目　录

上篇　地球物理反演基础

第一章　概述 …… (2)
一、地球物理反演的研究对象 …… (2)
二、地球物理反演的发展简史 …… (4)
三、地球物理反演的研究内容和方法 …… (7)
四、地球物理反演的计算机实现 …… (12)
第二章　算法基础 …… (15)
一、矩阵的范数 …… (15)
二、线性方程组和扰动分析 …… (18)
三、QR 分解和线性最小二乘问题 …… (24)
四、奇异值分解（SVD） …… (32)
五、奇异值分解 Fortran 子程序 …… (39)
第三章　广义反演方法 …… (45)
一、参数化、线性化和归一化 …… (45)
二、广义最小二乘法 …… (51)
三、广义线性反演 …… (55)
四、利用自然逆的广义反演方法 …… (62)
五、线性规划和联合线性反演 …… (67)
六、广义线性反演在位场处理解释中的应用 …… (74)
七、广义线性反演在地震处理解释中的应用 …… (79)
第四章　Backus-Gilbert 反演理论 …… (83)
一、连续地球物理模型的反问题 …… (83)
二、线性反问题解的非唯一性 …… (87)
三、非线性反问题的线性化 …… (91)
四、脉冲度准则 …… (96)
五、BG 展布准则 …… (100)
六、BG 折衷准则 …… (106)
第五章　地震层析成像 …… (112)
一、地震层析成像的数据采集 …… (112)
二、射线追踪方法 …… (117)
三、Radon 变换和连续图像重建 …… (125)
四、离散图像重建技术 …… (130)
五、广义 Radon 变换问题 …… (138)

六、衍射层析成像……（145）
七、逆散射地震层析成像……（153）
参考文献（上篇）……（166）

下篇　地球物理反演精义

第六章　应用地球物理中的正问题……（172）
一、应用地球物理学的基本方程式……（172）
二、用于波传播问题的有限差分法……（177）
三、有限元法简析……（179）
四、几种深入的波动问题……（183）
第七章　地球物理数据处理……（190）
一、数据与数据模型……（190）
二、最小二乘法数据处理……（194）
三、二维数字滤波……（199）
四、地球物理场的向下延拓……（204）
五、小波变换与多尺度分析……（209）
第八章　地球物理中的偏微分方程反演……（215）
一、广义解与正则化……（215）
二、最大熵法……（222）
三、地球物理线性反演……（226）
四、非线性迭代地震反演……（230）
五、电导率成像反演……（237）
第九章　非线性反演方法……（242）
一、迭代、分形和浑沌……（242）
二、地震道非线性浑沌反演……（246）
三、仿真淬火反演……（252）
四、遗传算法……（256）
五、人工神经网络的应用……（261）
结束语……（268）
参考文献（下篇）……（269）

Theory and Methods of Geophysical Inversion

By Dr. Yang. Wencai (Ph. D.)

Contents

Part One: Fundamentals of Geophysical Inversion

Chapter Ⅰ. Introduction to Geophysical Inversion …… (2)
1.1 The objects of geophysical inversion …… (2)
1.2 Development of geophysical inversion …… (4)
1.3 Contents and methodology …… (7)
1.4 Computer implementation of geophysical inversion …… (12)
Chapter Ⅱ. Matrix Computation …… (15)
2.1 The norm of matrices …… (15)
2.2 Linear equations and perturbation analysis …… (18)
2.3 The QR decomposition …… (24)
2.4 The singular value decomposition …… (32)
2.5 Fortran procedure for SVD …… (39)
Chapter Ⅲ. Generalized Inversion Methods …… (45)
3.1 Parameterization, linearization and normalization …… (45)
3.2 Generalized least squares methods …… (51)
3.3 Generalized linear inversion …… (55)
3.4 Generalized inverse by using singular value decomposition …… (62)
3.5 Linear programming and joint inversion …… (67)
3.6 Application to potential data inversion …… (74)
3.7 Application to seismic data inversion …… (79)
Chapter Ⅳ. Backus-Gilbert Inversions …… (83)
4.1 Inverse problems of continuous Earth parameters …… (83)
4.2 Non-uniqueness of inverse problems …… (87)
4.3 Linearization of nonlinear inverse problems …… (91)
4.4 The deltaness criterion …… (96)
4.5 The BG spread criterion …… (100)
4.6 The trade-off between resolution and variance …… (106)
Chapter Ⅴ. Seismic Tomography …… (112)

5.1 Data acquisition of seismic tomography …… (112)
5.2 The rays tracing …… (117)
5.3 The Radon transform and image reconstruction …… (125)
5.4 Discrete image reconstruction techniques …… (130)
5.5 The generalized Radon transform problems …… (138)
5.6 Diffraction seismic tomography …… (145)
5.7 Inverse scattering tomography …… (153)
References for Part I …… (166)

Part II: Advanced Geophysical Inversion

Chapter VI. Forward Problems in Applied Geophysics …… (172)
6.1 The common equation in geophysical inversion …… (172)
6.2 The finite difference method for wave propagation problems …… (177)
6.3 Introduction to the finite element method …… (179)
6.4 Special wave propagation problems …… (183)
Chapter VII. Geophysical Data Processing …… (190)
7.1 Geophysical data and data models …… (190)
7.2 The least squares methods in data processing …… (194)
7.3 Two-dimensional digital filters …… (199)
7.4 Downward continuation of geophysical fields …… (204)
7.5 The wavelet transform and multiscale analysis …… (209)
Chapter VIII. Inversion of Partial Differential Equations in Geophysics …… (215)
8.1 Generalized solutions and regularization …… (215)
8.2 The maximum entropy method …… (222)
8.3 Linear geophysical inversion …… (226)
8.4 Nonlinear iterative seismic inversion …… (230)
8.5 Imaging the electrical conductivity …… (237)
Chapter IX. Nonlinear Geophysical Inversion …… (242)
9.1 Iteration, Fractals and Chaos …… (242)
9.2 Nonlinear chaotic inversion of seismic traces …… (246)
9.3 The simulated annealing …… (252)
9.4 The genetic algorithm …… (256)
9.5 Application of the neural networks …… (261)
Afterword …… (268)
References for Part II …… (269)

上篇　地球物理反演基础

Part I Foundational Geophysical Inversion

地球物理学家面对人类赖以生存和发展的地球，她内部的奥秘无穷无尽。自50年代以来，地球物理观测仪器及其自动记录水平每隔10～15年提高一个级次，采集的信息量和计算机处理能力也相应提高一个级次。如果地球反演的水平也相应地提高一个层次，那么人类对地球内部组成与结构及其细节的认识，每隔30～50年，将有一次大的飞跃。

第一章 概 述

一、地球物理反演的研究对象

地球物理反演是在地球物理学中利用地球表面观测到的物理现象推测地球内部介质物理状态的空间变化及物性结构的一个分支。虽然地球物理学可分为固体地球物理学和勘探地球物理学两大方面，但这两方面在理论上都有一个共同的核心问题：如何根据地面上的观测信号推测地球内部与信号有关部位的物理状态，如物理性质、受力状态或热流密度分布等。这些问题就构成了地球物理反演的独特研究对象。具体来说，地球物理反演研究的是各种地球物理方法中反演问题共同的数学物理性质和解估计的构成和评价方法。它是从各个地球物理分支中抽象出来的新的边缘学科。

以地球内部为主要研究对象的地球物理研究过程一般可以用如图 1.1 的流程表示。反映地球内部物理过程的数据由数据获得系统收集，在勘探地球物理中数据获得系统不仅包括观测物理信号用的传感器和测量仪器，还可能包括激发物理信号的人工震源或场源，如可控震源、射线发生器等。数据获得是地球物理研究中最重要的环节之一，每当记录的地球物理数据的精度或信息量提高一个级次，就推动数据分析技术前进一大步，从而导致对地球内部结构的认识的精细程度提高一步。由于原始记录中包含着不同程度的干扰，也由于随着获得的数据量的不断增大而要求相应的数据管理和显示技术，原始记录的数据要先用计算机进行处理，以提高信噪比及提取与地质解释有关的某些特殊的信息。数据处理的结果一般还是描述物理现象的量，而不是描述地球内部物质物理状态的量，而反演计算的目的就是将物理现象的数据转化为地球内部物质物理状态变化的数据，为下一步地质解释提供依据。地质解释主要根据实验测定的岩石类型和物理性质之间的关系或者这方面的统计规律来推测地球内部不同岩石的分布区域以及相变带和突变带，并由此导出有关地球构造、动力和演化的佐证，或者得出有关矿产生成、储存和聚集的结论。由此不难看出，地球物理反演乃是地球物理研究过程中的关键环节，反演的准确度和分辨率实际上常常成为衡量地球物理研究成果的价值的关键指标。

地球物理反演又可分为单一地球物理现象的反演和多种地球物理现象的联合反演两大步骤。由于我们要推测的地球模型只有一个，它必须和地表观测到的所有物理现象保持一致，因此联合反演是地球物理数据分析的最理想工具。遗憾的是，由于各种技术上的困难和限制，适用于各种地球物理数据综合反演的模式和技术还不成熟，人们在大多数场合下还只能定性地对比单一地球物理数据反演结果的一致性。因此，这里讨论的反演方法主要是单一地球物理数据的反演方法，仅在第三章讨论广义反演时介绍联合线性反演的几种方法。

