

# スカラー値をとる行列の関数の微分の表現について

前川眞一

東京工業大学大学院 社会理工学研究科 人間行動システム専攻

040830,0901,02,05,06,07,14,15,1621,22,23,24,25,26,27,28,30,1004,05 2008. 9. 11

## 1 はじめに

スカラー値をとるベクトルや行列の関数のベクトルや行列による微分に関しては、Nel(1980)によると、1920年代より、数多くの論文や本が出版されているが、その利点は、微分の結果を再度ベクトルや行列として表現することにより、要素ごとの微分結果からは得にくい知見を簡単に得られることにある。たとえば、正方行列  $S$  を階数 1 の行列で最小 2 乗的に近似することを考える。そのためには、 $x$  を列ベクトルとして、 $S$  を  $xx'$  で近似すればよいが、最小 2 乗基準を行列とベクトルの要素で書くと

$$\text{RSS} = \sum_{i,j} (s_{ij} - x_i x_j)^2 \quad (1)$$

となる。したがって、 $S$  が  $3 \times 3$  の行列の場合、RSS を  $3 \times 1$  のベクトル  $x$  の各要素で偏微分したものは

$$\frac{\partial \sum_{i=1}^n \sum_{j=1}^n (s_{ij} - x_i x_j)^2}{\partial x_q} = 2 \left( 2x_q \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i (s_{iq} + s_{qi}) \right) \quad (2)$$

すなわち

$$\begin{aligned} \frac{\partial \text{RSS}}{\partial x_1} &= -4x_1^3 + 4x_1(x_2^2 + x_3^2 - s_{11}) - 2x_2(s_{12} + s_{21}) - 2x_3(s_{13} + s_{31}) \\ \frac{\partial \text{RSS}}{\partial x_2} &= -4x_2^3 + 4x_2(x_1^2 + x_3^2 - s_{22}) - 2x_1(s_{12} + s_{21}) - 2x_3(s_{23} + s_{32}) \\ \frac{\partial \text{RSS}}{\partial x_3} &= -4x_3^3 + 4x_3(x_1^2 + x_2^2 - s_{33}) - 2x_1(s_{13} + s_{31}) - 2x_2(s_{23} + s_{32}) \end{aligned} \quad (3)$$

となるが、これを縦に並べたものが、次節に示すように、スカラー関数 RSS のベクトル  $x$  による微分  $\frac{\partial \text{RSS}}{\partial x}$  に対応するものとなる。しかし、この表現は、あくまで元の行列やベクトルの要素を用いた表現であり、これを行列やベクトルそのものを用いて

$$\frac{\partial \text{RSS}}{\partial x} = \frac{\partial \text{tr}((S - xx')'(S - xx'))}{\partial x} = -2((S + S')x - 2(x'x)x) \quad (4)$$

と表現されることを見いだすのはかなり困難である。しかし、もしも (4) 式の表現が直接得られれば、RSS を最小とする  $x$  はその長さを固有値の平方根になるように正規化した固有ベクトルであることが瞬時に見て取れることになる。

この例から明らかのように、ベクトルや行列による微分を使うことにより、最小 2 乗基準を最小とするパラメタ行列や、正規分布の尤度関数を最大とするパラメタ行列を、その要素に戻ることなく、簡単に導くことが可能となる。スカラー値をとるベクトルや行列の関数に限定した理由は、スカラー関数の微分結果は、(4) 式が示す様に、元のベクトル・行列 ( $x$ ) と同じサイズのベクトル・行列として表現可能であり、その構造を把握しやすいことにある。ベクトル値を返す関数のベクトルによる偏微分は行列となり、行列値を返す関数のベクトル・行列に

よる微分は、何らかの工夫をしない限り、ベクトル・行列としては表現不可能である。行列値を返す関数のベクトル・行列による微分は、たとえば尤度関数の2次微分として、推定量の標準誤差を導出する際に必要となるが、Magnus & Neudecker(1988)によれば、*vec* 関数を用いてベクトルの形に変形し、結果を行列の形で示すという方針が良いとされている。しかし、本稿では、2次微分の問題は取り扱わない。

本稿の目的は、スカラー値をとるベクトルや行列の関数（ベクトルの2次形式、Hadamard積、要素の冪乗や逆行列、一般逆行列を含む行列のトレース、行列式、および固有値等）のベクトル・行列による微分公式を導出し、それらを、効率の良い形で計算機プログラムの関数として実装することにある。プログラム作成の目的は、実用場面での利用のみならず、行列の微分を学ぼうとする者の教育場面での利用も考慮している。プログラム作成に際しては、一般性を失わない限りの必要最低限な公式のみを実装するという方針を用いた。

本稿の構成としては、まず次節で実装に必要な公式を列挙し、その後、公式の導出の詳細を第3節に示した。また、プログラムの例を第4節に示し、最後に、いくつかの結果を付録として掲載した。

## 2 ベクトル・行列による微分の基本公式

いま、 $\mathbf{X} = \{x_{ij}\}$  を  $a \times b$  の行列とし、 $\mathbf{A}, \mathbf{B}$  等を  $\mathbf{X}$  との乗算が可能な定数行列とする。

いま、 $f(\mathbf{X})$  を行列からスカラーへの関数とした場合に、 $f(\mathbf{X})$  を行列  $\mathbf{X}$  の各要素で偏微分し、その行列の形に並べた  $a \times b$  の行列を、 $f(\mathbf{X})$  の  $\mathbf{X}$  による微分行列  $\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}}$  と定義する。

### 定義 2.1 (スカラー関数の行列による微分)

$$\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} = \left\{ \frac{\partial f}{\partial x_{ij}} \right\}, \quad i = 1, 2, \dots, a, \quad j = 1, 2, \dots, b \quad (5)$$

なお、列ベクトル  $\mathbf{x}$  による微分は、 $\mathbf{x}$  を  $a \times 1$  の行列として取り扱い

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \left\{ \frac{\partial f}{\partial x_i} \right\}, \quad i = 1, 2, \dots, a \quad (6)$$

を意味しているものとする。

以下に、微分を求める際に必要となるいくつかの公式を挙げるが、本稿では、それらを、基本公式、特定公式、変形公式と、連鎖律・乗法公式の4つのセットに分割している。基本公式とは、行列表現の単純化とスカラー関数の微分に際する一般的な公式、特定公式とは、関数  $f(\mathbf{X})$  に含まれる行列の関数の特徴（たとえば、逆行列を含む、行列式を含む等）に注目し、その最も単純な形を示したものである。また、変形公式は、行列演算（特にトレースや転置演算）の規則を用いて、与えられた関数  $f(\mathbf{X})$  を特定公式の形に同値変形するための規則であり、連鎖律・乗法公式とは、変形公式では不十分な場合に、 $f(\mathbf{X})$  を特定公式の形に変形するための規則である。

### 2.1 基本公式

- a1 行列表現の単純化

公式を適用する前に、なるべく、 $f$  の引数として与えられる行列表現を単純にしておく  
と便利である。たとえば、行列のスカラー（-1 を含む）倍は行列の積、トレースや行列  
式また逆行列等の外に出す、分配の法則を用いて、複数次項の和と行列の積を、行列の積  
の和に分解する、複数次項のトレースは複数のトレースの和に分解する、複数次項の転置は  
複数の転置の和に置き換える、といった作業である。なお、この作業は、特定公式の数  
を最小限に保つために必要な作業である。

以下に、その際に必要となる行列の演算規則を示す。ただし、 $A, B$  等は行列、 $a, b$   
等は列ベクトルであり、全て、以下に示す演算が可能なサイズであるとする。また、 $\gamma$   
はスカラーである。 $A \odot B$  は Hadamar 積（行列の要素ごとの積）を、（行列の要素ご  
との積）を、 $A \otimes B$  は Kronecker 積を表し、要素の冪乗を  $A^{(p)}$  として以下のように  
定義する。

$$A \odot B = \{a_{ij}b_{ij}\}, \quad A \otimes B = \{a_{ij}B\}, \quad A^{(p)} = \{a_{ij}^p\} \quad (7)$$

また、 $\mathbf{1}, I$  はそれぞれ、全てが 1 の列ベクトルと単位行列を表す。

#### 転置の演算規則

$$\gamma' = \gamma, \quad (A')' = A, \quad (A^{-1})' = A'^{-1} \quad (8)$$

$$(A+B)' = A' + B', \quad (AB)' = B'A', \quad \text{diag}(A)' = \text{diag}(A) \quad (9)$$

$$(A \odot B)' = A' \odot B' = B' \odot A', \quad (A^{(p)})' = (A')^{(p)} \quad (10)$$

#### トレースの演算規則

$$\text{tr}(\gamma) = \gamma, \quad \text{tr}(\gamma A) = \gamma \text{tr}(A) \quad (11)$$

$$\text{tr}(A) = \text{tr}(A'), \quad \text{tr}(A+B) = \text{tr}(A) + \text{tr}(B) \quad (12)$$

$$\text{tr}(AB) = \text{tr}(BA) = \text{tr}(B'A') = \text{tr}(A'B') \quad (13)$$

$$x'Ax = \text{tr}(x'Ax) = \text{tr}(Axx')$$

#### 行列式の演算規則

$$|\gamma| = \gamma, \quad |\gamma A| = \gamma|A|, \quad |I| = 1 \quad (15)$$

$$|A'| = |A|, \quad |AB| = |A||B| \quad (16)$$

$$|A^{-1}| = \frac{1}{|A|}, \quad |-A| = (-1)^{\dim(A)}|A| \quad (17)$$

#### その他の演算規則

$$(\gamma A)B = \gamma AB, \quad (\gamma A)^{-1} = \frac{1}{\gamma}A^{-1} \quad (18)$$

$$A \odot B = B \odot A, \quad A \odot I = \text{diag}(A), \quad I \odot A = \text{diag}(A) \quad (19)$$

$$A^{(p)} \odot A^{(q)} = A^{(p+q)}, \quad A^{(p)} \odot B^{(p)} = (A \odot B)^{(p)} \quad (20)$$

$$\text{tr}(A'(B \odot C)) = \text{tr}((A' \odot B')C), \quad \text{vec}(AXB) = (B' \otimes A)\text{vec}(X) \quad (21)$$

- a2 一般微分公式

スカラー値をとる行列の関数の行列に関する微分に際しては、その定義から明らかなように、通常的一般微分公式が適用できる。以下にその主なものを挙げる。ただし、 $f, f_1, f_2$  は引数として行列を持ちスカラー値を返す関数、 $f_3$  は引数としてスカラーを持ちスカラー値を返す関数である。

$$\frac{\partial a f(\mathbf{X})^b}{\partial \mathbf{X}} = ab f(\mathbf{X})^{b-1} \frac{\partial f(\mathbf{X})}{\partial \mathbf{X}}, \quad \text{ただし } a, b \text{ は定数} \quad (22)$$

$$\frac{\partial (f_1(\mathbf{X}) + f_2(\mathbf{X}))}{\partial \mathbf{X}} = \frac{\partial f_1(\mathbf{X})}{\partial \mathbf{X}} + \frac{\partial f_2(\mathbf{X})}{\partial \mathbf{X}} \quad (23)$$

$$\frac{\partial \exp(f(\mathbf{X}))}{\partial \mathbf{X}} = \exp(f(\mathbf{X})) \frac{\partial f(\mathbf{X})}{\partial \mathbf{X}}, \quad \frac{\partial \log f(\mathbf{X})}{\partial \mathbf{X}} = \frac{1}{f(\mathbf{X})} \frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} \quad (24)$$

$$\frac{\partial f_1(\mathbf{X}) f_2(\mathbf{X})}{\partial \mathbf{X}} = \frac{\partial f_1(\mathbf{X})}{\partial \mathbf{X}} f_2(\mathbf{X}) + f_1(\mathbf{X}) \frac{\partial f_2(\mathbf{X})}{\partial \mathbf{X}} \quad (25)$$

$$\frac{\partial f_3(f_2(\mathbf{X}))}{\partial \mathbf{X}} = \frac{\partial f_3(f_2)}{\partial f_2} \frac{\partial f_2(\mathbf{X})}{\partial \mathbf{X}} \quad (26)$$

- D1 転置行列による微分

$$\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}'} = \left( \frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} \right)' \quad (27)$$

(5) 式の定義から、転置行列で微分したものはもとの行列で微分したものを転置したものである。なお、このバリエーションとして、

$$\frac{\partial f(\mathbf{X}')}{\partial \mathbf{X}} = \left( \frac{\partial f(\mathbf{X}')}{\partial \mathbf{X}'} \right)' \quad (28)$$

があるが、これは、 $\frac{\partial f(\mathbf{X}')}{\partial \mathbf{X}}$  の結果は、 $\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}}$  の結果を  $\mathbf{R}$  としたときに、 $\mathbf{R}'$  に含まれる  $\mathbf{X}$  と  $\mathbf{X}'$  を交換したものととして得られることを示している。

- D2 ベクトル化された行列による微分

$$\frac{\partial f(\mathbf{X})}{\partial \text{vec}(\mathbf{X})} = \text{vec} \left( \frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} \right) \quad (29)$$

ただし、 $\text{vec}(\mathbf{A})$  は、行列  $\mathbf{A}$  の各列を縦に並べて長い列ベクトルを作る関数である。上式の左辺は、 $f$  を、行列  $\mathbf{X}$  の関数ではなく、 $\mathbf{X}$  をベクトル化したものである  $\text{vec}(\mathbf{X})$  の関数と見なし、 $f$  の  $\text{vec}(\mathbf{X})$  に関する微分をとったものであるが、これは定義から、 $\mathbf{X}$  に関する微分を  $\text{vec}(\cdot)$  演算してベクトル化したものとなる。

なお、微分結果が  $\mathbf{X}$  の 1 次式  $\mathbf{AXB}$  の場合、

$$\text{vec}(\mathbf{AXB}) = (\mathbf{B}' \otimes \mathbf{A}) \text{vec}(\mathbf{X}) \quad (30)$$

