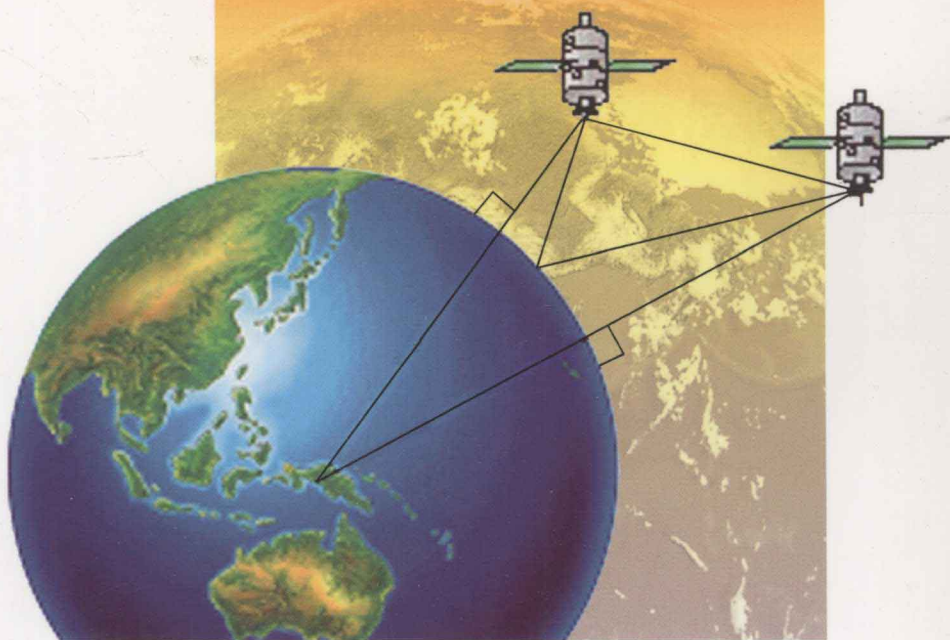


方向控制最小二乘法 理论

许君一 卿熙宏 刘国林 著



Direction Control Least Square Theory



测绘出版社

方向控制最小二乘法理论

Directional Control Least Squares Theory

许君一 卿熙宏 刘国林 著

测绘出版社

· 北京 ·

©许君一 卿熙宏 刘国林 2010

所有权利(含信息网络传播权)保留,未经许可,不得以任何方式使用。

内 容 简 介

本书在总结近年来非线性测量平差最新成果的基础上,用黎曼空间和流形的观点论述了新的线性和非线性最小二乘理论。主要包括:分析学基础、线性代数、张量与微分几何的基本概念、概率统计基础知识,基于方向控制的非线性最小二乘法的常用基本算法、非线性度量理论、非线性最小二乘平差神经网络方法等。

本书侧重方向控制的非线性平差理论,叙述力求深入浅出。可作为测绘工程专业、遥感科学与技术专业高年级大学生和研究生学习“现代测量误差理论与数据处理”的教学参考书,亦可供有关测绘与各类工程专业的教师、科研和工程技术人员参考。

图书在版编目(CIP)数据

方向控制最小二乘法理论/许君一,卿熙宏,刘国林
著. —北京:测绘出版社, 2010. 8
ISBN 978-7-5030-2086-5

I. ①方… II. ①许… ②卿… ③刘… III. ①最小
二乘法平差—研究 IV. ①P207

中国版本图书馆 CIP 数据核字(2010)第 145878 号

责任编辑	杨蓬莲	封面设计	李 伟	责任校对	董玉珍 李 艳
出版发行	测绘出版社				
地 址	北京市西城区三里河路 50 号			电 话	010—68531160(营销)
邮政编码	100045				010—68531609(门市)
电子信箱	smp@sinomaps.com			网 址	www.sinomaps.com
印 刷	北京金吉士印刷有限责任公司			经 销	新华书店
成品规格	184mm×260mm				
印 张	9.5			字 数	240 千字
版 次	2010 年 9 月第 1 版			印 次	2010 年 9 月第 1 次印刷
印 数	0001—1000			定 价	28.00 元

书 号 ISBN 978-7-5030-2086-5/P·474

如有印装质量问题,请与我社联系调换。

前 言

自从勒让德和高斯提出最小二乘法以来,线性最小二乘理论始终是平差的理论基础。非线性科学理论和非线性数学解算方法的最大特点是多学科的交叉。对于非线性模型和病态问题,是否存在更有效和统一的处理手段,始终是理论工作者和工程技术人员所追求的目标。针对非线性最小二乘平差问题,P. Vanick 的微分几何与张量代数的观点和 G. Blaha 和 P. J. G. Teunissen 的几何观点,对非线性最小二乘平差问题都是里程碑式的进步。

全书共分 7 章。第 1 章作为绪论,简要总结了非线性误差处理的基本概况;第 2 章介绍了基础知识;第 3 章介绍了最小二乘法基本理论;第 4 章介绍了方向控制最小二乘参数平差;第 5 章介绍了方向控制平差模型非线性度量;第 6 章介绍了非线性最小二乘测量平差神经网络方法;第 7 章介绍了最小二乘法的方向核空间问题。

本书能够顺利完成,与山东科技大学测绘科学与工程学院院长卢秀山教授的大力支持、鼓励和无私帮助是分不开的,在此表示由衷的敬意和感谢。在本书的写作过程中郑作亚副教授提供了许多宝贵的信息和建议,其中,与波动有关的最小二乘法问题就来源于与郑作亚副教授的有关讨论,在此表示感谢;感谢我的同事以及给予我支持和帮助的所有人员。感谢本书引用的参考文献的作者。感谢测绘出版社编辑们的辛勤工作。没有山东科技大学测绘科学与工程学院领导的支持、没有山东科技大学测绘科学与工程学院遥感科学与技术系同事们的鼓励和测绘出版社老师们的无私奉献,该书的出版是不可能的。