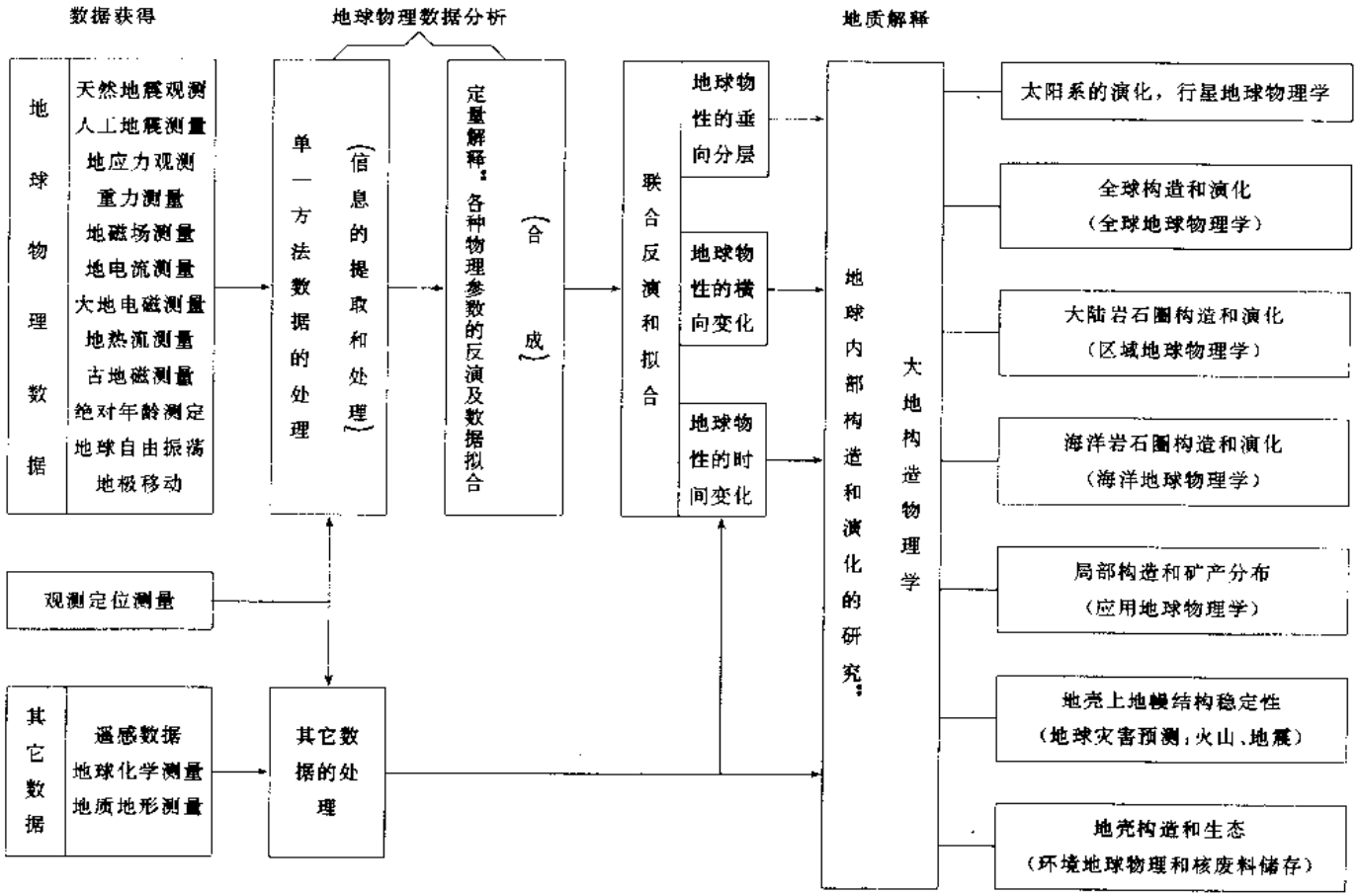


图 1.1 地球物理学研究的一般流程

二、地球物理反演的发展简史

如果把天然地震的震源参数的推测也归入地球物理反演的范围，那么我国东汉的科学家张衡应是从事地球物理反演的先驱。牛顿根据万有引力定律推测地球密度，开尔文(Lord Kelvin)研究地球的弹性和热传导，都是早期地球物理反演研究的范例。

地球物理学中反演问题的研究，在早期主要来自科学家对地球内部构造的好奇心，并受到物理学和地质学发展的推动。1880年美国学者J. A. Ewing等人发明了近代地震仪，此后地震记录的分析便提到日程上来。到了本世纪初，地震、地磁和重力的数据都有了一定程度的积累❶，地球物理数据的分析问题逐渐受到了重视，而定量的分析地球物理数据便导致了对地球物理反演问题的研究。有人认为[15]首先用于地球物理研究并取得反演问题数学解的是Herglotz于1907年提出的地震走时数据的反演。1909年A. Mohorovičic发现莫霍面，1912年Beno Gutenbeg发现古登堡面，以及1935年Inge Lehmann发现地球外核和内核的分界面。这些地学发展史上的辉煌事件，都对地球物理反演学术思想的形成和发展起过巨大的推动作用，至今还值得我们借鉴。

到了50年代，人们对地球内部的分层结构和各层的物理状态已经有了大致的了解，编制了地震波速、密度和磁化率随球状地球径向变化的分层模型，所用的反演方法主要是试错法和拟合法。此后，球状地球的分层模型在地震和地电、地磁数据的解释上再度建立功勋，确定了上地幔低速和高电导层的存在和深度，导致了岩石圈和软流圈概念的建立。

虽然应用物理学的方法来勘探矿产的思想萌芽于19世纪，但是勘探地球物理方法的研究和试验却是在第一次世界大战前后伴随着现代大工业的发展而兴起的。1926年在美国奥克拉荷马洲的沉积盆地上根据反射地震记录的解释布置的钻孔第一次打到了工业油流，证实了物探在找矿勘探中巨大的经济效益，促进了应用地球物理学的迅速发展。第二次世界大战以后，电子管地震仪，高精度石英丝重力仪和航空磁力仪的发明使勘探地球物理中的数据采集系统发生了质的变化，数据分析技术也随之走向复杂，因此有关勘探方法的地球物理学科逐渐从固体地球物理学中分离出来，自成体系。然而在50年代以前，由于计算工具的限制，对反演问题的研究远不如正演问题的研究那么受到重视。这时用于解释推断的工具主要为用于地震走时曲线或位场曲线的特征法(曲面斜率、特征点等)，与正演结果对比的选择法和量板法(即简化计算过程用的试错法，如电测深曲线量板)，以及用观测信号能量来推测场源参数的积分法等。

1947年第一台电子数字计算机的诞生标志着科学技术的一场革命，对地球物理数据分析产生了巨大的推动作用。50年代地球物理中的正演问题已广泛地使用了电子计算机，而与正演计算关系紧密的试错法和拟合法也随之转用计算机来实现。到了60年代，人们开始利用计算机高速运算的特性试验用计算机代替解释者自动修改地球模型的参数，实现所谓的计算机自动解释(称为自动拟合法或最优化选择法)。然而，在1967年Backus和Gilbert发表他们三篇经典性的论文以前，适用于各种地球物理方法的反演理论还没有形成，反演研

❶ 近代重力测量开始于1735年Pierre Bouguer率领的安第斯考察队。古代地磁测量在中国公元720年已有记录，J. Michell于1750年开始用磁秤测出磁场强度，1832年高斯提出了地磁场强度向量的测量方法。

究还只是作为正演研究的自然延伸分散在单一的地球物理分支之中。

总结 1970 年以前地球物理反演的研究过程，其主要特点可概括如下：

(1)地球模型在数学上多用被二次曲面分隔的均匀各向同性介质模型。

(2)由于对上述模型正问题的解析解(古典解)一般可以找到，反问题不需要直接与偏微分方程打交道，反演计算只用微积分或古典积分方程等数学工具。

(3)地球物理数据被认为是无限的、完整的、精确的或者只含可忽略的误差，在大多数情况下把它与根据假定模型正演计算取得的人工合成数据等同起来。

(4)虽然意识到反问题解的非唯一性，但没有对解非唯一的程度和后果作深入研究，评价反演结果的主要准则是用推测的地球模型求得的合成数据与实测数据的拟合好坏。

(5)所用到计算技术只涉及到初等数值分析，如数值微积分，最小二乘法解超定方程组等。在 60 年代后期开始用到快速 Fourier 变换和高速褶积。

60 年代是地质科学和计算机科学发生革命性飞跃的时期。电子计算机从第一代跃变为第三代，计算技术也获得了迅速发展，各种适用于高速计算机的新算法不断涌现。60 年代初开始的国际"上地幔研究"计划，要求研究地壳和上地幔的不均匀性，这就使基于二次曲面分割的地球模型不能满足新的要求。与此同时，地球物理勘探的仪器逐渐走向数字化，数据获得的总量和精度提高了一个级次。随着对复杂构造和地层岩性圈闭等精细地层结构圈定的需要，对反演计算结果的分辨率的要求也大大提高。因此，反演理论和方法的研究开始受到极大重视。在 1967～1970 年期间，美国地球物理学家 Backus 和应用数学家 Gilbert 连续发表了三篇关于地球物理反演的文章[17—19]，为建立统一的地球物理反演理论和方法奠定了基础。他们的主要贡献有：

(1)地球模型是客观地球在人们头脑中的反映，可以有无穷多个。在数学上地球模型可以用有限个有序的函数集合表示，与希尔伯特空间的元对应，这种希尔伯特空间称为模型空间。

(2)地球物理数据是有限个不精确的实数组成的集合，如果认为它们可以通过地球模型计算出来，则与模型空间相对应，也可以用希尔伯特空间的元表示地球物理数据集。

(3)这样定义的地球物理模型和数据的联系常常可以用有限个泛函方程式表示，反问题可以归纳为泛函方程组的求解。然而，由于零空间的存在和数据的有限性，反问题的解具有高度的非唯一性，即古典解一般不存在。

(4)对于精确的地球物理数据，地球物理反问题的古典解虽然可能不唯一，但是解的某种平均是唯一的，可以利用微扰法等数学工具构成某种迭代格式逐次逼近而求出满足规定准则的广义解。这样获得的解估计的分辨本领和精度不可能同时达到最高。

(5)在反问题研究中，重要的不是构造出一种广义的解估计，而是要对各种可能的解估计进行评价，这是由于反问题解的非唯一性所决定的，评价解估计的准则应该是在分辨本领和精度之间取合理的折衷，而不是实测和计算数据之间的拟合差最小。

Backus 和 Gilbert 的反演理论讨论的是模型为连续的情况，因此总是导致欠定的方程组，不便于在计算机上作快速计算。与 BG 理论对应的离散模型的情况下的反演在 1972 年由美国的 Wiggins 和英国的 Jackson 等人先后提出[20,21]，这就是广义反演方法。经过 Parker 等人的介绍和推广[22]，BG 反演理论和方法在 70 年代后期逐渐普及，并在北美许多大学的地球物理专业开设了地球物理反演方面的研究生课程，标志着它已经成为地球物理学中相

对独立的一个分支。

BG 反演理论的一个不足之处，是未能直接从运动方程和本构方程本身出发来提出反演问题。自 60 年代中期以来，偏微分方程的理论有了很大的发展，出现了拟微分算子和 Fourier 积分算子的理论。在另一方面，变系数波动方程中系数项或源项的反问题在物理学、医学、天文学和地学等不同领域及工程技术中不约而同地被提了出来，这些问题大都与不可直接观测的物体的成像问题有关。70 年代初期英国学者 G. Honsfield 研制出第一台医用 CT 机以及他和美国物理学家 A. M. Cormack 共同获得了 1979 年度生理学和医学的诺贝尔奖金，大大推动了有关不可见物体层析成像的研究热潮，激发了地球科学家们对变系数偏微分方程反问题及层析成像技术的巨大兴趣。由于 70 年代开展的国际“地球动力学计划”未能解决板块运动的驱动力等问题，地球科学家们认为 80 年代应该进一步研究地球内部的精细结构，以便为揭示地球内部物质运动的动力和地球演化提供可靠的佐证。因此，在 80 年代开展的国际地球岩石圈研究计划中，地学层析成像反演技术受到了特别关注，而且取得了令人鼓舞的成果。例如，美国哈佛大学和加州理工学院应用地震层析成像技术取得了揭示地球内部精细结构的层析图[23]，为板块运动的热对流学说提供了证据。由地震学得到的传统的地球模型仅给出了波速在不同深度上的平均值，只反映地球波速径向的平均变化；而层析图还同时揭示了波速的横向变化，并可由波速的横向变化了解地幔密度和温度的横向变化，即低速异常与低密度、高温的地幔物质对应。由层析图上地幔波速环状分布模式可推测地幔内热流的位置，从而间接地显示出地幔的对流图形。由于地震层析成像初步成果引人注目，1984 年美国 40 多个学术机构建立了地震学合作研究机构 IRIS(Incorporated Research Institutions for Seismology)，并取得了美国政府的大量资助。1984 年在美国亚特兰大召开的 54 届 SEG 年会上，也首次开辟了层析成像技术在油气勘探开发方面应用的专题讨论。总的来说，偏微分方程反演和地震层析成像的研究已把地球物理反演这门学科推向一个新的阶段。虽然这个领域中取得的研究成果还是初步的，理论还不系统，方法也不成熟，但是已为今后地球物理反演的发展勾划出一个重要的方向。有人认为地震层析成像的研究将导致地球科学的第二次革命，虽然还有不少人对此持慎重态度，但至少说明地球物理反演成像在今后的一段时期内将继续处于地学发展的前沿方向上。

