第3卷 第1期
1983年 1月

数学研究与评论
JOURNAL OF MATHEMATICAL
RESEARCH AND EXPOSITION

Vol.3 No.1
Jan. 1983

矩阵微分——推导精确分布的工具*

张尧庭

(武汉大学)

在推导统计量的精确分布时，矩阵微分是一个既初等又方便的工具，这一工具是由许宝騄先生首先提出的，在30年代末期，用它导出了著名的正态样本协差阵特征根的联合分布。后来，1947年在北卡罗林纳（Noth Carolina）大学讲授多元分析一课时，他系统地发展了这一工具，得到了很多结果，有些结果后来也一直没有正式发表过。在Hotelling教授的倡议下，由W. L. Deemer和I. Olkin整理发表了其中矩阵微分的纯代数部分，刊印在1951年的Biometrika杂志上。这篇文章（见[1]）不断地被广泛引用。就从I. Olkin在50年代所发表的一些结果来看，明显地有这一工具的影响。现在试图通过一些复杂的分布的推导来介绍这一工具。

先引入一些记号。假定X是$n\times p$的矩阵，X的元素记为x_{ij}，$i=1,2,\cdots,n$，$j=1,2,\cdots,p$；$f(X)$是np个变元x_{ij}的函数；从多元函数全微分的定义，有关系式

$$dX \triangleq (dx_{ij}) = \begin{pmatrix} dx_{11} & dx_{12} & \cdots & dx_{1p} \\ dx_{21} & dx_{22} & \cdots & dx_{2p} \\ \vdots & \vdots & & \vdots \\ dx_{n1} & dx_{n2} & \cdots & dx_{np} \end{pmatrix},$$

$$\frac{\partial f}{\partial X} \triangleq \left(\frac{\partial f}{\partial x_{ij}}\right) = \begin{pmatrix} \frac{\partial f}{\partial x_{11}} & \frac{\partial f}{\partial x_{12}} & \cdots & \frac{\partial f}{\partial x_{1p}} \\ \frac{\partial f}{\partial x_{21}} & \frac{\partial f}{\partial x_{22}} & \cdots & \frac{\partial f}{\partial x_{2p}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f}{\partial x_{n1}} & \frac{\partial f}{\partial x_{n2}} & \cdots & \frac{\partial f}{\partial x_{np}} \end{pmatrix},$$

$$df(X) = \sum_{i=1}^{n}\sum_{j=1}^{p}\frac{\partial f}{\partial x_{ij}}dx_{ij},$$

也即为

$$df(X) = tr\left(\frac{\partial f}{\partial X}\right)' dX, \tag{1}$$

注意：A'表示矩阵A的转置阵，而不表示A阵的微商。当$p=1$时，用x表示$n\times 1$的向量。当$f(x)$是n个函数组成的向量时，$f(x)=(f_1(x), f_2(x),\cdots,f_n(x))'$，将每一个$f_i(x)$看作(1)中的$f$，用公式(1)，写成矩阵的形式，就有

$$df = \begin{pmatrix} df_1 \\ df_2 \\ \vdots \\ df_n \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{pmatrix} \begin{pmatrix} dx_1 \\ dx_2 \\ \vdots \\ dx_n \end{pmatrix},$$

* 1982年3月15日收到.

上式可写成

$$df = \frac{\partial f}{\partial x} dx, \tag{2}$$

dx 的系数矩阵 $\frac{\partial f}{\partial x}$ 的行列式正好是积分变元变换将 $f \to x$ 时相应的雅可比行列式。（2）式告诉我们，不论 $f \to x$ 时函数形式如何复杂，但从微分向量 df，dx 来看，它们之间只是一个线性的关系，用 y_* 表示 df，x_* 表示 dx，（2）式就是 $y_* = Ax_*$，$|A|$ 就是雅可比行列式 $\left|\frac{\partial f}{\partial x}\right|$。只是这一线性变换的系数矩阵是变元 x 的函数，它随 x 的改变而变化。从（2）式来看，只要先处理好线性变换的雅可比行列式，非线性变换相应的行列式是好求的。这一结论对于矩阵 X 也是适用的，只需把矩阵 X 看作一个 np 维的向量就可以了。

