



21世纪全国本科院校电气信息类 **创新型** 应用人才培养规划教材

生物医学数据分析 及其MATLAB实现

尚志刚 张建华 编 著
陈 栋 李国文



北京大学出版社
PEKING UNIVERSITY PRESS



2015年度中国生物医学工程学会学术年会论文集

生物医学数据融合新

DEMATLAB 4.0



中国生物医学工程学会
2015年度学术年会论文集

21 世纪全国本科院校电气信息类创新型应用人才培养规划教材

生物医学数据分析 及其 MATLAB 实现

尚志刚 张建华 编著
陈 栋 李国文



北京大学出版社
PEKING UNIVERSITY PRESS

内 容 简 介

本书系统地介绍了生物医学数据分析的相关基础知识和主要的分析方法,并通过 MATLAB 对大量的生物医学数据分析的实例进行实现。全书共分为 7 章:第 1 章绪论;第 2 章 MATLAB 应用基础;第 3 章医用数学基础及其 MATLAB 实现;第 4 章多变量分析;第 5 章时间序列分析;第 6 章生物信息学基础;第 7 章生物医学数据的检索方法。全书各章都配有思考题和习题,并在出版社相应网站上附有参考答案和各章例题的 MATLAB 源程序。

本书可作为医科本科生、信息处理、生物医学工程等相关专业的高年级本科生和研究生的教学参考用书,也可供相关的科技人员阅读和参考。

图书在版编目(CIP)数据

生物医学数据分析及其 MATLAB 实现/尚志刚,张建华,陈栋,李国文编著. —北京:北京大学出版社, 2009.1

(21 世纪全国本科院校电气信息类创新型应用人才培养规划教材)

ISBN 978-7-301-14472-5

I. 生… II. ①尚…②张…③陈…④李… III. ①数据处理-计算机应用-生物医学工程-高等学校-教材 ②计算机辅助计算-软件包, MATLAB-应用-生物医学工程-高等学校-教材 IV. R318-39

中国版本图书馆 CIP 数据核字(2009)第 191107 号

书 名:生物医学数据分析及其 MATLAB 实现

著作责任者:尚志刚 张建华 陈 栋 李国文 编著

策划编辑:李娉婷 李 虎

责任编辑:李娉婷

标准书号:ISBN 978-7-301-14472-5/TP·0980

出 版 者:北京大学出版社

地 址:北京市海淀区成府路 205 号 100871

网 址:<http://www.pup.cn> <http://www.pup6.com>

电 话:邮购部 62752015 发行部 62750672 编辑部 62755127 出版部 62754962

电子邮箱:pup_6@163.com

印 刷 者:河北滦县鑫华书刊印刷厂

发 行 者:北京大学出版社

经 销 者:新华书店

787 毫米×1092 毫米 16 开本 13 印张 298 千字

2009 年 1 月第 1 版 2009 年 1 月第 1 次印刷

定 价:25.00 元

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有 侵权必究

举报电话:010-62752024

电子邮箱:fd@pup.pku.edu.cn

前 言

现代医学发展的一大趋势是从定性研究走向定量研究，随着数学化发展的进程，运用数学理论方法来描述医学实际问题，并且揭示其质和量的相互关系的内在规律已成为新的研究趋势。面对日益激增的生物医学海量数据，生物医学工作者面临如何从大量的实践中获得精确的数学模型，进而实现从定性到定量的描述问题；以及如何选择和使用合理有效的处理方法来提取特征信息，进而用于临床的诊断和治疗，这些问题都对现代医学的发展非常重要，也具有巨大的应用需求。理工科学生的数学基础相对医科学生来说比较扎实，但对于医学数据代表的实际意义却缺乏敏感认识，而对医科学生而言，当需要从海量医学数据中提取感兴趣的信息时往往感到无从下手。MATLAB 是一种易学易用的高效科研编程软件环境，在全世界广泛流行，其中大量的工具箱可以方便地实现各种数据的分析算法，便于医学工作者掌握。编写本书的目的在于为广大医科学生介绍如何利用工具软件 MATLAB 对医学数据进行处理分析，同时也为对医学信息处理方面感兴趣的工科学生介绍医学数据处理的特殊性和常规方法。