与固体地球物理学中层析成像研究几乎同步，应用地球物理领域在 80 年代初期也开始了偏微分方程反演与层析成像的研究。例如，海湾石油公司与美国加州大学合作从 80 年代开始秘密地进行利用反射数据重建地下速度结构的研究，并在 1984 年勘探地球物理(SEG)年会上首次公布了地震层析成像的研究成果，引起了轰动。地震波波速的提取不仅对于准确圈定地下构造的传统地震勘探是非常重要的，而且波速的变化可以指示非背斜型的地层岩性圈闭，而这些类型的圈闭已成为现今油气勘探的主要目标。地震层析成像技术的发展不仅适应了油气勘探的要求，而且渗透到以提高采油率为目标的油藏工程的全过程。在油田开发方面，开采方案的设计、油田开采的监测和强化采油是提高产量和降低成本的关键环节。这些环节都要求准确地了解地下岩层的孔隙率、饱和度、渗透率和温度，而地震波速和品质因子与地层的这些物理参数有密切的关系。以强化采油为例，对稠油的水采和火采要求准确控制地层中注水面或火焰前沿的位置，这是油藏工程中一个极其困难的问题。由于地层注水或燃烧后其波速变化可达 20%至 70%，因此地震波速的层析成像在强化采油时可以起到有效的监测作用。目前，为油藏工程服务而不是以找油为目标的地震方法已从地震勘探中分离出

来，形成一个新的地震学分支——开发地震学(Production Seismology)，而地震层析成像就是开发地震学中的主要方法之一。在采油过程中应用的跨孔电磁层析成像技术近年来也有很大发展。可以预计能源和工业上的需求将促进地球物理反演的新分支——层析成像逐渐走向完善和实用。

80年代以来偏微分方程反问题研究进一步深化了反演的内涵。反演的实践证明，由于正问题拟微分算子具有不同程度的奇异性，反问题求解过程同时制造人为的假象。因此地球物理反演研究面对着相互矛盾的两个方面：一方面要从数据中提取尽可能多的信息，如最大限度地提高反演地球模型的分辨率；另一方面要尽可能压制难以避免的人造假象。因此，反演的含义应该为研究从数据中提取尽可能多信息的同时控制人造假象的理论与方法。原苏联数学家吉洪诺夫很早就意识到这个问题，曾建立了一套正则化理论用于反问题研究。在BG理论中最佳折衷的思想也是正则化理论对地球物理反演的发展。但是，这些方法都属于调和反演面临矛盾的方法，还没有深入去解决这一矛盾。

进入90年代，非线性理论在自然科学的各个领域都成为研究前沿；标志着自然科学一场新的革命的来临。大多数自然现象都是非线性的，地球物理反问题也是这样。实际计算表明，非线性反演要比线性反演复杂得多，而决不像Backus和Gilbert当初想象的那样，只要用线性化加迭代就能解决。实际上，非线性迭代本身就在不断地制造人造假象，把有用信息转化为熵。虽然非线性反演理论框架的建立仍要相当长的时间，但目前已有几个研究方向上取得了初步的成果，本书将有专门章节进行讨论。

我国关于地球物理反演理论和成像技术的研究起步较晚，但是BG反演方法在70年代就开始介绍到中国，并用于解决我国某些地球物理数据分析问题。我国著名的老科学家傅承义、冯康等都十分重视反演问题的研究[1]。80年代初，我国的大学已经开始设立了地球物理反演的研究生课程，而且多次召开了关于数学物理反演的学术会议，一个关于地球物理反演成像的新的热潮已经兴起。近年来，我国的地球物理学家和数学家们已为地球物理反演方法的发展作出了应有的贡献。

三、地球物理反演的研究内容和方法

地球物理反演的出发点是描述地球内发生的物理现象满足的数学物理方程。数学家称为偏微分方程，物理学称之为运动方程或场方程。在物理学家眼中，运动方程是一个物理体系中物理过程的数学描述，这里"过程"指物质状态的变化。实际上，对运动方程的研究和证实是物理学方法论的核心。典型的物理学研究方法是，对某个系统在某个时刻的状态列出运动方程，通过求解运动方程计算系统中表示状态的变量是怎样随时间变化的，然后得到这一物理过程的某些预测数据；如果这些数据与实验取得的数据相符，则证实了他们的假说和运动方程，产生了新理论；如果预测数据和实验数据不符，则反过来修改运动方程和假说，然后进行下一个循环。典型的物理学研究方法见图1.2。

对于地球物理学家来说，他们习惯用的研究方法和物理学的方法只有一点不同，即他们全盘接受在物理学中已经证实的运动方程(包括稳定场方程，下同)，而只修改介质函数和边

[1] 见1982～1985年期间中国数学年会及地球物理学术讨论会的报告。

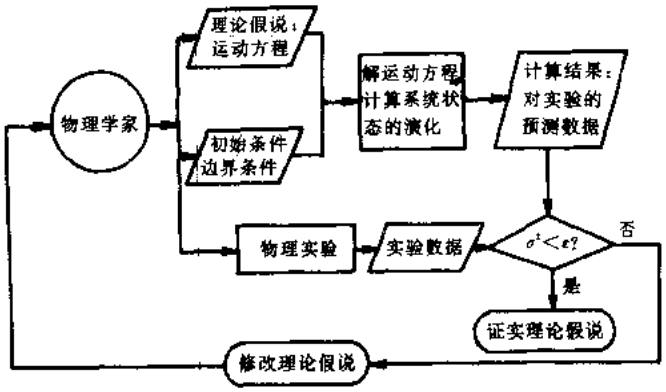


图 1.2　典型的物理学研究方法示意图

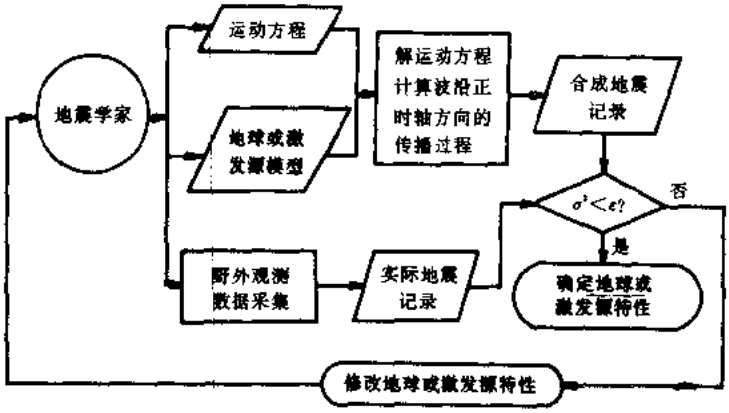


图 1.3　典型的地震学研究方法示意图

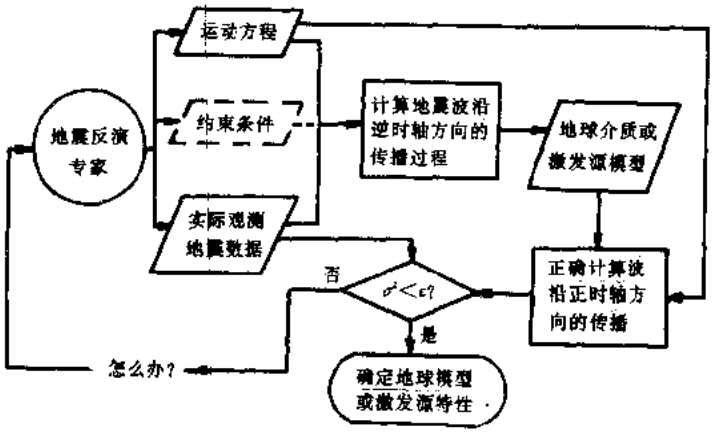


图 1.4　反演过程的示意图

界条件(地球模型)或初始条件(激发源特性)。以地震学为例,示如图 1.3 的典型研究方法可称为反演计算的试错法。对比图 1.2 和图 1.3 可清楚地看到研究方法的相似性。

在地球物理反演研究中,采用的方法和上述流程不一样。以地震反演为例,这时反演专家已知的运动方程和实际地震数据,这两者都不是可以随便修改的。利用波动方程来研究的不再是波向前的传播(过程的进化),而是波沿时间轴的后退(逆时轴传播),所期望的是通过波后退过程中地下质点振动状态的变化给出关于介质性质或激发源特性的消息(图 1.4)。为了证明反演的结果是否正确,可用图 1.3 所示的正演计算来合成地震图,并把它与实际地震数据比较。如果两者相符说明反演结果可靠,但是如果两者差别很大怎么办?难道允许反演专家去修改运动方程或实际数据(它们是反演系统的输入)吗?

对于这样的反演系统,至少有两个问题使物理学家产生怀疑。第一,反演涉及到的波的逆时传播是物理不可实现的。第二,系统的不封闭不能保证研究的深入,而且可能导致混乱(如错误地怀疑观测数据而布置更多实验)。这些问题的存在决定了地球物理反演研究内容和方法的特殊性。

要利用台站上接收的地震波传播的某种后果来研究发生在以前时刻的波的传播特性或激发过程,必须假定地震波沿时轴的正向和反向的传播是对称的。这一隐含的假定对数学家也许是无所谓的,但是物理学家们却心怀疑虑,因为时间具有方向性。根据热力学第二定律,如果在一个封闭的房间内放一小瓶香水,若干小时后房间内会充满宏观上均匀分布的香水分子,这是一个产生熵的不可逆过程,因为均匀分布香水分子的房间绝不会发生香水逐渐汇聚回小瓶子中去的"逆扩散"过程。物理学关于热力学的时间方向不可逆和自然过程产生熵的知识,导致沿时间轴正演和反演的不对称结论。与扩散过程一样,波动过程沿时间轴也是不对称的,在炮点的爆破产生了地面振动,如果人工发生相同的振动却不能导致振动能量集中回到炮点而激发爆破。从物理学和信息论分析,从源出发的地震波传播过程是消灭宏观信息而产生熵的过程,在传播过程中宏观信息不断地转化为微观信息,而微观信息又不断地被微扰所消耗(从而不断再生出熵增加的条件),因此利用波传播的"结果"不可能完全恢复波的"源",所谓的"逆时传播"是物理不可实现的。

当然,也有的物理学家认为,时间的方向性来自于孤立的封闭系统的假定实际上并不存在。他们认为,如果把封闭的物理体系扩大到整个宇宙,在测不准原理的规定限度内对宇宙给出完全微观的描述,则时间的两个方向就不会有质的差别,描述对于时间的正演和反演就将是对称的。这个观点看来有一定道理,但是对于无界的宇宙进行完全微观的描述是不现实的。

数学家们可以争辩说,逆时物理过程即使是物理不可实现的,但也可以用计算机进行"仿真"。以香水扩散过程为例,如果可以把小瓶子内每个分子扩散的轨迹用摄象机拍成电影,然后把电影片子倒过来放,就可以看到宏观上均匀分布的香水分子是怎样"逆扩散"回到小瓶子里去。类似地我们可以把地震波传播的计算机模拟(如波动方程地震合成)的计算步骤一步一步地颠倒过来,那么只要计算机的精度高到舍入误差可以忽略的程度,也可以把地面振动的传播能量逆向追回到源点,实现地震波逆时传播的"仿真"。这表明物理上不可能实现的自然过程具有数学上的可模拟性。然而,对于这一类物理不可实现的过程的模拟在数学上常常归结为所谓的不适定问题,即初始条件或边界条件有微小的误差会导致解答发散的一类数学问题,它的古典解即使存在,也难以准确地计算出来。这就使地球物理反演方法通

