新世纪经济学论丛

复杂数据下经济建模与诊断研究

赵进文 ◉ 著

科学出版社
www.sciencep.com

新世纪经济学论丛

复杂数据下经济建模与诊断研究

赵进文　著

（国家社会科学基金资助项目，批准号：02BTJ002）

科　学　出　版　社
北　京

内 容 简 介

本书主要介绍了统计与经济计量建模研究中的最新理论研究与应用成果，尤其侧重于非经典建模理论中的模型诊断、影响分析、稳健分析、非线性分析、向量自回归模型、多元响应回归模型、分类数据模型、单位根分析、协整理论与误差修正模型、随机脉冲响应分析、方差分量分解理论等的研究与创新。

本书的理论性与实践性皆很强，可作为从事经济建模等领域的研究者的参考书使用。

图书在版编目(CIP)数据

复杂数据下经济建模与诊断研究/赵进文著.—北京:科学出版社,2004
(新世纪经济学论丛)
ISBN 7-03-012528-2

Ⅰ.复…　Ⅱ.赵…　Ⅲ.经济模型　Ⅳ.F224.0

中国版本图书馆 CIP 数据核字(2003)第 117629 号

责任编辑:陈　亮/责任校对:钟　洋
责任印制:安春生/封面设计:耕者设计工作室

科学出版社 出版
北京东黄城根北街16号
邮政编码:100717
http://www.sciencep.com
丽源印刷厂 印刷
科学出版社发行　各地新华书店经销
*
2004年2月第 一 版　开本:A5(890×1240)
2004年2月第一次印刷　印张:8 1/8
印数:1—2 000　字数:246 000

定价:22.00元

(如有印装质量问题,我社负责调换〈新欣〉)

前　言

统计诊断问题的研究，最早可以追溯到 18 世纪 70 年代中期 Bernoulli D(1777)的文章，尽管它的正式发表是在1961年。19世纪中期，以 Airy G B, Gould B A J, Wintlock J 和 Stone E J 为代表的天文学家一直对观察到的天文异常值数据给予极大关注，但直到 19 世纪后期及 20 世纪 30 年代，此类问题才再次引起统计学家的兴趣。Newcomb S(1886)研究了如何组合观察值以获得最佳拟合效果的一般理论。McMay A T (1935)讨论了极端点与抽样分布之间的关系。此后，有关统计诊断问题的研究文献呈直线上升趋势。然而，真正作为统计学中一门新兴的、独立学科分支出现，则是 20 世纪 70 年代中期的事。到了 20 世纪 80 年代，统计诊断学已发展成为统计学中最富有生机和活力的前沿性学科之一，并相继出版了该领域若干经典性学术专著，例如：Belsley D A, Kuh E 和 Welsch(1980)的 *Regression Diagnostics*, Huber, P(1981)的 *Robust Statistic*, 以及 Cook R D 和 Weisberg S(1982)的 *Residuals and Influence in Regression* 等等。这些著作的出版，为统计诊断学的进一步发展奠定了坚实的理论基础。

20 世纪 80 年代后期，统计诊断学研究热点开始转向非线性模型、多元统计模型、广义模型、随机效应模型及非线性固定约束与随机约束模型等复杂问题的影响评价与诊断技术的发展和完善。例如，Hossain A 和 Naik D N(1989)讨论了多元统计模型中的统计诊断问题等等。进入 20 世纪 90 年代后，统计诊断的研究更是如火如荼。人们从更深层次、更广范围、更多角度，对统计模型及与之相关的数据质量问题进行了深入探讨，并取得了丰硕的研究成果。在非平稳、非线性时间序列分析的诊断与应用方面的成果尤其引起统计学家与经济计量学家的极大关注。例如，Philip H F 和 Niels H(1994)首先研究了 AO 异常对单位根和协整检验的影响；Regina C C, Marcio G P G 和 Pierre P(1999)探讨了突发性政府干预对单位根过程检验的影响，并具体应用于巴西金融危机的研究等等。