本书通过大量的实例，帮助读者熟悉生物医学数据处理的数学背景知识，了解生物医学数据及信号的特点，掌握生物医学数据处理的常规方法以及生物医学数据库的检索方法，了解生物信息学的基本知识和发展趋势，并学会用数学工具软件 MATLAB 来完成对实际生物医学数据的处理分析。

由于生物医学数据分析所涉及的数学知识多、应用领域广，书中很多方法只是进行了简单的介绍，并未深入，希望能起到带读者入门的目的。限于作者自身的知识和经验，书中一定有许多不足之处，诚恳地希望读者给予指正和提出宝贵的意见。

编 著 者

2008 年 11 月

E-mail: zhigang_shang@zzu.edu.cn

目 录

第 1 章 绪论.....	1	3.2 医学统计学实例分析.....	61
1.1 生物医学数据的来源和特点.....	1	3.2.1 统计描述.....	62
1.2 生物医学数据分析概述.....	3	3.2.2 统计推论.....	71
1.3 MATLAB 在生物医学数据分析中 的应用价值.....	5	3.2.3 回归分析.....	82
本章小结.....	6	本章小结.....	90
思考题与习题.....	6	思考题与习题.....	90
第 2 章 MATLAB 应用基础.....	7	第 4 章 多变量分析.....	92
2.1 MATLAB 简介.....	9	4.1 多元回归分析.....	93
2.1.1 MATLAB 的特点.....	9	4.2 主成分分析.....	98
2.1.2 MATLAB 7.x 的开发环境.....	11	4.3 判别分析.....	105
2.1.3 MATLAB 的帮助系统.....	14	4.4 聚类分析.....	111
2.2 MATLAB 编程基础.....	15	本章小结.....	120
2.2.1 数据类型.....	16	思考题与习题.....	120
2.2.2 变量、运算符和表达式.....	17	第 5 章 时间序列分析.....	122
2.2.3 矩阵元素的定位.....	20	5.1 时间序列分析概述.....	123
2.3 MATLAB 程序设计.....	22	5.1.1 基本概念.....	124
2.3.1 M 文件.....	22	5.1.2 时间序列的分类.....	125
2.3.2 函数.....	23	5.1.3 时间序列的组合成分.....	128
2.3.3 程序流程控制.....	24	5.2 时间序列分析方法.....	129
2.4 MATLAB 数据可视化.....	27	5.2.1 时间序列分析的基本思想.....	129
2.4.1 二维图形的绘制.....	27	5.2.2 确定型时间序列的分析 方法.....	130
2.4.2 三维图形的绘制.....	31	5.2.3 随机型时间序列的分析 方法.....	138
2.4.3 体数据的可视化.....	36	5.3 时间序列分析 MATLAB 示例 ——脑电信号分析.....	145
本章小结.....	39	本章小结.....	149
思考题与习题.....	39	思考题与习题.....	149
第 3 章 医用数学基础及其 MATLAB 实现.....	41	第 6 章 生物信息学基础.....	151
3.1 医用高等数学实例分析.....	43	6.1 生物信息学概述.....	153
3.1.1 函数分析.....	43	6.1.1 生物信息学产生的背景.....	153
3.1.2 微分方程在生物医学中 的应用.....	52	6.1.2 什么是生物信息学.....	154



6.1.3 生物信息学的研究目标 和内容	155
6.1.4 生物信息学的研究方法	157
6.2 生物信息学中的重要概念	158
6.3 生物信息学研究中的算法	162
6.4 Bioinformatics 工具箱介绍	163
6.4.1 概述	163
6.4.2 Bioinformatics Toolbox 2.0 的特点及功能	164
6.5 Bioinformatics 工具箱使用示例	169
6.5.1 人类线粒体基因序列 的分析	169
6.5.2 两条基因序列的比对	176
本章小结	184
思考题与习题	184

第 7 章 生物医学数据的检索方法	185
7.1 数据检索的基础知识	185
7.1.1 布尔逻辑检索	185
7.1.2 截词检索	187
7.1.3 限定检索	187
7.2 生物医学数据库检索方法	188
7.2.1 中国医学生物信息网	188
7.2.2 美国国立卫生院	188
7.2.3 美国国立医学图书馆	189
7.2.4 世界卫生组织	189
7.2.5 生物医学网	190
7.3 生物医学数据的网络检索方法	191
本章小结	195
思考题与习题	195
参考文献	197