の関係を用いれば、結果を  $\text{vec}(\mathbf{X})$  を用いて表現することも可能である。

- D3 対角行列による微分

$$\frac{\partial f(\mathbf{X})}{\partial \text{diag}(\mathbf{X})} = \text{diag}\left(\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}}\right) \quad (31)$$

$$\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} = \text{diag}\left(\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}}\right) \quad \text{ただし } \mathbf{X} \text{ は対角行列} \quad (32)$$

ただし、 $\text{diag}(\mathbf{A})$  は行列  $\mathbf{A}$  の非対角要素を全て 0 とおく関数である。

行列  $\mathbf{X}$  の対角成分のみによる微分は、定義により、 $\mathbf{X}$  による微分の対角部分となる。 $\mathbf{X}$  が、元来、対角行列の場合は、その非対角要素は 0 であるため、(5) 式の非対角部分はゼロとなる。したがって、対角行列であることを気にせずに微分し、その結果を対角化すればよい。

- D4 対称行列による微分

$$\begin{aligned} \frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} &= \mathbf{R}, \quad \mathbf{X} \text{ が無制約の場合} \\ &= \mathbf{R} + \mathbf{R}' - \text{diag}(\mathbf{R}), \quad \mathbf{X} \text{ が対称の場合} \end{aligned} \quad (33)$$

$\mathbf{X}$  行列が対称行列の場合には、対応する非対角要素が等しいという制約を課しながら微分することになる。その結果、制約のない行列による微分の結果を  $\mathbf{R}$  とすると、 $\mathbf{X}$  が対称な場合には上式となる。なお、 $\mathbf{X}$  行列の要素が一般的な等号制約、たとえば  $x_{ij} = x_{kl}$  を持つ場合、制約なしの微分結果を  $\mathbf{R}$  行列とすれば、制約付きの結果は、 $\mathbf{R}$  行列の  $ij$  並びに  $kl$  要素を共に  $a_{ij} + a_{kl}$  で置き換えたものとなる。また、 $\mathbf{X}$  のある特定要素が定数  $c$  (たとえば  $c=0$ ) という制約は、 $\mathbf{R}$  のその部分を 0 と置いたものとなる。

## 2.2 特定公式

本稿で取り扱う特定公式は以下に挙げたものである。以下に示す結果は、全て、行列  $\mathbf{X}$  の要素間に制約がない場合である。制約が付いている場合には、以下の結果に、たとえば (33) 式等を用いることになる。なお、 $\mathbf{A}$  や  $\mathbf{B}$  は  $\mathbf{X}$  の要素を含まない行列である。

- S1  $f$  が  $\mathbf{X}$  を含まない場合

$$\frac{\partial f(\mathbf{A})}{\partial \mathbf{X}} = \mathbf{O} \quad (34)$$

- S2  $f$  が  $\mathbf{X}$  を一つだけ含むトレースの場合

$$\frac{\partial \text{tr}(\mathbf{X}\mathbf{A})}{\partial \mathbf{X}} = \mathbf{A}' \quad (35)$$

- S3.1  $f$  が  $\mathbf{X}^{-1}$  を一つだけ含むトレースの場合

$$\frac{\partial \text{tr}(\mathbf{X}^{-1}\mathbf{A})}{\partial \mathbf{X}} = -(\mathbf{X}^{-1}\mathbf{A}\mathbf{X}^{-1})' \quad (36)$$

- S3.2  $f$  が  $\mathbf{X}^-$  を一つだけ含むトレースの場合

$$\frac{\partial \operatorname{tr}(\mathbf{X}^- \mathbf{A})}{\partial \mathbf{X}} = -(\mathbf{X}^- \mathbf{A} \mathbf{X}^-)' \quad (37)$$

ただし、 $\mathbf{X}^-$  は、正方行列  $\mathbf{X}$  中の、最大の線形独立な部分をその逆行列で置き換え、残りの部分を 0 とおいた形の一般逆行列である。この形の一般逆行列は、g12 inverse または反射型 (reflective inverse) と呼ばれ、たとえば、スweep演算子 (Goodnight, 1979) を用いて簡単に求めることが可能である。(Moor-Penrose 逆行列  $\mathbf{X}^+$  はこの形ではない。)

- S3.3  $f$  が  $\mathbf{X}^+$  を一つだけ含むトレースの場合

$$\frac{\partial \operatorname{tr}(\mathbf{X}^+ \mathbf{A})}{\partial \mathbf{X}} = -(\mathbf{X}^+ \mathbf{A} \mathbf{X}^+)' + \mathbf{A} \mathbf{X}^+ (\mathbf{X}^+)' - \mathbf{X} \mathbf{X}^+ \mathbf{A} \mathbf{X}^+ (\mathbf{X}^+)' \quad (38)$$

$$+ (\mathbf{X}^+)' \mathbf{X}^+ \mathbf{A} - (\mathbf{X}^+)' \mathbf{X}^+ \mathbf{A} \mathbf{X}^+ \mathbf{X} \quad (39)$$

- S4  $f$  が  $\mathbf{X}$  の行列式の場合

$$\frac{\partial |\mathbf{X}|}{\partial \mathbf{X}} = |\mathbf{X}| (\mathbf{X}')^{-1} \quad (40)$$

- S5  $f$  が  $\mathbf{X}$  の固有値の場合

$$\frac{\partial \lambda(\mathbf{X})}{\partial \mathbf{X}} = \frac{1}{q'p} p q' \quad (41)$$

ただし、 $\lambda, p, q$  は、それぞれ、 $\mathbf{X}$  の固有値並びにそれに対応する左右の固有ベクトル ( $\mathbf{X}p = \lambda p$ ,  $q' \mathbf{X} = \lambda q'$ ) である。(  $\mathbf{X}$  が対称行列の場合は  $p=q$  となるが、この場合は、上記の結果に対称行列に関する微分の公式 (33) 式 を適用する必要がある。)

- S6  $f$  が  $\mathbf{X}$  の Hadamar 積を一つだけ含むトレースの場合

$$\frac{\partial \operatorname{tr}((\mathbf{X} \odot \mathbf{A}) \mathbf{B})}{\partial \mathbf{X}} = (\mathbf{A} \odot \mathbf{B}') \quad (42)$$

なお、この形は、 $\mathbf{X} \odot \mathbf{I} = \operatorname{diag}(\mathbf{X})$  の関係を考慮すれば、 $\frac{\partial \operatorname{tr}(\operatorname{diag}(\mathbf{X}) \mathbf{B})}{\partial \mathbf{X}}$  を含んでいる。

- S7  $f$  が  $\mathbf{X}$  の 要素の冪乗行列を一つだけ含むトレースの場合

$$\frac{\partial \operatorname{tr}(\mathbf{X}^{(p)} \mathbf{B})}{\partial \mathbf{X}} = p \mathbf{X}^{(p-1)} \odot \mathbf{B}' \quad (43)$$

## 2.3 変形公式

与えられた関数  $f$  が特定公式の形には含まれない場合、トレースや行列式の演算規則を組み合わせるにより、 $f$  の形を特定公式に含まれる形に同値変形していくことが必要となる。この作業により、かなり多くの関数の微分を求めることが出来る。たとえば、付録に示した結果のうち、右辺に複数の表現を含むものは、変形公式の適用により得られたものである。

変形公式としては、後に示す連鎖律や乗法公式の形から、以下に示すトレースを含む形のみを考えておけば十分である。なお、 $A$  と  $B$  は  $X$  を含まない定数行列の一つ以上の積であるが、 $Y, y_1, y_2$  は  $X$  を含んでも良い。また、アスタリスクを付けたものは、無くても良い。

なお、変形公式は、一意に定まるものではなく、ここでは、第 4 節 に述べる、コンピュータへの実装に適したものを表示した。

- G1 ベクトルと行列の積

$$\frac{\partial y_1' Y y_2}{\partial x} = \frac{\partial \text{tr}(Y y_2 y_1')}{\partial x} \quad (44)$$

- G2 トレースの中に少なくとも一つの  $X$  を含む形

$$\frac{\partial \text{tr}(A X Y^*)}{\partial X} = \frac{\partial \text{tr}(X Y A)}{\partial X} \quad (45)$$

- G3 トレースの中にひとつの  $X'$  を含む形

$$\frac{\partial \text{tr}(A^* X' B^*)}{\partial X} = \frac{\partial \text{tr}(X (B A)')}{\partial X} \quad (46)$$

- G4.1 トレースの中に  $X$  の関数の逆行列をひとつだけ含む形

$$\frac{\partial \text{tr}(A F(X)^{-1} B^*)}{\partial X} = \frac{\partial \text{tr}(F(X)^{-1} B A)}{\partial X} \quad (47)$$

- G4.2, 4.3 トレースの中にトレースの中に  $X$  の関数の一般逆行列や Moor-Penrose をひとつだけ含む形  
これは上式の  $X^{-1}$  をそれぞれ  $X^-$  や  $X^+$  で置き換えたものとなる。

- G5.1 トレースの中に  $X$  の Hadamar 積をひとつだけ含む形

$$\frac{\partial \text{tr}(A(X \odot C) D^*)}{\partial X} = \frac{\partial \text{tr}((X \odot C) D A)}{\partial X} \quad (48)$$

$$\frac{\partial \text{tr}(A(B \odot X) D^*)}{\partial X} = \frac{\partial \text{tr}((X \odot B) D A)}{\partial X} \quad (49)$$

- G5.2 トレースの中に  $X'$  の Hadamar 積をひとつだけ含む形

$$\frac{\partial \text{tr}((X' \odot C) D^*)}{\partial X} = \frac{\partial \text{tr}((X \odot C') D')}{\partial X} \quad (50)$$

- G5.3 トレースの中に  $X$  の関数の Hadamar 積をひとつだけ含む形

$$\frac{\partial \text{tr}((B \odot F(X)) C^*)}{\partial X} = \frac{\partial \text{tr}((F(X) \odot B) C)}{\partial X} \quad (51)$$

- G6.1 トレースの中に  $X$  の関数の要素の冪乗行列をひとつだけ含む形

$$\frac{\partial \text{tr}(A F(X)^{(p)} B^*)}{\partial X} = \frac{\partial \text{tr}(F(X)^{(p)} B A)}{\partial X} \quad (52)$$

- G6.2 トレースの中に  $X$  の関数の転置の要素の冪乗行列をひとつだけ含む形

$$\frac{\partial \text{tr}((F(X))^{(p)} B^*)}{\partial X} = \frac{\partial \text{tr}(F(X)^{(p)} B')}{\partial X} \quad (53)$$

- G7 トレースの中に少なくともひとつの  $X$  の関数の対角化を含む形

$$\frac{\partial \text{tr}(A^* \text{diag}(F(X)) Y^*)}{\partial X} = \frac{\partial \text{tr}((F(X) \odot I) Y A)}{\partial X} \quad (54)$$

なお、これらの公式は、上記の順に優先度が高いものとする。すなわち、たとえ (45) 式の  $Y$  が  $X'$  を含んでいたとしても、(46) 式ではなく (45) 式を適用する。

## 2.4 連鎖律・乗法公式

上記の変形公式を用いても、特定公式の形に変形が出来ない場合には、以下に示す連鎖律ならびに乗法公式を用いる必要が生じる。なお、 $F$  や  $G$  は、 $F(X)$  や  $G(X)$  を表しているが、 $X$  の関数であることを無視することを意味する。

- C1 連鎖律

$F = F(X)$  を  $X$  の関数とする。もしも、 $f(X)$  を  $F$  の関数と見なした場合には、特定公式が利用できるなら、以下の連鎖律を利用することが出来る。

$$\frac{\partial f(F(X))}{\partial X} = \frac{\partial \operatorname{tr} \left( \left[ \frac{\partial f(F)}{\partial F} \right]'_c F(X) \right)}{\partial X} \quad (55)$$

ただし、 $\left[ \frac{\partial f(F)}{\partial F} \right]'_c$  は、 $f$  を  $F$  の要素で偏微分したものを、 $X$  を含まない定数行列のように取り扱うことを示している。したがって、特定公式のそれぞれに対応する連鎖律が存在することになる。

なお、たとえ  $F$  が対称行列でも、 $\frac{\partial f(F)}{\partial F}$  の評価に際しては、対称性を無視して適用するものとする。

- P1 トレーズの微分に関する乗法公式 1

$f(X)$  が  $X$  の関数の積のトレーズの場合、乗法公式を使うと、それを、二つのトレーズの和に分解することが出来る。

$$\frac{\partial \operatorname{tr}(F(X)G(X))}{\partial X} = \frac{\partial \operatorname{tr}(F(X)G_c)}{\partial X} + \frac{\partial \operatorname{tr}(F_c G(X))}{\partial X} \quad (56)$$

ただし、 $F = F(X)$  と  $G = G(X)$  は行列  $X$  の関数として定義される行列で、 $\frac{\partial \operatorname{tr}(F_c G(X))}{\partial X}$  や  $\frac{\partial \operatorname{tr}(F(X)G_c)}{\partial X}$  は、それぞれ、 $F$  や  $G$  を  $X$  の関数ではなくあたかも定数のようにして微分することを表している。なお、たとえ  $F$  や  $G$  が対称行列でも、 $\frac{\partial f(F)}{\partial F}$  や  $\frac{\partial f(G)}{\partial F}$  の評価に際しては、対称性を無視して適用するものとする。

- P2 トレーズの微分に関する乗法公式 2

同様に、 $f(X)$  が  $X$  の関数の Hadamar 積のトレーズの場合、以下の乗法公式が利用できる。

$$\frac{\partial \operatorname{tr}((F(X) \odot G(X))B)}{\partial X} = \frac{\partial \operatorname{tr}((F(X) \odot G_c)B)}{\partial X} + \frac{\partial \operatorname{tr}((F_c \odot G(X))B)}{\partial X} \quad (57)$$