常都不依靠古典解，而只能根据某种准则找广义的解估计并且还要对解估计进行评价。在这里我们用“解估计”一词来表示反演计算的结果，它的含义是在某种准则或附加条件下对地球模型的一种猜测，而不是古典意义下数学问题的解，这是由反演问题解的非唯一性决定的。

让我们回到反演系统不封闭的问题上来，看看造成不封闭的原因是否与理论上的不完善性有关。为此，我们不妨把自己打扮成一个福尔摩斯式的侦探，到图 1.4 的反演流程中去找问题的症结。假定数学家们已经完满地解决了对称时轴上的逆时传播的“仿真”计算，那么造成反演失败的环节就只能出在：①运动方程；②地球物理数据；③介质模型或激发模型；以及④上述三方面的总体配合上。

首先让我们回顾一下与地球物理学有关的一些数学物理方程，它们大都可以用二阶偏微分方程表示。这些数学物理方程是由基本物理定律，例如质量守恒定律、牛顿运动定律和能量守恒定律等，加上由实验取得的本构关系推导出来的。应该注意到，在导出这些方程时我们对地球介质的性质作出了一定的假设，因此在用它们描述实际地球内发生的物理过程时带有一定的近似性。例如，我们假设地球物质是完全弹性体，因此可以用完全弹性体的本构关系即广义虎克定律来描述应力和应变的关系，从而推导出弹性波方程。但是，实际地球内的岩石和其它地内物质并非严格的完全弹性体，而是成分复杂的多相物质组成的非完全弹性体，这样弹性波方程就只能是地球内地震波传播过程的近似描述，而不是精确的描述。如果用弹性波方程作反演还太复杂，还必须加上其它假设条件进一步把它简化为声波方程，则作为反演出发点的方程本身的近似程度便更差一些。这种由于反演使用的基本方程的近似性有可能造成反演系统的不封闭。按理说，为了避免由于这种原因造成的反演的失败，应该尽量采用准确和复杂的数学物理方程作为设计反演算法的基础。然而，这样做不仅会导致反演算法的复杂化和计算机实现的困难，更重要的是可能使反问题的提法本身发生错误，因为方程变得复杂之后反问题的解可能不存在或者更加不唯一。一般来说，准确的方程涉及更多的未知地球参数，而作为边界条件的已知数据却总是有限的，对反演来说我们只能采用与数据中的信息量相匹配的适当近似的方程，而不是最准确的数学物理方程。由此可见，当发现反演系统不封闭时必须首先研究反问题的提法是否合适及基本方程本身的近似程度，并把它和已知地球物理数据的性质联系起来。

在进行地球物理反演时各类地球物理数据的物理意义是很清楚的，但它们在数学上的准确描述就不那么容易了。笼统地说，它们是含有地球信息的有限实数集，或者说以观测函数表示的地球信息和地质噪音在某种形式上的组合。理想的数据集可定义为观测函数在给定测点集上的取样数据集，与观测函数的关系可以用取样定理表示。这里假定取样是无限稠密的，而且是不带任何误差的，这就是数学家们对待地球物理数据的常用方法。

上述关于地球物理数据的定义对于正演问题是合适的，因为它们这时是指定的输出，应与实验数据拟合。在反演计算中，实验数据是系统的输入，有限性和不确定性是实验数据的属性。Backus 和 Gilbert 考虑到数据的有限性，把数据定义为“可以通过规定运算得到的有限实数集”，其中“规定的运算”可以由运动方程导出，因此数据实际上可用泛函方程组表示。虽然 BG 理论比“数据是观测函数的取样集”走前了一步，导出了一些关于数据有限性引起解的非唯一性定理，但数据中的不确定性仍然没有准确的描述。最常见的描述这一不确定性的方法是把数据视为确定成分（观测函数的取样）和噪音（高斯分布的随机变量）的和，并对噪

103156

音的统计性质予以规定。遗憾的是，这种"噪音模型"对地震反演不适用。例如当利用纵波和声波方程进行反演时，作为干扰的面波、横波和转换波都是有序的而不是随机的。对于波动方程反演来说还有一个极其头痛的问题，就是地震数据在频率域只分布在一个窄带之内，说明反散射等反演方法要求输入的许多信息在地震数据中根本不存在。数据中信息量的不足常常是导致地球物理反演发生困难的根源❶。因此，对地球物理数据的准确描述，它的信息量和自由度的估计也是地球物理反演研究的内容。

地球模型是反演计算的目标，也是数学家和地球物理学家常有争议的概念。地球模型一般指描述地下介质物理性质的函数集，它一般出现在运动方程的系数项或源项。对于天然地震研究也可以指激发特性，它一般出现在运动方程的源项。如果我们想用孕震区的微震来推测地下应力状态的变化，则这时震源过程的模型要比三维地壳模型复杂得多，能否用一个有序的函数集描写震源过程似乎还有问题，因为这时对应的正演问题还不十分清楚，我们以下只讨论介质模型的情况。这时，Backus 和 Gilbert 对地球模型是"由有限个函数组成的有序函数集"的定义仍然是可行的，问题是规定介质函数的具体性质。

从连续介质物理学的概念出发，地球模型最好用"逐块"光滑的间断函数集来描述，允许空间中存在某个间断面，介质函数 $m(r)$ 在间断点 r_d 处在有限范围内取不定值，而在相邻间断面内为连续光滑函数，这样的介质函数实际上是广义函数。间断型介质函数的复杂性不仅在于它在间断点处取不定值，还在于存在场函数没有定义的奇点。例如，在不同间断面的交点上波场 $f(r,t)$ 没有定义，因为这时 f 要满足相互矛盾的边界条件。在间断面不光滑的地方，由于界面的法线向量取不定值场函数也没有定义，也是场方程的不可去奇点。对于反射地震法来说，场的奇点是最重要的绕射源，与地层的断裂、尖灭和不整合等构造有关。然而，由于反问题的困难是目前大多数反演方法还没有用广义函数来描述地球介质，这也是导致反演系统不封闭的原因之一。

对于目前讨论得最多的反射地震法来说，介质函数是纵波速度（或声阻抗）$C(r)$，它是声波方程（或纵波方程）中的系数项，可用上面说的逐块连续的间断函数表示。令 S_0 为观测面，$u(r_0,t)$ 为其上的观测函数，$r_0\in S_0$，则反问题的提法可能有以下几种：

①给定关于 $u(r_0,t)$ 的有限数据集，根据运动方程确定函数 $C(r)$，即层析成像。

②给定关于 $u(r_0,t)$ 的有限数据集确定 $C(r)$ 的间断点集 $\{r'\}$。这类反问题可称为介质奇性反问题。当间断面为单一的封闭曲面时，它相当于确定边界形状的反问题，即轮廓成像。

③根据 S_0 面上的观测函数同时求研究区域内的介质函数 $C(r)$ 和波场函数 $u(r,t)$。这种反问题的提法可称外推反问题或"偏移反演"。

如上所述，这样提出的反问题不能保证反演系统封闭，即取得可以与数据拟合的解估计，其原因可能出在运动方程过于简化，地球物理数据的不足和地球模型的描述不够准确。或者从总体上看，运动方程要求的数据和模型与反演使用的不一致。一种常用的改进地球物

---

❶ 根据香农（C. E. Shannon）的信息论，信息被定义为对物理体系作统计描述的一种性质，单位为比特（bit），一比特的信息是两个相等的可能性之间决定其中一个需要的信息量，而一个体系的信息量等于对该体系完备的统计描述进行编码所需要的二进位数的最少位数。信息服从于守恒定律，即一个体系的信息与熵的和保持恒定，等于该体系在给定条件下所能达到的最大信息量或最大熵：

$$I+H=I_{\max}=H_{\max}$$

式中 I 表示信息量，H 为熵。信息是对物理体系的有组织程度（确定性）的度量，而熵是对系统的无序程度（不确定性）的度量，因此熵就是负的宏观信息。

理反演效果的方法是增加输入的信息量。如果附加的信息量来自于其它地球物理方法或地质观测的数据，我们所进行的是联合反演。如果附加的信息量来自于我们对地球模型的某些先验知识，则我们要进行的是加约束的反演。但是，约束条件来自于人们的认识，未免带有主观性，而联合反演虽然要求取得更多的数据，但避免了主观认识出错的可能性。

除钻孔资料及测井数据外，在地球物理反演中常用的约束条件有以下几种：①根据物理知识给出的宽约束，如密度和速度为正值，地球模型为有限能量等等；②以不等式给出的地球参数的紧约束，如某沉积盆地内岩石速度的变化范围或变化率等等；③对模型参数统计特性的约束（如相关函数，概率分布类型）；④对模型参数的加权平均值或线性组合所加的约束。按照 BG 理论，在许多情况下反问题的解不唯一，但是解估计的平均却是唯一的，因此对解估计的线性组合加约束是很有道理的；⑤吉洪诺夫正则方法，即对解估计本身及其偏导数加上一定的约束，例如对一阶偏导数的平方和（光滑度）加限制。后两种加约束的方法在正则化理论中都有详细讨论。后面的三种约束一般是数学家提出来的，数学家们关心的常常是用什么样的约束条件能恰如其分地把不适定问题的允许解集限制在紧集之中，化为条件适定问题求解，但这种恰如其分的约束不一定与物理观念一致。看来，要想找出数学家和物理学家都满意的符合客观的约束条件，需要在概念和数学工具上对约束有深一层的认识和升华。

综上所述，地球物理反演的研究方法是在对地球数据、模型和约束条件作准确的数学描述的基础上，根据数学物理方程导出能在计算机上快速实现的准确稳定算法，以进行某种物理不可实现过程的计算机“仿真”。因此，地球物理反演研究的内容不仅涉及到地球物理学、应用数学、而且还与信息科学、计算机科学的发展密切相关。

四、地球物理反演的计算机实现

在研究实际地球物理问题的时候，本书讨论的所有地球物理反演方法都要依靠现代电子计算机才能实现。因此，地球物理反演也是介于地球物理学和计算机科学之间的一门边缘学科，它的发展直接受到计算机科学技术水平的影响。用于固体地球物理学的反演方法最终是通过编制科学计算程序来实现的，这些程序本身一般不具有商业价值。然而，在地球物理勘探中反演方法的计算机实现是具有商业价值的软件，它和软件科学有更密切的关系，在算法设计、程序设计和质量控制（QC）方面有更高的要求。

数学公式的建立既不意味着解答存在，也不意味着求解的实用方法已经完成。例如，克莱姆法则以明晰的方式表示了线性代数方程组的解集的数学公式，但是没有提供任何求解线性方程组的实际方法。算法设计的中心就在于充分利用计算机的优点来减少求解的不可能性，同时尽量提高求解的效率和精度。自 50 年代以来，“数值分析”这门科学分支进入了与计算机科学结合的新纪元，促进了各种高速高精度算法不断涌现，而且建立了新的误差分析理论。由于在计算机中数的位数总是有限的，舍入误差不可避免，这种误差不仅影响到计算精度，指数传播的误差还造成了新的问题——算法的不稳定性。计算速度、精度和稳定性是在算法设计时要首先考虑的三个主要因素。

用计算机进行数值分析的核心思想是递推和逐步逼近，它们充分发挥了现代数字计算机的特点。利用先前的计算结果计算当前项的递推公式常常是简化程序和加速计算的关键。例如，在地震数据处理中常用的莱文生（Levinson）递推就是托布里兹（Toeplitz）矩阵求逆和