绪论

引言

福尔摩斯：“数据！数据！数据！没有黏土，怎么能做成砖？”
——Doly A C 爵士

“在表面上看上去是发生着偶然性的地方，其实这种偶然性本身总是服从于内部隐藏着
的规律性。全部的问题就在于发现这种规律。”
——恩格斯

“工欲善其事，必先利其器。”
——论语

1.1 生物医学数据的来源和特点

1. 生物医学数据的来源

生物医学领域中不断地产生着大量的数据，即各种形式的人体生理信息通过生物医学传感器和各种测量技术被检测出来。用来表征人体生理状态的信息，大体上可以分为物理量、化学量和生物量 3 部分。物理量包括各器官的生物电信息、磁信息、压力、振动、速度、流量、温度和形态等；化学量包括 O₂、CO₂、H₂O、K⁺、Na⁺、Ca²⁺等各种化学成分的浓度以及多种化学参量；生物量包括酶、抗原、抗体、激素、神经递质、DNA、RNA 和蛋白质等。表 1-1 为诊断时经常需要测量的人体生理信息。

表 1-1 人体生理信息的种类

形 态	身高、骨长、心脏、胃、肾等器官的几何形状、血管直径等
振 动	心音、肠鸣音、呼吸音、血管音等
压 力	血压、心内压、颅内压、胸腔内压、眼内压等
速 度	血流速度、排尿速度、神经传导速度等
流 量	血流量、呼吸流量、尿流量等
温 度	体表温度、口腔温度、血液温度、直肠温度、其他脏器温度等
生 物 电	细胞电位、脑电、心电、胃电、肌电等



(续表)

生物磁	心磁、脑磁、胃磁等
物理性质参数	组织密度、对 X 射线的衰减系数、体内氢核分布、对超声波的声阻抗等
化学成分	O_2 、 CO_2 、 H_2O 、 N_2 、 NH_3 、 K^+ 、 Na^+ 、 Ca^{2+} 等
生物量	酶、抗原、抗体、激素、神经递质、DNA、RNA 和蛋白质等

2. 生物医学数据的特点

要对生物医学数据进行分析 and 处理, 首先要对数据的特点有所了解。

综合来说, 生物医学数据具有以下特点。

1) 多样性

生物医学数据的多样性主要体现在两个方面。

(1) 数据来源的多样性。

生物医学数据是通过对人体各个层次上的生物医学信息进行检测而获得的, 包括对携带有生物结构和特性信息的物理量、化学量和生物量进行的检测。例如, 生理电信号、磁信号, 光信号和声波振动等的检测; 分子水平上的信号检测: 基因序列的测序、血糖、激素、毒素、药物分子等浓度水平的检测; 利用 X 射线、NMR 信号、超声、 γ 射线检测的生理参数等。

(2) 数据格式的多样性。

由于医学数据是从医学影像、实验数据以及医生与病人的交流中获得的, 所以原始的医学数据具有多种形式。医学数据包括二维影像(如 SPECT)、时序信号(如 ECG)、纯数据(如体征参数、化验结果)、文字(如病人的身份记录、症状描述、检测和诊断结果的文字表述)等。

2) 不完整性

生物医学数据的搜集和处理过程经常相互脱节。搜集以治疗或认识机理为直接目的, 而处理以寻找某种疾病的一般规律为目的, 因此搜集的信息可能无法涵盖研究需要的所有的信息。此外, 人为因素也可能导致数据记录的偏差和残缺, 并且许多生物医学数据的表达和记录本身也具有不确定性和模糊性。

3) 随机性

由于缺乏对人体各种生理信号产生机制的全面了解, 加之个体差异的普遍存在和测量过程中不可避免的测量误差, 大部分生物医学数据都表现出高度的随机性, 所以通过对个体的分析难以得出正确的结论, 必须在统计意义上进行分析。

另外, 大部分生物医学数据都源自于对生物医学信号的采集。因此, 有必要对生物医学信号的特点进行简要介绍。

生物医学信号的主要特性是时变的和非线性的。这主要是由生物体是一个开放的、时变的、非线性的系统而决定的, 另外人的身高、体重、心率、血压等生理参数也都是时变的。而经典的信号处理技术主要处理的是线性和非时变问题, 因为一般认为生命系统的时变过程是非奇异的、确定性的、相对缓慢的过程, 所以近似认为在相当短的时间内, 生命系统是非时变的, 同时也近似认为在相当短的时空限度内, 生物医学信号和系统具有线性