在我国，有关统计诊断理论的系统研究起步于20世纪80年代初。陈希孺院士与王松桂教授率先将统计诊断的理论系统地介绍到国内，并持之以恒地进行了卓有成效的研究，1987年出版了该领域国内第一本学术专著：《近代回归分析——原理、方法及应用》，为培养我国在该领域开展研究、追赶世界前沿的学术梯队做出了奠基性的重大贡献。韦博成教授在国家自然科学基金资助下先后于1989年、1991年出版了至今仍是国内最具权威性的学术专著：《近代非线性回归分析》与《统计诊断引论》（与鲁国斌、史建清合著）。这些专著的出版，系统地总结了20世纪90年代前有关统计诊断研究的最新发展。此外，著者于2000年出版了《经济计量诊断学》，在国内率先将统计诊断理论系统地延伸到经济计量学领域，在许多方面丰富和发展了模型的诊断技术，具体应用于经济学不同领域的研究。吴喜之教授、王学仁教授等一批国内知名学者在统计模型的诊断研究方面，为我国赢得了荣誉。

针对不同的统计模型，国内外统计学家已发展了数量可观的统计诊断技术与方法。例如，Cook距离、广义Cook距离、WK统计量、AP统计量、信息比统计量、协方差比统计量、置信域体积比统计量、似然距离、统计曲率型统计量、Kullback-Leibler距离法、Score统计量、Box-Tiao方法、Chaloner-Brant方法、复共线性强影响点的Walker诊断法等。著者还提出了复共线性强影响点的主成分诊断法、检验强影响点的T诊断法、序列相关检验与异方差检验的诊断技术以及模型结构变动的诊断技术等。这些方法与技术各有其应用环境与条件，互为补充，相辅相成。然而，这些方法也存在许多不足，甚至是严重的缺陷，尤其对于高维复杂数据及多向分类数据模型下的诊断问题，其效果并不尽如人意。此外，这些方法大多局限于常规统计数据模型，对于新近发展与兴起的非常规统计问题的建模与诊断，例如，平行数据、截面数据、方向数据、空间统计数据、相依数据、非平稳非线性非对称数据等情况的影响度量、诊断技术及稳健建模理论等，还有待深入开展研究，系统构建其理论分析框架，并有效地应用于各种实际问题。

2002年，全国哲学社会科学规划办公室将“复杂数据的统计诊断方法及其应用”列入国家社会科学基金资助项目，著者有幸获得资助（批准号：02BTJ002），从而能够顺利地开展相关问题的研究。本书是该项目研究的最终成果，主要介绍最新的理论进展、方法创新、应用建

模实践等，尤其侧重于非经典建模理论中的模型诊断、影响分析、稳健理论、非线性分析、向量自回归模型、多元响应回归模型、分类数据模型、分布滞后模型、单位根检验、协整理论与误差修正模型、随机脉冲响应分析、方差分量分解理论等的研究与创新。

全书分两大部分共 14 章。第一部分为理论与方法研究，包括 9 章，主要介绍著者的最新研究成果；第二部分为建模与应用研究，包括 5 章，重点介绍理论与方法在中国经济问题中的最新应用与实践。除此之外，我们在中国利率市场化模型构建、稳健货币政策操作规则的创立、财政政策与货币政策实施效果的模型评价、人民币汇率形成机制、经济结构变动分析等方面取得了十分理想的阶段性应用研究成果，但由于篇幅的限制，未能在本书中体现。对此，我们将以论文和研究报告的形式予以发表，并继续开展深入细致的研究，择时集结出版。

在本书付梓出版之际，首先感谢全国哲学社会科学规划办公室对本项研究课题的大力资助，以及国内同行专家的积极支持与充分信任，并对东北财经大学各级领导、各个部门的全力配合表示衷心的感谢。在此，我要特别感谢东北财经大学邱东教授、艾洪德教授、马国强教授、郭连成教授、蒋萍教授，国家统计局贺铿副局长，中国科学院研究生院陈希孺院士，清华大学李子奈教授，中国人民大学袁卫教授、赵彦云教授，厦门大学曾五一教授，上海财经大学张尧庭教授，西南财经大学庞皓教授，东南大学韦博成教授，北京工业大学王松桂教授，天津财经学院周逸江教授，南开大学谷书堂教授、陈宗胜教授、柳欣教授，中国社会科学院汪同三研究员等专家的大力帮助。