反褶积的心脏。利用迭代过程进行逐步逼近常常可以求出数学上认为不可解问题的近似解，缩小求解的不可能性。在地球物理反演的算法研究阶段，现代数值分析的方法和技术起了主要作用。

在找出适应计算机特点的快速稳定算法之后，就要用到计算机软件工程方面的一些方法和程序设计技术忠实地实现算法。在勘探地球物理中，还要求把反演计算过程标准化，使它成为具有商业价值的商品——应用软件。在这时计算机软件科学中的研究方法起到相当重要的作用。自 60 年代以来地震数据处理解释已被视为计算机应用软件行业的一部分，搞应用地球物理反演的人员还必须熟悉应用软件系统设计方面的要求和技术。

现代计算机软件设计原先多采用结构程序和模块程序设计，其基本思想相似于封建社会中的等级制度。比较复杂的程序设计目前多采取多层分级模块结构，在软件设计时首先作出输入-输出过程的分级图 HIPO(Hierarchy Input-Process-Output)。这种由相对独立的模块组成的软件不仅可以使软件生产工业化，而且便于软件系统中各模块的修改和不断更新，同时并不影响整个软件系统的运转。图 1.5 给出了一个利用模块技术设计应用软件系统的例子：反射法地震勘探数据处理软件系统的部分 HIPO 图。图中的模块分为五级，控制作业的程序一般在高级别的模块中，而操作程序和服务性程序占据着大部分“基层”模块。等级制度的作用表现在高级别的模块可以任意调用低级别的模块，而不能由低级模块调用高级模块。

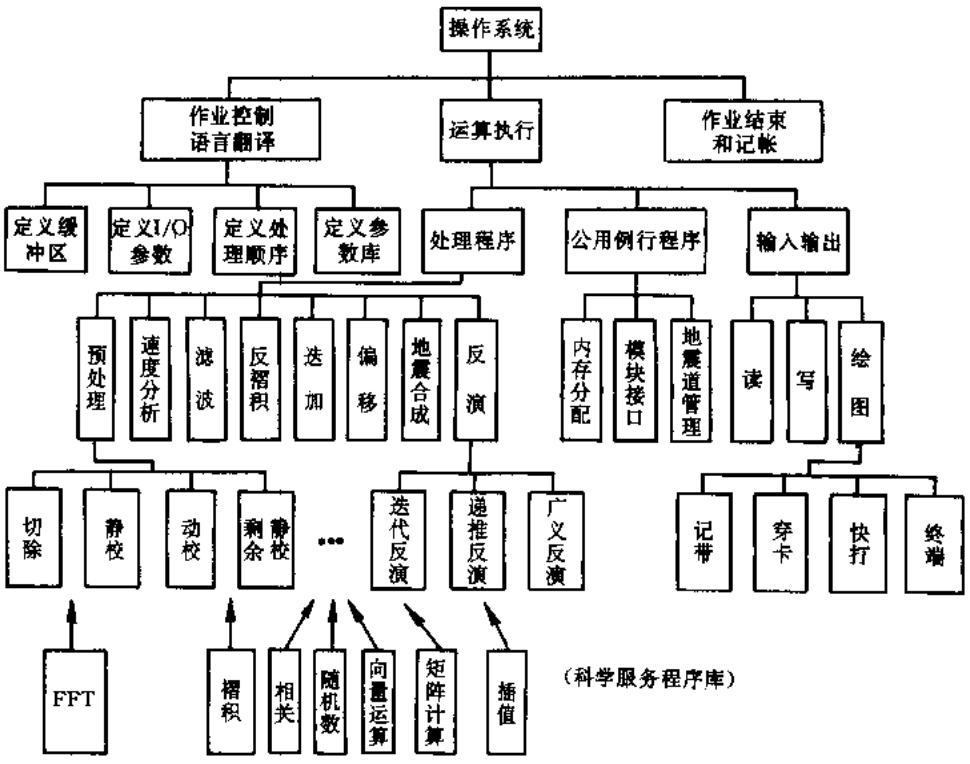


图 1.5　模块程序设计的例子：反射地震数据处理的一种软件系统

除了程序语言之外，在大型计算机的操作系统或大型专用软件系统中还规定了专门的作业控制语言(JCL)，它是程序员和操作员通过操作系统和计算机中央处理器联络的专用工具，是计算机用户需要掌握的工具之一。为了把作业控制语言化为低级机器语言，软件系

统常包含有翻译程序，这就是图 1.5 中翻译模块的功能之一。

上述多层覆盖的 HIPO 软件设计技术的缺点是灵活性差，而且层次的繁复给予软件的更新增加了不少麻烦。近年来，模块矩阵的设计思想受到广泛的重视。模块矩阵指的是以模块单元组成的阵列，通过建立一种软件平台，这些模块单元可以沿任意的次序和组合被调用。目前，大部分地震处理的商业软件都采用了这种结构设计。

应用软件的质量要从以下八个方面综合考虑，在设计应用软件时也要同时考虑这八个方面：

1. 计算精度。

2. 效率。

3. 稳定性。

4. 适用范围和韧性(Robustness)。韧性指的是数据和问题偏离于数学模型的原始假设时计算结果的精度和计算过程的稳定性变坏的速度的倒数。

5. 使用的难易程度。软件包应该给用户提供足够的提示和帮助，使外行的用户也能在计算机的帮助下使用它；同时尽量利用计算机的逻辑处理功能来减少用户的操作，并在出错时告诉用户处理的方法。这个指标有时也被称为计算的友善程度。

6. 合法性(Validation)。指问题和结果的定义是否准确，算法的细节是否正确合理，程序是否忠实地执行了算法，对典型问题的计算结果是否与解析公式结果相符，等等。

7. 兼容性：能用在多少不同的机器环境中。

8. 档案的完整性，软件的档案中最少应包括以下各项文件：(1)源程序，(2)控制过程及有关的作业控制语句规定，(3)输入输出(I/O)规定，(4)HIPO 图及流程图，(5)计算速度和精度试验结果，(6)试验所用的数据，(7)使用说明书。

进入 80 年代之后，向量计算机和便于人机对话的解释工作站已进入了计算机市场，为地球物理数据处理提供了更快更灵活的硬件工具。向量计算机特别适用于象地震勘探那样多道收录的数据集的处理；人机工作站又特别适合于大量地球物理数据的编辑、加工和显示，并可以方便地插入人们的先验知识，以提高地球物理解释的水平。实践表明，新的计算机硬件的出现不仅为地球物理反演方法的实现提供了新的可能性，也为发展反演软件提供了新的市场。由于计算机技术发展极快，许多过去认为是不现实的反演方法可能很快就会变为现实，因此，在地球物理反演方法研究中还必须有长远的眼光，使地球物理反演方法紧跟计算机技术不断地更新换代。

第二章　算法基础

由于反演计算一般是在计算机上实现的，而90%的反演方法最后要归结为线性问题上机计算，因此，有关矩阵运算的数值分析方法非常重要。通常，矩阵运算包括以下七方面的内容[24~25]：(1)线性方程组求解；(2)矩阵求逆和行列式问题；(3)特征值问题；(4)线性最小二乘问题；(5)广义逆；(6)奇异值分解；(7)算法的稳定性和扰动分析。本章我们将讨论除特征值问题外以后各章要用到的内容。算法设计方面要考虑的主要因素为精度、速度和内存要求，其中最主要的为精度，因此扰动分析也是算法研究的一个重点。下面首先从复习矩阵的范数开始进行算法讨论。

一、矩阵的范数

在本章中，我们用小写的罗马字母表示向量，如 $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c},\cdots\cdots$，用大写罗马字母表示矩阵，如 $\boldsymbol{A},\boldsymbol{B},\boldsymbol{C},\cdots\cdots$，本章中标量用小写希腊字母表示。

因此，
$$\boldsymbol{A}=[\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n]=(\alpha_{ij})$$
其中 i 是行号，j 是列号。

空间用大写花体字母表示，如 $\mathscr{R}^n$ 表示 n 维欧几里德空间。$\boldsymbol{A}^{\mathrm{T}}$ 表示 $\boldsymbol{A}$ 的转置。$\boldsymbol{A}^{\mathrm{H}}$ 表示复数矩阵 $\boldsymbol{A}$ 的复共轭转置，$\boldsymbol{A}^{\mathrm{H}}=\overline{\boldsymbol{A}}^{\mathrm{T}}=(\overline{\boldsymbol{A}^{\mathrm{T}}})$。$\boldsymbol{A}^{-1}$表示方阵 $\boldsymbol{A}$ 的逆矩阵。$\det(\boldsymbol{A})$表示方阵 $\boldsymbol{A}$ 的行列式$|\boldsymbol{A}|$，$\mathrm{Rank}(\boldsymbol{A})$表示矩阵 $\boldsymbol{A}$ 的秩。

按照上述规定，各种特殊矩阵可表为：

$\boldsymbol{A}=\boldsymbol{A}^{\mathrm{T}}$，$\boldsymbol{A}$ 为对称矩阵；

$\boldsymbol{A}=\boldsymbol{A}^{\mathrm{H}}$，$\boldsymbol{A}$ 为厄米特(Hermitian)矩阵；

$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}=\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}}=\boldsymbol{I}$，$\boldsymbol{Q}$ 为正交矩阵，其中 $\boldsymbol{I}$ 为单位矩阵。

如果 $\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{Q}=\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{H}}=\boldsymbol{I}$，则 $\boldsymbol{Q}$ 为酉阵(Unitary)。

在地球物理数据处理中常见的还有 Toeplitz 矩阵，它的主对角线元素皆为 α_0，上面的次对角线元素皆为 α_{-1}，而下面的次对角线元素为 α_1 等等，满足 $\alpha_{ij}=\alpha_{i+1,j+1}$。以 Toeplitz 矩阵为系数矩阵的 n 阶线性方程组可以用 Levinson 递推算法求解，运算次数为 n^2 拍。Toeplitz 矩阵在描述层状介质中的数学物理方法中起重要作用，在高分辨率频谱分析和数字滤波中应用非常广泛。

下面要简单地复习一下向量和矩阵的范数的有关定义，因为范数的概念在线性运算中是非常重要的。范数是一个不为负的数，它是向量或矩阵的模的一种简单的表示。范数概念来源于一个数的绝对值，数的绝对值满足以下规则：

$$\left.\begin{aligned}&\text{(i)}\ |\alpha|\geqslant 0;\ |\alpha|=0,\text{当且仅当}\ \alpha=0\\&\text{(ii)}\ |\alpha\beta|=|\alpha||\beta|\\&\text{(iii)}\ |\alpha+\beta|\leqslant|\alpha|+|\beta|\end{aligned}\right\}\tag{2.1}$$

将绝对值的概念推广到向量的范数，则向量的范数应具有与(2.1)对应的三个性质：

$$\left.\begin{array}{l}\text{(i)}\ \|\boldsymbol{v}\|\geqslant 0;\ \|\boldsymbol{v}\|=0,\text{如果而且仅对}\quad \boldsymbol{v}=\boldsymbol{0}(\text{零向量})\\ \text{(ii)}\ \|\alpha\boldsymbol{v}\|=|\alpha|\cdot\|\boldsymbol{v}\|\\ \text{(iii)}\ \|\boldsymbol{u}+\boldsymbol{v}\|\leqslant\|\boldsymbol{u}\|+\|\boldsymbol{v}\|\end{array}\right\}\qquad (2.2)$$