特征,所以可以采用经典的信号处理技术来近似处理。

生物医学信号的特点还表现为信号非常微弱、频率低、信噪比(Signal to Noise Ratio, SNR)特别低,这些都表明生物医学信号中的很多反映真实生理或病理状态的信息是潜在的、不明显的,所以必须通过特殊的数据处理和分析方法把内在的信息提取出来。当然这一特点也决定了生物医学数据分析的难度和复杂性,所以处理方法就必须要考虑信号的特点和实际情况中的各种联系。

1.2 生物医学数据分析概述

由传感器获得的是人体生理状态的原始信息,具有不同的特点。在这些信息中不仅含有有价值的信息,还混杂有干扰信息。为了把有用的潜在信息从原始信息中提取出来,就需要根据各种信息的特征进行处理和加工。

如何从大量数据中去发现和理解隐藏在其中的内在规律以及变量间的相互关系,进而为疾病的诊断、治疗提供依据和决策支持,并且加深对生命现象的认识和理解,是进行生物医学数据分析的目的。

在刚刚过去的 20 世纪中,生物医学的研究由于数学思想和方法的介入而发生了深刻的变化。数学推动了人们探索生命世界的步伐,使得现代生物学和医学的面貌为之改观。

如果没有傅里叶变换和 Radon 变换,CT 与 MRI 的发明是不可想象的。

如果没有统计学习的各种理论和方法,DNA 序列对于人类永远是无法破译的天书。

.....

计算机数值诊断(专家系统)是数学方法应用于医学的一个典型例子。其原理是利用数学的信息理论、数据处理技术和计算机,对疾病患者的症状表现、各种化验检验指标进行数学加工分析,做出疾病的定量诊断结果。它与临床诊断不同,临床诊断是医生根据自己的经验和理论知识的推理做出最有可能的判断,诊断的准确性与医生本人的经验和知识水平有着直接的关系。而数值诊断则不然,它依赖于大量的历史诊断记录和对这些资料的数学处理方式。已诊断病例越多,症状资料越详细,数学处理方法越得当,就越能得到较确切的诊断结果。近年来,人工神经网络、模糊数学、支持矢量机等新方法都应用于计算机数值诊断领域。

流行病学的数学模型也是数据分析方法应用的重要实例。例如当今最严重的流行病——艾滋病(AIDS)。在艾滋病的防治工作中有两个重要问题:一是要估计过去、现在和将来造成人类免疫缺损的 HIV 病毒的传播率,以便预测将来艾滋病例的数目;二是要明确 HIV 病毒在各种易感人群之间的传播方式,以便防止进一步传染,从而规划医疗卫生资源的优化组织。因此,建立艾滋病的数学模型成为一项重要工作,在建模过程中,研究人员要把各种数学方法和统计技术结合起来,可以采用时间序列分析方法来对该问题进行建模。

对现代生命科学的深入研究已离不开数学的支持,数学化已是生命科学发展的历史必然。事实上,在许多场合,数学已不再单纯是一种辅助性的工具,而已经成为解决许多重大问题的关键性的思想与方法。



因此,生物医学数据分析,就是在考虑生物医学数据特点的情况下,根据不同的分析和应用目的而采用不同的数学分析方法,建立研究对象的数学模型,对所采集到的生物医学信号进行分析、辨认、解释、分类、显示。其研究目的一是对生物体系的结构与功能的研究,二是协助对疾病的诊断和治疗。

从数学的角度来看,生物医学领域常用的数据分析方法包括时间序列分析、回归分析、多变量分析以及统计学习。对信息的分析可以在时域和频域中进行,其中包括信息的频谱、能量谱、功率谱、相关和卷积分析等。这些分析理论和分析方法目前都比较成熟。

