$  が成り立つ。

これらのうち難しいのは、実は (3) のみである。後は簡単に示すことができるのでこの講義ノートでは省略し、(3) だけを問題にしたい\*<sup>1</sup>。このためには、次の結果を先に示して

---

\*<sup>1</sup> ちなみに、(3) は実数での同じような不等式、

$$|a + b| \leq |a| + |b|$$

おくと便利である。

$$|x \cdot y| \leq \|x\| \|y\|$$

これは極めて重要な不等式であり、**Cauchy-Schwarz** の不等式という名前がついている。最初に、これが成り立っていれば (3) を示すことができるというのを確認しておこう。それは、

$$\begin{aligned} \|x + y\|^2 &= (x + y) \cdot (x + y) \\ &= x \cdot x + 2x \cdot y + y \cdot y \\ &= \|x\|^2 + 2x \cdot y + \|y\|^2 \\ &\leq \|x\|^2 + 2\|x\| \|y\| + \|y\|^2 \\ &= (\|x\| + \|y\|)^2 \end{aligned}$$

から、両辺の二乗を取り除いて得ることができる。間にひとつだけはさまった不等号のところ Cauchy-Schwarz の不等式を使っていることに注意。

後は Cauchy-Schwarz の不等式を示すだけである。このためには、次の関数

$$f(t) = \|tx + y\|^2$$

を考える。右辺はノルムの定義からどんな  $t$  に対しても 0 以上である。一方で、(5)(6)(7) などを使って計算すると、

$$\begin{aligned} f(t) &= \|tx + y\|^2 = (tx + y) \cdot (tx + y) \\ &= \|x\|^2 t^2 + 2(x \cdot y)t + \|y\|^2 \\ &= at^2 + bt + c \end{aligned}$$

となる。ただし  $a = \|x\|^2, b = 2x \cdot y, c = \|y\|^2$  である。ここで  $a = 0$  であるとすれば  $x = 0$  なので、 $x \cdot y$  と  $\|x\| \|y\|$  は両方 0 になり、よって Cauchy-Schwarz の不等式が成り立つ。よって我々は  $a > 0$  であるときだけを考えれば十分である。

さて、 $f(t)$  は上で書いたように二次の多項式であり、さらに  $f(t) \geq 0$  がすべての  $t$  について成り立つ。よって前期で示したように、この式の判別式  $D = b^2 - 4ac$  は 0 以下である。計算すると、

$$D/4 = (x \cdot y)^2 - \|x\|^2 \|y\|^2 \leq 0$$

ということになる。ここから、

$$(x \cdot y)^2 \leq \|x\|^2 \|y\|^2$$

---

があって、これも三角不等式と呼ばれる。この証明は簡単で、単に  $a$  と  $b$  の符号が  $+$  のときと  $-$  のときで場合分けして示すだけでよい。しかし、ベクトルではそう簡単にはいかない。

がわかり、両辺の正の平方根を取ることで、

$$|x \cdot y| \leq \|x\| \|y\|$$

がわかる。これで示せた。

注意：実は、この証明は内積の性質 (i)-(iv) だけしか使っておらず、ベクトルの正体が  $(x_1, \dots, x_n)$  であるという事実を一切、用いていない。したがって、たとえば  $[0, 1]$  区間上で定義された連続な実数値関数  $f, g$  に対して

$$f \cdot g = \int_0^1 f(x)g(x)dx$$

と定義して、これが (i)-(iv) の条件を満たすことを確かめれば、そこからただちにこの場合の Cauchy-Schwarz の不等式

$$\left| \int_0^1 f(x)g(x)dx \right| \leq \sqrt{\int_0^1 (f(x))^2 dx} \sqrt{\int_0^1 (g(x))^2 dx}$$

を得ることができる。このように、この不等式の応用範囲は極めて広い。

問題4：内積とノルムの性質 (1)(2)(i)(ii)(iii)(iv) を示せ。

・行列

$n \times m$  のボックスに数が詰まっているものを行列 (matrix) と呼ぶ。たとえば、

$$\begin{pmatrix} 2 & 3 & 1 \\ 1 & 0 & -5 \end{pmatrix}$$

などは  $2 \times 3$  の行列である。横に抜き出したのを行 (row)、縦に抜き出したのを列 (column) と呼ぶ。たとえば上の行列の二列目は

$$\begin{pmatrix} 3 \\ 0 \end{pmatrix}$$

である。A が行列のとき、 $i$  行  $j$  列の要素は  $a_{ij}$  と書くことにする。

$n \times m$  の行列 A と  $m \times r$  の行列 B はかけ算ができる。このかけ算  $C = AB$  は、

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \dots + a_{im}b_{mj}$$

を  $i$  行  $j$  列の要素として持つ行列になる。たとえば、

$$A = \begin{pmatrix} 2 & 3 & 1 \\ 1 & 0 & -5 \end{pmatrix}, B = \begin{pmatrix} 1 & 0 & 3 \\ 0 & 4 & -1 \\ 2 & 0 & 0 \end{pmatrix}$$