ただし、 $F = F(X)$  と  $G = G(X)$  は行列  $X$  の関数として定義される行列で、 $\frac{\partial \operatorname{tr}(F_c \odot G(X))}{\partial X}$  や  $\frac{\partial \operatorname{tr}(F(X) \odot G_c)}{\partial X}$  は、それぞれ、 $F$  や  $G$  を  $X$  の関数ではなくあたかも定数のようにして微分することを表している。なお、たとえ  $F$  や  $G$  が対称行列でも、 $\frac{\partial f(F)}{\partial F}$  や  $\frac{\partial f(G)}{\partial F}$  の評価に際しては、対称性を無視して適用するものとする。

### 連鎖律・乗法公式の例

たとえば、 $f$  が  $X$  を含む行列  $F = F(X)$  の逆行列のトレースとして定義されている場合、(55) 式と (36) 式より、

$$\begin{aligned} \frac{\partial \operatorname{tr}(F(X)^{-1}A)}{\partial X} &= \frac{\partial \operatorname{tr} \left( \left[ \frac{\partial \operatorname{tr}(F^{-1}A)}{\partial F} \right]'_c F(X) \right)}{\partial X} \\ &= - \frac{\partial \operatorname{tr} \left( \left[ (F^{-1}AF^{-1})' \right]'_c F(X) \right)}{\partial X} \end{aligned} \quad (58)$$

$$= - \frac{\partial \operatorname{tr} \left( \left[ F^{-1}AF^{-1} \right]'_c F(X) \right)}{\partial X} \quad (59)$$

となる。したがって、たとえば、 $\frac{\partial \operatorname{tr}((BXC)^{-1}A)}{\partial X}$  を求める場合には、 $F(X) = BXC$  と置いて、(59) 式と (45) 式、並びに (35) 式を以下のように使うことになる。

$$\begin{aligned} \frac{\partial \operatorname{tr}((BXC)^{-1})}{\partial X} &= - \frac{\partial \operatorname{tr} \left( \left[ (BXC)^{-1}A(BXC)^{-1} \right]'_c BXC \right)}{\partial X} \\ &= - \frac{\partial \operatorname{tr} \left( XC \left[ (BXC)^{-1}A(BXC)^{-1} \right]'_c B \right)}{\partial X} \\ &= - \left( C \left[ (BXC)^{-1}A(BXC)^{-1} \right]'_c B \right)' \\ &= -B' \left( (BXC)^{-1}A(BXC)^{-1} \right)' C' \end{aligned} \quad (60)$$

その他の関数の場合も同様である。

また、トレースの乗法公式に関しては、たとえば、 $\frac{\partial \operatorname{tr}(X'AX)}{\partial X}$  を求める場合には、(56) 式と (46) 式、(45) 式、ならびに (35) 式を、 $F = F(X) = X'$ 、 $G = G(X) = AX$  として、以下のように使うことになる。

$$\begin{aligned} \frac{\partial \operatorname{tr}(X'AX)}{\partial X} &= \frac{\partial \operatorname{tr}(X'(AX)_c)}{\partial X} + \frac{\partial \operatorname{tr}(X'_cAX)}{\partial X} \\ &= \frac{\partial \operatorname{tr}(X(AX)'_c)}{\partial X} + \frac{\partial \operatorname{tr}(XX'_cA)}{\partial X} \\ &= (AX)_c + (X'_cA)' = (A + A')X \end{aligned} \quad (61)$$

### 3 公式の導出

#### 3.1 特定公式の導出

行列もしくはスカラーを返す行列  $X$  の関数を  $G = G(X)$  とする。本稿では、 $G$  の形として、逆行列、一般逆行列、Moor-Penrose 逆行列、行列式、固有値、Hadamard 積、要素の冪乗を考えているが、それらの特定公式は、 $G$  が行列の場合は  $\frac{\partial \operatorname{tr}(G(X)B)}{\partial X}$  という形、スカラー関数の場合は  $\frac{\partial G(X)}{\partial X}$  という形で与えられている。本稿では、主に、Magnus & Neudecker(1988) や Harville(1997) を参考にして特定公式を導いたが、ここでは、これらの特定公式に対応するものは、以下に示すような行列関数の行列の要素による偏微分公式という形で示されている。ただ、それらの公式は、一般性を考慮して、 $G$  を直接行列  $X$  の関数と

見なすのではなく、 $F = F(\mathbf{X})$  という、 $\mathbf{X}$  の行列関数  $F$  の関数とした場合の公式が示されている。しかし、本稿では、特定公式としては最も簡単な形を提示し、複雑な場合は連鎖律を用いる方針をとっているため、これらの公式そのままの形では利用しない。

本節では、これらの公式から、特定公式を導いた過程について説明する。また、それ以外の特定公式（Hadamard 積並びに要素の冪乗）の導出は、第 3.1.4 節以降に述べる。

### 3.1.1 Magnus & Neudecker 並びに Harville の公式

以下に、 $F(\mathbf{X}) = \mathbf{X}$  と置いた場合の、行列関数  $G$  の要素を  $x_{ij}$  で偏微分した際の公式を示す。出展は Magnus & Neudecker(1988) を MN で、Harville(1997) を H で示し、章番号. 節番号. 式番号の形で示した。

- s3.1  $f$  が  $\mathbf{X}^{-1}$  をひとつだけ含むトレースの場合  
H15.8.15, MN8.4.1

$$G(\mathbf{X}) = \mathbf{X}^{-1} \quad (62)$$

$$\frac{\partial \mathbf{X}^{-1}}{\partial x_{ij}} = -\mathbf{X}^{-1} \frac{\partial \mathbf{X}}{\partial x_{ij}} \mathbf{X}^{-1} \quad (63)$$

- s3.2  $f$  が  $\mathbf{X}^{-}$  をひとつだけ含むトレースの場合  
H15.10.1

$$G(\mathbf{X}) = \mathbf{X}^{-} \quad (64)$$

$$\frac{\partial G(\mathbf{X})}{\partial x_{ij}} = \frac{\partial \mathbf{X}^{-}}{\partial x_{ij}} = -\mathbf{X}^{-} \frac{\partial \mathbf{X}}{\partial x_{ij}} \mathbf{X}^{-} \quad (65)$$

- s3.2  $f$  が  $\mathbf{X}^{-}$  (g12) をひとつだけ含むトレースの場合  
H15.10.2

$$G(\mathbf{X}) = \mathbf{X} \mathbf{X}^{-} \mathbf{X} \quad (66)$$

$$\frac{\partial G(\mathbf{X})}{\partial x_{ij}} = \frac{\partial \mathbf{X} \mathbf{X}^{-} \mathbf{X}}{\partial x_{ij}} = -\mathbf{X} \mathbf{X}^{-} \frac{\partial \mathbf{X}}{\partial x_{ij}} \mathbf{X}^{-} \mathbf{X} \quad (67)$$

- s3.3  $f$  が  $\mathbf{X}^{+}$  をひとつだけ含むトレースの場合  
H20.8.1, MN8.5.10

$$G(\mathbf{X}) = \mathbf{X}^{+} \quad (68)$$

$$\begin{aligned} \frac{\partial G(\mathbf{X})}{\partial x_{ij}} &= \frac{\partial \mathbf{X}^{+}}{\partial x_{ij}} = \left( \frac{\partial \mathbf{X}^{+}}{\partial x_{ij}} \right)' \mathbf{X}^{+} \mathbf{X}^{+'} - \mathbf{X}^{+'} \frac{\partial \mathbf{X}^{+}}{\partial x_{ij}} \mathbf{X}^{+'} + \mathbf{X}^{+'} \mathbf{X}^{+} \left( \frac{\partial \mathbf{X}^{+}}{\partial x_{ij}} \right)' \\ &\quad - \mathbf{X} \mathbf{X}^{+} \left( \frac{\partial \mathbf{X}^{+}}{\partial x_{ij}} \right)' \mathbf{X}^{+} \mathbf{X}^{+'} - \mathbf{X}^{+'} \mathbf{X}^{+} \left( \frac{\partial \mathbf{X}^{+}}{\partial x_{ij}} \right)' \mathbf{X}^{+} \mathbf{X} \end{aligned} \quad (69)$$

- s4  $f$  が  $|\mathbf{X}|$  の場合  
H15.8.5, MN8.3.2

$$G(\mathbf{X}) = |\mathbf{X}| \quad (70)$$

$$\frac{\partial G(\mathbf{X})}{\partial x_{ij}} = \frac{\partial |\mathbf{X}|}{\partial x_{ij}} = |\mathbf{X}| \operatorname{tr} \left( \mathbf{X}^{-1} \frac{\partial \mathbf{X}}{\partial x_{ij}} \right) \quad (71)$$

- s5  $f$  が  $X$  の固有値の場合  
H21.15.2, MN8.9.3

$$G(X) = \lambda(X) \quad (72)$$

$$\frac{\partial G(X)}{\partial x_{ij}} = \frac{\partial \lambda(X)}{\partial x_{ij}} = \mathbf{q}' \frac{\partial X}{\partial x_{ij}} \mathbf{p} \quad (73)$$

ただし、 $\mathbf{p}$  は  $X$  の  $\lambda$  に対応する固有ベクトル、また、 $\mathbf{q}$  は  $X'$  の  $\lambda$  に対応する固有ベクトルである。

### 3.1.2 行列関数の $X$ の要素による微分

$F = F(X)$  ならびに  $G = G(X)$  を行列  $X$  に対して行列  $F$  並びに  $G$  を返す関数とする。また、 $A$  と  $B$  を  $X$  とは関係のない行列とする。

行列  $F$  の各要素を  $x_{ij}$  で偏微分し、行列  $F$  の形に並べたものを

$$\frac{\partial F}{\partial x_{ij}} = \left\{ \frac{\partial f_{pq}}{\partial x_{ij}} \right\} \quad (74)$$

とし、行列関数  $F$  の  $X$  の要素による微分と呼ぶ。なお、ベクトル  $\mathbf{x}$  の要素に関する微分に関しては、 $\mathbf{x}$  を列のサイズが 1 の行列として取り扱う。

たとえば、 $F(X) = XA$  の場合、上式は、

$$\frac{\partial F}{\partial x_{ij}} = \frac{\partial XA}{\partial x_{ij}} = \left\{ \frac{\partial [XA]_{pq}}{\partial x_{ij}} \right\} = \left\{ \frac{\partial \sum_k x_{pk} a_{kq}}{\partial x_{ij}} \right\} \quad (75)$$

を表している。

この定義から、 $F$  が  $X$  に無関係の場合には

$$\frac{\partial A}{\partial x_{ij}} = O \quad (76)$$

であり、

$$\frac{\partial F'}{\partial x_{ij}} = \left( \frac{\partial F}{\partial x_{ij}} \right)' \quad (77)$$

である。また、 $F(X) = X$  の場合、

$$\frac{\partial F}{\partial x_{ij}} = \frac{\partial X}{\partial x_{ij}} = \mathbf{u}_i \mathbf{u}'_j \quad (78)$$

となる。ただし、 $\mathbf{u}_i$  は、ゼロベクトルの第  $i$  要素を 1 と置いたのもである。すなわち、(78) 式に示される行列は、 $X$  と同じサイズのゼロ行列の第  $ij$  要素を 1 と置いたものとなっている。また、 $X$  が対称行列の場合は

$$\frac{\partial F}{\partial x_{ij}} = \frac{\partial X}{\partial x_{ij}} = \mathbf{u}_i \mathbf{u}'_j + \mathbf{u}_j \mathbf{u}'_i \quad (79)$$

となる。

行列関数の  $X$  の要素による微分に関しては、加法公式

$$\frac{\partial aF + bG}{\partial x_{ij}} = a \frac{\partial F}{\partial x_{ij}} + b \frac{\partial G}{\partial x_{ij}}, \quad \text{ただし、} a, b \text{ は定数} \quad (80)$$

ならびに、乗法公式

$$\frac{\partial \mathbf{FG}}{\partial x_{ij}} = \mathbf{F} \frac{\partial \mathbf{G}}{\partial x_{ij}} + \frac{\partial \mathbf{F}}{\partial x_{ij}} \mathbf{G} \quad (81)$$

が成り立つ。なぜなら、(80) 式は行列の微分の定義から明らかであり、また (81) 式は

$$\frac{\partial [\mathbf{FG}]_{pq}}{\partial x_{ij}} = \frac{\partial \sum_k f_{pk} g_{kq}}{\partial x_{ij}} \quad (82)$$

$$= \sum_k \frac{\partial f_{pk} g_{kq}}{\partial x_{ij}} = \sum_k \left( f_{pk} \frac{\partial g_{kq}}{\partial x_{ij}} + \frac{\partial f_{pk}}{\partial x_{ij}} g_{kq} \right) \quad (83)$$

$$= \sum_k f_{pk} \frac{\partial g_{kq}}{\partial x_{ij}} + \sum_k \frac{\partial f_{pk}}{\partial x_{ij}} g_{kq} \quad (84)$$

$$= \sum_k f_{pk} \left[ \frac{\partial \mathbf{G}}{\partial x_{ij}} \right]_{kq} + \sum_k \left[ \frac{\partial \mathbf{F}}{\partial x_{ij}} \right]_{pk} g_{kq} \quad (85)$$

$$= \left[ \mathbf{F} \frac{\partial \mathbf{G}}{\partial x_{ij}} \right]_{pq} + \left[ \frac{\partial \mathbf{F}}{\partial x_{ij}} \mathbf{G} \right]_{pq} \quad (86)$$

$$= \left[ \mathbf{F} \frac{\partial \mathbf{G}}{\partial x_{ij}} + \frac{\partial \mathbf{F}}{\partial x_{ij}} \mathbf{G} \right]_{pq} \quad (87)$$

の関係が、 $\frac{\partial \mathbf{FG}}{\partial x_{ij}}$  の第  $pq$  要素が、行列  $\mathbf{F} \frac{\partial \mathbf{G}}{\partial x_{ij}} + \frac{\partial \mathbf{F}}{\partial x_{ij}} \mathbf{G}$  の第  $pq$  要素に対応していることを示しているからである。