从数据分析的角度来看,被测的生理信息可以分为两大类:确定性信息和非确定性信息。确定性信息是指可以用明确的数学关系式来描述的信息,具有精确可预测的特点。只要掌握了规律,就可以由过去值准确地预测未来值,而且只要条件相同,信息的取值便相同。例如血压、血流、呼吸、心电等生理参数基本上可视为确定性信息。非确定性信息则不能用明确的数学关系来描述,即使在完全相同的条件下,多次测量的结果也不相同,即信息的出现不服从确定性规律,即由过去值不能准确地预测未来值。由于非确定信息是随机出现的,因而一般也称之为随机信息。随机信息因人而异、因时而异,难以从少量的观测结果中总结信息的特征和规律。

需要强调的是,没有绝对的确定性信息或随机信息。由于人体中的每个子系统(如循环系统、神经系统等)都是一个极其复杂的物理过程和化学过程的综合体,而且各个子系统之间存在着有机的交织和相互影响,因此生理信息中往往既有确定性的成分又有随机性的成分,而且随机性的噪声背景都很强,相当复杂。因而为了分析方便,才大体上将其划分为确定性信息和随机信息,以便于用不同的方法来其特征进行分析。

从数据分析的目的来看,生物医学数据分析主要有两种:一种是为了提取不同变量之间的联系或因果关系,可用于阐述疾病的病因和病症间的关系或进行分类,如回归分析、多变量分析等;另一种是为了确定某一变量自身随时间变化的规律,可用于对未来发展趋势的预测或施加控制,如时间序列分析。在有些情况下,这两种分析目的都要兼顾,例如多道脑电的分析。

下面从信号处理的角度来介绍生物医学数据分析。

噪声和干扰信号总是伴随着生物医学信号,如心电、肌电信号总是伴随着因肢体和精神紧张等带来的假象,且常有较强的工频干扰;诱发脑电信号总是伴随着较强的自发脑电信号;超声回波信号中往往伴随着其他发射杂波。此外,信号中的无用成分也应该视为检测中的干扰。生物医学信号处理技术即是研究从被检测的湮没在干扰和噪声中的生物医学信号中提取有用的生物医学信息的方法。这些方法包括强噪声背景下微弱生理信号的动态提取、多道生理信号的同步观察与处理、生理信号的时间-频率表示、自适应处理、人工神经网络和混沌与分形理论的应用、医学专家系统等。生物医学信息处理技术已经在生物医学测量、医学成像和医学仪器等领域等到广泛的应用。例如,累加平均技术已成功地用于诱发脑电、希式束电和心室晚电位等微弱信号的提取;在心电和脑电的体表检测中采用计算机进行多道信号的同步观察与处理,并推求原始信号源的活动(即逆问题求解);在生理信号的数据压缩中开始引入人工神经网络方法;在医药学特别是中国传统医学中的医学



专家系统已在发挥实际效用。

从广义程度上来看,医学图像处理技术也属于对生物医学数据的分析。医学图像处理技术是对已获得的图像进行分析、识别、分割、解释、分类以及三维重建与显示等处理,目的是增强或提取其特征信息。由于医学图像具有直观、形象和信息量丰富的特点,而且便于观测和储存,因而发展十分迅速,在现代医学临床诊断中已占有越来越重要的地位。

医学图像处理的方法很多。基于如临床、解剖学、成像技术和统计学等知识的综合运用,使图像处理技术发展迅速。此外,模糊处理、人工神经网络等技术在医学图像处理中已受到高度重视,三维图像显示技术亦是重要的发展方向。

总之,生物医学信息处理技术的研究领域广泛,但大多数尚处在发展之中,并存在大量前沿课题,均需继续加强系统的深入研究,扩大其实用价值,而数据分析方法在其中大有用武之地。

1.3 MATLAB 在生物医学数据分析中的应用价值

由于在生物医学数据分析中涉及了很多方面的数学知识,同时还要将这些分析方法变成计算机能够处理的算法,因此没有好的工具是难于完成这项任务的,尤其令医学背景的人望而生畏。

MATLAB 为解决这一问题提供了有效的手段。MATLAB 是一种用于算法开发、数据可视化、数据分析以及数值计算的高级技术计算语言和交互式环境。使用 MATLAB,用户可以比使用传统的编程语言(如 C、C++ 和 Fortran)更快地解决技术计算问题。作为全球最大的工程、科学计算的软件平台, MATLAB 提供了大量由各领域专家编写的工具箱,同时具有强大的数值计算和图形显示功能,它目前已广泛应用于数学计算、信号处理和通信、图像处理、计算生物学、自动控制、生物信息学、统计和时间序列分析等诸多领域的教学和实际应用中。