とすれば、

$$AB = \begin{pmatrix} 4 & 12 & 3 \\ -9 & 0 & 3 \end{pmatrix}$$

となる。

特に重要なのが、 $n \times m$  行列  $A$  と  $m$  次元ベクトル  $x$  のかけ算である。やってみるとわかるが、 $A$  の  $i$  列目のベクトルを  $a_i$  とすれば、

$$Ax = x_1 a_1 + \dots + x_m a_m$$

と分解できる。したがって、特に  $\|x\| \leq 1$  の場合、

$$\|Ax\| \leq \|a_1\| + \|a_2\| + \dots + \|a_m\|$$

がわかる。ところがこのとき、任意のベクトル  $x \neq 0$  に対して、 $y = \frac{1}{\|x\|}x$  とすれば、

$$Ax = \|x\|Ay$$

なので、

$$\|Ax\| \leq \|x\|[\|a_1\| + \dots + \|a_m\|]$$

である。当然ながら、この式は  $x = 0$  のときも成り立つので、すべての  $x$  に対して

$$\|Ax\| \leq M\|x\|$$

となる数  $M$  の存在がわかった。このような  $M$  の最小値を  $\|A\|$  と書く。この関数  $A \mapsto \|A\|$  は実はノルムであることが知られているが、ここでは証明しない。

行列の理論はものすごくいろいろあるが、本講義で使う知識はさしあたりここまでである。

#### ・全微分可能性

関数  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$  を考える。この関数  $f$  と点  $x \in \mathbb{R}^n$  に対して、ある  $m \times n$  行列  $A$  が、

$$\lim_{v \rightarrow 0} \frac{\|f(x+v) - f(x) - Av\|}{\|v\|} = 0$$

という式を満たすとき、 $f$  は  $x$  において全微分可能である、と言う。また、行列  $A$  のことを  $f$  の  $x$  における全微分の値と呼んで、これを  $Df(x)$  と書く。

全微分可能な関数は偏微分可能であり、さらに

$$Df(x) = (f_{x_1}(x), f_{x_2}(x), \dots, f_{x_n}(x))$$

となることが知られている。これは非常に簡単に示せる。実際、たとえば  $Df(x) = A$  だったとして、その第一列を  $a_1$  とし、 $v = (h, 0, \dots, 0)$  とすれば、三角不等式から

$$\begin{aligned} & \left\| \frac{f(x_1 + h, x_2, \dots, x_n) - f(x_1, x_2, \dots, x_n)}{h} - a_1 \right\| \\ &= \frac{\|f(x + v) - f(x) - Av\|}{\|v\|} \\ &\rightarrow 0 \end{aligned}$$

となるので、正しい。

逆に  $f$  が偏微分可能だからと言って、全微分可能になるとは限らない。しかし次は成り立つ。

**定理 1** :  $f$  が  $x$  の近くで連続微分可能（つまり、偏導関数が連続）ならば、 $f$  は  $x$  において全微分可能である。

証明 :  $f$  が実数値関数のときのみ示す。実際のところ、 $f$  がベクトル値関数であっても、その全微分可能性は、 $f$  の各座標ごとの全微分可能性と同値であることは簡単に示せるので、これでよい。

さて、 $f$  が実際に  $x$  の近くで連続微分可能であったとしよう。そして、 $v = (v_1, \dots, v_n)$  とし、

$$\begin{aligned} g_1(y_1) &= f(y_1, x_2 + v_2, \dots, x_n + v_n), \\ g_2(y_2) &= f(x_1, y_2, x_3 + v_3, \dots, x_n + v_n), \\ &\dots \\ g_n(y_n) &= f(x_1, \dots, x_{n-1}, y_n) \end{aligned}$$

とする。すると定義から

$$g'_i(y_i) = f_{x_i}(x_1, \dots, x_{i-1}, y_i, x_{i+1} + v_{i+1}, \dots, x_n + v_n)$$