また、上式の  $\mathbf{F}$  や  $\mathbf{G}$  のどちらかが  $\mathbf{X}$  を含まない場合には、(76) 式より

$$\frac{\partial \mathbf{FB}}{\partial x_{ij}} = \frac{\partial \mathbf{F}}{\partial x_{ij}} \mathbf{B}, \quad \frac{\partial \mathbf{AF}}{\partial x_{ij}} = \mathbf{A} \frac{\partial \mathbf{F}}{\partial x_{ij}}, \quad \frac{\partial \mathbf{AFB}}{\partial x_{ij}} = \mathbf{A} \frac{\partial \mathbf{F}}{\partial x_{ij}} \mathbf{B} \quad (88)$$

であり、(78) 式を用いれば、

$$\frac{\partial \mathbf{XB}}{\partial x_{ij}} = \frac{\partial \mathbf{X}}{\partial x_{ij}} \mathbf{B} = u_i u_j' \mathbf{B} \quad (89)$$

である。

多くの  $\mathbf{G}$  の場合、(63) 式 - (69) 式を見ても明らかのように、 $\frac{\partial \mathbf{G}}{\partial x_{ij}}$  は  $\frac{\partial \mathbf{X}}{\partial x_{ij}}$  の一次式

$$\frac{\partial \mathbf{G}}{\partial x_{ij}} = \sum_k \left( \mathbf{H}_{k1}(\mathbf{X}) \frac{\partial \mathbf{X}}{\partial x_{ij}} \mathbf{H}_{k2}(\mathbf{X}) + \mathbf{H}_{k3}(\mathbf{X}) \left( \frac{\partial \mathbf{X}}{\partial x_{ij}} \right)' \mathbf{H}_{k4}(\mathbf{X}) \right) \quad (90)$$

の形で表される。ただし、 $\mathbf{H}_{k1}(\mathbf{X})$  等は、 $\mathbf{X}$  の関数行列もしくは単位行列である。

### 3.1.3 $\frac{\partial \text{tr}(\mathbf{G}(\mathbf{X})\mathbf{B})}{\partial \mathbf{X}}$ の導出

トレースは線形演算であるため、

$$\frac{\partial \text{tr}(\mathbf{G})}{\partial x_{ij}} = \frac{\partial \sum_p g_{pp}}{\partial x_{ij}} = \sum_p \frac{\partial g_{pp}}{\partial x_{ij}} \quad (91)$$

$$= \sum_p \left[ \frac{\partial \mathbf{G}}{\partial x_{ij}} \right]_{pp} = \text{tr} \left( \frac{\partial \mathbf{G}}{\partial x_{ij}} \right) \quad (92)$$

となる。したがって、(88) 式より、

$$\frac{\partial \operatorname{tr}(\mathbf{G}(\mathbf{X})\mathbf{B})}{\partial x_{ij}} = \operatorname{tr}\left(\frac{\partial \mathbf{G}(\mathbf{X})\mathbf{B}}{\partial x_{ij}}\right) = \operatorname{tr}\left(\frac{\partial \mathbf{G}(\mathbf{X})}{\partial x_{ij}}\mathbf{B}\right) \quad (93)$$

である。

したがって、(90) 式の形を仮定して (78) 式を用いれば、(93) 式は

$$\begin{aligned} \frac{\partial \operatorname{tr}(\mathbf{G}(\mathbf{X})\mathbf{B})}{\partial x_{ij}} &= \operatorname{tr}\left(\sum_k \left(\mathbf{H}_{k1}(\mathbf{X})\mathbf{u}_i\mathbf{u}'_j\mathbf{H}_{k2}(\mathbf{X}) + \mathbf{H}_{k3}(\mathbf{X})(\mathbf{u}_i\mathbf{u}'_j)'\mathbf{H}_{k4}(\mathbf{X})\right)\mathbf{B}\right) \quad (94) \\ &= \operatorname{tr}\left(\mathbf{u}'_j\sum_k(\mathbf{H}_{k2}(\mathbf{X})\mathbf{B}\mathbf{H}_{k1}(\mathbf{X}))\mathbf{u}_i + \mathbf{u}'_i\sum_k(\mathbf{H}_{k4}(\mathbf{X})\mathbf{B}\mathbf{H}_{k3}(\mathbf{X}))\mathbf{u}_j\right) \\ &= \mathbf{u}'_j\sum_k(\mathbf{H}_{k2}(\mathbf{X})\mathbf{B}\mathbf{H}_{k1}(\mathbf{X}))\mathbf{u}_i + \mathbf{u}'_i\sum_k(\mathbf{H}_{k4}(\mathbf{X})\mathbf{B}\mathbf{H}_{k3}(\mathbf{X}))\mathbf{u}_j \quad (95) \end{aligned}$$

となる。しかし、 $\frac{\partial \operatorname{tr}(\mathbf{G}(\mathbf{X})\mathbf{B})}{\partial x_{ij}}$  は行列  $\frac{\partial \operatorname{tr}(\mathbf{G}(\mathbf{X})\mathbf{B})}{\partial \mathbf{X}}$  の第  $ij$  要素であり、また、 $\mathbf{u}'_i\mathbf{A}\mathbf{u}_j$  が行列  $\mathbf{A}$  の第  $ij$  要素を、 $\mathbf{u}'_j\mathbf{A}\mathbf{u}_i$  が行列  $\mathbf{A}$  の第  $ji$  要素を抜き出すことから、上式は

$$\frac{\partial \operatorname{tr}(\mathbf{G}(\mathbf{X})\mathbf{B})}{\partial \mathbf{X}} = \left(\sum_k(\mathbf{H}_{k2}(\mathbf{X})\mathbf{B}\mathbf{H}_{k1}(\mathbf{X}))\right)' + \sum_k(\mathbf{H}_{k4}(\mathbf{X})\mathbf{B}\mathbf{H}_{k3}(\mathbf{X})) \quad (96)$$

を示している。

したがって、 $\frac{\partial \mathbf{G}(\mathbf{X})}{\partial x_{ij}}$  が (90) 式という形で与えられていれば、それを用いて、 $\frac{\partial \operatorname{tr}(\mathbf{G}(\mathbf{X})\mathbf{B})}{\partial \mathbf{X}}$  を求めることが出来る。

たとえば、 $\mathbf{G}(\mathbf{X}) = \mathbf{X}$  の場合、(78) 式より、

$$\frac{\partial \operatorname{tr}(\mathbf{G}(\mathbf{X})\mathbf{B})}{\partial x_{ij}} = \frac{\partial \operatorname{tr}(\mathbf{X}\mathbf{B})}{\partial x_{ij}} = \operatorname{tr}\left(\frac{\partial \mathbf{X}}{\partial x_{ij}}\mathbf{B}\right) = \operatorname{tr}(\mathbf{u}_i\mathbf{u}'_j\mathbf{B}) = \mathbf{u}'_j\mathbf{B}\mathbf{u}_i \quad (97)$$

となるから (同じ形が (89) 式にも示されている。)

$$\frac{\partial \operatorname{tr}(\mathbf{X}\mathbf{B})}{\partial \mathbf{X}} = \mathbf{B}' \quad (98)$$

となる。また、 $\frac{\partial \operatorname{tr}(\mathbf{X}^{-1}\mathbf{B})}{\partial \mathbf{X}}$  を求める場合は、(63) 式より、

$$\frac{\partial \operatorname{tr}(\mathbf{X}^{-1}\mathbf{B})}{\partial x_{ij}} = \operatorname{tr}\left(\frac{\partial \mathbf{X}^{-1}}{\partial x_{ij}}\mathbf{B}\right) = \operatorname{tr}\left(\left(-\mathbf{X}^{-1}\frac{\partial \mathbf{X}}{\partial x_{ij}}\mathbf{X}^{-1}\right)\mathbf{B}\right) \quad (99)$$

$$= \operatorname{tr}\left(\left(-\mathbf{X}^{-1}\mathbf{u}_i\mathbf{u}'_j\mathbf{X}^{-1}\right)\mathbf{B}\right) \quad (100)$$

$$= -\mathbf{u}'_j\mathbf{X}^{-1}\mathbf{B}\mathbf{X}^{-1}\mathbf{u}_i \quad (101)$$

となるため、

$$\frac{\partial \operatorname{tr}(\mathbf{X}^{-1}\mathbf{B})}{\partial \mathbf{X}} = -\left(\mathbf{X}^{-1}\mathbf{B}\mathbf{X}^{-1}\right)' \quad (102)$$

を得る。(65) 式や (69) 式に関しても同様である。

また、 $\mathbf{G}$  が、行列式や固有値のように、スカラーを返す関数  $g$  の場合は、 $\mathbf{B} = 1$  と置いて、

$$\frac{\partial g(\mathbf{X})}{\partial x_{ij}} = \frac{\partial \operatorname{tr}(g(\mathbf{X}))}{\partial x_{ij}} \quad (103)$$

となるため、(96) 式をそのまま用いることが出来る。ただし、この場合、 $\mathbf{H}_{k1}, \mathbf{H}_{k,3}$  は行ベクトル、 $\mathbf{H}_{k2}, \mathbf{H}_{k,4}$  列ベクトルとなる。

### 3.1.4 その他の特定公式

- $\frac{\partial \text{tr}((\mathbf{X} \odot \mathbf{A})\mathbf{B})}{\partial \mathbf{X}}$  の導出

ここでは (96) 式 を用いて  $G(\mathbf{X}) = \mathbf{X} \odot \mathbf{A}$  の場合の特定公式を導く。

Hadamard 積の  $x_{ij}$  による偏微分は

$$\frac{\partial (\mathbf{X} \odot \mathbf{A})}{\partial x_{ij}} = (\mathbf{u}_i \mathbf{u}'_j) \odot \mathbf{A} \quad (104)$$

である。したがって、乗法公式 (88) 式 より

$$\frac{\partial (\mathbf{X} \odot \mathbf{A})\mathbf{B}}{\partial x_{ij}} = ((\mathbf{u}_i \mathbf{u}'_j) \odot \mathbf{A})\mathbf{B} \quad (105)$$

となる。しかし、

$$\text{tr}(\mathbf{A}'(\mathbf{B} \odot \mathbf{C})) = \text{tr}((\mathbf{A}' \odot \mathbf{B}')\mathbf{C}) \quad (106)$$

の関係が知られている。(Magnus & Neudecker, 1988, Theorem 3.7) したがって、

$$\frac{\partial \text{tr}((\mathbf{X} \odot \mathbf{A})\mathbf{B})}{\partial x_{ij}} = \text{tr}(((\mathbf{u}_i \mathbf{u}'_j) \odot \mathbf{A})\mathbf{B}) \quad (107)$$

$$= \text{tr}(\mathbf{B}'((\mathbf{u}_i \mathbf{u}'_j) \odot \mathbf{A}')) = \text{tr}(\mathbf{B}'(\mathbf{A}' \odot (\mathbf{u}_i \mathbf{u}'_j)')) \quad (108)$$

$$= \text{tr}((\mathbf{B}' \odot \mathbf{A})(\mathbf{u}_i \mathbf{u}'_j)') = \mathbf{u}'_i(\mathbf{B}' \odot \mathbf{A})\mathbf{u}_j \quad (109)$$

$$= \mathbf{u}'_i(\mathbf{A} \odot \mathbf{B}')\mathbf{u}_j \quad (110)$$

したがって、

$$\frac{\partial \text{tr}((\mathbf{X} \odot \mathbf{A})\mathbf{B})}{\partial \mathbf{X}} = \mathbf{A} \odot \mathbf{B}' \quad (111)$$

である。特に、 $\mathbf{B} = \mathbf{I}$  の場合は

$$\frac{\partial \text{tr}((\mathbf{X} \odot \mathbf{A})\mathbf{B})}{\partial \mathbf{X}} = \text{diag}(\mathbf{A}) \quad (112)$$

であり、 $\mathbf{A} = \mathbf{I}$  の場合は

$$\frac{\partial \text{tr}((\mathbf{X} \odot \mathbf{I})\mathbf{B})}{\partial \mathbf{X}} = \frac{\partial \text{tr}(\text{diag}(\mathbf{X})\mathbf{B})}{\partial \mathbf{X}} = \text{diag}(\mathbf{B}) \quad (113)$$

である。

- $\frac{\partial \text{tr}(\mathbf{X}^{(p)}\mathbf{B})}{\partial \mathbf{X}}$  の導出

ここでは (96) 式 を用いて  $G(\mathbf{X}) = \mathbf{X}^{(p)}$  の場合の特定公式を導く。

この場合、

$$\frac{\partial \mathbf{X}^{(p)}}{\partial x_{ij}} = p(\mathbf{u}_i \mathbf{u}'_j) \odot \mathbf{X}^{(p-1)} \quad (114)$$

であるから、Hadamard 積の場合と同様に

$$\frac{\partial \text{tr}(\mathbf{X}^{(p)}\mathbf{B})}{\partial \mathbf{X}} = p \mathbf{X}^{(p-1)} \odot \mathbf{B}' \quad (115)$$

を得る。

- $\frac{\partial \text{tr}(\text{diag}(\mathbf{X})\mathbf{B})}{\partial \mathbf{X}}$  の導出

ここでは  $\mathbf{G}(\mathbf{X}) = \text{diag}(\mathbf{X})$  の場合の特定公式を導く。

この場合、

$$\frac{\partial \mathbf{G}}{\partial x_{ij}} = \frac{\partial \text{diag}(\mathbf{X})}{\partial x_{ij}} = \text{diag}(\mathbf{u}_i \mathbf{u}'_j) \quad (116)$$

であるから、(93) 式より

$$\frac{\partial \text{tr}(\text{diag}(\mathbf{X})\mathbf{B})}{\partial x_{ij}} = \text{tr}(\text{diag}(\mathbf{u}_i \mathbf{u}'_j)\mathbf{B}) = \begin{cases} b_{ii}, & i = j \\ 0, & i \neq j \end{cases} \quad (117)$$

となる。もしくは、トレースを書き下すと

$$\text{tr}(\text{diag}(\mathbf{X})\mathbf{B}) = \sum_i x_{ii} b_{ii} \quad (118)$$