在课题研究的过程中，得到了国际知名专家的大力支持与鼓励，他们是：诺贝尔经济学奖得主、美国宾夕法尼亚大学 Klein L R 教授，经济学大师、哈佛大学 Jorgensen D W 教授，美国联邦储备委员会资深经济学家 Ericsson N R 教授，著名经济计量学家、美国南加利福尼亚大学肖政（Cheng Hsiao）教授以及威斯康星大学 Durlauf S 教授等。在此，一并表示感谢。

最后，我要诚挚地感谢我的妻子邵国莉女士的鼎力协助与全心呵护。她承担了全部书稿的文字录入与复杂枯燥的数学公式排版工作，并在勤勉工作的同时操持繁重的家务，使我能全身心地投入科研与教学工作。

值得指出的是，由于时间仓促及本人学术水平所限，书中难免有错漏与不妥之处，敬请各界专家、同仁批评、指正！

赵进文

2003年8月于东财园

目　　录

第1章 预备知识

不言而喻,进行模型定量化研究离不开数学作为分析工具。本书以统计模型与经济计量学模型为研究对象,更离不开坚实的数学基础。为使阅读更顺畅,研究更深入,有必要将一些重要的定理、公式、结论做一预备性介绍。这些预备知识在一般教科书中并不容易找到,但它们是进行宏观、微观建模分析研究的工具。

1.1 矩阵代数

1.1.1 分块矩阵与"和式求逆"公式

设矩阵 A 的分块形式为

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \tag{1.1.1}$$

并假定以下公式中所涉及的逆矩阵均存在。

设 A 为方阵,则 A 的行列式 $|A|$ 可表示为

$$\begin{aligned} |A| = \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} &= |A_{11}| \cdot |A_{22} - A_{21}A_{11}^{-1}A_{12}| \\ &= |A_{22}| \cdot |A_{11} - A_{12}A_{22}^{-1}A_{21}| \end{aligned} \tag{1.1.2}$$

特别地,若取 $A_{11} = A, A_{12} = M, A_{22} = I, A_{21} = N$,则有

$$|A - MN| = |A| \cdot |I - NA^{-1}M| \tag{1.1.3}$$

矩阵 A 的逆矩阵可表示为

$$\begin{aligned} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1} &= \begin{pmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}A_{22.1}^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}A_{22.1}^{-1} \\ -A_{22.1}^{-1}A_{21}A_{11}^{-1} & A_{22.1}^{-1} \end{pmatrix} \\ &= \begin{pmatrix} A_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} A_{11}^{-1}A_{12} \\ -I \end{pmatrix} A_{22.1}^{-1} (A_{21}A_{11}^{-1}, -I) \end{aligned} \tag{1.1.4}$$

或者

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1} = \begin{pmatrix} A_{11.2}^{-1} & -A_{11.2}^{-1}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}A_{11.2}^{-1} & A_{22}^{-1}+A_{22}^{-1}A_{21}A_{11.2}^{-1}A_{12}A_{22}^{-1} \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 0 \\ 0 & A_{22}^{-1} \end{pmatrix} + \begin{pmatrix} -I \\ A_{22}^{-1}A_{21} \end{pmatrix} A_{11.2}^{-1}(-I, A_{12}A_{22}^{-1}) \tag{1.1.5}$$