该书的出版得到国家自然科学基金青年科学基金项目《基于组合约束优化的海上 GPS 浮标网络定位算法研究》的资助(项目编号:40904001)。

由于作者水平有限,经验不足,因此书中难免有不妥和谬误之处,恳请同行专家及广大读者不吝赐教。

作者

2009 年 12 月

目 录

第 1 章 绪 论	1
§ 1.1 非线性误差处理概论	1
§ 1.2 研究的目的、内容及方法	3
第 2 章 基础知识	4
§ 2.1 分析学基础	4
§ 2.2 变分	4
§ 2.3 矩阵与张量代数运算	6
§ 2.4 概率统计基础	20
§ 2.5 微分几何基础	32
§ 2.6 神经网络基础	40
§ 2.7 Hilbert-Huang 方法	42
第 3 章 最小二乘法基本理论	44
§ 3.1 最小二乘法与测量平差	44
§ 3.2 非线性参数平差的线性化法	59
§ 3.3 非线性参数平差的近似直接解法	62
§ 3.4 张量几何与最小二乘平差	66
§ 3.5 非线性最小二乘平差的泛函型法	69
第 4 章 方向控制最小二乘参数平差	75
§ 4.1 黎曼空间与线性最小二乘法	75
§ 4.2 黎曼空间与非线性最小二乘法	88
第 5 章 方向控制平差模型非线性度量	106
§ 5.1 概 述	106
§ 5.2 曲率度量的定义	106
§ 5.3 曲率立体阵与曲率度量公式的简化	108
§ 5.4 带权的非线性强度的曲率度量公式	111
§ 5.5 非线性平差模型的非线性诊断	114
§ 5.6 非线性平差模型的非线性度量——微分几何法	115
§ 5.7 非线性对参数估计及残差的影响	118

第 6 章 非线性最小二乘测量平差神经网络方法	121
§ 6.1 神经网络	121
§ 6.2 非线性最小二乘平差的神经网络方法	122
§ 6.3 神经网络模型及其稳定性分析	123
§ 6.4 自适应递推最小二乘法与神经网络	124
 第 7 章 最小二乘法的方向核空间问题	130
§ 7.1 方向核空间	130
§ 7.2 波动干扰下的方向核空间	134
§ 7.3 Morse 函数与方向核空间	139
 参考文献	141
 后 记	144

第1章 绪 论

§ 1.1 非线性误差处理概论

在相当长的时间里,传统测量领域都使用线性误差处理方式来进行误差处理。但是人们发现有许多时候处理的结果与实际情况存在较大差异,因为人们发现在地学领域非线性是一种普遍存在现象,而线性只是一种特例。于是人们开始尝试利用非线性的理论和方法来解决测绘中数据处理,比如在常规大地测量中的测边网、测角网以及边角同测网的条件平差和参数平差模型中都是非线性的条件方程式和误差方程式(陈俊勇等,2000,2007)。在大地测量网的优化设计问题中会涉及更多的非线性优化设计数学模型(李德仁,1998;Gunte,1998;李桂苓等,2001)。另外,在联合平差求解和参数估计中更需要进行非线性处理方法的研究(魏子卿,2003;周忠谟等,1999)。在变形观测数据处理中非线性模型能更好地揭示和反映变形的客观规律,也更能与实际变形相吻合(靳奉祥等,1995;兰孝奇等,2005;李德仁,1998)。

随着测绘领域高新技术的应用,特别是在高度发展的全球定位系统(GPS)和地理信息系统(GIS)中,许多非线性的误差理论和数据处理问题,如GPS中的基线向量的平差解算、网平差等的数据处理、坐标转换等,GIS中的栅格处理和边坡问题等的处理,成为测绘科技工作者不可回避和亟待解决的问题(程鹏飞等,2007;陶本藻,2000;王继周等,2003)。但是,由于过去处理非线性的数值计算方法和计算工具的限制以及非线性理论的发展不完善,在处理非线性问题时一直沿用泰勒级数展开并取至一次项,略去二次以上的高次项。实际上,在变形观测数据处理中,将非线性问题转化为线性模型来处理,不能客观反映实际变化过程,这是经过我们改造进行了取舍后的一种“似真”结果,并不能完全真实反映变形的客观实际变化情况(黄立人,2001)。从某种程度上说,“似真”处理影响了我们对客观世界的正确认识。因此,应采用非线性的数据处理办法来处理非线性数据,以试图完整准确描述客观的实际变化情况。特别是在一些变形观测和与时间相关联的空间测绘问题,在可能情况下对其客观的发展趋势做出较为准确预报(陶本藻,2000)。如果此意图得以实现,那么对各种变形监测必将带来实质性进展。因为随着社会的不断进步和发展,人们对各种变形的监测要求和精度都提出了较高的要求,甚至大到板块运动、地壳变形,小到局部地区水位变化、地表沉陷等,人们都希望能够从时间到具体空间上给出更精确的变形预测和分析(栾元重等,2002;朱文耀等,2000)。

笔者试图提供用非线性平差模型的估计理论,提出基于方向控制最小二乘原理以及基于神经网络平差方法的非线性化方法,发展非线性平差模型的非线性处理方法。以非线性的方法处理非线性的测量观测值,目的是获得更精确、更可靠的测量观测成果。在新算法研究中,通过试验来验证新算法的可靠性、精确性、客观真实性,为测绘成果的非线性处理模式提出更为科学、有效、快捷的新方法。

非线性科学理论和非线性数学解算方法的提出是20世纪自然科学的重大基础理论课题。非线性模型的研究最早始于20世纪60年代初期,在各门以非线性为特征的分支学科,如耗散

结构理论、分形理论、混沌理论、突变论、协同学等理论的基础上综合、交叉形成了一门新兴学科——非线性科学。但在 20 世纪 80 年代以前该学科发展一直比较缓慢,直到 1980 年,加拿大统计学家 Bates 和 Watts 引入曲率度量后,非线性模型理论才得到新的进展(Bates et al, 1980)。非线性科学在许多学科得到较为广泛的应用。据不完全统计,全国近 6 年来各个学科发表的非线性方面的研究论文 1 万余篇。由此可见非线性理论在近一段时期发展之迅速,研究成果之多可见一斑。1998 年到 2008 年的 10 年间,国内发表的非线性最小二乘法方面的文章还不足百篇,而发表有关 GPS 的非线性最小二乘算法方面的文章仅有 20 余篇。这也从另一个方面说明了具有 200 多年发展历史的最小二乘法和后来发展起来的非线性最小二乘算法在理论创新上的难度。