使用 MATLAB,研究者的工作效率可能会有成百上千倍的提高,使得研究工作真正是在分析数据、寻找规律或验证思路,而不是在编程。使用 MATLAB 大大简化了学习和研究中的编程量,比如:使用 C 或者 Fortran 编写一个优化的程序,一般需要几百行或者几千行的程序,并且在首次写程序时还可能需要大量的时间来调试这个程序,当这个程序能够运行时,可能已花费掉一个星期或者更长的时间。而下一次在进行另一种优化方法的运算时,需要同样的时间工作,也就是说,使用这些语言编程时,大量时间花在了编程上,而不是在研究人员应当做的思考工作上,这样大大降低了工作效率。如果使用 MATLAB 编程,一个优化程序只需要 10 行以下的程序,因此基本不会出现错误,这样在 1 分钟左右就完成了编程,并且马上就可以运行看到结果。如果想改变优化算法,也只需要对优化的函数进行修改就可以了。

另外,采用 MATLAB 的一大好处是其具有强大的数据可视化功能,这在复杂的生物医学数据分析领域尤其重要,数据及其分析结果可以通过图形或图像方式直观显示,这样可以帮助使用者加深对抽象的数理方程的理解,有助于发现问题和解决问题。使用



MATLAB 处理数据的过程直观方便,简单的命令就可以完成繁琐的计算,大大地激发了使用者的研究兴趣。

统计学分析是生物医学数据分析中最常用的方法,与流行的 SPSS、SAS 等统计软件相比, MATLAB 毫不逊色,不仅能完成各种统计功能,还能做出非常漂亮的图表,其功能强大,非常灵活,但是仅仅用 MATLAB 来做统计分析简直是大材小用。用 MATLAB 还可以根据研究目的将统计工具箱与其他多种工具箱组合使用,然后按照自己的研究思路对数据进行任意操作,进行数学建模分析。

在生物医学数据分析中用到的各种方法几乎在 MATLAB 及其工具箱中都有现成的函数,只要在学习中搞清楚各种分析方法的思想、作用和应用场合,在处理具体问题时就可以利用 MATLAB 选择最适合的算法来实现,这相当于善于使用工具而不是制造工具。

在生物医学数据分析中常用的 MATLAB 工具箱有:

- (1) 统计工具箱(Statistics Toolbox)。
- (2) 信号处理工具箱(Signal Processing Toolbox)。
- (3) 图像处理工具箱(Image Processing Toolbox)。
- (4) 优化工具箱(Optimization Toolbox)。
- (5) 符号数学工具箱(Symbolic Math Toolbox)。
- (6) 曲线拟合工具箱(Curve Fitting Toolbox)。
- (7) 生物信息学工具箱(Bioinformatics Toolbox)。
- (8) 人工神经网络工具箱(Neural Network Toolbox)。

本章小结

本章介绍生物医学数据的特点、生物医学数据分析的目的和常用方法,以及为什么要采用 MATLAB 作为工具进行生物医学数据分析。

思考题与习题

1. 生物医学数据分析的目的是什么?
2. 怎样理解生物医学数据中的确定性信息和随机信息?