假定 $n \times p$ 的随机矩阵 X 的联合分布密度是 $p(X'X)$（例如正态分布就是一个重要的特例），现在要求 $X'X$ 的特征根的联合分布。$X'X$ 是一个非负定的矩阵，它的特征根共有 p 个，记它为 $\lambda_1, \cdots \lambda_p$，且不妨假定 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_p$。从矩阵的知识可知，存在正交阵 Γ 使

$$X'X = \Gamma \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_p \end{pmatrix} \Gamma'. \tag{3}$$

于是从 X 的已知的分布密度出发，要求出 $\lambda_1, \cdots \lambda_p$ 的联合分布，要解决两个问题：

（i）从 X 的分布导出 $X'X$ 的分布。这里 X 有 np 个变量，而 $X'X$ 只有 $\frac{1}{2}p(p+1)$ 个变量。

（ii）从 $X'X$ 的分布导出 $\lambda_1, \cdots, \lambda_p$ 的分布。这里 $X'X$ 有 $\frac{1}{2}p(p+1)$ 个变量，而 $\lambda_1, \cdots, \lambda_p$ 只有 p 个变量。

记 $A = X'X = (a_{ij})$ 后，a_{ij} 都是 x_{ij} 的函数，λ_i 都是 a_{ij} 的函数，但是从 $X \to A$，从 $A \to \lambda = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_p \end{pmatrix}$，都不是能直接利用（2）式来求出雅可比行列式的。都需要寻求合适的辅助变量，引入这些辅助变量后，就可以利用（2）。

假定 $n \geqslant p$，且 $|X'X| = 0$ 的概率为 0，即有 $\mathrm{rk}X = p$（X 的秩为 p）的概率是 1。利用矩阵分解的知识，在 $\mathrm{rk}X = p$ 条件上，有

$$\underset{n \times p}{X} = \underset{n \times p}{\Gamma} \underset{p \times p}{T}, \quad \Gamma\Gamma' = I_p, \quad T = (0^{t_{ij}}), \tag{4}$$

即 Γ 是 p 个标准正交列向量组成的，T 是一个 $p \times p$ 的上三角矩阵。分析一下（4）式两边的变元数，左边 X 中有 np 个，右边 Γ 中有 np 个但从标准正交条件知道只有 $np - \frac{p}{2}(p+1)$ 个可以自由变，T 中有 $\frac{p}{2}(p+1)$ 个变元。因此，从自由变元的数目来看，左右两边是相等的。因此只要（4）式相应的变换是 1—1 的（也即 X 唯一地分解为正交阵 Γ 与三角阵 T），那么从微分的性质可知

$$dX = \Gamma dT + (d\Gamma)T,$$

dX 也只是 dT 与 $d\Gamma$ 的线性变换。但 $d\Gamma$ 是一 $n \times p$ 的阵，并不是自由变量的微分组成的，这需要设法克服的（可参看[2]）。然而，这也可以直接用正交变换，逐步降维，归纳地求出（可参看[3]）。不论从那一途经，都可以求得 A 的分布。从 A 的分布去导出 λ 的分

布，就涉及到（3）所相应的矩阵分解的唯一性．若限制了Γ的主对角元均为正，则（3）的分解就是唯一的．此时

$$dA=d(X'X)=(d\Gamma)\Lambda\Gamma'+\Gamma(d\Lambda)\Gamma'+\Gamma\Lambda(d\Gamma)',$$

式中 $\Lambda=\begin{pmatrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & \lambda_p\end{pmatrix}$．对上式乘以$\Gamma'$和$\Gamma$，就得（注意$\Gamma\Gamma'=\Gamma'\Gamma=I$）：

$$\Gamma'(dA)\Gamma=\Gamma'(d\Gamma)\Lambda+d\Lambda+\Lambda(d\Gamma)'\Gamma. \tag{5}$$