随着测绘科学和技术手段的发展使测绘观测值具有多源、多类型、多精度等特点,因而寻找处理具有多源、多类型、多精度特点的观测数据的有效算法引起测绘科学理论工作者的重视(陶华学等,2003;李述山等,2005)。测绘科学领域中普遍存在着非线性数学问题,如常规大地测量中的测边网、测角网以及边角同测网的条件平差和参数平差模型中,都是非线性的条件方程式和误差方程式。在大地测量网的优化设计问题中,也会涉及非常多的非线性优化设计数学模型。另外,在联合平差求解和参数估计中,更需要进行非线性处理方法的研究。在变形观测数据处理中,非线性模型能更好地揭示和反映变形的客观规律,也更能与实际变形相吻合。特别是目前高度发展的高新技术全球定位系统(GPS)和地理信息系统(GIS)中许多数据处理问题已经对非线性的误差理论和非线性数学描述提出了前所未有的要求(如 GPS 中的基线向量的平差解算、网平差等的数据处理、坐标转换等,GIS 中的栅格处理和边坡问题等的处理)。人们迫切需要找到有效的用非线性数学方法来解决测绘过程中遇到的非线性问题(刘大杰等,1999;魏子卿,1992;王宇平等,1997)。

过去之所以非线性理论在应用方面受到限制,特别是在测绘科学方面应用不多,这是因为非线性理论本身还不够完善。但自从 1980 年 Bates 和 Watts 又从微分几何的观点出发定义了模型的固有曲率和参数效应曲率,同时引进了与方向向量无关,仅与模型及其参数有关的曲率立体阵概念,为非线性程度的度量奠定了基础,解决了一部分非线性模型线性化近似后不能得到令人满意结果的问题,使非线性理论的实际应用更加完善和广泛。在非线性的测量平差模型中对参数进行估计之前,我们应对其非线性程度作出明确的判断。目前,国内外主要有 7 种非线性程度的度量指标:①最大固有曲率(maximum inherent curvature);②最大参数效应曲率(maximum parameter effect curvature);③二阶余项(second-order remainder term);④参数偏差(parameter deviation);⑤残差偏差(residual deviation);⑥类权 K 因子;⑦百分偏差(percentage deviation)。从实际应用的角度看,陶本藻先生推荐使用前两种(陶本藻,1998)。1997 年,王新洲教授对非线性模型线性化时所引起的参数偏差和随机误差的大小进行了假设检验(王新洲,1997),从而得出了非线性模型允许线性化的曲率度量准则,还进一步给出了非线性模型线性近似的容许曲率;1998 年,刘国林教授提出了非线性平差模型加权曲率度量的两类准则并以此作为可否线性化的判别依据。这些新研究成果的取得为把它们应用到测量数据处理方面提供了理论上的可能性。在过去相当长的时间里测绘学科研究误差传播也仅仅是研究线性的传播规律。但这要求未知数近似值与真值要充分接近且非线性模型的非线性程度较弱才能进行线性化。而实际上在很多情况下这种条件是难以满足的。因此线性化非线性函数模型必然影响到其真实性和可靠性。这种线性化处理的前提是观测值与其真值的偏离以及

所选未知数的近似值与其真值的偏离要相对较小。一般情况下,只有在模型非线性程度较弱的情况下才能得到较为满意的结果。否则,非线性模型的线性化不仅仅会影响到非线性模型反映事物的真实性,而且还会影响平差成果的精确性。所以,用非线性平差模型的估计理论突破传统的线性化方法,发展非线性平差模型的非线性处理方法以获得更精确、更可靠的观测成果是测绘工作者面对的技术问题,也是一种责任。

把非线性科学研究取得的最新成果和相应的数学算法引进测绘科学中加以运用、发展和完善,已有相当多科研工作者在这一领域做了许多有益的尝试和突出的成绩。山东科技大学陶华学教授长期以来一直致力于从事广义非线性动态最小二乘法理论与应用方面的研究,已经有许多研究成果和理论在实际测量工作中得以应用,并在处理多源、多维、多时态、多类型、多精度数据处理方面有独到的理论见解(陶华学等,1998,2003);刘国林教授在非线性的理论应用于测量平差中做了大量工作(刘国林等,1998,1999);韩晓东教授将非线性理论应用于GPS定位的研究(韩晓东,1999,2000);宁伟教授、郭金运教授运用非线性数学方法研究了测量数据的处理(宁伟等,2004;郭金运等,2002)。还有其他一些学者在这方面也进行了不少尝试和研究(彭军还等,2007;Pope,1974)。这些研究成果对开展该领域研究工作打下了良好的基础,也为本书的研究提供了有价值的参考和借鉴。

§ 1.2 研究的目的、内容及方法

正如前面所述,在实际测绘过程中遇到的最多的是非线性问题。过去我们出于计算的方便或者受数学解算的限制,大多数时候不得不把非线性的问题转化为线性问题来处理。但这样处理以后虽然使计算工作得到简化,计算速度大为提高,可我们也为此付出相当的代价。这样的处理使一些数据处理出现了失真,计算所得的结果并没有完全反映客观真实。如果使用这样计算出的参数来做预测,将使预测的结果偏离客观实际的发展趋势。而且随着预测时间周期的加长,失真的可能性也在不断地加大。一般情况下,这种加大不是呈线性的加大,而是呈几何级数的迅速增加。笔者采用非线性的数据处理办法来处理这些数据,以期完整准确客观地反映变形的实际变化情况。如果这些得以实现,那么对各种高精度动态变形监测有较大的推动作用。

非线性最小二乘问题是典型的非线性优化问题,所以可以利用非线性优化的有关方法来解决非线性最小二乘测量平差问题。笔者在前人研究成果的基础上,以最小二乘计算方法原理为基点,以最小二乘方法、非线性参数平差理论为基础,以 Hilbert-Huang 方法中的经验模式分解(Empirical Model Decomposition, EMD)为手段,提出了基于几何方向控制的非线性最小二乘法算法、EMD 频谱分解的非线性最小二乘法,基于张量模型的非线性最小二乘法及基于神经网络的非线性最小二乘法。这些方法在测量数据处理方面具有一定的理论价值和应用价值。这些方法也可以是基于几何方向控制的非线性最小二乘法算法。

第2章 基础知识

§2.1 分析学基础

泛函是把定义在函数域上的矢量函数空间映射到实数上的映射,即,函数的函数。泛函中的定义域是函数空间,其中的“点”是一个函数。下面两个定理的进一步说明,可参阅“高等数学引论”(华罗庚,1979)。