である。よって平均値の定理から、 $0 \leq \theta_i \leq 1$  となるある数  $\theta_i$  について

$$\begin{aligned} f(x+v) - f(x) &= \sum_{i=1}^n [g_i(x_i + v_i) - g_i(x_i)] \\ &= \sum_{i=1}^n v_i g'_i(x_i + \theta_i v_i) \\ &= \sum_{i=1}^n v_i f_{x_i}(x_1, \dots, x_{i-1}, x_i + \theta_i v_i, x_{i+1} + v_{i+1}, \dots, x_n + v_n) \\ &\equiv Bv \end{aligned}$$

である。ただし  $B$  はその  $i$  列目が  $f_{x_i}(x_1, \dots, x_{i-1}, x_i + \theta_i v_i, x_{i+1} + v_{i+1}, \dots, x_n + v_n)$  と一致するような横ベクトル（つまり、 $1 \times n$  行列）である。そこで横ベクトル  $A$  を、その  $i$  列目が  $f_{x_i}(x)$  と一致するようなものとすれば、

$$\begin{aligned} 0 &\leq \frac{1}{\|v\|} |f(x+v) - f(x) - Av| \\ &= \frac{1}{\|v\|} |Bv - Av| \\ &= \frac{1}{\|v\|} |(B-A)v| \\ &\leq \frac{1}{\|v\|} \|B-A\| \|v\| = \|B-A\| \end{aligned}$$

となる。ここで不等号は Cauchy-Schwarz の不等式によった ( $(B-A)v$  はベクトル  $B-A$  と  $v$  の内積に等しい)。  $\theta_i$  は 1 より小さいので、 $f_x, f_y$  が連続であるという仮定から、この右辺  $\|B-A\|$  は  $v \rightarrow 0$  のときに 0 に収束する。したがって、0 とこれにはさまれた項である

$$\frac{1}{\|v\|} |f(x+v) - f(x) - Av|$$

も 0 に収束し、よって  $A = Df(x)$  である。これで証明が終わる。 ■

#### ・合成関数の微分の公式、証明

一変数で、合成関数の微分の公式とは、 $h(x) = f(g(x))$  であるとき、 $h'(x) = f'(g(x))g'(x)$  となるという公式だった。今回はより一般の場合の公式を与える。つまり、 $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$  とし、 $g: \mathbb{R}^m \rightarrow \mathbb{R}^k$  とする。すると  $\ell(x) = (f \circ g)(x)$  とすれば、 $\ell: \mathbb{R}^m \rightarrow \mathbb{R}^m$  である。これについて、次の定理が成り立つ。

定理2 (chain rule) :  $f$  が  $g(x)$  で全微分可能、 $g$  が  $x$  で全微分可能ならば、 $\ell$  は  $x$  で全微分可能であり、

$$D\ell(x) = Df(g(x))Dg(x)$$

が成り立つ。

$n = m = k = 1$  ならば、ここからただちに一変数の合成関数の微分の公式  $(f \circ g(x))' = f'(g(x))g'(x)$  が出てくる。また  $n = k = 1, m = 2$  ならば、

$$(f(g_1(x), g_2(x)))' = f_{x_1}(g_1(x), g_2(x))g_1'(x) + f_{x_2}(g_1(x), g_2(x))g_2'(x)$$

が出てくる。後で述べるように、この公式を  $f(x_1, x_2) = x_1x_2$  に適用することで、容易にかけ算の微分の公式を得ることができる。

一変数の合成関数の微分の公式は  $f, g$  が微分可能さえあれば適用することができたが、二変数の合成関数の微分の公式は  $f, g$  が偏微分可能である、というだけでは出てこない。 $f, g$  は全微分可能でないといけないのである。これが、この公式を難解にしているひとつの原因である。ただし、さきほど証明したことから、連続微分可能であれば全微分可能であるため、この仮定はそこまで強いものではない。

証明に入る前に、これを使った例をいくつか挙げておこう。

例1 :  $f(x, y) = xy$  とする。このとき  $f_x(x, y) = y, f_y(x, y) = x$  である。したがって、上の定理を適用すれば、

$$(f(g(z), h(z)))' = f_x(g(z), h(z))g'(z) + f_y(g(z), h(z))h'(z) = h(z)g'(z) + g(z)h'(z)$$

が成り立つ。これは明らかにかけ算の微分の公式である。

例2 :  $f(x, y) = x^n y$  とし、 $f(x, y) = 1$  という等高線を考える。この式は、

$$f(x, y) = 1 \Leftrightarrow xy = 1 \Leftrightarrow y = \frac{1}{x^n} = x^{-n}$$