为了方便，dA 代以 A_*，$d\Lambda$ 代以 Λ_*，$\Gamma'(d\Gamma)$ 代以 R，于是有

$$\Gamma'A_*\Gamma=R\Lambda+\Lambda_*+\Lambda R'.$$

注意到 $O=dI=d(\Gamma'\Gamma)=(d\Gamma')\Gamma+\Gamma'(d\Gamma)=R'+R$，也即 $R'=-R$，R 是一反对称阵．代入上式，即得

$$\Gamma'A_*\Gamma=R\Lambda-\Lambda R+\Lambda_*. \tag{6}$$

$A_*\to\Gamma'A_*\Gamma$ 是一线性变换，从 $\Gamma'A_*\Gamma\to(R,\Lambda_*)$ 也是线性的，将(6)式左端记为 $B=(b_{ij})$，$B'=B$．从（6）就有（$R=(r_{ij})$）：

$$\begin{cases}b_{ii}=\partial r_{ii}+d\lambda_i,\\ b_{ij}=(\lambda_i-\lambda_j)r_{ij}\quad(i\neq j),\end{cases}$$

于是 $b_{11},b_{12},\cdots,b_{1p},b_{22},\cdots,b_{2p},\cdots,b_{pp}$ 对 $r_{12},\cdots,r_{1p},r_{23},\cdots,r_{2p},\cdots,r_{pp},d\lambda_1,\cdots,d\lambda_p$ 的雅可比行列式从上式就可以立即求得，它就是 $a^p\prod\limits_{i<j}(\lambda_i-\lambda_j)$．注意到 A_* 到 B 相应的雅可比行列式的绝对值为 1，而 R 对 Γ 中自由变元的微分所相应的变换仅仅是 Γ 中元素的函数，与特征根 $\lambda_1,\cdots,\lambda_p$ 无关，这样就可以知道 $\lambda_1,\cdots,\lambda_p$ 的联合密度具有的形式仅仅是 $p(X'X)$ 中 $X'X$ 用 λ_i 来表示的部份以及 $\prod\limits_{i<j}(\lambda_i-\lambda_j)$ 的乘积．例如对正态分布 $p(X'X)=ce^{-\frac{1}{2}\operatorname{tr}X'X}=ce^{-\frac{1}{2}\sum_{i=1}^{p}\lambda_i}$，则 $\lambda_1,\cdots,\lambda_p$ 的密度就是

$$be^{-\frac{1}{2}\sum_{i=1}^{p}\lambda_i}\qquad\prod_{i<j}(\lambda_i-\lambda_j)\qquad\lambda_1>\lambda_2>\cdots>\lambda_p>0,$$

其中 b 是一正则化的常数．

从上面的讨论可以看出，这个方法可以统一地获得许多矩阵的特征根的联合分布，而且推导中主要涉及的仅仅是矩阵的分解与各种矩阵之间线性变换相应的雅可比行列式，本质上这些推导都只是一些线性代数的演算．

现在来考察典型相关系数的联合分布的问题．假定 A 是一 $p+q$ 阶的正定随机阵，将 A 分块

$$A=\begin{array}{cc}\begin{pmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{pmatrix} & \begin{array}{l}p\\ q\end{array}\\ \begin{array}{cc}p & \quad q\end{array} & \end{array},$$