定理 2.1 (费马定理) 设 $f(x)$ 在 (a, b) 内有定义且在 $c \in (a, b)$ 上取极大(或极小)值。如果存在有限导数 $f'(c)$ ($|f'(c)| < +\infty$), 则一定有 $f'(c) = 0$ 。

定理 2.2 (达布定理) 设 $f(x)$ 在 $[a, b]$ 内有有限导数, 则 $f'(x)$ 至少有一次取介于 $f'(a)$ 和 $f'(b)$ 之间的每一个值。

定理 2.1 建立了极值与导数之间的必要关系。定理 2.2 给出了函数导数在区间取值的一种“介值”特性。令 $A = \{y \mid y = f'(x), x \in [a, b]\}$, $B = \{y \mid y = f'(x) \text{ 介于 } f'(a) \text{ 与 } f'(b) \text{ 之间}, x \in [a, b]\}$ 。根据定理 2.2, 有 $B \subseteq A$ 。如果 $f'(a) = f'(b) = 0$, 根据定理 2.2, 至少有一点 $c \in [a, b]$, 使得 $f'(c) = 0$ 。

极值与存在性条件:

对于多元函数 $f(x)$, $f: R^n \mapsto R^1$, 如果在存在 x^* 的邻域 $\delta(x^*)$ 中, 使得

$$f(x^*) \leq f(x) \text{ (或 } f(x^*) \geq f(x)), \forall x \in \delta(x^*) \quad (2-1)$$

则称 $f(x)$ 有极值, x^* 被称为局部极小点(或局部极大点)。如果不等式是严格的, 则称 x^* 为严格局部极小点(或严格局部极大点)。存在局部极值的必要条件是 $\nabla f(x^*) = 0$, $\nabla f(x) = (\partial f(x)/\partial x_1, \partial f(x)/\partial x_2, \dots, \partial f(x)/\partial x_n)^T$ 。存在局部极值的充分条件是 $x^T H(x^*) x > 0$, 其中 $H(x) = (\partial^2 f(x)/\partial x_i \partial x_j)$, 称 $H(x)$ 为海塞(Hessian)矩阵。

§2.2 变分

变分法是研究求泛函极值的方法。

定义 2.1 函数空间上点 x 的增量 $\delta x = x - x_0$ 称为点 x 的变分。

定义 2.2 令 $\delta x = x - x_0$ 。当 $\max |\delta x| \rightarrow 0$ 时, 泛函 $f(x)$ 在点 x_0 附近的增量 $\delta f(x_0) = f(x_0) \delta x$ 称为泛函 $f(x)$ 的变分。

一般情况下, 点 x 本身是函数。为了区别于普通的微分 dx , 这里采用了变分符号 δ 。对函数 x , 变分可理解为是变差, 也可以理解为是全微分。由变分定义知道, 变分是对泛函全微分或全变差的另一种称呼。由此我们不难得到有关变分的一些定理。

定义 2.3 如果泛函 $f(x)$ 在点 x_0 附近的值都不大于(小于) $f(x_0)$, 则说泛函 $f(x)$ 在点 x_0 达到极大(极小)值。

定理 2.3 如果具有变分的泛函 $f(x)$ 在点 x_0 有极大(极小)值, 则 $\delta f(x_0) = 0$ 。

一般泛函的极值问题中一类特殊的极值问题是变分问题。变分问题中,泛函定义域是特殊的定义域。其特殊之处在于,矢量函数空间点的分量是函数、函数自变量及其各阶导数构成,如 $(t, x, x', x'', \dots, x^{(n)})$ 。当涉及更多的函数的时候,用于变分计算的泛函的定义域由各个不同函数及其它们的各阶导数构成。如,考虑一维情形下的泛函极值问题。令 $v(x) = \int_a^b F(t, x, x') dt, x = x(t), t \in [a, b], F$ 三次可微;则 $\delta v = 0$ 的一个必要条件是满足欧拉方程,即

$$F_x - \frac{dF_{x'}}{dt} = F_x - F_{x'} - F_{x'x'}x' - F_{x'x''}x'' = 0 \quad (2-2)$$

把满足欧拉方程的解叫做极值曲线。所以,泛函的极值一定是极值曲线。但是,反过来未必成立。下面列出具有不同泛函定义域的变分问题。变分问题往往需要所谓的边值条件,如 $x(a) = x_0, x(b) = x_1$ 。如果我们得到了满足边值条件的极值曲线,则得到了泛函极值的一个解。

1. 一维变分

针对一维的不同情形,给出对应的欧拉方程。

(1) 如果 $F = F(t)$, 则在极值曲线上 $F'(t) = 0$ 。实际上,这个时候的极值曲线退化为一个实数轴上的点。这种情形下欧拉方程不存在。

(2) 如果 $F = F(x)$ 或 $F = F(t, x)$, 则在极值曲线上 $F_x = 0$ 。

(3) 如果 $F = F(x')$, 则在极值曲线上 $F_{x'} = 0$ 。

(4) 如果 $F = F(t, x')$, 则在极值曲线上 $F_{x'} + F_{x'x''}x'' = 0$ 。

(5) 如果 $F = F(x, x')$, 则在极值曲线上 $F_x - F_{x'x'}x' - F_{x'x''}x'' = 0$ 。

2. 高阶问题

令

$$v(x) = \int_a^b F(t, x, x', \dots, x^{(n)}) dt, x = x(t), t \in [a, b]$$

F 是 $n+2$ 次可微。则 $\delta v = 0$ 的一个必要条件满足欧拉-卜阿松方程,即

$$F_x + \sum_{i=1}^n (-1)^i \frac{d^i F_{x^{(i)}}}{dt^i} = 0 \quad (2-3)$$

推而广之,令

$$v(x) = \int_a^b F(t, x_1, x'_1, \dots, x_1^{(n_1)}, x_2, x'_2, \dots, x_2^{(n_2)}, \dots, x_m, x'_m, \dots, x_m^{(n_m)}) dt, x_k = x_k(t), t \in [a, b]$$

F 是 n_k+2 次可微, $k=1, 2, \dots, m$ 。则 $\delta v = 0$ 的一个必要条件满足欧拉-卜阿松方程,也即

$$F_{x_k} + \sum_{i=1}^{n_k} (-1)^i \frac{d^i F_{x_k^{(i)}}}{dt^i} = 0 \quad (k=1, 2, \dots, m) \quad (2-4)$$