其中第三性质即是大家熟知的三角形不等式。

容易证明，用下述方法定义的向量范数具有(2.2)所表示的性质。设

$$\boldsymbol{x}=(\alpha_1,\alpha_2,\cdots,\alpha_n)^{\mathrm{T}}$$

定义范数一

$$\|\boldsymbol{x}\|_1=\sum_{i=1}^{n}|\alpha_i| \qquad (2.3)$$

范数二

$$\|\boldsymbol{x}\|_2=\left(\sum_{i=1}^{n}|\alpha_i|^2\right)^{\frac{1}{2}}=(\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x})^{\frac{1}{2}} \qquad (2.4)$$

而对于 $\rho=1,2,3\cdots\cdots$

$$\|\boldsymbol{x}\|_\rho=\left(\sum_{i=1}^{n}|\alpha_i|^\rho\right)^{\frac{1}{\rho}} \qquad (2.5)$$

如果 $\rho=\infty$，则规定

$$\|\boldsymbol{x}\|_\infty=\max_i|\alpha_i| \qquad (2.6)$$

在证明由(2.4)式定义的范数二满足三角形不等式

$$\|\boldsymbol{x}+\boldsymbol{y}\|_2\leqslant\|\boldsymbol{x}\|_2+\|\boldsymbol{y}\|_2$$

时，要用到 Gauchy-Schwarz 不等式

$$|\boldsymbol{x}^{\mathrm{H}}\boldsymbol{y}|\leqslant\|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2 \qquad (2.7)$$

此不等式的证明如下。

当向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 为线性相关时有

$$\alpha\boldsymbol{x}+\beta\boldsymbol{y}=0,\text{其中 }\alpha\neq 0\text{ 或 }\beta\neq 0$$

或写为

$$\boldsymbol{y}=\gamma\boldsymbol{x},\gamma\neq 0$$

因为

$$\|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2=\|\boldsymbol{x}\|_2|\gamma|\|\boldsymbol{x}\|_2=|\gamma|\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}$$

因此

$$|\boldsymbol{x}^{\mathrm{H}}\boldsymbol{y}|=|\gamma||\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}|=|\gamma|\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}=\|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2$$

在一般情况下，由范数的性质 1，有

$$0\leqslant\|\boldsymbol{x}-\alpha\boldsymbol{y}\|_2^2$$

其中 α 为任意的数。取 $\alpha=\|\boldsymbol{x}\|_2^2/(\boldsymbol{x}^{\mathrm{H}}\boldsymbol{y})$，代入上式

$$\begin{aligned}0&\leqslant\left\|\boldsymbol{x}-\frac{\|\boldsymbol{x}\|_2^2\boldsymbol{y}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{y}}\right\|_2^2\\&=\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}-2\boldsymbol{x}^{\mathrm{H}}\frac{\|\boldsymbol{x}\|_2^2\boldsymbol{y}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{y}}+\frac{\|\boldsymbol{x}\|_2^4\boldsymbol{y}^{\mathrm{H}}\boldsymbol{y}}{|\boldsymbol{x}^{\mathrm{H}}\boldsymbol{y}|^2}\\&=\|\boldsymbol{x}\|_2^2-2\|\boldsymbol{x}\|_2^2+\frac{\|\boldsymbol{x}\|_2^4\|\boldsymbol{y}\|_2^2}{|\boldsymbol{x}^{\mathrm{H}}\boldsymbol{y}|^2}\\&=-\|\boldsymbol{x}\|_2^2+\frac{\|\boldsymbol{x}\|_2^4\|\boldsymbol{y}\|_2^2}{|\boldsymbol{x}^{\mathrm{H}}\boldsymbol{y}|^2}\end{aligned}$$

在不等式中提出 $\|\boldsymbol{x}\|_2^2$ 因子有

$$\frac{\|x\|_2^2\|y\|_2^2}{|x^Hy|^2}-1\geqslant 0$$

于是不等式(2.7)得到证明。注意，上面的证明同时指出当向量 x 和 y 线性相关时 Gauchy-Schwarz 不等式取等号，这一点在本章第四节证明奇异值分解定理时要用到。

矩阵的范数常用的有 Frobenius 范数和从属(subordinate)范数两种。矩阵的 Frobenius 范数(F 范数)定义为

$$\|A\|_F=\Big(\sum_{i,j}|\alpha_{i,j}|^F\Big)^{1/F} \tag{2.8}$$

它具有(2.2)所规定的向量范数的三个性质。在实际中用得最多的是取 $F=2$ 的情况，即令

$$\|A\|_F=\Big(\sum_{i,j}|\alpha_{i,j}|^2\Big)^{\frac{1}{2}} \tag{2.9}$$

在科学和工程问题中，矩阵 A 经常与一个向量 x 一起运算，即构成一个新向量 Ax。那么，向量 Ax 的模比 x 的模大多少，必定与矩阵 A 的范数有关，于是，这就引出了所谓从属于某一向量 x 的矩阵的范数的概念。我们可以定义矩阵从属于向量 x 的范数 ρ 为

$$\|A\|_\rho=\max_{x\neq 0}\frac{\|Ax\|_\rho}{\|x\|_\rho} \tag{2.10}$$

其中向量的范数由(2.5)式规定。可以证明，这样规定的矩阵的从属范数满足以下性质：

$$\left.\begin{aligned}&\text{(i)}\ \|A\|\geqslant 0;\ \|A\|=0,\text{如果而且仅对 } A=0\text{(零矩阵)}\\&\text{(ii)}\ \|\alpha A\|=|\alpha|\cdot\|A\|\\&\text{(iii)}\ \|A+B\|\leqslant\|A\|+\|B\|\\&\text{(iv)}\ \|AB\|\leqslant\|A\|\cdot\|B\|\end{aligned}\right\} \tag{2.11}$$

作为一个练习，我们先来证明(2.11)中的特殊性质(iv)的一个特例，即

$$\|Ax\|_\rho\leqslant\|A\|_\rho\cdot\|x\|_\rho, \tag{2.12}$$

由定义对于所有非零向量 x 满足

$$\frac{\|Ax\|_\rho}{\|x\|_\rho}\leqslant\max_{x\neq 0}\frac{\|Ax\|_\rho}{\|x\|_\rho}=\|A\|_\rho,$$

说明从属范数满足向量范数的一个特性，即满足不等式(2.12)。注意到 $\|x\|_\rho^{-1}$ 是一个非负的数，由(2.11)中的性质(ii)有

$$\|A\|_\rho=\max_{x\neq 0}\left\|\frac{Ax}{\|x\|_\rho}\right\|_\rho \tag{2.13}$$

设新向量 $y=x/\|x\|_\rho$，易见 $\|y\|_\rho=1$，则得从属范数的另一个表示形式为

$$\|A\|_\rho=\max_{x\neq 0}\left\|\frac{Ax}{\|x\|_\rho}\right\|_\rho=\max_{\|y\|_\rho=1}\|Ay\|_\rho \tag{2.14}$$

现在就不难证明矩阵从属范数的特有性质，即(2.11)中的性质(iv)了。因为利用(2.12)式有

$$\begin{aligned}\|AB\|_\rho&=\max_{\|y\|_\rho=1}\|ABy\|_\rho\leqslant\max_{\|y\|_\rho=1}\|A\|_\rho\|By\|_\rho\\&=\|A\|_\rho\max_{\|y\|_\rho=1}\|By\|_\rho=\|A\|_\rho\cdot\|B\|_\rho\end{aligned}$$

上面介绍了矩阵的 F 范数和从属范数，那么在实际中如何计算它们呢？对从属范数一，设

$$A=[a_1,a_2,\cdots,a_n]$$

则由(2.13)有

$$\| \boldsymbol{A} \|_1 = \max_{\| \boldsymbol{x} \|_1 = 1} \| \boldsymbol{Ax} \|_1 = \max_{j=1,2,\cdots,n} \| \boldsymbol{a}_j \|_1 \tag{2.15}$$

注意上边利用了 $\| \boldsymbol{x} \|_1 = 1$ 的条件。对于 F 范数一的平方，有

$$\| \boldsymbol{A} \|_F^2 = \sum_i \sum_j |\alpha_{ij}|^2 = \sum_j \| \boldsymbol{a}_j \|_2^2 = \mathrm{tr}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})$$

其中"tr→trace"(迹)表示矩阵对角线元素的总和。对于范数二，可用正交矩阵 $\boldsymbol{Q}$ 把一个对称矩阵 $\boldsymbol{A}$ 变为对角矩阵 $\boldsymbol{Q}^{-1}\boldsymbol{AQ}$ 来计算范数，因为正交变换不改变矩阵的范数。如设 $\boldsymbol{Q}$ 为酉阵，$\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{Q} = \boldsymbol{QQ}^{\mathrm{H}} = \boldsymbol{I}$，由从属范数定义

$$\| \boldsymbol{A} \|_2 = \max_{x \neq 0} \frac{\| \boldsymbol{Ax} \|_2}{\| \boldsymbol{x} \|_2}$$

$$\| \boldsymbol{QA} \|_2 = \max_{x \neq 0} \frac{\| \boldsymbol{QAx} \|_2}{\| \boldsymbol{x} \|_2}$$

由于

$$\| \boldsymbol{QAx} \|_2^2 = (\boldsymbol{QAx})^{\mathrm{H}}(\boldsymbol{QAx}) = \boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{Ax} = \| \boldsymbol{Ax} \|_2^2$$

则有

$$\| \boldsymbol{QA} \|_2 = \max_{x \neq 0} \frac{\| \boldsymbol{Ax} \|_2}{\| \boldsymbol{x} \|_2} = \| \boldsymbol{A} \|_2$$

说明正交变换不改变矩阵的范数。

二、线性方程组和扰动分析

(一)特殊系数矩阵的方程组

给定非奇异的 $n \times n$ 矩阵 $\boldsymbol{A}$ 和 $n \times 1$ 向量 $\boldsymbol{b}$，要求线性代数方程组的解

$$\boldsymbol{Ax} = \boldsymbol{b} \tag{2.16}$$

在下列情况下是十分容易的：

(1)$\boldsymbol{A} = \boldsymbol{I}$ (单位矩阵)，

(2)$\boldsymbol{A} = \boldsymbol{D}$ (对角阵)，

(3)$\boldsymbol{A} = \boldsymbol{Q}$ (正交矩阵)，$\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$

(4)$\boldsymbol{A} = \boldsymbol{L}$ (三角阵)

当 $\boldsymbol{A} = \boldsymbol{L}$ 时方程 $\boldsymbol{Lx} = \boldsymbol{b}$ 的解法在大多数分解型的线性算法中都要用到，因此有必要注记一下。此时将方程(2.16)写为

$$\begin{bmatrix} \lambda_{11} & 0 & \cdots & 0 \\ \lambda_{21} & \lambda_{22} & \cdots & 0 \\ \vdots & & & \vdots \\ \lambda_{n1} & \lambda_{n2} & \cdots & \lambda_{nn} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} \tag{2.17}$$

它的求解过程为

外循环

$$\begin{aligned} \eta_1 &= \beta_1/\lambda_{11} \\ \eta_2 &= (\beta_2 - \lambda_{21}\eta_1)/\lambda_{22} \\ &\vdots \\ \eta_i &= (\beta_i - \lambda_{21}\eta_1 - \cdots - \lambda_{i,i-1}\eta_{i-1})/\lambda_{ii} \end{aligned}$$