第2章

MATLAB 应用基础



本章知识结构





引 言

在科学和工程领域中,大量的数值计算、分析任务需要计算机来处理。在 20 世纪 50 年代,即使最简单的矩阵相加程序也需要用机器语言编写大量源代码并通过穿孔纸带来运行。1957 年推出的 Fortran 是第一种高级编程语言,其专为数值计算而设计。在 Fortran 中求解形如 $Ax=b$ 这样的线性方程可能需要编写 1 000 行左右的源代码,并且还很容易出错。因此,如果能把高质量的通用的数值分析算法进行集中编写,以“库”或子程序的形式发布,供为解决不同任务的人自由使用,将会极大地提高效率,减少重复性编程工作和提高程序的准确性。1976 年发布的用于特征根计算的软件包 EISPACK 和 1979 年发布的用于线性问题求解的软件包 LINPACK 就是 Fortran 数值计算函数库的典型代表。作为这两种软件包的主要开发者之一,美国 New Mexico 大学的 Cleve Moler 教授觉得编写 Fortran 源码、编译、调试、执行、修改源码、编译等这套流程仍过于繁琐。为使学生更容易使用 LINPACK,帮助使用者将注意力集中于待解决的问题本身而非程序编写问题,1978 年,他以 Fortran 语言为底层代码编写了 MATLAB,名称由 MATrix LABoratory(矩阵实验室)这两个单词的前 3 个字母组合而成,包含 80 个函数。由于 MATLAB 具有交互式界面、易于进行输入输出、数值计算功能强大,而且使用免费,所以很快在工程和科学领域流行起来,尤其受到对计算机编程和算法不熟悉、但因工作需要大量涉及数值计算任务的控制工程领域人员的欢迎。自动控制工程师 Jack Little 在 Cleve Moler 教授于斯坦福大学访问期间的一次演讲中第一次见到 MATLAB 后,认为这种使用方便、数值分析能力强大的软件市场潜力巨大,因而积极推进 MATLAB 的商业化。当 20 世纪 80 年代初期 IBM 推出第一台个人计算机(PC)时,Jack Little 敏锐地意识到结合 MATLAB 和 PC 进行技术计算的可能性,因此他和同事 Steve Bangert 一起用更具效率的 C 语言对 MATLAB 进行重新编写,并加入 M 文件、工具箱和更多功能强大的特性。1984 年,Cleve Moler、Jack Little 和 Steve Bangert 一起在加州建立了 MathWorks 公司,正式把 MATLAB 推向市场。

目前,Jack Little 担任 MathWorks 公司的 CEO,Cleve Moler 为公司首席科学家,MATLAB 的最新版本是 7.6(R2008a),包含超过 8 000 多个函数和多种工具箱,已经发展成为适合多学科、多种工作平台的功能强大的大型软件,正版用户超过百万。在 Google 中输入 MATLAB 进行搜索,约有 17 900 000 项查询结果。正如 MathWorks 公司主页(www.mathworks.com)上商标图片中所写(图 2.1),MATLAB 已成为加速工程技术和科学研究步伐的有力工具。



图 2.1 MathWorks 公司的商标

“工欲善其事,必先利其器”,本章将介绍 MATLAB 的主要特性,学习 MATLAB 的使用方法,通过实例掌握 MATLAB 的编程方法和数据的可视化。



2.1 MATLAB 简介

MATLAB 是一种用于算法开发、数据可视化、数据分析以及数值计算的高级技术计算语言和交互式环境。使用 MATLAB, 用户可以比使用传统的编程语言(如 C、C++ 和 Fortran)更快地解决技术计算问题。作为全球最大的工程、科学计算的软件平台, MATLAB 提供了大量由各领域专家编写的工具箱, 同时具有强大的数值计算和图形显示功能, 目前已广泛应用于数学计算、信号处理和通信、图像处理、计算生物学、自动控制、生物信息学、统计和时间序列分析等诸多领域。在欧美和国内许多高校, MATLAB 已经成为线性代数、自动控制理论、数理统计、数字信号处理、时间序列分析、动态系统仿真等高级课程的基本教学工具, 并且成为攻读学位的大学生、硕士生、博士生必须掌握的基本技能。在国际学术界, MATLAB 已经被确认为准确、可靠的科学计算标准软件。在许多国际一流学术刊物上(尤其是信息科学刊物), 都可以看到 MATLAB 的应用。

2.1.1 MATLAB 的特点

1. 使用方便, 帮助全面

MATLAB 具有友好的工作平台和编程环境, 既可以作为计算器使用, 也可以作为编程语言。随着 MATLAB 的不断升级, 其用户界面也越来越精致, 更加接近 Windows 的标准界面, 人机交互性更强, 操作更简单。同时, MATLAB 提供了完整的联机查询、帮助系统、应用示例(Examples 和 Demos), 极大地方便了用户的学习和使用。简单的编程环境提供了比较完备的调试系统, 程序不必经过编译就可以直接运行, 而且能够及时地报告出现的错误及进行出错原因分析。用户使用 MATLAB 时, 可以利用其丰富的库函数避开繁杂的子程序编程任务, 压缩了一切不必要的编程工作。库函数都由本领域的专家编写并经过正确性校验和优化, 因而保证了函数的可靠性。