其中，$A_{11.2}=A_{11}-A_{12}A_{22}^{-1}A_{21}$，　$A_{22.1}=A_{22}-A_{21}A_{11}^{-1}A_{12}$。

假定$|A_{11}|\neq 0$，$|A_{22}|\neq 0$，则由以上两个四块求逆公式，可以推出极其重要的矩阵和式求逆公式。比较(1.1.4)与(1.1.5)式左上角元素，有

$$(A_{11}-A_{12}A_{22}^{-1}A_{21})^{-1}=A_{11}^{-1}+A_{11}^{-1}A_{12}A_{22.1}^{-1}A_{21}A_{11}^{-1}$$

又在此公式中重新记：$A_{11}=A$，$A_{12}=-M$，$A_{22}=I$，$A_{21}=N$，则有如下矩阵和式求逆公式：

矩阵和式求逆公式　设 A 为 n 阶方阵，M、N 分别为 $n\times p$ 和 $p\times n$ 阶矩阵，并假定以下有关的逆矩阵均存在，则有

$$(A+MN)^{-1}=A^{-1}-A^{-1}M(I+NA^{-1}M)^{-1}NA^{-1} \tag{1.1.6}$$

此公式在模型诊断与影响分析中有广泛的应用。由式(1.1.6)还可得到以下两个常用公式：

$$\begin{aligned}&(X^TX-X_1^TX_1)^{-1}\\ &=(X^TX)^{-1}+(X^TX)^{-1}X_1^T(I-X_1(X^TX)^{-1}X_1^T)^{-1}X_1(X^TX)^{-1}\end{aligned} \tag{1.1.7}$$

$$(I-A)^{-1}=I+(I-A)^{-1}A=I+A(I-A)^{-1} \tag{1.1.8}$$

其中，X 为 $n\times p$ 阶矩阵，X_1 为 $m\times p$ 阶矩阵，$m<n$。

1.1.2　投影矩阵与二次投影公式

设 X 为 $n\times p$ 阶矩阵，其列向量在欧氏空间 R^n 中生成的线性空间记为$L(X)$。$L(X)$空间上的正交投影矩阵(简称投影阵)记为 P_X，在不致引起混淆时亦记为 P。$L(X)$的正交补空间上的投影阵记为 Q_X，在不致引起混淆时亦记为 Q。正交投影阵 P_X 满足下列关系：

$$P_X^T=P_X=P_X^2=X(X^TX)^{-1}X^T, Q_X=I-P_X \tag{1.1.9}$$

其中，$(X^TX)^{-1}$表示矩阵 X^TX 的广义逆矩阵，在今后的讨论中大多假定 X^TX 可逆。

今设 X 可写成分块形式 $X=(X_1,X_2)$,X_1 和 X_2 分别为 $n\times k$ 和 $n\times(p-k)$阶矩阵。若 $X_1^TX_2=0$,则 $L(X_1)$和 $L(X_2)$正交。此时,有

$$L(X)=L(X_1)\oplus L(X_2)$$

$$P_X=P_{X_1}+P_{X_2}$$

其中,$\oplus$表示线性空间的直和。若 $X_1^TX_2\neq0$,则可把 $L(X_2)$进行适当“旋转”,使之与 $L(X_1)$垂直。显而易见,下式成立:

$$L(X_1,X_2)=L(X_1,P_{X_1}X_2+Q_{X_1}X_2)=L(X_1,Q_{X_1}X_2)$$

再由 $X_1^T(Q_{X_1}X_2)=0$,有

$$L(X)=L(X_1)\oplus L(Q_{X_1}X_2)$$

由此可得如下重要的二次投影公式:

二次投影公式 设 $n\times p$ 阶矩阵 X 可分块为 $X=(X_1,X_2)$,则投影阵 P_X 可表示为

$$P_X=P_{X_1}+P_{Q_{X_1}X_2}=P_{X_1}+P_{(1-P_{X_1})X_2} \tag{1.1.10}$$

此公式表示,在线性空间 $L(X)$上的投影可以分成两次来进行:首先求出在 $L(X_1)$上的投影,然后再求其正交补 $L(Q_{X_1}X_2)$上的投影,后者由 $Q_{X_1}X_2$ 生成。

1.1.3 对称矩阵的谱分解

设 A 为n 阶对称矩阵,$\mathrm{rank}(A)=k\leqslant n$,又设 $\lambda_1,\lambda_2,\cdots,\lambda_k$ 为A 的非零特征值,$\gamma_1,\gamma_2,\cdots,\gamma_k$ 为相应的单位特征向量。记

$$\Lambda=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_k),\quad \Gamma=(\gamma_1,\gamma_2,\cdots,\gamma_k)$$