3. 高维变分

三维空间的一个典型变分问题可写为

$$v(z(x, y)) = \iint_{\Omega} F\left(x, y, z, \frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}\right) dx dy$$

令 $p = \frac{\partial z}{\partial x}, q = \frac{\partial z}{\partial y}$ 。则有变分方程

$$F_z - \frac{\partial}{\partial x} \{F_p\} - \frac{\partial}{\partial y} \{F_q\} = 0 \quad (2-5)$$

其中,

$$\frac{\partial}{\partial x} \{F_p\} = F_{px} + F_{pz} \frac{\partial z}{\partial x} + F_{pp} \frac{\partial p}{\partial x} + F_{pq} \frac{\partial q}{\partial x}, \quad \frac{\partial}{\partial y} \{F_q\} = F_{qy} + F_{qz} \frac{\partial z}{\partial y} + F_{qp} \frac{\partial p}{\partial y} + F_{qq} \frac{\partial q}{\partial y}$$

令

$$v(z(x_1, x_2, \dots, x_n)) = \iiint_{\Omega} \dots \int F(x_1, x_2, \dots, x_n, z, p_1, p_2, \dots, p_n) dx_1 dx_2 \dots dx_n, \quad (x_1, x_2, \dots,$$

$$x_n)^T \in \mathbf{R}^n, p_i = \frac{\partial z}{\partial x_i} \quad (i = 1, 2, \dots, n)$$

则有变分方程

$$F_z - \sum_{i=1}^n \frac{\partial}{\partial x_i} \{F_{p_i}\} = 0 \quad (2-6)$$

其中,

$$\frac{\partial}{\partial x_i} \{F_{p_i}\} = F_{p_i x_i} + F_{p_i z} \frac{\partial z}{\partial x_i} + F_{p_i p_i} \frac{\partial p_i}{\partial x_i} + \sum_{j=1, j \neq i}^n F_{p_i p_j} \frac{\partial p_j}{\partial x_i}$$

4. 等式约束变分

对高维变分问题,如果有下面的等式约束 $\varphi_i = 0 (i = 1, 2, \dots, m)$, 则

$$\tilde{F}_z - \sum_{i=1}^n \frac{\partial}{\partial x_i} \{\tilde{F}_{p_i}\} = 0 \quad (2-7)$$

其中, $\tilde{F}(x_1, x_2, \dots, x_n, z, p_1, p_2, \dots, p_n) = F + \sum_{i=1}^m \alpha_i(x_1, x_2, \dots, x_n) \varphi_i$ 。

有关变分学方面的进一步论述,可参阅 Л. Э. 艾利斯哥尔兹 1981 年出版的《变分学》。

§ 2.3 矩阵与张量代数运算

矩阵本质上是一种对复杂对象的符号化表示。根据记载,“矩阵”一词是由西尔维斯特 (James Joseph Sylvester) 于 1848 年引入到线性方程组计算中的。常见的矩阵有二维矩阵,也有多维矩阵,如立体矩阵(刘国林, 2002; De Moor, 1993)。

2.3.1 矩阵

定义 2.4 设 V 是复数域上的线性空间。对 $\forall x, y \in V$, 如果关系 $\langle *, * \rangle$ 满足:

$$\langle x, y \rangle = \overline{\langle y, x \rangle}, \quad \bar{x} \text{ 表示 } x \text{ 的共轭复数};$$

$$\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle, \quad \forall z \in V;$$

$$\langle \alpha x, \beta y \rangle = \alpha \bar{\beta} \langle x, y \rangle, \quad \forall \alpha, \beta \in \mathbf{C}, \mathbf{C} \text{ 为复数域};$$

$$\langle x, x \rangle \geq 0, \text{ 当且仅当 } x = \mathbf{0} \text{ 时 } \langle x, x \rangle = 0。$$

则称 $\langle *, * \rangle$ 是定义在复数域 $\mathbf{C}$ 上的向量空间的内积(被称为埃尔米特内积, Hermite), 具有这样内积的向量空间 V 称为酉空间。

如果把数域 $\mathbf{C}$ 定义为实数域, 则变成欧氏空间。欧氏空间的向量长度和夹角为

$$\text{向量长度 } |x| = \sqrt{\langle x, x \rangle}, \quad \text{夹角 } \cos \angle(x, y) = \cos \theta = \frac{\langle x, y \rangle}{|x| |y|} \quad (2-8)$$

酉空间的向量长度和夹角为

$$\text{向量长度 } |\mathbf{x}| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}, \text{ 夹角 } \cos^2 \angle(\mathbf{x}, \mathbf{y}) = \cos^2 \theta = \frac{\langle \mathbf{x}, \mathbf{y} \rangle \langle \mathbf{y}, \mathbf{x} \rangle}{|\mathbf{x}|^2 |\mathbf{y}|^2} \quad (2-9)$$

从以上定义可知, 夹角与两个向量的相关系数等价。向量长度与下面将给出的 L_2 范数是一致的。

定义 2.5 nm 个数 a_{ij} ($i = 1, 2, \dots, m; j = 1, 2, \dots, n$) 排成 m 行 n 列的表, 即

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{nm} \end{bmatrix} \text{ (本书采用), 或 } \mathbf{A} = \left\| \begin{array}{ccc} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{nm} \end{array} \right\| \text{ (其他表示方法)}$$

叫做 $m \times n$ 矩阵, 记为 $\mathbf{A}_{m \times n} = (a_{ij})_{m \times n}$ 。叫做 $\mathbf{A}_{m \times n}$ 中第 i 行, 第 j 列上的元素。当 $n = m$ 时, $\mathbf{A} = \mathbf{A}_n$ 又叫做 n 级(阶)矩阵, 也叫做 n 级(阶)方阵。