となるため、

$$\frac{\partial \text{tr}(\text{diag}(\mathbf{X})\mathbf{B})}{\partial x_{ij}} = \begin{cases} b_{ii}, & i = j \\ 0, & i \neq j \end{cases} \quad (119)$$

となるが、これから

$$\frac{\partial \text{tr}(\text{diag}(\mathbf{X})\mathbf{B})}{\partial \mathbf{X}} = \text{diag}(\mathbf{B}) \quad (120)$$

となる。

しかし、本稿ではこの形はと特定公式としては取り扱わずに、(111) 式を代用した。

- Kronecker 積に関して

$\mathbf{G}(\mathbf{X}) = \mathbf{X} \otimes \mathbf{A}$  と置いた場合、

$$\frac{\partial \mathbf{X} \otimes \mathbf{A}}{\partial x_{ij}} = (\mathbf{u}_i \mathbf{u}'_j) \otimes \mathbf{A} \quad (121)$$

であり、

$$\frac{\partial \text{tr}(\mathbf{X} \otimes \mathbf{A})\mathbf{B}}{\partial x_{ij}} = ((\mathbf{u}_i \mathbf{u}'_j) \otimes \mathbf{A})\mathbf{B} \quad (122)$$

となるが、この形は (90) 式の形ではないため、 $\frac{\partial \text{tr}((\mathbf{X} \otimes \mathbf{A})\mathbf{B})}{\partial \mathbf{X}}$  を導くことは出来ない。ただ、 $\mathbf{X}$  と  $\mathbf{A}$  が共に正方行列の場合は、

$$\text{tr}(\mathbf{A} \otimes \mathbf{B}) = \text{tr}(\mathbf{A})\text{tr}(\mathbf{B}), \quad |\mathbf{AB}| = |\mathbf{A}|^p |\mathbf{B}|^q \quad (123)$$

ただし、 $p, q$  は  $\mathbf{A}, \mathbf{B}$  のサイズ、という関係があるため、積の微分の一般公式 (25) 式を用いれば  $\frac{\partial \text{tr}((\mathbf{X} \otimes \mathbf{A})\mathbf{B})}{\partial \mathbf{X}}$  や  $\frac{\partial |\mathbf{X} \otimes \mathbf{A}|}{\partial \mathbf{X}}$  を求めることが出来る。

\*\*\*\*\*

怪しい。 $((\mathbf{u}_i \mathbf{u}'_j) \otimes \mathbf{A})\mathbf{B}$  が何かの  $ij$  要素であるかが分かればよい。

\*\*\*\*\*

### 3.2 連鎖律 (55) 式ならびに乗法公式 (56) 式の導出

連鎖律 (55) 式は、以下に示す通常ののスカラーによる微分の連鎖律のうち、 $\frac{\partial f(\mathbf{F})}{\partial u_{k,l}}$  の部分がもはや  $\mathbf{X}$  の関数と見なされていないこと、ならびに、それを行列の微分の定義を用いて  $\frac{\partial f(\mathbf{F})}{\partial \mathbf{F}}$  の  $k, l$  要素と見なし、対応する要素の積和をトレースを用いて書き直したものである。

$$\frac{\partial f(\mathbf{G}(\mathbf{X}))}{\partial x_{ij}} = \sum_{k,l} \frac{\partial f(\mathbf{F})}{\partial u_{k,l}} \frac{\partial u_{k,l}}{\partial x_{ij}} \quad (124)$$

$$= \frac{\partial \sum_{k,l} \left( \frac{\partial f(\mathbf{F})}{\partial u_{k,l}} \right)_c u_{k,l}}{\partial x_{ij}} \quad (125)$$

$$= \frac{\partial \operatorname{tr} \left( \left[ \frac{\partial f(\mathbf{F})}{\partial \mathbf{F}} \right]'_c \mathbf{F} \right)}{\partial x_{ij}} \quad (126)$$

なお、この導出から、たとえ  $\mathbf{F}$  行列が対称であっても、 $\frac{\partial f(\mathbf{F})}{\partial \mathbf{F}}$  の評価に際してその事実を用いなくても良いことが解る。

乗法公式 (56) 式の導出は以下のように行う。スカラー値を持つ関数  $\operatorname{tr}(\mathbf{F}\mathbf{G})$  の  $x_{ij}$  に関する微分は、(92) 式と (81) 式より、

$$\frac{\partial \operatorname{tr}(\mathbf{F}\mathbf{G})}{\partial x_{ij}} = \operatorname{tr} \left( \frac{\partial \mathbf{F}\mathbf{G}}{\partial x_{ij}} \right) \quad (127)$$

$$= \operatorname{tr} \left( \mathbf{F} \frac{\partial \mathbf{G}}{\partial x_{ij}} + \frac{\partial \mathbf{F}}{\partial x_{ij}} \mathbf{G} \right) = \operatorname{tr} \left( \mathbf{F} \frac{\partial \mathbf{G}}{\partial x_{ij}} \right) + \operatorname{tr} \left( \frac{\partial \mathbf{F}}{\partial x_{ij}} \mathbf{G} \right) \quad (128)$$

$$= \frac{\partial \operatorname{tr}(\mathbf{F}_c \mathbf{G})}{\partial x_{ij}} + \frac{\partial \operatorname{tr}(\mathbf{F}\mathbf{G}_c)}{\partial x_{ij}} \quad (129)$$

となる。なお、上記の導出に際しては、(93) 式より導かれる

$$\frac{\partial \operatorname{tr}(\mathbf{G}(\mathbf{X})\mathbf{B})}{\partial x_{ij}} = \operatorname{tr} \left( \frac{\partial \mathbf{G}(\mathbf{X})}{\partial x_{ij}} \mathbf{B} \right) = \operatorname{tr} \left( \mathbf{B} \frac{\partial \mathbf{G}(\mathbf{X})}{\partial x_{ij}} \right) \quad (130)$$

の関係を利用した。(56) 式は、上式の  $\mathbf{F}$  を  $\mathbf{F}$ ,  $\mathbf{G}$  を  $\mathbf{G}$  と置いたものとなっている。

Hadamard 積の乗法公式 (57) 式に関しては、その要素の偏微分が<sup>\*</sup>

$$\left[ \frac{\partial \mathbf{F} \odot \mathbf{G}}{\partial x_{ij}} \right]_{pq} = \frac{\partial f_{pq} g_{pq}}{\partial x_{ij}} = f_{pq} \frac{\partial g_{pq}}{\partial x_{ij}} + \frac{\partial f_{pq}}{\partial x_{ij}} g_{pq} \quad (131)$$

したがって、

$$\frac{\partial \mathbf{F} \odot \mathbf{G}}{\partial x_{ij}} = \mathbf{F} \odot \frac{\partial \mathbf{G}}{\partial x_{ij}} + \frac{\partial \mathbf{F}}{\partial x_{ij}} \odot \mathbf{G} \quad (132)$$

となるため、(56) 式の導出と同様の議論により

$$\operatorname{tr} \left( \frac{\partial \mathbf{F} \odot \mathbf{G}}{\partial x_{ij}} \right) = \operatorname{tr} \left( \mathbf{F} \odot \frac{\partial \mathbf{G}}{\partial x_{ij}} \right) + \operatorname{tr} \left( \frac{\partial \mathbf{F}}{\partial x_{ij}} \odot \mathbf{G} \right) \quad (133)$$

$$= \frac{\partial \operatorname{tr}(\mathbf{F}_c \odot \mathbf{G})}{\partial x_{ij}} + \frac{\partial \operatorname{tr}(\mathbf{F} \odot \mathbf{G}_c)}{\partial x_{ij}} \quad (134)$$

となる。

## 4 Mathematica への実装

### 4.1 関数 mD の作成

本稿で導いた公式を計算機に実装する際に、もっとも簡単な方法は、それらの公式を独立した関数として定義していく方法であるが、このようなプログラムの作成には polymorphic な関数の利用が便利である。Polymorphic な関数とは、その引数のパターンによって関数の動作が異なる関数のことをいう。Polymorphic な関数定義を行えば、たとえば、 $f(a+b)$  と  $f(a b)$  の二つの関数を定義し、 $f(a+b)$  の場合には  $a+b$  の平方根を返し  $f(a b)$  の場合には  $a b$  の2乗を返す、という関数を定義することが可能となる。これは、Mathematica においては

```
Clear[f];
f[a_+b_]:=Sqrt[a+b];
f[a_ b_]:= (a b)^2;
```

と書くことができる。

スカラー関数  $\exp$  のベクトルもしくは行列  $X$  による微分の公式をプログラムする際には、まず、メインとなる関数を  $mD[\exp,X]$  とし、それをおよそ以下のように定義する。

```
mD[exp_,X_]:=Module[{B1,B2,B3},
(* Simplify the input. *)
B1=matSimplify[exp];
(* Differentiate the input. *)
B2=mD0[B1,X];
(* Simplify the result. *)
B3=matSimplify[B2];
(* Output the result. *)
B3
]; (* end of mD *)
```

この関数は、まず与えられた式  $\exp$  を関数  $matSimplify$  を用いて単純化し、その後、 $mD0[\exp,X]$  という形で個別公式を呼び出し、その結果をまた単純化して出力するという形式を持っている。なお、Mathematica の関数は、 $;$  が付いていない式の値を戻り値として返すため、 $B3$  が出力されることになる。また、 $(* comment *)$  はコメントであり、 $Module[]$  は、関数の定義の範囲と、その中で使用されるローカルな変数の宣言を行うものである。なお、 $matSimplify$  は、たとえば、 $(A')'$  や  $AI$  等を単純化するための自作の関数であるが、複数の項のトレースを複数のトレースに変換する操作等も行うものである。

なお、本節に示す関数は、その引数が行列やベクトルであることを認識していない。したがって、実行に際しては、 $mD[\exp,X]$  への引数のうち、少なくとも  $X$  が、実行時には未定義であることが必要があり、そのためには、 $Clear$  関数を用いて  $X$  を未定義にすればよい。また、 $\exp$  に含まれる行列（たとえば  $A, B, C$  等）が、実行時に未定義であれば、結果は  $A, B, C$  等を用いた形で表示され、すでにその要素が定義されていれば、結果はその要素を用いて表示される。（本節後半の Mathematica による結果のチェックの項を参照のこと。）

前節までに示した個別の公式は  $mD0[\exp,X]$  という形で定義したが、これが polymorphic な関数となっており、与えられた引数  $\exp$  の形により、様々な公式に対応するようになっている。

たとえば、特定公式の一つ  $S2$  (35) 式はおよそ以下のような形、

```
mD0[Tr[Dot[X_,A_:Identity]],X_]:=Transpose[A] /; FreeQ[A,X]
```

をしているが、これは  $Tr[X.A]$  が引数として呼ばれ、 $A$  が  $X$  を持たないときに  $Transpose[A]$  を返す関数である。なお、Mathematica において、行列のトレースは  $Tr$  という関数で、転

置行列は Transpose という関数で表現される。また、行列の掛け算は A.B もしくは Dot[A,B] と書かれる。なお、関数の引数の中にある :Identity は、行列 A のデフォルトの値であり、この関数が mD0[Tr[X],X] という形で呼ばれたときに対応している。そして、関数の最後にある /; FreeQ[A,X] は、行列 A が X を含んでいない場合にのみ関数が戻り値 Transpose[A] を返すことを意味している。もしもこの条件が満たされない場合、この関数は、自分自身 mD0[Tr[Dot[X,A]],X] を戻り値として返すことになるが、たとえば、mD0[Tr[X.Transpose[X],X] のような場合である。この場合、Mathematica は引数のパターンが当てはまる関数 mD0 を探し、見つければそれを実行することになる。(この場合は、トレースの乗法公式が適用されることになる。)

変形公式の一つ G2 (45) 式は

```
mD0[Tr[Dot[A_,X_,B_:Identity]],X_]:=mD0[Tr[X.B.A],X] /; FreeQ[A,X]
```

という形をしているが、これは A, B が、共に X を含まない場合に mD0[Tr[X.B.A],X] を返す関数である。関数 mD0 が、引数の形を変え、再帰的に呼ばれていることに注意されたい。なお、再帰的定義の終了条件は、特定公式を与える関数となっている。

最後に、トレースの乗法公式 P1 (56) 式は

```
mD0[Tr[Dot[A_,B_]],X_]:=Module[{Q,Q9,Q8},
  If[!FreeQ[A,X] && !FreeQ[B,X],
    Q=mD0[Tr[Dot[Q9,B]],X] + mD0[Tr[Dot[A,Q8]],X];
  ];
  Q=Q /. {Q9 -> A, Q8 -> B };
  Q /; !FreeQ[A,X] && !FreeQ[B,X]
];
```

という形をしている。これは、A, B が共に X を含む場合に、A を Q9 で、B を Q8 で置き換えて X を含まない形にした後に、mD0[Tr[Dot[Q9,B]],X] と mD0[Tr[Dot[A,Q8]],X] を実行し、その結果に含まれる Q9 を A に、Q8 を B に置き換えたものを返す関数である。なお、Q -> {rules} という表現は Q に {rules} という変換規則を適用することを表しているが、変換規則は、lhs を rhs で置き換えることを表す lhs :> rhs という形をコンマで繋いだものとなっている。また、If[condition,exp1,exp2] は、condition が真の場合に exp1 が、偽の場合に exp2 が実行される関数であり、偽の部分は省略することが可能である。

このように、スカラー関数のベクトル・行列による微分の公式を Mathematica でプログラムすることは、基本的には、それぞれの公式を対応する Mathematica の関数に置き換えるだけでよく、簡単である。なお、polymorphic な関数が複数ある場合、その適用順序は、関数の定義された順であるが、第 2 節の順で全ての関数を定義していけばよい。気をつけなくてはならないのは連鎖律をなるべく後で定義をすることで、それらが、必要もないのに実行されることを防ぐことである。