则由 $A\gamma_i=\lambda_i\gamma_i,i=1,2,\cdots,k$,可得 $A\Gamma=\Gamma\Lambda$。即有

$$A=\Gamma\Lambda\Gamma^T,\qquad \Gamma^T\Gamma=I_k \tag{1.1.11}$$

其中,I_k 为k 阶单位矩阵。称(1.1.11)为 A 的**谱分解**。若 A 为n 阶非负对称矩阵,则可设 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_k\geqslant0$,于是可将 A 表示为

$$A=\Gamma\Lambda^{\frac{1}{2}}\Gamma^T\Gamma\Lambda^{\frac{1}{2}}\Gamma^T=B^2$$

$$B=\Gamma\Lambda^{\frac{1}{2}}\Gamma^T$$

$$\Lambda^{\frac{1}{2}}=\mathrm{diag}(\sqrt{\lambda_1},\sqrt{\lambda_2},\cdots,\sqrt{\lambda_k})$$

通常,记 $B=A^{\frac{1}{2}}$。进一步,若 A 为正定矩阵,则 $B=A^{\frac{1}{2}}$亦为正定矩

阵。又 $A^{-1}=(B^{-1})^2$,因此,通常记 $B^{-1}=A^{-\frac{1}{2}}$。

1.1.4 任意矩阵的奇异值分解

设 A 为 $m\times n$ 阶矩阵,则 A^TA 为 n 阶对称矩阵。对 A^TA 施行谱分解,则存在正交矩阵 U,满足

$$A^TA = U\mathrm{diag}(\lambda_1^2,\lambda_2^2,\cdots,\lambda_n^2)U^T$$

其中,$\lambda_1^2,\lambda_2^2,\cdots,\lambda_n^2$ 为 A^TA 顺序特征值 $\lambda_1^2\geqslant\lambda_2^2\geqslant\cdots\geqslant\lambda_n^2$。又设 $\mathrm{rank}(A)=r$,则有

$$\lambda_1^2\geqslant\lambda_2^2\geqslant\cdots\geqslant\lambda_r^2>0,\quad \lambda_{r+1}^2=\cdots=\lambda_n^2=0$$

记 $d_i=(\lambda_i^2)^{\frac{1}{2}},i=1,\cdots,r$,并称之为 A 的奇异值。令 $D_1=\mathrm{diag}(d_1,\cdots,d_r)$,$D=\begin{bmatrix}D_1 & 0\\ 0 & 0\end{bmatrix}$,并将 U 作分块:$U=(\underset{r}{U_1}\ \vdots\ U_2)$,则有

$$A^TA = U_1\mathrm{diag}(\lambda_1^2,\lambda_2^2,\cdots,\lambda_r^2)U_1^T\Rightarrow D_1^{-1}U_1^TA^TAU_1D_1^{-1} = I_r$$

记 $V_1\triangleq(v_1,\cdots,v_r)\triangleq AU_1D_1^{-1}$,则 V_1 是 $m\times r$ 阶列正交矩阵,且有

$$\begin{aligned}AA^TV_1 &= AA^TAU_1D_1^{-1} = AU_1D_1^2U_1^TU_1D_1^{-1}\\ &= AU_1D_1^{-1}D_1^2 = V_1D_1^2\end{aligned}$$

可见,$v_1,\cdots,v_r$ 是 AA^T 的相应于 $\lambda_1^2,\cdots,\lambda_r^2$ 的正交规范化特征向量集。由 V_1 的定义知

$$V_1D_1U_1^T = AU_1U_1^T = A(I-U_2U_2^T) = A$$

将上式改写为

$$A = VDU^T \tag{1.1.12}$$

其中,V,U 分别为 m,n 阶正交矩阵,D 是以 A 的奇异值为主对角元,其余元素均为零的 $m\times n$ 阶矩阵。称上式为矩阵 A 的**奇异值分解**。