内循环

其中 $i=2,3,\cdots,n$。将 η_i 赋值到 β_i 的内存单元中去，则(2.17)求解的方法可表为

$$\begin{array}{l} i=1,2,\cdots,n \\ \quad j=1,2,\cdots,i-1,(i>1) \\ \quad\quad \beta_i := \beta_i-\lambda_{ij}\beta_j \quad (\text{内循环}) \\ \quad \beta_i := \beta_i/\lambda_{ii} \quad (\text{外循环}) \end{array} \tag{2.18}$$

其中 := 表示赋值号。

现在来估计算法的计算工作量。一般规定，在计算机中做一次乘法和一次加法，外加两次存取数叫做“一拍”(a flop)：1 拍=1 次乘法+1 次加法+2 次存取。

由(2.18)可见，在(2.18)的内循环中每循环一次正好为一拍，因此内循环的计算量为 $(i-1)$拍。由于外循环要做$(n-1)$次除法($i=1$ 时算入内循环中)，总的计算量为

$$1+2+\cdots+(n-1)=n(n-1)/2\simeq n^2/2(\text{拍})。$$

(二)Cholesky 分解

当(2.16)中系数矩阵 $\boldsymbol{A}$ 为对称正定时，即

$$\boldsymbol{A}=\boldsymbol{A}^{\mathrm{T}},\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{b}>0(\boldsymbol{x}\neq\boldsymbol{0})$$

时，$\boldsymbol{A}$ 可以分解为某一矩阵 $\boldsymbol{B}$ 和它的转置 $\boldsymbol{B}^{\mathrm{T}}$ 的乘积。实际上，令 $\boldsymbol{A}=\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}$，有

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{x}=(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{x})=\|\boldsymbol{B}^{\mathrm{T}}\boldsymbol{x}\|_2^2>0,$$

这是因为范数大于零，只有在 $\boldsymbol{x}$ 为零向量时才等于零。其实，可以证明，$\boldsymbol{A}$ 为对称正定时 $\boldsymbol{B}$ 为下三角阵 $\boldsymbol{L}$，Cholesky 分解就是首先把 $\boldsymbol{A}$ 分解为 $\boldsymbol{L}\boldsymbol{L}^{\mathrm{T}}$。由于此时(2.16)变为

$$\boldsymbol{L}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{x}=\boldsymbol{b}$$

令 $\boldsymbol{y}=\boldsymbol{L}^{\mathrm{T}}\boldsymbol{x}$，我们可以先求方程 $\boldsymbol{L}\boldsymbol{y}=\boldsymbol{b}$ 的解，解法即为上节讲过的回代。然后，再求 $\boldsymbol{L}^{\mathrm{T}}\boldsymbol{x}=\boldsymbol{y}$ 的解。于是，对称正定矩阵 $\boldsymbol{A}$ 的方程组(2.16)的解法可归纳为以下步骤：

(1)将 $\boldsymbol{A}$ 分解为三角阵 $\boldsymbol{L}$：$\boldsymbol{A}=\boldsymbol{L}\boldsymbol{L}^{\mathrm{T}}$

(2)用回代法解方程　　$\boldsymbol{L}\boldsymbol{y}=\boldsymbol{b}$

(3)用回代法解方程　　$\boldsymbol{L}^{\mathrm{T}}\boldsymbol{x}=\boldsymbol{y}$

(4)求余数向量 $\boldsymbol{s}=\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}$ 以确定精度。

下面就以 3×3 的对称正定矩阵为例，说明 Cholesky 分解的过程：

把

$$\boldsymbol{A}=\boldsymbol{L}\boldsymbol{L}^{\mathrm{T}} \tag{2.19}$$

写为

$$\begin{bmatrix}\alpha_{11}&\alpha_{12}&\alpha_{13}\\ \alpha_{21}&\alpha_{22}&\alpha_{23}\\ \alpha_{31}&\alpha_{32}&\alpha_{33}\end{bmatrix}=\begin{bmatrix}\lambda_{11}&&\\ \lambda_{21}&\lambda_{22}&\\ \lambda_{31}&\lambda_{32}&\lambda_{33}\end{bmatrix}\begin{bmatrix}\lambda_{11}&\lambda_{21}&\lambda_{31}\\ &\lambda_{22}&\lambda_{32}\\ &&\lambda_{33}\end{bmatrix}$$

第一步，由 $\boldsymbol{A}$ 的第一行得

$$\lambda_{11}^2=\alpha_{11},\lambda_{11}\lambda_{21}=\alpha_{12},\lambda_{11}\lambda_{31}=\alpha_{13}$$

即

$$\lambda_{11}=(\alpha_{11})^{1/2}$$
$$\lambda_{21}=\alpha_{12}/\lambda_{11}$$
$$\lambda_{31}=\alpha_{13}/\lambda_{11}$$

第二步，由 $\boldsymbol{A}$ 的第二行得

$$\lambda_{21}^2+\lambda_{22}^2=\alpha_{22}$$

$$\lambda_{21}\lambda_{31}+\lambda_{22}\lambda_{32}=\alpha_{23}$$

$$\lambda_{22}=(\alpha_{22}-\lambda_{21}^2)^{1/2}$$

$$\lambda_{32}=(\alpha_{23}-\lambda_{21}\lambda_{31})/\lambda_{22}$$

第三步，由 A 的第三行得

$$\lambda_{31}^2+\lambda_{32}^2+\lambda_{33}^2=\alpha_{33}$$

即

$$\lambda_{33}=(\alpha_{33}-\lambda_{31}^2-\lambda_{32}^2)^{1/2}$$

由于 A 是对称矩阵，作 Cholesky 分解时只要用到 A 对角线以上的元素 α_{ij}。

对于 $n\times n$ 的矩阵 A，可用数学归纳法求得 Cholesky 分解的算式为

对 $i=j$

$$\lambda_{jj}=\left[\alpha_{jj}-\sum_{k=1}^{j-1}\lambda_{jk}^2\right]^{1/2} \tag{2.20}$$

对 $i=j+1,j+2,\cdots,n$

$$\lambda_{ij}=\frac{1}{\lambda_{jj}}\left[\alpha_{ji}-\sum_{k=1}^{j-1}\lambda_{ik}\lambda_{jk}\right] \tag{2.21}$$

其中 $k=1,\cdots,j-1$ 为内循环，$i=j,j+1,\cdots,n$ 为中循环，而 $j=1,2,\cdots,n$ 为外循环，三重循环意味着计算量正比于 n^3。由于(2.20)和(2.21)实际上是递推公式，即用已求得的 λ_{ij} 来算新的 $\lambda_{kl}(k\geqslant i,l\geqslant j)$，实际计算量为 $N^3/6$ 拍，为高斯消去法的一半。把 Cholesky 分解表为以下过程，便不难看出这一点：

```
j = 1,2,…,n
    i = j
        sum = 0
        k = 1,2,…,j − 1
        sum = sum + λ_ik^2
    λ_ij := (α_ji − sum)^(1/2)
    i = j + 1,…,n
        sum = 0
        k = 1,…,j − 1
        sum = sum + λ_ik λ_jk
    λ_ij := (α_ji − sum)/λ_jj
```

由以上分析可见，Cholesky 分解法的特点是：

(1)除非系数矩阵为对角线，三角阵或 Toeplitz 矩阵，它在求解线性方程组时是最快的。

(2)舍入误差很小。

(3)如果矩阵不对称则(2.20)中的开方结果变为复数，此法的上述优点不再存在。实际上由于系数矩阵在计算机实现时可能变得不完全对称，单精度的 Cholesky 分解有可能失败，采用双精度计算常常可以避免出现这个麻烦。

(三)LU 分解

高斯消去法等效于将系数矩阵 A 分解为一个下三角阵 L 和一个上三角阵 U：

$$A=LU$$

但是，不带主元素的高斯消去法在数值计算中是不稳定的，例如，在一个十进制九位的计算机中对含矩阵

$$A=\begin{pmatrix}-10^{-10} & 1\\ 1 & 1\end{pmatrix}$$

进行高斯消去时，相当于用一个下三角阵去乘 A：

$$\underset{L^{-1}}{\begin{pmatrix}1 & 0\\ 10^{10} & 1\end{pmatrix}}\underset{A}{\begin{pmatrix}-10^{-10} & 1\\ 1 & 1\end{pmatrix}}=\underset{U}{\begin{pmatrix}-10^{-10} & 1\\ 0 & 1+10^{10}\end{pmatrix}}$$

对上述机器而言，在存 $1+10^{10}$这个数时，就把原始矩阵中十分重要的信息 $a_{22}=1$ 去掉了。因为这时实际存入机器的矩阵 U 为

$$\tilde{U}=\begin{pmatrix}-10^{-10} & 1\\ 0 & 10^{10}\end{pmatrix}$$

而

$$L\tilde{U}=\begin{pmatrix}1 & 0\\ -10^{+10} & 1\end{pmatrix}\begin{pmatrix}-10^{-10} & 1\\ 0 & 10^{10}\end{pmatrix}=\begin{pmatrix}-10^{10} & 1\\ 1 & 0\end{pmatrix}$$

比较矩阵 A 可知用直接消去法求 A 的逆是不稳定的。

作行主元素消去法时先把 A 的行号对调，这相当于把 A 乘以矩阵 P：

$$PA=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}-10^{-10} & 1\\ 1 & 1\end{pmatrix}=\begin{pmatrix}1 & 1\\ -10^{-10} & 1\end{pmatrix}$$

然后，把 PA 作三角阵分解，即用下三角阵 L^{-1}去乘 PA，把它化为上三角阵 U：

$$\underset{L^{-1}}{\begin{pmatrix}1 & 0\\ 10^{-10} & 1\end{pmatrix}}\underset{PA}{\begin{pmatrix}1 & 1\\ -10^{-10} & 1\end{pmatrix}}=\underset{U}{\begin{pmatrix}1 & 1\\ 0 & 1+10^{-10}\end{pmatrix}}$$

此时实际上存入计算机的矩阵为

$$\tilde{U}=\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}$$

其中舍入误差为 10^{-10}。对九位浮点运算来说，对矩阵 A 作 LU 分解的精度也是 10^{-10}，因为

$$L\tilde{U}=\begin{pmatrix}1 & 0\\ -10^{-10} & 1\end{pmatrix}\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}=\begin{pmatrix}1 & 1\\ -10^{-10} & 1-10^{-10}\end{pmatrix}$$

它与(PA)的元的最大误差小于等于 10^{-10}。通过此例，我们知道直接消去法的精度很低，在大多数场合下至少要使用半主元素消去法求线性方程组的解。

下面来估计高斯消去法的计算速度。记矩阵 A 的 LU 分解为以下形式

$$\underset{A}{\begin{pmatrix}\alpha_{11} & \cdots & \alpha_{1n}\\ \vdots & & \\ \alpha_{n1} & \cdots & \alpha_{nn}\end{pmatrix}}=\underset{L}{\begin{pmatrix}1 & & & \\ \lambda_{21} & 1 & & \\ \vdots & \ddots & \ddots & \\ \lambda_{n1} & \cdots & \lambda_{n,n-1} & 1\end{pmatrix}}\underset{U}{\begin{pmatrix}\mu_{11} & \cdots & & \mu_{1n}\\ & \mu_{22} & \cdots & \mu_{2n}\\ & & \ddots & \vdots\\ & & & \mu_{nn}\end{pmatrix}}$$

为了节省内存，把 L 和 U 的元写回矩阵 A 中去成为

$$\begin{pmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1n} \\ \lambda_{21} & \mu_{22} & \cdots & \mu_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ \lambda_{n1} & \cdots & \lambda_{n,n-1} & \mu_{nn} \end{pmatrix}$$