たとえば、上記の順に関数を 4 つ定義し、mD[Tr[X.A.X], X] を実行してみると、以下のようステップを踏み Transpose[A.X] + Transpose[X.A] が結果として帰ってくる。すなわち、 $\frac{\partial \text{tr}(XAX)}{\partial X} = (AX)' + (XA)'$  である。

1. 与えられた式の形は、特定公式 S2 にも変形公式 G2 にも当てはまらないので乗法公式 P1 が適用され mD0x[Tr[X.Q8], X] + mD0x[Tr[Q9.A.X], X] が得られる。
2. 乗法公式の結果の第 2 項 mD0x[Tr[Q9.A.X], X] の部分に変形公式 G2 が適用され mD0x[Tr[X.Identity.Q9.A], X] を得る。
3. mD0x[Tr[X.Identity.Q9.A], X] が特定公式 S2 により Transpose[Identity.Q9.A] となる。

4. 乗法公式の結果の第 1 項  $mD0x[\text{Tr}[X.Q8], X]$  の部分が  
特定公式 S2 により  $\text{Transpose}[Q8]$  となる。
5. 乗法公式の結果が  $\text{Transpose}[Q8] + \text{Transpose}[\text{Identity}.Q9.A]$  となる。
6. 乗法公式の結果の中の Q8, Q9 がそれぞれ  $AX$  と  $X$  に置き換えられ、  
 $\text{Transpose}[A.X] + \text{Transpose}[\text{Identity}.X.A]$  を得る。
7. この結果を単純化して最終的に  $\text{Transpose}[A.X] + \text{Transpose}[X.A]$  を得る。

なお、実際のプログラミングに当たっては、`prinrules` という変数の値が 1 の場合には使用された公式名を、また、その値が 2 の場合には各公式の適用結果を出力するようになってい  
る。また、Mathematica の出力は、通常の出力 (Standard Form) に関数 `TraditionalForm` を  
作用させることにより、たとえば  $\text{Inverse}[X].A.\text{Transpose}[X]$  は  $X^{-1}.A.X^T$  と出力させること  
が可能である。そのためには `mD[exp,var] // TraditionalForm` と入力すればよい。

図 1 と図 2 に、関数 `mD` を用いて重み付き最小 2 乗基準の回帰係数行列に関する偏微分

$$\frac{\partial \text{tr}((Y - XB)'W(Y - XB)G)}{\partial B} \quad (135)$$

と、因子分析の尤度に基づく基準関数の因子パターン行列に関する偏微分

$$\frac{\partial \text{tr}S(\Lambda\Phi\Lambda' + \Psi)^{-1} - \log |S(\Lambda\Phi\Lambda' + \Psi)^{-1}|}{\partial \Lambda} \quad (136)$$

を示した。最小 2 乗基準の方は 15 ステップで、最尤解の方は 22 ステップで解を得ている。  
図中、Rules Used: として列挙されているものが利用された公式であり、そこに表示されてい  
る括弧の中の数字の直後の英記号が本文中の公式に付与された英記号とほぼ対応している。ま  
た、後半部分に、括弧の中の数字とともに示されているのは、各公式の適用結果である。

なお、いずれの解も、式中に現れる行列の対称性を考慮すれば、よく知られている形に変形  
することが可能である。

## 4.2 結果のチェック

次に、公式の正当性のチェックを Mathematica を用いて行う方法について簡単に触れる。  
Mathematica の偏微分を与える関数 `D[exp,var]` は、`exp` に行列  $X$  の関数として定義されるス  
カラーを、また `var` に行列  $X$  に第  $ij$  要素を与えると、 $\frac{\partial \text{exp}}{\partial x_{ij}}$  を返してくるが、これは  $\frac{\partial \text{exp}}{\partial X}$   
の第  $ij$  要素に他ならない。しかも、`exp` の定義に用いられている行列が、数値ではなく記号  
的に定義されていれば、返ってくる結果もこれらの記号を用いた代数演算の結果となる。この  
ことを利用して、関数 `mD` のチェックをすることが出来る。

たとえば、`mD[Tr[Inverse[Transpose[X].B.X].C], X]` は、結果として

$$\frac{\partial \text{tr}((X'BX)^{-1}C)}{\partial X} = -BX(X'BX)^{-1}C(X'BX)^{-1} - B'X(X'BX)^{-1'}C'(X'BX)^{-1'} \quad (137)$$

を与えるが、この正当性を  $2 \times 2$  の行列を用いてチェックするためには以下のような Math-  
ematica のプログラムを作成すればよい。

```
(* 変数のクリア *)
Clear[X,B,x,b,c];
```

```

(* 未定義な変数を用いて微分されるべき関数を定義する。 *)
exp = Tr[Inverse[Transpose[X].B.X].C];

(* 要素のシンボリックな定義 (要素そのものは未定義) *)
X1={{Subscript[x,1,1],Subscript[x,1,2]},{Subscript[x,2,1],Subscript[x,2,2]}};
B1={{Subscript[b,1,1],Subscript[b,1,2]},{Subscript[b,2,1],Subscript[b,2,2]}};
C1={{Subscript[c,1,1],Subscript[c,1,2]},{Subscript[c,2,1],Subscript[c,2,2]}};

(* 変換規則 *)
rule={X:>X1, B:>B1, C:>C1};

(* 未定義な X, B, C を用いて mD を実行しその結果を mDres とする。 *)
mDres=mD[exp,X];

(* mDres に現れる行列 B,C,X を B1, C1, X1 で置き換える。 *)
mDres1=mDres /. rule;

(* exp に含まれる行列 B,C,X を B1, C1, X1 で置き換える。 *)
exp1=exp /. rule;

(* 要素が定義された exp1 を関数 D を用いて x_11 で偏微分する。 *)
Dres11=D[exp1,Subscript[x,1,1]];

(* mDres1 の第 1,1 要素と Dres11 とを比較する。 *)
Simplify[Dres11] === Simplify[mDres1[[1, 1]]]

```

なお、Mathematica において、行列はリストとして表現され、要素を中括弧で囲んだ形で定義される。また、Subscript[a,i,j] とは変数 a に添字 i,j をつけるための関数、mDres1[[1,1]] とは、行列 mDres1 の第 1,1 要素を示すものであり、lhs === rhs という表現は二つが数学的に等しければ True を返すものである。したがって、この結果が True を返してくれば、関数 mD は、少なくともその 11 要素に関しては正しい結果を返してきたことになる。上記の手続きを全ての X の要素に関して繰り返せばよい。途中結果を確認したい場合は、プログラム中のセミコロンを適宜はずすと、その式の結果が出力される。

なお、上記の例では、行列の固有値、固有ベクトル、Hadamard 積、要素の冪乗、一般逆行列が出てきていないが、固有値や固有ベクトルを含む場合には、それを計算し、その中の一つを選択した上で mDres に現れる固有値・固有ベクトルの表現  $p, q, \lambda$  を置き換える必要がある。また、Hadamard 積は Mathematica では、通常の積を表す関数 Times に置き換え、要素の冪乗は Power に置き換える必要が生じる。また、Moor-Penrose 逆行列は、mD の中では Ginv と表現されるが、それは PseudInverse に置き換えなければならない。また、一般逆行列は、mD の中では ginv と表現されるが、Mathematica でそれを求めるためには、たとえばスweep演算子等を用いて自前の関数を作成する必要が生じる。また、diag に関しても自作する必要がある。

最後に、蛇足ではあるが、Mathematica の偏微分を与える関数 D[exp,var] は、exp に行列を与えると、その各要素をスカラーである var で微分した形、すなわち、第 2 節 に示した、行列関数の X の要素に関する微分を行うことも出来る。したがって、第 2 節 に示した行列関数の X の要素に関する微分の正当性を確かめることも出来る。ただし、当然ではあるが、Mathematica は微分結果を要素で表現するため、それを行列の表現に並べ替えなければならない。

## 5 結語

本稿では、スカラー値を返すベクトルや行列の関数をベクトルや行列で微分するための公式を導き、それを Mathematica のプログラムとして実装した。筆者は、1970 年代に Schönemann (1965) を手に入れて以来、機会があれば行列による微分を用いているが、本稿の出版が、多くの読者にその機会を提供することを望んでいる。なお、行列のトレースの微分は、一見、行列そのものの微分より一般性に欠けると考えられるが、

$$\frac{\partial \operatorname{tr}((\mathbf{Y} - \mathbf{G}(\mathbf{X}))' \mathbf{W}(\mathbf{Y} - \mathbf{G}(\mathbf{X})) \mathbf{G})}{\partial \mathbf{X}} \quad (138)$$

を用いることにより、 $\mathbf{Y}$  行列に  $\mathbf{G}(\mathbf{X})$  行列を当てはめる重み付き最小 2 乗基準が表現できる。また、

$$\frac{\partial \operatorname{tr}(\mathbf{G}(\mathbf{X}) \mathbf{u}_q \mathbf{u}_p')}{\partial \mathbf{X}} = \frac{\partial \mathbf{u}_p' \mathbf{G}(\mathbf{X}) \mathbf{u}_q}{\partial \mathbf{X}} = \frac{\partial \operatorname{tr}([\mathbf{G}(\mathbf{X})]_{pq})}{\partial \mathbf{X}} \quad (139)$$

を求めることにより、行列  $\mathbf{G}(\mathbf{X})$  の第  $pq$  要素の行列による微分を、

$$\frac{\partial \operatorname{tr}(\mathbf{G}(\mathbf{X}) \mathbf{1} \mathbf{1}')}{\partial \mathbf{X}} = \frac{\partial \mathbf{1}' \mathbf{G}(\mathbf{X}) \mathbf{1}}{\partial \mathbf{X}} \quad (140)$$

により、行列  $\mathbf{G}(\mathbf{X})$  の要素の総和の微分を求めることが出来る。

スカラー関数の行列による微分は、いったんその規則を公式という形で決めてしまえば、後は単なる規則の機械的な適用でその結果を出すことが出来る。ただ、関数が少し複雑になると、適用する規則の数が多くなり、紙と鉛筆でその導出を行う際に間違いを犯す可能性が高くなる。その点、今回開発したプログラムが役に立つことを期待している。

今回のプログラムの作成に当たっては、どのような関数を与えられ場合にも解が求まる最小の公式セットを実装することをその基本方針とした。したがって、テキストによく現れる

$$\frac{\partial \operatorname{tr}(\mathbf{X}' \mathbf{X})}{\partial \mathbf{X}} = 2\mathbf{X}, \quad \frac{\partial |\mathbf{A} \mathbf{X} \mathbf{A}|}{\partial \mathbf{X}} = |\mathbf{A} \mathbf{X} \mathbf{A}| \mathbf{A}' \left( (\mathbf{A} \mathbf{X} \mathbf{B})^{-1} \right)' \mathbf{B}' \quad (141)$$

等は、特定公式として採用しなかった。この基本方針のもとでは、上記の二つの微分は、トレースの乗法公式や連鎖律を用いて導かれるべきものとなる。この基本方針の利点は、プログラムステップ数が減ることはもちろんであるが、それと同時に、何がもっとも基本的な公式であるかを再確認できたことである。なお、計算時間のことを考えれば、出来るだけ多くの公式を特定公式として、すなわち、関数の再呼び出しの終了条件として、準備しておくことが望ましいが、その数にも限界がある。しかも、この方向突き詰めていけば、結局、全ての考えられる場合を特定公式として準備するという芸のないプログラムができあがってしまうことになる。ただ、紙と鉛筆を使って微分を求める場合には、ある程度の関数の微分が分かっていることが望ましいため、付録として、行列  $\mathbf{X}$  が二つまで含まれる場合のいくつかの関数に関して、その微分をまとめて掲載した。全ての結果は、関数  $mD$  を用いて導出し、前節で述べた方法で、 $2 \times 2$  の行列を用いてチェック済みである。計算上のスピードを重視するなら、これらの結果を特定公式として用いることも考えられる。

最後に今後の課題について言及する。本稿であつかったのは、スカラー関数の微分のみであった。これは、最小 2 乗基準や尤度関数の極値を求める際に必要であるが、推定されたパラメタの標準誤差を評価する際には、2 次微分が必要となる。これは、行列の行列による微分となり、本稿で扱った公式の範囲外のものである。Magnus & Neudecker (1988) によれば、行列の行列による微分は、それぞれを  $vec$  関数でベクトルの形に変形し、結果を行列の形で示すという方針が良いとされているが、今後、機会があれば、それを実現するプログラムを開発していきたいと思っている。

## 参考文献

- Harville, D.A. (1997) *Matrix Algebra from a Statistician's Perspective* Springer-Verlag New York, Inc.
- Goodnight, J.H. (1979) Tutorial on the SWEEP operator. *The American Statistician*, 33, 3, 149-158.
- Magnus, J.R. & Neudecker, H. (1988) *Matrix Differential Calculus with Applications in Statistics and Econometrics* John Wiley & Sons Ltd.
- Nel, D. G. (1980) On Matrix Differentiation in Statistics *South African Journal* 14, 137-193
- Schönemann, P.H. (1965) On The Formal Differentiation of Traces and Determinants. *Research Memorandum* 27, 1-45 The Psychometric Lab, University of North Carolina, Chapel Hill.