A_{11},A_{22} 分别为 p,q 阶的正定子阵．假定 A 的联合密度是已知的函数 $g(A)$，如何去求 $A_{11}^{-1}A_{12}A_{22}^{-1}A_{21}$ 的特征根的联合分布密度．此时 $A_{11}^{-1}A_{12}A_{22}^{-1}A_{21}$ 并不是一个对称阵，但它的全部非 0 特征根与 $A_{11}^{-\frac{1}{2}}A_{12}A_{22}^{-1}A_{21}A_{11}^{-\frac{1}{2}}$ 是相同的，而后者是一正定阵（只要 $p\leqslant q$）．然而要求出 $A_{11}^{-\frac{1}{2}}A_{12}A_{22}^{-1}A_{21}A_{11}^{-\frac{1}{2}}$ 的联合分布，也是不容易的，这可以参阅［2］．另一种方法可以直接从矩阵唯一分解的公式来推导相应的分布．若对 A_{11} 和 A_{22} 分别施行非退化的矩阵变换：

$$A_{11}\to \underset{p\times p}{G}A_{11}G',\ A_{22}\to \underset{q\times q}{H}A_{22}H'$$
$$|G|\neq 0,\qquad |H|\neq 0,$$

它们的正定性是不会改变的．这也就是相当于对A施行形如$\begin{pmatrix} G & O \\ O & H \end{pmatrix}$的变换，即：

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}\to\begin{pmatrix} G & O \\ O & H \end{pmatrix}\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}\begin{pmatrix} G' & O \\ O & H' \end{pmatrix}=\begin{pmatrix} GA_{11}G' & GA_{12}H' \\ HA_{21}G' & HA_{22}H' \end{pmatrix},$$

而 $(GA_{11}G')^{-1}GA_{12}H'(HA_{22}H')^{-1}HA_{21}G'=(G')^{-1}A_{11}^{-1}A_{12}A_{22}^{-1}A_{21}G'$，注意到$|G|\neq 0$，可见变换以后相应的 $(G')^{-1}A_{11}^{-1}A_{12}A_{22}^{-1}A_{21}G'$ 与 $A_{11}^{-1}A_{12}A_{22}^{-1}A_{21}$ 有相同的全部特征根。从矩阵变换的知识知道，对正定阵A，存在G,H使

$$\begin{pmatrix} G & O \\ O & H \end{pmatrix}A\begin{pmatrix} G' & O \\ O & H' \end{pmatrix}=\begin{pmatrix} I_p & D(\Gamma) & O \\ D(\Gamma) & & I_q \\ O & & \end{pmatrix},$$

其中 $D(\Gamma)=\begin{pmatrix} \Gamma_1 & & O \\ & \ddots & \\ O & & \Gamma_p \end{pmatrix}$，易见 $\Gamma_1^2,\Gamma_2^2,\cdots,\Gamma_p^2$ 就是 $A_{11}^{-1}A_{12}A_{22}^{-1}A_{21}$ 的全部特征根．上式也可写为

$$A=\begin{pmatrix} P & O \\ O & Q \end{pmatrix}\begin{pmatrix} I_p & D(\Gamma) & O \\ D(\Gamma) & & I_q \\ O & & \end{pmatrix}\begin{pmatrix} P' & O \\ O & Q' \end{pmatrix}. \tag{7}$$

比较(7)式两边的自由变量的数目，左端A中有$\frac{1}{2}(p+q)(p+q+1)$个，右端有 p^2+q^2+p 个，记 $\delta=q-p$ 后，就有

$$p^2+q^2+p-\frac{1}{2}(p+q)(p+q+1)=\frac{1}{2}\delta(\delta-1),$$

可见 (7) 式相应的分解是不可能唯一的．为了达到这一目的，将Q再分块：

$$Q=\begin{array}{c} \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}\begin{matrix} p \\ \delta \end{matrix} \\ \begin{matrix} p & \quad\delta \end{matrix} \end{array}.$$

事实上，有关系式

$$A=\begin{pmatrix} P & O & O \\ O & Q_{11} & Q_{12}T \\ O & Q_{21} & Q_{22}T \end{pmatrix}\begin{pmatrix} I_p & D(\Gamma) & O \\ D(\Gamma) & I_p & O \\ O & O & I_\delta \end{pmatrix}\begin{pmatrix} P' & O & O \\ O & Q_{11}' & Q_{21}' \\ O & T'Q_{12}' & T'Q_{22}' \end{pmatrix}, \tag{8}$$