数学到处都是最基本的四则运算: 加、减、乘、除。矩阵也不例外地存在这样的运算。当然, 矩阵的四则运算是在一定限制下的运算。矩阵的加减运算只在大小一样的矩阵之间可能发生。另外, 矩阵之间的元素必须是可进行某种“加减”运算的。两个矩阵满足了这些条件, 就可以进行矩阵加减运算了。所以, 当矩阵 $\mathbf{A}_{m \times n}$ 的行列个数分别与矩阵 $\mathbf{B}_{m \times n}$ 的行列个数相同时可定义两个矩阵之间的加减运算。此时, 矩阵加减法为对应元素的加减运算 $\mathbf{A} \pm \mathbf{B} = (a_{ij} \pm b_{ij})$ 。同样, 可定义矩阵的乘法。矩阵乘法为 $\mathbf{AB} = (c_{ij})$, $c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}$, 其中 $\mathbf{A} = (a_{ij})_{m \times n}$, $\mathbf{B} = (b_{ij})_{n \times q}$ 。那么, 什么是所谓的“除法”运算呢? 在矩阵运算中, 把矩阵的“除法”运算称为矩阵的逆运算, 其结果被称为矩阵的逆。一般情况下两个矩阵乘法不存在交换率, 所以, 通过两个矩阵乘法表示的矩阵的逆, 相对于另一个矩阵分为左逆和右逆。若 $\mathbf{A}_{m \times n}$, $\mathbf{B}_{n \times m}$, $\mathbf{AB} = \mathbf{E}$, 其中 $\mathbf{E}$ 的对角线上的元素为 1, 其他元素为 0 的单位矩阵, 则称 $\mathbf{B}$ 为 $\mathbf{A}$ 的右逆矩阵, $\mathbf{A}$ 为 $\mathbf{B}$ 的左逆矩阵; 若 $n = m$, 称 $\mathbf{B}$ 为 $\mathbf{A}$ 的逆矩阵, $\mathbf{A}$ 为 $\mathbf{B}$ 的逆矩阵, 也称为 $\mathbf{A}, \mathbf{B}$ 互逆; $\mathbf{A}$ 的逆用 $\mathbf{A}^{-1}$ 表示。

如果把一个矩阵分块成许多子矩阵, 这种表示方法为矩阵的分块表示。每个分块子矩阵占据被分块的矩阵的一个方形区域的元素。

定义 2.6 矩阵 $\mathbf{A}$ 的伴随矩阵为 $\mathbf{A}$ 矩阵的转置共轭矩阵, $\mathbf{A}^* = \bar{\mathbf{A}}^T$ 。

注意, 这里“伴随矩阵”一词与大学线性代数中的“伴随矩阵”不是同一个概念。设 $\mathbf{C}^{m \times n}$ 为实数域或复数域上的 $m \times n$ 矩阵构成的线性空间。如果没有特别说明就认定为实数域。 $\|\cdot\|_F$ 表示 Frobenius 范数: $\|\mathbf{A}\|_F^2 = \sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2$, 其中 $|a_{ij}|^2 = a_{ij} \bar{a}_{ij}$, $|a_{ij}|$ 为复数模, $\bar{a}_{ij}$ 为 a_{ij} 的复共轭。 $\|\cdot\|_F$ 也有其他名称, 如 Euclid 范数、Schur 范数或 Hilbert-Schmidt 范数。如果酉空间的 $\forall \mathbf{x}, \mathbf{y} \in \mathbf{V}$ 为

$$\mathbf{x} = (x_1, x_2, \dots, x_n), \mathbf{y} = (y_1, y_2, \dots, y_n)$$

则

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i \bar{y}_i, \langle \mathbf{x}, \mathbf{x} \rangle = \sum_{i=1}^n x_i \bar{x}_i = \sum_{i=1}^n |x_i|^2 = \|\mathbf{x}\|_F^2 \quad (2-10)$$

对于矩阵, 也有对应的长度和夹角。对实数域上的 $\forall \mathbf{A}, \mathbf{B} \in \mathbf{C}^{m \times n}$, 它们的长度和夹角为

$$\text{矩阵长度 } |\mathbf{A}|_F = \sqrt{\langle \mathbf{A}, \mathbf{A} \rangle} = \|\mathbf{A}\|_F, \text{ 夹角 } \cos \angle(\mathbf{A}, \mathbf{B}) = \cos \theta = \frac{\langle \mathbf{A}, \mathbf{B} \rangle}{|\mathbf{A}|_F |\mathbf{B}|_F}$$

这里, 长度、行列式和绝对值的符号都使用了相同的形式。当可能发生混淆时, 本书将对其注明。

$$\mathbf{A} = \left(\begin{array}{cccccccc} a_{11} & \dots & & \dots & & \dots & & \dots & a_{1n} \\ \dots & \dots & & \dots & & \dots & & \dots & \dots \\ \hline & & a_{i_1 j_1} & & a_{i_1 j_2} & & a_{i_1 j_k} & & \\ \dots & \dots & & \dots & & \dots & \dots & \dots & \dots \\ \hline & & a_{i_2 j_1} & & a_{i_2 j_2} & & a_{i_2 j_k} & & \\ \dots & \dots & & \dots & & \dots & \dots & \dots & \dots \\ \hline & & a_{i_k j_1} & & a_{i_k j_2} & & a_{i_k j_k} & & \\ \dots & \dots & & \dots & & \dots & & \dots & \dots \\ \hline a_{m1} & \dots & & \dots & & \dots & & \dots & a_{nm} \end{array} \right), \mathbf{M} = \left(\begin{array}{cccc} a_{i_1 j_1} & a_{i_1 j_2} & \dots & a_{i_1 j_k} \\ a_{i_2 j_1} & a_{i_2 j_2} & \dots & a_{i_2 j_k} \\ \vdots & \vdots & & \vdots \\ a_{i_k j_1} & a_{i_k j_2} & \dots & a_{i_k j_k} \end{array} \right)$$

正交变换是等距变换,有的书籍也称其为等积变换。在矩阵论中,正交变换的矩阵为 $\mathbf{A}\mathbf{A}^T = \mathbf{E}$, 正交酉变换的酉矩阵为 $\mathbf{A}\mathbf{A}^* = \mathbf{E}$ 。如果矩阵 $\mathbf{A}$ 是正定(半正定)的, 则 $\mathbf{A}$ 可分解为 $\mathbf{A} = \mathbf{B}^T \mathbf{B}$ (复数域上 $\mathbf{A} = \mathbf{B}^* \mathbf{B}$) 且秩 $\mathbf{B} = \text{秩 } \mathbf{A}$ 。下面设 $\mathbf{p}, \mathbf{h}$ 和 $\mathbf{q}$ 都是向量。