1.1.5 矩阵的向量化运算和 Kronecker 乘积

设 A 为 $n\times m$ 阶矩阵,则 A 的向量化运算为

$$\mathrm{Vec}(A) = (A_1^T,A_2^T,\cdots,A_m^T)^T$$

其中,$A_1,A_2,\cdots,A_m$ 表示 A 的 m 个 n 维列向量。矩阵 A 的向量化亦称为 A 的拉直,亦记作 $\vec{A}$。它与矩阵的 Kronecker 乘积有密切的关系。

设 A 为 $n\times m$ 阶矩阵，B 为 $p\times q$ 阶矩阵，A 与 B 的 Kronecker 乘积或叉积记为 $A\otimes B$，它是一个 $np\times mq$ 阶矩阵，定义为

$$A\otimes B=(a_{ij}B)=\begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1m}B \\ a_{21}B & a_{22}B & \cdots & a_{2m}B \\ \vdots & \vdots & & \vdots \\ a_{n1}B & a_{n2}B & \cdots & a_{nm}B \end{bmatrix}$$

Kronecker 乘积除了满足通常的结合律、分配律外，还满足如下关系：

$$(A_1\otimes B_1)(A_2\otimes B_2)=(A_1A_2)\otimes(B_1B_2)$$

$$(A\otimes B)^T=A^T\otimes B^T$$

$$(A\otimes B)^{-1}=A^{-1}\otimes B^{-1}$$

$$\mathrm{tr}(A\otimes B)=(\mathrm{tr}A)(\mathrm{tr}B)$$

若 $A\geqslant 0, B\geqslant 0$，则 $A\otimes B\geqslant 0$。

以下性质经常用到，应予特别注意：

$$\mathrm{Vec}(ABC)=(C^T\otimes A)\mathrm{Vec}(B) \tag{1.1.13}$$

$$\begin{aligned}\mathrm{Vec}(AB)&=(I\otimes A)\mathrm{Vec}(B)\\ &=(B\otimes I)\mathrm{Vec}(A)\end{aligned} \tag{1.1.14}$$

1.1.6 矩阵微商

矩阵微商在模型参数估计和性能评价中有重要应用，它是普通函数导数（微商）的推广。

先介绍矩阵对标量的微商。设 $Y=(y_{ij}(t))$ 为 $p\times q$ 阶矩阵，并且，它的元素为 t 的函数。现定义矩阵对标量的微商为

$$\frac{\partial Y}{\partial t}=\left(\frac{\partial y_{ij}(t)}{\partial t}\right)_{p\times q}=\begin{bmatrix} \frac{\partial y_{11}}{\partial t} & \cdots & \frac{\partial y_{1q}}{\partial t} \\ \vdots & & \vdots \\ \frac{\partial y_{p1}}{\partial t} & \cdots & \frac{\partial y_{pq}}{\partial t} \end{bmatrix}$$

可以证明，下述关系式成立：

(1) $\dfrac{\partial(X+Y)}{\partial t}=\dfrac{\partial X}{\partial t}+\dfrac{\partial Y}{\partial t}$；

(2) $\dfrac{\partial(XY)}{\partial t}=\dfrac{\partial X}{\partial t}Y+X\dfrac{\partial Y}{\partial t}$；

(3) $\dfrac{\partial(X\otimes Y)}{\partial t}=\dfrac{\partial X}{\partial t}\otimes Y+X\otimes\dfrac{\partial Y}{\partial t}$;

(4) $\left(\dfrac{\partial X}{\partial t}\right)^T=\dfrac{\partial X^T}{\partial t}$;

(5) $\dfrac{\partial X}{\partial x_{ij}}=E_{ij}$,式中 E_{ij}表示第(i,j)元素为1,其余元素均为0的矩阵;

(6) $\dfrac{\partial AXB}{\partial x_{ij}}=AE_{ij}B$。

其次,我们介绍矩阵标量函数对矩阵的微商。设 $y=f(X)$是 $m\times n$ 矩阵 X 的函数,则 y 对 X 的微商定义为

$$\frac{\partial y}{\partial X}=\left(\frac{\partial y}{\partial x_{ij}}\right)_{m\times n}=\begin{pmatrix}\dfrac{\partial y}{\partial x_{11}} & \cdots & \dfrac{\partial y}{\partial x_{1n}}\\ \vdots & & \vdots\\ \dfrac{\partial y}{\partial x_{m1}} & \cdots & \dfrac{\partial y}{\partial x_{mn}}\end{pmatrix}$$