其中T是$\delta\times\delta$的正交阵．我们若选T使$Q_{22}T$成为一个上三角阵，即$Q_{22}T=(0\nabla)$，这是容易做到的，由于 $|Q_{22}|\neq 0$，还可要求$Q_{22}T$的主对角元素均大于0（这只要用施密特正交化方法就可以得到）．因此我们可以限制（7）式中的Q为

$$Q=\begin{array}{c} \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & U \end{pmatrix}\begin{matrix} p \\ \delta \end{matrix} \\ \begin{matrix} p & \quad\delta \end{matrix} \end{array}, \tag{9}$$

其中U是一$\delta\times\delta$的上三角矩阵，且U的主对角元素 u_{ii} 均大于0．利用(9)式，再比较(7)式两边的自由变元的个数，容易判断它们是相等的，并且可以进一步证明这样的分解是唯

一的。于是，剩下的问题是如何利用(7)，(9)两式来求出相应于这一变换的雅可比行列式。

经过一系列的演算（参看[1]中的有关部份）。可以求出上述矩阵的变换式（7），(9)相应的雅可比行列式是

$$Q^{p+q}|P|^{q+1}|Q|^{p+1}\left(\prod_{i=1}^{\delta}u_{ii}^{i-1}\right)\left(\prod_{i=1}^{p}\Gamma_i^{\delta}\right)\left(\prod_{1\leq i<j\leq p}(\Gamma_i^2-\Gamma_j^2)\right), \tag{10}$$

当 $q=p$ 或 $q=p+1$ 时，式中含 u_{ii} 这一项不出现。于是利用 A 的联合分布密度 $g(A)$ 的表达式和（10）的雅可比，就不难求得 $\Gamma_1,\Gamma_2,\cdots,\Gamma_p$ 的联合分布。

近年来，发现不少统计问题与矩阵的奇异值有密切的关系，因此寻求矩阵 X 的奇异值的联合密度也是很有意义的。假定 X 是一 $n\times p$ 的随机矩阵，它的秩概率为 1 地等于 p，则从奇异值分解式

$$\underset{n\times p}{X}=\underset{n\times p}{U}\begin{pmatrix}\sigma_1 & & 0\\ & \sigma_2 & \\ & & \ddots \\ 0 & & \sigma_p\end{pmatrix}\underset{p\times p}{V}, \tag{11}$$

$$\text{其中}\quad U'U=I_p=V'V=VV',$$

可知当 $\sigma_1>\sigma_2>\cdots>\sigma_p>0$ 时，对 U，V 作适当的限制后，(11）式相应的分解是唯一的。比较（11）式两端的自由变元的数目，左端是 np 个，右端是 $np-\frac{1}{2}p(p+1)+p+\frac{1}{2}p(p-1)=np$ 个。因此，可以利用矩阵微分的工具来导出相应的雅可比行列式，这样来获得奇异值 $\sigma_1>\sigma_2>\cdots\sigma_p>0$ 的联合分布密度。另一种方法是注意到关系式：$\sigma_i=\sqrt{\lambda_i}$，$\lambda_1>\lambda_2>\cdots>\lambda_p>0$ 是 $X'X$ 的全部特征根，而 $X'X$ 的特征根的联合分布前面已经讨论过了，于是只消作一个很简单的变换，就可以获得 $\sigma_1,\cdots,\sigma_p$ 的联合分布。