定义 2.19 称 p 是 h 在 q 上的正交投影,若 $\exists t \in \mathbb{R}^1 - \{0\}$ 使得 $p = qt$ 且 $\langle h - p, q \rangle = 0$ 。

定义 2.19 中, $p = qt$ 为向量平行及伸缩后重合条件, $\langle h - p, q \rangle = 0$ 为 p 到 h 的向量 $h - p$ 与 q 正交的条件, p 是 h 在 q 上正交投影点。 h 在 $h - p$ 与 q 方向上可唯一分解为 $h - p$ 与 p 。

2.3.2 广义逆矩阵

1920 年 Moore 首先引进了广义逆矩阵的概念。1955 年, Penrose 用明确的形式给出了广义逆矩阵的定义。广义逆与解线性方程组求解有密切关系。广义逆在最小二乘法和最优优化问题中也有着重要的应用。下面着重介绍广义逆的定义和基本的性质。

定义 2.20 设 $A \in \mathbb{C}^{m \times n}$, 若 $\exists G \in \mathbb{C}^{n \times m}$ 满足如下 4 个方程或部分方程

$$\textcircled{1} AGA = A; \textcircled{2} GAG = G; \textcircled{3} (GA)^* = GA; \textcircled{4} (AG)^* = AG$$

则称 G 为 A 的广义逆矩阵, 简称广义逆。

设 $A \in \mathbb{C}^{m \times n}$ 。Moore-Penrose 广义逆有 15 种。下面给出了常用的几种。更多的有关论述可参阅参考文献[34]。

满足①的减号逆 G 记为 A^- , 则

$$AA^-A = A, A^*(A^-)^*A^* = A^*, \text{rank}(A^-) \geq \text{rank}(A)$$

对任意矩阵减号逆存在但不一定唯一。

满足①、②的自反减号逆 G 记为 A_{r}^- , 则

$$AA_{\text{r}}^-A = A, A_{\text{r}}^-AA_{\text{r}}^- = A_{\text{r}}^-, A^*(A_{\text{r}}^-)^*A^* = A^*, (A_{\text{r}}^-)^- = A$$

若 A 是高矩阵(行满秩), 则 A 的右逆为

$$A_{\text{R}}^{-1} = A^T(AA^T)^{-1}$$

如果 A^T 是高矩阵(列满秩), A^T 的右逆为

$$(A^T)_{\text{R}}^{-1} = A(A^TA)^{-1}$$

则 A 的左逆为

$$A_{\text{L}}^{-1} = ((A^T)_{\text{R}}^{-1})^T = (A^TA)^{-1}A^T。$$

左逆和右逆满足定义 2.20 的 4 个方程, 所以, 左逆和右逆是特殊的广义逆。行满秩矩阵 A 的右逆的一般表达式为

$$A_{\text{R}}^{-1} = VA^T(AVA^T)^{-1}$$

其中, V 是使得 $\text{rank}(AVA^T) = \text{rank}(A)$ 成立的任意 n 阶矩阵。同理, 列满秩矩阵 A 的左逆的一般表达式为

$$A_{\text{L}}^{-1} = (A^TUA)^{-1}A^TU$$

其中, U 是使得 $\text{rank}(A^TUA) = \text{rank}(A)$ 成立的任意 m 阶矩阵。

满足定义 2.20 中①、③的最小范数广义逆 G 记为 A_m^- 。所以, 行满秩和列满秩时的广义逆也是最小范数广义逆。根据最小范数广义逆定义

$$A_m^-AA^* = (A_m^-A)^*A^* = (AA_m^-A)^* = A^*$$

如果存在矩阵 G 使得 $GAA^* = A^*$, 则

$$GAA^*G^* = A^*G^*, GA(GA)^* = (GA)^*, GA = (GA)^*$$

进一步

$$(GA)^*A^* = A^*, AGA = A$$

所以, $GAA^* = A^*$ 与最小范数广义逆的条件等价。

满足定义 2.20 中①、④的最小二乘广义逆 G 记为 A^- 。满足定义 2.20 中①至④的加号逆 (也称伪逆, Moore-Penrose 逆) G 记为 A^+ 。其中:

A^+ 唯一, $(A^+)^+ = (A^+)^*$, $A^+ = (A^* A)^+ A^* = A^* (A A^*)^+ = A^* (A^* A)^+ = A^+ (A^*)^+ = A^+ (A^+)^*$ 。另外, $\text{rank}(A) = \text{rank}(A^+) = \text{rank}(A^+ A) = \text{rank}(A A^+)$ 。

2.3.3 立体阵

三维数组 $X = (X_{ij})(t = 1, 2, \dots, n; i = 1, 2, \dots, p; j = 1, 2, \dots, q)$ 称为一个立体矩阵, 简称立体阵。立体阵 X 中, 下标 t, i, j 分别表示立体阵的面、行和列。一个立体阵常有两种理解方法和表达形式:

(1) $X = (X_{ij}), X_{ij} = (X_{1ij}, \dots, X_{nij})^T (i = 1, 2, \dots, p; j = 1, 2, \dots, q)$ 为向量。

(2) $X = (X_t)(t = 1, 2, \dots, n), X_t$ 为 $p \times q$ 矩阵 $X_t = (X_{tij})(i = 1, 2, \dots, p; j = 1, 2, \dots, q)$ 。此时 t 为一个固定的值, 称 X_t 为立体阵的第 t 个面。

令 $V = (V_{tij}), A = (A_{ik})$, 立体阵 X 左乘 $V = AX$ 定义为 $V_{tij} = \sum_{k=1}^p A_{ik} X_{tkj}$, 其对应的维数关系为

$$\underset{n \times r \times q}{V} = \underset{r \times p}{A} \underset{n \times p \times q}{X}$$

令 $H = (H_{ij}), B = (B_{kj})$, 则立体阵 X 右乘 $H = XB$ 定义为 $H_{ij} = \sum_{k=1}^p X_{ik} B_{kj}$, 其对应的维数关系为

$$\underset{n \times p \times r}{H} = \underset{n \times p \times q}{X} \underset{q \times r}{B}$$