2. 编程语言简单易学

MATLAB 作为一种以矩阵计算为基础的解释性高级编程语言, 包含控制流语句、函数、数据结构、输入/输出和面向对象编程功能。用户既可以在命令窗口中将输入语句与执行命令同步, 也可以先编写好一个较大的复杂的应用程序(M 文件)后再一起运行。由于用户可以交互式地获得程序运行结果, 故其也被称为“演草纸式”的科学计算语言。MATLAB 编程语法限制不严格、程序设计自由度大, 更加符合科技人员对数学表达式的书写格式。由于 MATLAB 由 C 语言编写, 因此语法特征与 C 语言相似, 而且更加简单。丰富的运算符和库函数可以使程序简洁高效, 使之更利于非计算机专业的科技人员使用。程序的可移植性好, 基本上不用做修改就可以在各种型号的计算机和操作系统上运行。

3. 科学计算能力强大

MATLAB 本身提供的函数库包括三角运算、矩阵运算、微分方程及偏微分方程求解、统计分析和快速傅里叶变换等数百个工程和科学计算常用的函数, 可以方便地实现用户



所需的各种计算功能。函数中所使用的算法都是科研和工程计算中的最新研究成果,而且经过了各种优化和容错处理,在计算要求相同的情况下,使用 MATLAB 的编程工作量会大大减少,同时保证了程序计算结果的可靠性。另外, MATLAB 还具有良好的符号计算功能。

4. 数据可视化功能强大

MATLAB 自产生之日起就具有方便的数据可视化功能,可将向量和矩阵用图形的形式表现出来,并且可以对图形进行定制、标注和打印。高层次的作图包括二维和三维数据的可视化、图像处理、动画和表达式作图,可以提供图形的光照处理、色度处理以及四维数据的表现等其他数据可视化软件所没有的功能。另外, MATLAB 还具有较强的编辑图形用户界面(GUI)的能力。

5. 功能强大的工具箱

工具箱(Toolbox)是单独提供的专用 MATLAB 函数集,这些文件把 MATLAB 的环境扩展到解决特殊类型的问题上。工具箱可分为功能性工具箱和学科工具箱。功能工具箱用来扩充 MATLAB 的符号计算、可视化建模仿真、文字处理及实时控制等功能;学科工具箱是专业性比较强的工具箱。MATLAB 对许多专门的领域都开发了功能强大的模块集和工具箱。一般来说,它们都是由特定领域的专家开发的,用户可以直接使用工具箱学习、应用和评估不同的方法而不需要自己编写代码。目前, MATLAB 已经把工具箱延伸到了科学研究和工程应用的诸多领域,诸如数据采集、数据库接口、概率统计、样条拟合、优化算法、偏微分方程求解、神经网络、小波分析、信号处理、图像处理、系统辨识、控制系统设计、LMI 控制、鲁棒控制、模型预测、模糊逻辑、金融分析、地图工具、非线性控制设计、实时快速原型及半物理仿真、嵌入式系统开发、定点仿真、DSP 与通信、电力系统仿真等。工具箱的使用和扩展是 MATLAB 的重要特色之一。

6. 高度的适应性和开放性

开放性使 MATLAB 广受用户欢迎。除内部函数外,所有 MATLAB 的核心部分和各种工具箱里的 M 文件都是可读和可修改的 ASCII 文件,用户可以直接对其进行复制、修改或加入自己编写的程序语句来生成新的函数。MATLAB 的工具箱可以任意增减,任何人也可以自己生成 MATLAB 工具箱并将其扩充到系统中。

7. 实用的程序接口和发布平台

MATLAB 可以利用 MATLAB 编译器和 C/C++ 数学库和图形库,将自己的 MATLAB 程序自动转换为独立于 MATLAB 运行的 C 和 C++ 代码。同时支持 C、Fortran、C++、Java 等编程语言,可以直接调用。另外, MATLAB 网页服务程序还容许在 Web 应用中使用自己的 MATLAB 数学和图形程序。

基于以上特点, MATLAB 在许多领域得到广泛应用,成为科研和工程技术人员进行科学研究、工程计算时的首选工具。与其他高级程序相比, MATLAB 程序的执行速度较慢,但它为用户提供了大量解决问题的“工具”,帮助用户专注于待解决问题的本身,快