上面的这些讨论使我们对矩阵的分解与求雅可比行列式之间的关系会有深刻的印象，但对于如何能使非线性的变换实质上线性化了，不易体会，除非自己认真地逐个计算前面介绍的那些公式。下面再用一个例子来具体说明这一点。矩阵微分在形式上和普通的微分可以完全一样有相应的各种公式，这一点是很方便就可以证明的，只要注意遇到乘法时不要随便改动顺序，因此就有

$$d(XY)=(dX)Y+X(dY) \tag{12}$$

等等这一类的公式。现在来看 $\underset{n\times n}{Y}=\underset{n\times n}{X}^{-1}$ 这一变换所相应的雅可比行列式应怎样来求。这是一个非线性的变换。由于有关系式

$$XY=I_n,$$

上式两端同时求微分，就得，

$$(dX)Y+X(dY)=0,$$

$$\therefore\quad dY=-X^{-1}(dX)Y=-X^{-1}(dX)X^{-1}.$$

上式就告诉我们：阵 dY 的元素是阵 dX 元素的线性函数，用 Y_* 代 dY，X_* 代 dX，就得

$$Y_*=-X^{-1}X_*X^{-1}. \tag{13}$$

容易看出 $\underset{n\times n}{Y_*}=\underset{n\times n}{A}\ \underset{n\times n}{X_*}\ \underset{n\times n}{B}$ 相应的雅可比行列式是 $|A|^n|B|^n$。这只消先求 $\tilde{Y}=A\tilde{X}$，$Y_*=\tilde{Y}B$ 相应的雅可比行列式，然后用复合函数微分的公式立即就可得到。因此变换（13）相应的雅可比就是

$$|X|^{-2n},$$

它也就是 $Y=X^{-1}$ 这一变换所相应的雅可比行列式．如果我们限制考虑 $Y=\underset{n\times n}{X^{-1}}$，$X$ 是一对称阵，即 $X'=X$，此时 X 与 Y 都只有 $\frac{n}{2}(n+1)$ 个变元．虽然（13）式依然是成立的，但由于 Y_*，X_*，X 均为对称阵，虽然 Y_* 的元素仍然是 X_* 的元素的线性组合，雅可比行列式不能由变换 $\tilde{Y}=A\tilde{X}, Y_*=\tilde{Y}B$ 复合来获得．如果将 Y_* 中的自由元素 $\frac{n}{2}(n+1)$ 个逐个写出，然后计算 $\frac{n}{2}(n+1)$ 阶的行列式，这就比较麻烦．此时就考虑 $Y'=Y$，$X'=X$，且 $Y=AXA'$，$|A|\neq 0$ 的变换．利用非退化阵可以表成一系列初等变换矩阵的乘积．因此只消考虑 A 是初等变换阵，这样就把问题大大地简化了，很方便就可以得到 $Y=AXA'$，$|A|\neq 0$ 的雅可比是 $|A|^{n+1}$．因此当 $Y=X^{-1}$，$X'=X$ 时，相应的雅可比行列式就是 $|X|^{-(n+1)}$．

从这些讨论中，不难看出矩阵微分这一工具确实是很方便的，它能充分发挥矩阵运算的一些特点来简化证明计算．因此，这一工具是学习数理统计的人值得注意的．如要掌握它也并不困难，只要钻研［1］，并在此基础上，做一做［4］中有关的一些习题，就可以在一定程度上应用这一工具了．

参 考 文 献

［1］ Deemer, Walter L. aud Ingram Olkin (1951), The Jacobians of certain matrix transformations useful in multivariate analysis, *Biomertrika*, 38, 345—367.

［2］ 翁世沛（1981），不变分布及其导出分布（同济大学研究生毕业论文）.

［3］ 水电八局，武汉大学应用数学研究室（1978），关于某些统计量分布的推导，武汉大学学报（自然科学版）1978 年.

［4］ Narayan C. Giri (1977), Multivariate Statistical Inference, Academic Press.

Matrix Differentials

By Zhang Yaoding (张尧庭)

Abstract

In this paper, we introduce the method of matrix differentials, by which we can derive some exact distributions; In 1947, Prof. P. L. Hsu gave lectures in multivariate analysis at the University of North Carolina, in which he developed the matrix differential technigue for finding Jacobians of some matrix transformations. In 1951, W. L. Deemer and I. Olkin prepared an expository paper giving a systematic development with all proofs. By this method, we got some exact distributions, which appeared in [2] and other papers.