令 d 为 p 维列向量, X 为 $n \times p \times p$ 阶, 则二次型

$$d^T X d = (d^T X_1 d, \dots, d^T X_n d)^T \quad (2-11)$$

表示一个 n 维列向量。

1. 矩阵与立体阵的方括号乘法

令 $V = (V_{sj})$ 。定义矩阵与立体阵的左方括号乘法 $V = [A][X]$ 为 $V_{sj} = \sum_{t=1}^n A_{st} X_{tij}$, 其维数关系为

$$\underset{m \times p \times q}{V} = \underset{m \times n}{[A]} \underset{n \times p \times q}{[X]}$$

令 $Y = (Y_{sj})$ 。定义矩阵与立体阵的右方括号乘法 $Y = [X][B]$ 为 $Y_{sj} = \sum_{t=1}^n X_{tij} B_{ts}$, 其维数关系为

$$\underset{m \times p \times q}{Y} = \underset{n \times p \times q}{[X]} \underset{n \times m}{[B]}$$

2. 立体阵的迹

阶为 $n \times p \times p$ 的立体阵 X 的迹 $\text{tr}[X]$ 定义为 $\text{tr}[X] = (\text{tr} X_1, \dots, \text{tr} X_n)^T$, $\text{tr} X_i$ 为矩阵 X_i 的迹。 $\text{tr}[X]$ 是一个 n 维向量。

3. 立体阵的性质

设 X, Y 表示立体阵, $A, B, \dots, L, M$ 表示矩阵, $a, b, c, d, \dots$ 表示向量。在下面的计算中设它们的维数满足相应计算的要求, 则立体阵有如下性质。

(1) $[E][X] = X, E$ 为单位矩阵;

(2) $[\lambda A][X] = [A][\lambda X], \lambda$ 为实数;

- (3) $[A+B][X] = [A][X] + [B][X]$;
- (4) $[A][X+Y] = [A][X] + [A][Y]$;
- (5) $[A][X] = [X][A^T]$;
- (6) $[A][LXM] = L[A][X]M$;
- (7) $[AB][X] = [A][B][X]$;
- (8) $[Aa][(Bb)(Cc)^T] = B[A][[a][bc^T]]C^T$;
- (9) $d^T X d = (X d)d = \sum_i \sum_j d_i d_j X_{ij}$;
- (10) $A(d^T X d) = [A][d^T X d] = d^T [A][X]d$;
- (11) $\text{tr}[AX] = \text{tr}[XA]$;
- (12) $A \text{tr}[X] = \text{tr}[[A][X]]$;
- (13) 若 $n \times p \times p$ 阶立体阵 X 的每个面都是对称矩阵时, 则 $(d^T X)^T = X d$ 。

2.3.4 矩阵 Kronecker 乘积与张量代数

定义 2.21 令所有数对 (i_1, i_2) (其中 $1 \leq i_1 \leq m_1, 1 \leq i_2 \leq m_2$) 在某种确定的次序下被编号为 $s_1, s_2, \dots, s_{m_1 m_2}$; 同理, 所有数对 (j_1, j_2) (其中 $1 \leq j_1 \leq n_1, 1 \leq j_2 \leq n_2$) 在某种确定的次序下被编号为 $t_1, t_2, \dots, t_{n_1 n_2}$ 。关于 $m_1 \times n_1$ 矩阵 A 和 $m_2 \times n_2$ 矩阵 B 的 Kronecker 乘积 (或称直接积、直积), 是指 $f: A, B \mapsto C$ 的一个 $m_1 m_2 \times n_1 n_2$ 矩阵 C , 记为 $C = A \otimes B$ 。它是由矩阵 A 和 B 的元素的所有可能的乘积适当排列而成, 即矩阵 C 的元素 c_{ij} 为 $c_{ij} = a_{i_1 j_1} b_{i_2 j_2}, (i_1, i_2) = s_i, (j_1, j_2) = t_j$ 。

定义 2.22 对 $\forall A \in C^{m \times n}$ 和 $\forall B \in C^{p \times q}$, 称乘积 $A \otimes B = (a_{ij} B)_{mp \times nq}$ 为右 Kronecker 乘积; 自然, 也有左 Kronecker 乘积 $A \otimes B = (A b_{ij})_{mp \times nq}$ 。在本书中称 Kronecker 乘积是指右 Kronecker 乘积。令 $A^{\otimes 1} = A$ 并定义 $A^{\otimes (k+1)} = A \otimes A^{\otimes k}$, 则 $A^{\otimes k}$ 表示 k 个 A 连续进行 Kronecker 乘积。对于若干不同矩阵的 Kronecker 的连续乘积的表示方法为 $A_1 \otimes A_2 \otimes \dots \otimes A_n = \bigotimes_{k=1}^n A_k$ 。

Kronecker 乘积中的“乘积运算”也有广义形式, 如在 1994 年齐东旭著的《分形及其计算机生成》一书中定义了 5 种广义模运算和逻辑运算的 Kronecker 乘积:

- (1) $A \otimes B = (a_{ij} B)_{mp \times nq} \pmod{N}$;
- (2) $A \otimes B = (a_{ij} \mid B)_{mp \times nq}$;
- (3) $A \otimes B = (a_{ij} \vee B)_{mp \times nq}$;
- (4) $A \otimes B = (a_{ij} \wedge B)_{mp \times nq}$;
- (5) $A \otimes B = (a_{ij} \& B)_{mp \times nq}$ 。

定义 2.23 $\text{Vec}(A) = (a_{1,1}, a_{2,1}, \dots, a_{m,1}, a_{1,2}, a_{2,2}, \dots, a_{m,2}, \dots, a_{m,n})^T, \text{Vec}(A) \in C^m$ 为 $A \in C^{m \times n}$ 的相伴向量 (按列垂直摆放形成的向量), 也称为列压栈。显然, Vec 变换不改变元素的值。

定义 2.24 令 B, C 分别由 $m \times d$ 个分块矩阵 $B_{ij} \in C^{p \times q}$ 和 $d \times n$ 个分块矩阵 $C_{ij} \in C^{q \times t}$ 组成, 即 $B = (B_{ij})_{mp \times dq}, C = (C_{ij})_{ds \times nt}$ 。定义映射关系

$$f: B, C \mapsto A, A = (A_{ij}) \in C^{mps \times nqt}, A_{ij} = \sum_{k=1}^d B_{ik} \otimes C_{kj}$$

称这种关系为强 Kronecker 乘积, 表示为 $A = B \circ C$ 。