## 付録 いくつかの結果

以下に、無制約な  $\mathbf{X}$  行列が 2 個まで含まれる場合の例を示す。トレースの演算規則を用いて生成される同値な結果のうち、 $\mathbf{X}$  や  $\mathbf{X}'$  の出現位置が質的に異なるもののみを列挙した。ここに示した例よりも単純なケースは、 $\mathbf{A}, \mathbf{B}, \mathbf{C}$  等を単位行列とおくことによって得られる。なお、 $\mathbf{X}$  が対称行列の場合には、以下の結果に (33) 式を適用すればよい。

$$\begin{aligned} \frac{\partial \operatorname{tr}(\mathbf{AXB})}{\partial \mathbf{X}} &= \frac{\partial \operatorname{tr}(\mathbf{BAX})}{\partial \mathbf{X}} = \frac{\partial \operatorname{tr}(\mathbf{XBA})}{\partial \mathbf{X}} = \frac{\partial \operatorname{tr}(\mathbf{B}'\mathbf{X}'\mathbf{A}')}{\partial \mathbf{X}} \\ &= \frac{\partial \operatorname{tr}(\mathbf{A}'\mathbf{B}'\mathbf{X}')}{\partial \mathbf{X}} = \frac{\partial \operatorname{tr}(\mathbf{X}'\mathbf{A}'\mathbf{B}')}{\partial \mathbf{X}} = \mathbf{A}'\mathbf{B}' \end{aligned} \quad (142)$$

$$\begin{aligned} \frac{\partial \operatorname{tr}(\mathbf{AX}'\mathbf{BXC})}{\partial \mathbf{X}} &= \frac{\partial \operatorname{tr}(\mathbf{CAX}'\mathbf{BX})}{\partial \mathbf{X}} = \frac{\partial \operatorname{tr}(\mathbf{XCAX}'\mathbf{B})}{\partial \mathbf{X}} \\ &= \frac{\partial \operatorname{tr}(\mathbf{BXCAX}')}{\partial \mathbf{X}} = \frac{\partial \operatorname{tr}(\mathbf{X}'\mathbf{BXC}\mathbf{A})}{\partial \mathbf{X}} \\ &= \mathbf{BXC}\mathbf{A} + \mathbf{B}'\mathbf{X}\mathbf{A}'\mathbf{C}' \end{aligned} \quad (143)$$

$$\begin{aligned} \frac{\partial \operatorname{tr}(\mathbf{AXBXC})}{\partial \mathbf{X}} &= \frac{\partial \operatorname{tr}(\mathbf{CAXBX})}{\partial \mathbf{X}} = \frac{\partial \operatorname{tr}(\mathbf{XCAXB})}{\partial \mathbf{X}} \\ &= \frac{\partial \operatorname{tr}(\mathbf{BXCAX})}{\partial \mathbf{X}} = \frac{\partial \operatorname{tr}(\mathbf{C}'\mathbf{X}'\mathbf{B}'\mathbf{X}'\mathbf{A}')}{\partial \mathbf{X}} \\ &= \frac{\partial \operatorname{tr}(\mathbf{A}'\mathbf{C}'\mathbf{X}'\mathbf{B}'\mathbf{X}')}{\partial \mathbf{X}} = \frac{\partial \operatorname{tr}(\mathbf{X}'\mathbf{A}'\mathbf{C}'\mathbf{X}'\mathbf{B}')}{\partial \mathbf{X}} \\ &= (\mathbf{BXC}\mathbf{A})' + (\mathbf{CAXB})' \end{aligned} \quad (144)$$

$$\begin{aligned} \frac{\partial \operatorname{tr}(\mathbf{A}(\mathbf{BXC})^{-1}\mathbf{D})}{\partial \mathbf{X}} &= \frac{\partial \operatorname{tr}(\mathbf{DA}(\mathbf{BXC})^{-1})}{\partial \mathbf{X}} = \frac{\partial \operatorname{tr}((\mathbf{BXC})^{-1}\mathbf{DA})}{\partial \mathbf{X}} \\ &= \frac{\partial \operatorname{tr}(\mathbf{D}'(\mathbf{C}'\mathbf{X}'\mathbf{B}')^{-1}\mathbf{A}')}{\partial \mathbf{X}} = \frac{\partial \operatorname{tr}(\mathbf{A}'\mathbf{D}'(\mathbf{C}'\mathbf{X}'\mathbf{B}')^{-1})}{\partial \mathbf{X}} \\ &= \frac{\partial \operatorname{tr}((\mathbf{C}'\mathbf{X}'\mathbf{B}')^{-1}\mathbf{A}'\mathbf{D}')}{\partial \mathbf{X}} \\ &= -(\mathbf{C}(\mathbf{BXC})^{-1}\mathbf{DA}(\mathbf{BXC})^{-1}\mathbf{B})' \end{aligned} \quad (145)$$

$$\begin{aligned} \frac{\partial \operatorname{tr}(\mathbf{A}(\mathbf{X}'\mathbf{BX})^{-1}\mathbf{C})}{\partial \mathbf{X}} &= \frac{\partial \operatorname{tr}(\mathbf{CA}(\mathbf{X}'\mathbf{BX})^{-1})}{\partial \mathbf{X}} = \frac{\partial \operatorname{tr}((\mathbf{X}'\mathbf{BX})^{-1}\mathbf{CA})}{\partial \mathbf{X}} \\ &= -\mathbf{BX}(\mathbf{X}'\mathbf{BX})^{-1}\mathbf{CA}(\mathbf{X}'\mathbf{BX})^{-1} \\ &\quad -\mathbf{B}'\mathbf{X}(\mathbf{X}'\mathbf{BX})^{-1}\mathbf{A}'\mathbf{C}'(\mathbf{X}'\mathbf{BX})^{-1} \end{aligned} \quad (146)$$

$$\begin{aligned}
\frac{\partial \operatorname{tr} \left( A(XBX')^{-1}C \right)}{\partial X} &= \frac{\partial \operatorname{tr} \left( CA(XBX')^{-1} \right)}{\partial X} = \frac{\partial \operatorname{tr} \left( (XBX')^{-1}CA \right)}{\partial X} \\
&= -(XBX')^{-1}CA(XBX')^{-1}XB \\
&\quad -(XBX')^{-1}A'C'(XBX')^{-1}XB' \quad (147)
\end{aligned}$$

$$\frac{\partial |AXB|}{\partial X} = \frac{\partial |B'X'A'|}{\partial X} = |AXB|A'(B'X'A')^{-1}B' \quad (148)$$

$$\begin{aligned}
\frac{\partial |AX'BXC|}{\partial X} &= |AX'BXC| \times \\
&\quad \left( BXC(AX'BXC)^{-1}A + B'XA'(AX'BXC)^{-1}C' \right) \quad (149)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial |AXBX'C|}{\partial X} &= |AXBX'C| \times \\
&\quad \left( C(AAXBX'C)^{-1}AXB + A'(AXBX'C)^{-1}C'XB' \right) \quad (150)
\end{aligned}$$

$$\frac{\partial |A(BXC)^{-1}D|}{\partial X} = -|A(BXC)^{-1}D| \times \quad (151)$$

$$\left( C(BXC)^{-1}D(A(BXC)^{-1}D)^{-1}ABXC^{-1}B \right)' \quad (152)$$

$$\frac{\partial \lambda(AXB)}{\partial X} = \frac{1}{q'p} A'pq'B' \quad (153)$$

$$\frac{\partial \lambda(AX'B)}{\partial X} = \frac{1}{q'p} Bpq'A \quad (154)$$

$$\frac{\partial \lambda(AX^{-1}B)}{\partial X} = \frac{1}{q'p} - X^{-1'} A'pq'B' X^{-1'} \quad (155)$$

$$\frac{\partial \lambda(AX'BXC)}{\partial X} = \frac{1}{q'p} (BXCpq'A + B'XA'pq'C') \quad (156)$$

$$\frac{\partial \lambda(AXBX'C)}{\partial X} = \frac{1}{q'p} (Cpq'AXB + A'pq'C'XB') \quad (157)$$

なお、上記の固有値の微分に際して、 $q$  ならびに  $p$  は、固有値を求めるべき行列、たとえば  $AXB$  の、その固有値に対応したの左右の固有ベクトルを表している。

$$\frac{\partial \operatorname{tr}(\mathbf{A} \operatorname{diag}(\mathbf{BXC})\mathbf{D})}{\partial \mathbf{X}} = (\mathbf{C} \operatorname{diag}(\mathbf{DA})\mathbf{B})' \quad (158)$$

$$\frac{\partial \operatorname{tr}(\mathbf{A} \operatorname{diag}(\mathbf{X}'\mathbf{BXC})\mathbf{C})}{\partial \mathbf{X}} = \mathbf{BX} \operatorname{diag}(\mathbf{CA}) + (\operatorname{diag}(\mathbf{CA})\mathbf{X}'\mathbf{B})' \quad (159)$$

$$\frac{\partial \operatorname{tr}(\mathbf{A} \operatorname{diag}(\mathbf{BX}^{-1}\mathbf{C})\mathbf{D})}{\partial \mathbf{X}} = -(\mathbf{X}^{-1}\mathbf{C} \operatorname{diag}(\mathbf{DA})\mathbf{BX}^{-1})' \quad (160)$$

$$\frac{\partial \operatorname{tr}(\mathbf{A} \operatorname{diag}(\mathbf{X}'\mathbf{B}^{-1}\mathbf{X})\mathbf{C})}{\partial \mathbf{X}} = -\mathbf{BX}(\mathbf{X}'\mathbf{BX})^{-1} \operatorname{diag}(\mathbf{CA})(\mathbf{X}'\mathbf{BX})^{-1} \quad (161)$$

$$-((\mathbf{X}'\mathbf{BX})^{-1} \operatorname{diag}(\mathbf{CA})(\mathbf{X}'\mathbf{BX})^{-1}\mathbf{X}'\mathbf{B})' \quad (162)$$

$$\frac{\partial \operatorname{tr}(\mathbf{A}(\mathbf{X} \odot \mathbf{B})\mathbf{C})}{\partial \mathbf{X}} = \mathbf{B} \odot (\mathbf{CA})' \quad (163)$$

$$\frac{\partial \operatorname{tr}(\mathbf{A}(\mathbf{X}' \odot \mathbf{B})\mathbf{C})}{\partial \mathbf{X}} = \mathbf{B}' \odot (\mathbf{CA})' \quad (164)$$

$$\frac{\partial \operatorname{tr}(\mathbf{A}((\mathbf{X}'\mathbf{BX}) \odot \mathbf{C})\mathbf{D})}{\partial \mathbf{X}} = \mathbf{BX}(\mathbf{C}' \odot (\mathbf{DA})) + ((\mathbf{C}' \odot (\mathbf{DA}))\mathbf{X}'\mathbf{B})' \quad (165)$$

$$\frac{\partial \operatorname{tr}(\mathbf{A}(\mathbf{X}^{-1} \odot \mathbf{B})\mathbf{C})}{\partial \mathbf{X}} = -(\mathbf{X}^{-1}(\mathbf{B}' \odot (\mathbf{CA}))\mathbf{X}^{-1})' \quad (166)$$

$$\frac{\partial \operatorname{tr}(\mathbf{A}\mathbf{X}^{(p)}\mathbf{B})}{\partial \mathbf{X}} = p\mathbf{X}^{(p-1)} \odot (\mathbf{BA})' \quad (167)$$

$$\frac{\partial \operatorname{tr}(\mathbf{A}(\mathbf{X}'\mathbf{BX})^{(p)}\mathbf{C})}{\partial \mathbf{X}} = p(\mathbf{BX}((\mathbf{X}'\mathbf{BX})^{(p-1)'} \odot (\mathbf{CA})) \quad (168)$$

$$+ (((\mathbf{X}'\mathbf{BX})^{(p-1)'} \odot (\mathbf{CA}))\mathbf{X}'\mathbf{B})') \quad (169)$$

$$\frac{\partial \operatorname{tr}(\mathbf{A}\mathbf{X}^{(p)-1}\mathbf{B})}{\partial \mathbf{X}} = -p\mathbf{X}^{(p-1)} \odot (\mathbf{X}^{(p)-1}\mathbf{BAX}^{(p)-1})' \quad (170)$$

Derivation of (61)

$$\mathbf{X}\mathbf{X}^{-1} = \mathbf{I} \quad (171)$$

$$\text{lefthandside} = \frac{\partial \mathbf{X}\mathbf{X}^{-1}}{\partial x_{ij}} = \mathbf{X} \frac{\partial \mathbf{X}^{-1}}{\partial x_{ij}} + \frac{\partial \mathbf{X}}{\partial x_{ij}} \mathbf{X}^{-1} \quad (172)$$

$$= \mathbf{O} \quad (173)$$

Therefore, by multiplying  $\mathbf{X}^{-1}$  from the left,

$$\mathbf{X}^{-1} \mathbf{X} \frac{\partial \mathbf{X}^{-1}}{\partial x_{ij}} + \mathbf{X}^{-1} \frac{\partial \mathbf{X}}{\partial x_{ij}} \mathbf{X}^{-1} = \frac{\partial \mathbf{X}^{-1}}{\partial x_{ij}} + \mathbf{X}^{-1} \frac{\partial \mathbf{X}}{\partial x_{ij}} \mathbf{X}^{-1} = \mathbf{O} \quad (174)$$

Derivation of (63)

$$\mathbf{X}\mathbf{X}^{-}\mathbf{X} = \mathbf{X} \quad \text{and} \quad \mathbf{X}^{-}\mathbf{X}\mathbf{X}^{-} = \mathbf{X}^{-} \quad (175)$$

$$\frac{\partial \mathbf{X}^{-}\mathbf{X}\mathbf{X}^{-}}{\partial x_{ij}} = \mathbf{X}^{-} \frac{\partial \mathbf{X}\mathbf{X}^{-}}{\partial x} + \frac{\partial \mathbf{X}^{-}}{\partial x_{ij}} \mathbf{X}\mathbf{X}^{-} \quad (176)$$

$$= \mathbf{X}^{-} \mathbf{X} \frac{\partial \mathbf{X}^{-}}{\partial x_{ij}} + \mathbf{X}^{-} \frac{\partial \mathbf{X}}{\partial x_{ij}} \mathbf{X}^{-} + \frac{\partial \mathbf{X}^{-}}{\partial x_{ij}} \mathbf{X}\mathbf{X}^{-} \quad (177)$$

$$= \frac{\partial \mathbf{X}^{-}}{\partial x_{ij}} \quad (178)$$

Multiply  $\mathbf{X}$  from the right and left

$$\mathbf{X}\mathbf{X}^{-}\mathbf{X} \frac{\partial \mathbf{X}^{-}}{\partial x_{ij}} \mathbf{X} + \mathbf{X}\mathbf{X}^{-} \frac{\partial \mathbf{X}}{\partial x_{ij}} \mathbf{X}^{-}\mathbf{X} + \mathbf{X} \frac{\partial \mathbf{X}^{-}}{\partial x_{ij}} \mathbf{X}\mathbf{X}^{-}\mathbf{X} = \mathbf{X} \frac{\partial \mathbf{X}^{-}}{\partial x_{ij}} \mathbf{X} \quad (179)$$