可以证明,如下事实成立:

(1) $\left(\dfrac{\partial f(X)}{\partial X}\right)^T=\dfrac{\partial f(X)}{\partial X^T}$;

(2) $\dfrac{\partial \mathrm{tr}(X)}{\partial X}=I$;

(3) $\dfrac{\partial \mathrm{tr}(AXB)}{\partial X}=A^TB^T$;

(4) $\dfrac{\partial \mathrm{tr}(AX)}{\partial X}=\begin{cases}A^T\\ A+A^T-\mathrm{diag}(a_{11},\cdots,a_{nn}),X=X^T\end{cases}$;

(5) $\dfrac{\partial x^TAx}{\partial x}=(A+A^T)x$,这里,$x$ 为列向量。

第三,我们介绍向量对向量的微商。设 x 为 n 维列向量,而

$$y=\begin{pmatrix}f_1(x)\\ \vdots\\ f_m(x)\end{pmatrix}=\begin{pmatrix}y_1\\ \vdots\\ y_m\end{pmatrix}$$

则 y 对 x 的微商定义为

$$\frac{\partial y^T}{\partial x} = \begin{pmatrix} \frac{\partial y^T}{\partial x_1} \\ \vdots \\ \frac{\partial y^T}{\partial x_n} \end{pmatrix} = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_1} \\ \vdots & & \vdots \\ \frac{\partial y_1}{\partial x_n} & \cdots & \frac{\partial y_m}{\partial x_n} \end{pmatrix}$$

不难证明,下述事实成立:

(1) 设 $y = Ax$,则$\frac{\partial y^T}{\partial x} = A^T$;

(2) 设 $Y = AxB$,则$\frac{\partial(\mathrm{Vec}(Y))^T}{\partial(\mathrm{Vec}(x))} = B \otimes A^T$。

最后,我们介绍矩阵的微分。设 $f(X)$是一个 $m \times n$ 矩阵 X 的函数,它的微分可仿照多元函数的微分定义如下:

$$\mathrm{d}f = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{\partial f}{\partial x_{ij}} \mathrm{d}x_{ij} = tr\left[\left(\frac{\partial f}{\partial X} \right)^T \mathrm{d}X \right]$$

其中,$\mathrm{d}X = (\mathrm{d}x_{ij})$为 $m \times n$ 矩阵。可见,它与多元函数 $f(X) = f(x_1, \cdots, x_n)$的微分形式:

$$\mathrm{d}f = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \mathrm{d}x_i = \frac{\partial f}{\partial x^T} \mathrm{d}x$$

十分相似,式中,$\mathrm{d}x = (\mathrm{d}x_1, \cdots, \mathrm{d}x_n)^T$。

矩阵微分满足如下性质:

(1) $\mathrm{d}(X + Y) = \mathrm{d}X + \mathrm{d}Y$;

(2) $\mathrm{d}(kX) = k\mathrm{d}X$,$k$ 为任意实数;

(3) $(\mathrm{d}X)^T = \mathrm{d}(X^T)$;

(4) $\mathrm{d}(\mathrm{tr}(X)) = \mathrm{tr}(\mathrm{d}X)$;

(5) $\mathrm{d}(XY) = (\mathrm{d}X)Y + X(\mathrm{d}Y)$;

(6) $\mathrm{d}(X \otimes Y) = (\mathrm{d}X) \otimes Y + X \otimes (\mathrm{d}Y)$。

如同普通函数的微分与导数关系一样,矩阵微商是两个矩阵微分的商。因此,可通过矩阵微分方便地求出矩阵的微商。

设 $f(X)$是一个 $m \times n$ 矩阵 X 的标量函数,且存在矩阵 A,使下式成立:

$$\mathrm{d}f = \mathrm{tr}(A^T \mathrm{d}X)$$

则有