と変形できる。そこで、 $g(x) = x^{-n}$  としよう。このとき、

$$f(x, g(x)) = 1$$

が常に成り立つため、

$$(f(x, g(x)))' = 0$$

である。一方で定理2から、

$$(f(x, g(x)))' = f_x(x, g(x))(x)' + f_y(x, g(x))g'(x) = f_x(x, g(x)) + f_y(x, g(x))g'(x)$$

であるため、

$$f_x(x, g(x)) + f_y(x, g(x))g'(x) = 0$$

が成り立っていないといけないことになる。この式を変形すると、

$$g'(x) = -\frac{f_x(x, g(x))}{f_y(x, g(x))}$$

である。  $f_x(x, y) = nx^{n-1}y$ ,  $f_y(x, y) = x^n$  だから、これを当てはめると、

$$g'(x) = \frac{-nx^{n-1}g(x)}{x^n} = -nx^{-n-1}$$

であることがわかる。こうして我々は、 $(x^{-n})' = -nx^{-n-1}$  という公式を得た。(ただし、実はこの証明では  $g$  の微分可能性が示されていない。これは、陰関数定理という極めて高等な定理を用いないと一般には証明できないので、普通このような計算は使わない。)

さて、ふたつの例で定理 2 の感覚はだいたいつかめたと思われるので、いよいよ定理の証明に入っていこう。このために、ひとつ補題を準備する。

**補題:**  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$  のとき、 $Df(x) = A$  であることと、以下の条件を満たす関数  $\varepsilon(v)$  が存在することは同値である。

- 1)  $\varepsilon(0) = 0$  である。
- 2)  $\varepsilon(v)$  は  $v = 0$  で連続である。
- 3) 次の等式

$$\|f(x+v) - f(x) - Av\| \leq \varepsilon(v)\|v\|$$

が成り立つ。

**補題の証明:** 補題のような  $\varepsilon(v)$  が存在すれば明らかに

$$\lim_{v \rightarrow 0} \frac{\|f(x+v) - f(x) - Av\|}{\|v\|} = \lim_{v \rightarrow 0} \varepsilon(v) = 0$$

が成り立つので、 $Df(x) = A$  である。逆に、 $Df(x) = A$  であるとき、

$$\varepsilon(v) = \begin{cases} 0 & \text{if } v = 0, \\ \frac{\|f(x+v) - f(x) - Av\|}{\|v\|} & \text{otherwise} \end{cases}$$

と定義すれば、明らかに 1)-3) が成り立つ。 ■

以上の補題により、まず

$$\begin{aligned}\varepsilon_1(w)\|w\| &\geq \|f(g(x) + w) - f(g(x)) - Df(g(x))w\| \\ \varepsilon_2(v)\|v\| &\geq \|g(x + v) - g(x) - Dg(x)v\|\end{aligned}$$

が成り立つような  $\varepsilon_1, \varepsilon_2$  という関数が存在することが示せた。ただし両方とも 0 で連続でそのときの値は 0 である。  $g(x + v) - g(x) = w$  と以下略記することにする。よって、

$$\begin{aligned}&\|\ell(x + v) - \ell(x) - Df(g(x))Dg(x)v\| \\ &= \|f(g(x + v)) - f(g(x)) - Df(g(x))Dg(x)v\| \\ &= \|f(g(x) + w) - f(g(x)) - Df(g(x))Dg(x)v\| \\ &\leq \|f(g(x) + w) - f(g(x)) - Df(g(x))w\| + \|Df(g(x))(g(x + v) - g(x) - Dg(x)v)\| \\ &\leq \varepsilon_1(w)\|w\| + \|Df(g(x))\|\varepsilon_2(v)\|v\| \\ &\leq \varepsilon_1(w)\|g(x + v) - g(x) - Dg(x)v\| + \varepsilon_1(w)\|Dg(x)v\| + \|Df(g(x))\|\varepsilon_2(v)\|v\| \\ &\leq [\varepsilon_1(w)\varepsilon_2(v) + \varepsilon_1(w)\|Dg(x)\| + \|Df(g(x))\|\varepsilon_2(v)]\|v\|\end{aligned}$$

となるので、

$$\varepsilon(v) = \varepsilon_1(w)\varepsilon_2(v) + \varepsilon_1(w)\|Dg(x)\| + \|Df(g(x))\|\varepsilon_2(v)$$

と定義すると、 $\varepsilon(v)$  は関数  $\ell$  について補題の 1)-3) の条件を満たす。よって補題から、

$$D\ell(x) = Df(g(x))Dg(x)$$

であることがわかって、証明が完成する。 ■