Therefore,

$$\mathbf{X} \frac{\partial \mathbf{X}^{-}}{\partial x_{ij}} \mathbf{X} + \mathbf{X}\mathbf{X}^{-} \frac{\partial \mathbf{X}}{\partial x_{ij}} \mathbf{X}^{-}\mathbf{X} + \mathbf{X} \frac{\partial \mathbf{X}^{-}}{\partial x_{ij}} \mathbf{X} = \mathbf{X} \frac{\partial \mathbf{X}^{-}}{\partial x_{ij}} \mathbf{X} \quad (180)$$

and

$$\mathbf{X} \frac{\partial \mathbf{X}^{-}}{\partial x_{ij}} \mathbf{X} = \mathbf{X}\mathbf{X}^{-} \frac{\partial \mathbf{X}}{\partial x_{ij}} \mathbf{X}^{-}\mathbf{X} \quad (181)$$

Or

$$\mathbf{X} \left( \frac{\partial \mathbf{X}^{-}}{\partial x_{ij}} \mathbf{X} - \mathbf{X}^{-} \frac{\partial \mathbf{X}}{\partial x_{ij}} \mathbf{X}^{-}\mathbf{X} \right) = \mathbf{X} \left( \frac{\partial \mathbf{X}^{-}}{\partial x_{ij}} - \mathbf{X}^{-} \frac{\partial \mathbf{X}}{\partial x_{ij}} \mathbf{X}^{-} \right) \mathbf{X} \quad (182)$$

In[636]= mD[tr[t[Y - X.B].W.(Y - X.B).V], B] // tf

Input = mD[ Tr[(Y - X.B)<sup>T</sup>.W.Y.V] - Tr[(Y - X.B)<sup>T</sup>.W.X.B.V], B]

Simplified Input =mD[ Tr[Y<sup>T</sup>.W.Y.V] - Tr[B<sup>T</sup>.X<sup>T</sup>.W.Y.V] - Tr[Y<sup>T</sup>.W.X.B.V] + Tr[B<sup>T</sup>.X<sup>T</sup>.W.X.B.V], B]

Rules Used : (1)A1((A+B).X) (2)S1(constant) (3)A1((A+B).X)

(4)A2(-A) (5)G3(A.t[X].B) (6)S2(tr[X.A])

(7)A1((A+B).X) (8)A2(-A) (9)G2(A.X.B)

(10)S2(tr[X.A]) (11)P1(tr[f.g]) (12)G2(A.X.B)

(13)S2(tr[X.A]) (14)G3(A.t[X].B) (15)S2(tr[X.A])

(1) mD0x (Tr[Y<sup>T</sup>.W.Y.V], B) + mD0x (-Tr[B<sup>T</sup>.X<sup>T</sup>.W.Y.V] - Tr[Y<sup>T</sup>.W.X.B.V] + Tr[B<sup>T</sup>.X<sup>T</sup>.W.X.B.V], B)

(2) Null

(3) mD0x (-Tr[B<sup>T</sup>.X<sup>T</sup>.W.Y.V], B) + mD0x (Tr[B<sup>T</sup>.X<sup>T</sup>.W.X.B.V] - Tr[Y<sup>T</sup>.W.X.B.V], B)

(4) -mD0x (Tr[B<sup>T</sup>.X<sup>T</sup>.W.Y.V], B)

(5) mD0x (Tr[B.V<sup>T</sup>.Y<sup>T</sup>.W<sup>T</sup>.X], B)

(6) (V<sup>T</sup>.Y<sup>T</sup>.W<sup>T</sup>.X)<sup>T</sup>

(7) mD0x (-Tr[Y<sup>T</sup>.W.X.B.V], B) + mD0x (Tr[B<sup>T</sup>.X<sup>T</sup>.W.X.B.V], B)

(8) -mD0x (Tr[Y<sup>T</sup>.W.X.B.V], B)

(9) mD0x (Tr[B.V.Y<sup>T</sup>.W.X], B)

(10) (V.Y<sup>T</sup>.W.X)<sup>T</sup>

(11) mD0x (Tr[B<sup>T</sup>.Q8\$7248], B) + mD0x (Tr[Q9\$7248.X<sup>T</sup>.W.X.B.V], B)

(12) mD0x (Tr[B.V.Q9\$7248.X<sup>T</sup>.W.X], B)

(13) (V.Q9\$7248.X<sup>T</sup>.W.X)<sup>T</sup>

(14) mD0x (Tr[B.Q8\$7248<sup>T</sup>], B)

(15) Q8\$7248<sup>TT</sup>

(15) (V.B<sup>T</sup>.X<sup>T</sup>.W.X)<sup>T</sup> + (X<sup>T</sup>.W.X.B.V)<sup>TT</sup>

Out[636]/TraditionalForm=

$-X^T.W.Y.V + X^T.W.X.B.V - (V.Y^T.W.X)^T + (V.B^T.X^T.W.X)^T$

```

In[526]=  $\Sigma = \Lambda \cdot \Xi \cdot t[\Lambda] + \Psi$ 
           $mD[\text{tr}[S \cdot \text{Inverse}[\Sigma]] - \text{Log}[\text{Det}[S \cdot \Sigma]], \Lambda] // \text{tf}$ 

Out[526]=  $\Psi + \Lambda \cdot \Xi \cdot \text{Transpose}[\Lambda]$ 

Input =  $mD[\text{Tr}[S \cdot (\Psi + \Lambda \cdot \Xi \cdot \Lambda^T)^{-1}] - \log(\{S \cdot (\Psi + \Lambda \cdot \Xi \cdot \Lambda^T)\}), \Lambda]$ 

Simplified Input =  $mD[\text{Tr}[S \cdot (\Psi + \Lambda \cdot \Xi \cdot \Lambda^T)^{-1}] - \log(\{S \cdot \Psi + S \cdot \Lambda \cdot \Xi \cdot \Lambda^T\}), \Lambda]$ 

Rules Used : (1)A1((A+B).X) (2)A2(-A) (3)A4(log f[X])
(4)C2(det) (5)A1((A+B).X) (6)S1(constant)
(7)G2(A.X.B) (8)P1(tr[f.g]) (9)G3(A.t[X].B)
(10)S2(tr[X.A]) (11)S2(tr[X.A]) (12)G4.1(A.inv[B].C)
(13)C1.1(inv) (14)A1((A+B).X) (15)S1(constant)
(16)G2(A.X.B) (17)P1(tr[f.g]) (18)G3(A.t[X].B)
(19)S2(tr[X.A]) (20)S2(tr[X.A])

(1)  $mD0x(-\log(\{S \cdot \Psi + S \cdot \Lambda \cdot \Xi \cdot \Lambda^T\}), \Lambda) + mD0x(\text{Tr}[S \cdot (\Psi + \Lambda \cdot \Xi \cdot \Lambda^T)^{-1}], \Lambda)$ 
(2)  $-mD0x(\log(\{S \cdot \Psi + S \cdot \Lambda \cdot \Xi \cdot \Lambda^T\}), \Lambda)$ 
(3)  $\frac{mD0x(\{S \cdot \Psi + S \cdot \Lambda \cdot \Xi \cdot \Lambda^T\}, \Lambda)}{\{S \cdot \Psi + S \cdot \Lambda \cdot \Xi \cdot \Lambda^T\}}$ 
(4)  $mD0x(\text{Tr}[(\{Q9\6920 \mid Q9\6920^{T-1}\}^T \cdot S \cdot \Psi) + \text{Tr}[(\{Q9\6920 \mid Q9\6920^{T-1}\}^T \cdot S \cdot \Lambda \cdot \Xi \cdot \Lambda^T)], \Lambda)$ 
(5)  $mD0x(\text{Tr}[(\{Q9\6920 \mid Q9\6920^{T-1}\}^T \cdot S \cdot \Psi), \Lambda] + mD0x(\text{Tr}[(\{Q9\6920 \mid Q9\6920^{T-1}\}^T \cdot S \cdot \Lambda \cdot \Xi \cdot \Lambda^T)], \Lambda)$ 
(6) Null
(7)  $mD0x(\text{Tr}[\Lambda \cdot \Xi \cdot \Lambda^T \cdot (\{Q9\6920 \mid Q9\6920^{T-1}\}^T \cdot S)], \Lambda)$ 
(8)  $mD0x(\text{Tr}[\Lambda \cdot Q8\6942], \Lambda) + mD0x(\text{Tr}[Q8\6942 \cdot \Xi \cdot \Lambda^T \cdot (\{Q9\6920 \mid Q9\6920^{T-1}\}^T \cdot S)], \Lambda)$ 
(9)  $mD0x(\text{Tr}[\Lambda \cdot \Xi^T \cdot Q9\6942^T \cdot S^T \cdot (\{Q9\6920 \mid Q9\6920^{T-1}\})], \Lambda)$ 
(10)  $(\Xi^T \cdot Q9\6942^T \cdot S^T \cdot (\{Q9\6920 \mid Q9\6920^{T-1}\})^T)^T$ 
(11)  $Q8\6942^T$ 
(11)  $(\Xi \cdot \Lambda^T \cdot (\{Q9\6920 \mid Q9\6920^{T-1}\}^T \cdot S))^T + (\Xi^T \cdot \Lambda^T \cdot S^T \cdot (\{Q9\6920 \mid Q9\6920^{T-1}\})^T)^T$ 
(11)  $(\Xi \cdot \Lambda^T \cdot (\{S \cdot \Psi + S \cdot \Lambda \cdot \Xi \cdot \Lambda^T\} \cdot (S \cdot \Psi + S \cdot \Lambda \cdot \Xi \cdot \Lambda^T)^{T-1})^T \cdot S)^T + (\Xi^T \cdot \Lambda^T \cdot S^T \cdot (\{S \cdot \Psi + S \cdot \Lambda \cdot \Xi \cdot \Lambda^T\} \cdot (S \cdot \Psi + S \cdot \Lambda \cdot \Xi \cdot \Lambda^T)^{T-1})^T)^T$ 
(12)  $mD0x(\text{Tr}[(\Psi + \Lambda \cdot \Xi \cdot \Lambda^T)^{-1} \cdot S], \Lambda)$ 
(13)  $-mD0x(\text{Tr}[Q9\6957^{-1} \cdot S \cdot Q9\6957^{-1} \cdot \Psi] + \text{Tr}[Q9\6957^{-1} \cdot S \cdot Q9\6957^{-1} \cdot \Lambda \cdot \Xi \cdot \Lambda^T]), \Lambda)$ 
(14)  $mD0x(\text{Tr}[Q9\6957^{-1} \cdot S \cdot Q9\6957^{-1} \cdot \Psi], \Lambda) + mD0x(\text{Tr}[Q9\6957^{-1} \cdot S \cdot Q9\6957^{-1} \cdot \Lambda \cdot \Xi \cdot \Lambda^T]), \Lambda)$ 
(15) Null
(16)  $mD0x(\text{Tr}[\Lambda \cdot \Xi \cdot \Lambda^T \cdot Q9\6957^{-1} \cdot S \cdot Q9\6957^{-1}], \Lambda)$ 
(17)  $mD0x(\text{Tr}[\Lambda \cdot Q8\6971], \Lambda) + mD0x(\text{Tr}[Q8\6971 \cdot \Xi \cdot \Lambda^T \cdot Q9\6957^{-1} \cdot S \cdot Q9\6957^{-1}], \Lambda)$ 
(18)  $mD0x(\text{Tr}[\Lambda \cdot \Xi^T \cdot Q9\6971^T \cdot Q9\6957^{-1T} \cdot S^T \cdot Q9\6957^{-1T}], \Lambda)$ 
(19)  $(\Xi^T \cdot Q9\6971^T \cdot Q9\6957^{-1T} \cdot S^T \cdot Q9\6957^{-1T})^T$ 
(20)  $Q8\6971^T$ 
(20)  $(\Xi \cdot \Lambda^T \cdot Q9\6957^{-1} \cdot S \cdot Q9\6957^{-1})^T + (\Xi^T \cdot \Lambda^T \cdot Q9\6957^{-1T} \cdot S^T \cdot Q9\6957^{-1T})^T$ 
(20)  $-(\Xi \cdot \Lambda^T \cdot (\Psi + \Lambda \cdot \Xi \cdot \Lambda^T)^{-1} \cdot S \cdot (\Psi + \Lambda \cdot \Xi \cdot \Lambda^T)^{-1})^T - (\Xi^T \cdot \Lambda^T \cdot (\Psi + \Lambda \cdot \Xi \cdot \Lambda^T)^{-1T} \cdot S^T \cdot (\Psi + \Lambda \cdot \Xi \cdot \Lambda^T)^{-1T})^T$ 

Out[527]//TraditionalForm=
 $-(S \cdot \Psi + S \cdot \Lambda \cdot \Phi \cdot \Lambda^T)^{-1} \cdot S \cdot \Lambda \cdot \Phi - (\Psi + \Lambda \cdot \Phi \cdot \Lambda^T)^{-1} \cdot S \cdot (\Psi + \Lambda \cdot \Phi \cdot \Lambda^T)^{-1} \cdot \Lambda \cdot \Phi -$ 
 $(\Phi \cdot \Lambda^T \cdot (S \cdot \Psi + S \cdot \Lambda \cdot \Phi \cdot \Lambda^T)^{-1} \cdot S)^T - (\Phi \cdot \Lambda^T \cdot (\Psi + \Lambda \cdot \Phi \cdot \Lambda^T)^{-1} \cdot S \cdot (\Psi + \Lambda \cdot \Phi \cdot \Lambda^T)^{-1})^T$ 

```

\*\*\*\*\*  
a matrix derivative of a scalar function of matrix/vector variables  
\*\*\*\*\*