由高斯消去法可知，这时执行 LU 分解的算法及计算量可以表示如下：

$j = 1,2,\cdots,n-1$　　外循环

↑ $i = j+1,\cdots,n$　　中循环$(n-j)^2$ 拍

$\alpha_{ij} := -\alpha_{ij}/\alpha_{jj}$

$k = j+1,\cdots,n$　　内循环

↑ $\alpha_{ik} := \alpha_{ik} + \alpha_{ij}\alpha_{jk}$　$(n-j)\times 1$ 拍

由此可见，LU 分解的总计算量为

$$\sum_{j=1}^{n-1}(n-j)^2 \simeq n^3/3(\text{拍})$$

为 Cholesky 分解的一倍。

半主元素高斯消去法的计算步骤可归纳如下：

(1)找出 P 使 PA 的对角线元素占优势(具有较大绝对值)，同时有

$$\boldsymbol{PAx=Pb} \tag{2.22}$$

(2)做三角阵分解　$\boldsymbol{PA=LU}$

(3)由于

$$\boldsymbol{LUx=Ly=Pb} \tag{2.23}$$

用回代法解方程

$$\boldsymbol{Ly=Pb} \tag{2.24}$$

(4)解方程

$$\boldsymbol{Ux=y} \tag{2.25}$$

(5)将 x 回代求误差向量 $s=Ax-b$，确定计算精度。

半主元素消去法解(2.16)的计算量可以估计如下，做 LU 分解的计算量为 $n^3/3$ 拍，解方程(2.24)和(2.25)各要 $n^2/2$ 拍，加上乘矩阵 P 的计算量，总的计算成本略大于 $n^3/3+n^2$。高斯主元素消去法的最大问题是计算精度强烈地依赖于矩阵 A 的特性而难以估计，因此求得的结果一般都需要回代检查。鉴于此法存在的缺陷，在应用它解线性方程组时必须考虑以下因素：

(1)如果系数矩阵 A 为对称正定则用 Cholesky 分解而不用消去法。如果 A 的对角线元素明显地占优势，则不用主元素，因为用它不仅不必要反而有损解的精度。

(2)如果要用主元素或半元素，必须肯定系数的相对大小反映了它们相对的重要性。注意在不同的系数反映不同物理性质的数量时，如在联合反演的场合，这种反映关系常常被破坏。

(3)如果用消去法不能提供足够精确的解答，首先要考虑应用 QR 分解等其它方法的可能性，然后再考虑用全主元素或双精度计算。

(四)线性方程组的扰动分析

在解方程组时解的误差不仅来自机器字长有限引起的舍入误差，还可能来自系数不准

确或数据不准确引起的误差。设(2.16)式中的系数矩阵 A 带有微小误差 δA，或数据 b 带有微小误差 δb，这时求得的解为 $\hat{x}=x+\delta x$ 也是带误差的。

通过分析以下两式

$$(A+\delta A)(x+\delta x)=b \tag{2.26}$$

$$A(x+\delta x)=b+\delta b \tag{2.27}$$

来推测 δA 及 δb 在多大程度上影响 $\hat{x}$ 的问题，叫做扰动分析(Perturbation analysis)。在地球物理反演问题中，研究扰动分析是很有意义的，因为在许多情况下系统矩阵 A 不可能精确地求得，例如，在进行最佳滤波时，A 的元素是由信号的自相关系数计算的，当然包含有观测信号中的误差，在波动方程反演中也有类似的情况。在实际进行扰动分析时，对 δA 或 δx 的每一个元素进行分析是不方便的，因此我们一般讨论 δA，δb 和 δx 的模，即 $\|\delta A\|$、$\|\delta b\|$ 及 $\|\delta x\|$，或它们的相对值 $\|\delta A\|/\|A\|$，$\|\delta b\|/\|b\|$ 和 $\|\delta x\|/\|x\|$。

在开始进行线性方程组的扰动分析之前，先来复习一下与此有关的方程组的病态和矩阵奇异的概念。在线性方程组中，如果某些系数或数据的微小变化会导致解的明显变化，则称此方程组为病态方程组，否则称方程组为良态方程组。由克莱姆法则可知，病态方程组的系数矩阵 A 比较接近奇异，即 $|\det(A)|$ 与最大系数 $|\alpha_{ij}|$ 的比值非常小，而逆矩阵 A^{-1} 中的元素与解相比却非常大。当 $|\det(A)|=0$ 或等于机器零时 A 变成奇异，即 $n\times n$ 矩阵 A 的秩小于 n(亏秩)，方程组无解。在某些情况下，可以用一种很直观的办法来分辨系数非稀疏矩阵 A 是否接近奇异，即比较它的对角线元素和最大的非对角线元素，如果前者的绝对值大得多，则 A 非奇异的可能性大，方程组在多数情况下是良态的。但是，这种直观的办法不是绝对可靠的，只有第四章讲的奇异值分解的办法才真正可靠。

举一个病态方程组的例子，设二阶方程组为

$$\left.\begin{aligned} 2\alpha_1+6.00001\alpha_2&=8.00001\\ 2\alpha_1+\quad\ 6\alpha_2&=8 \end{aligned}\right\} \tag{2.28}$$

显然，(2.28)的解为 $\alpha_1=\alpha_2=1$。如果第一式右边变为 8.00002，则新的解变为

$$\alpha_1=-2 \qquad \alpha_2=2$$

说明当数据变化仅为 $\|\delta b\|_2/\|b\|_2\simeq 10^{-6}$ 时，解的变化为 $\|\delta x\|_2/\|x\|_2=\sqrt{5}$。方程组(2.28)为病态的原因在几何上可以用它们代表的两条直线接近平行来说明。对于一般的方程组 $Ax=b$ 如何来定量地表示是否病态呢？这就是扰动分析要讨论的问题。

设数据 b 有微小变化时解 x 的变化为 δx，这时由(2.16)有

$$A(x+\delta x)=b+\delta b$$

$$\delta x=A^{-1}\delta b$$

根据从属范数的性质(2.12)式有

$$\|\delta x\|\leqslant\|A^{-1}\|\,\|\delta b\| \tag{2.29}$$

由于 $Ax=b$，有 $\|b\|\leqslant\|A\|\cdot\|x\|$，即

$$\frac{\|\delta x\|}{\|A\|\,\|x\|}\leqslant\frac{\|A^{-1}\|}{\|b\|}\|\delta b\|$$

因此

$$\frac{\|\delta x\|}{\|x\|}\leqslant\mathrm{cond}(A)\frac{\|\delta b\|}{\|b\|} \tag{2.30}$$

其中
$$\mathrm{cond}(A)=\|A\|\cdot\|A^{-1}\| \tag{2.31}$$
称为矩阵 A 对方程的解的条件数。显然，$\mathrm{cond}(A)$很小时方程组对 b 的扰动是良态；反之，$\mathrm{cond}(A)$很大说明方程组病态。

当矩阵 A 带有误差 δA 时，由(2.16)式有
$$(A+\delta A)(x+\delta x)=b$$
$$\delta x=-A^{-1}\delta A(x+\delta x)$$
因此
$$\|\delta x\|\leqslant\|A^{-1}\|\cdot\|\delta A\|\cdot\|x+\delta x\| \tag{2.32}$$
$$\frac{\|\delta x\|}{\|x+\delta x\|}\leqslant\mathrm{cond}(A)\frac{\|\delta A\|}{\|A\|} \tag{2.33}$$
由此可知，当矩阵 A 带有误差 δA 时解的相对误差仍然取决于条件数的大小。因此，条件数是系数矩阵是否接近奇异的一种定量表示。然而，当 $\|\delta x\|$ 很大时(2.33)式并不表示解的相对误差，而这时可用(2.32)式导出更精确的公式，即
$$\|\delta x\|\leqslant\|A^{-1}\|\cdot\|\delta A\|\cdot(\|x\|+\|\delta x\|)$$
$$\|\delta x\|(1-\|A^{-1}\|\|\delta A\|)\leqslant\|A^{-1}\|\|\delta A\|\|x\|$$
最后可得，当
$$\mathrm{cond}(A)\cdot\|\delta A\|/\|A\|<1\text{ 时有}$$
$$\frac{\|\delta x\|}{\|x\|}\leqslant\frac{\mathrm{cond}(A)}{1-\mathrm{cond}(A)\dfrac{\|\delta A\|}{\|A\|}}\cdot\frac{\|\delta A\|}{\|A\|} \tag{2.34}$$
由(2.34)可见，条件数对估计方程组的解的质量很重要。那么，如何计算条件数 $\mathrm{cond}(A)$呢？利用(2.31)式的定义求它要先求矩阵 A 的逆 A^{-1}，这往往是困难的，当 A 为奇异时是不可能的。利用奇异值分解算法求条件数的问题将在本章第四节中讨论。

三、QR 分解和线性最小二乘问题

QR 分解指的是把一个 $n\times m$ 矩阵 A 分解为一个正交矩阵 Q 和一个三角矩阵 R，由于 QR 分解算法具有较好的稳定性，在解所谓的线性最小二乘问题时特别适用。由于地球物理数据不可避免地带有误差，常常必须将误差的影响最小化，这就导致了最小二乘问题。QR 分解不仅是对于最小二乘问题的最强有力的工具，而且也是更复杂的算法(如下章将要讨论的奇异值分解算法)的基础。

作为 QR 分解的基础的正交三角变换原先是由 Householder 于 1957 年提出的[26]，因此它又被称为 Householder 变换。第一次用这个变换来解最小二乘问题的是 Golub(1965)[27]，他和 Basinger 一起公布了一个用 QR 算法解法方程组的程序。对方程的细节、舍入误差和扰动分析的讨论一直延续到七十年代后期，这不仅使它逐步成熟起来，而且使它在经济、工程和科学计算中取得了广泛的应用。

(一)非奇异矩阵的 QR 分解

通过下述的例子可以看到，用一个正交矩阵乘以一个向量可以使该向量的一个元素变为零。设 P 为一个 2×2 的正交矩阵：

$$\boldsymbol{P}=\begin{bmatrix}\gamma & \sigma\\ -\sigma & \gamma\end{bmatrix} \tag{2.35}$$

其中

$$\sigma^2+\gamma^2=1 \tag{2.36}$$

显然，$\sigma=\sin\theta$ 和 $\gamma=\cos\theta$ 满足(2.36)式，而且 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}=\boldsymbol{I}$。

假如有一个二阶向量 $\boldsymbol{x}=(\alpha,\beta)^{\mathrm{T}}$，则可以选取 γ 和 σ 的值使得

$$\boldsymbol{P}\boldsymbol{x}=\begin{bmatrix}\gamma & \sigma\\ -\sigma & \gamma\end{bmatrix}\begin{bmatrix}\alpha\\ \beta\end{bmatrix}=\begin{bmatrix}\rho\\ 0\end{bmatrix} \tag{2.37}$$

由(2.37)的第二个方程有

$$\gamma\beta=\sigma\alpha \tag{2.38}$$

代入(2.36)有

$$\gamma^2\beta^2=(1-\gamma^2)\alpha^2$$

移项后得

$$\gamma=\alpha/\sqrt{\alpha^2+\beta^2} \tag{2.39}$$

类似地由(2.38)得

$$\sigma=\gamma\beta/\alpha=\beta/\sqrt{\alpha^2+\beta^2} \tag{2.40}$$

由(2.37)的第一个方程和(2.39)、(2.40)可得

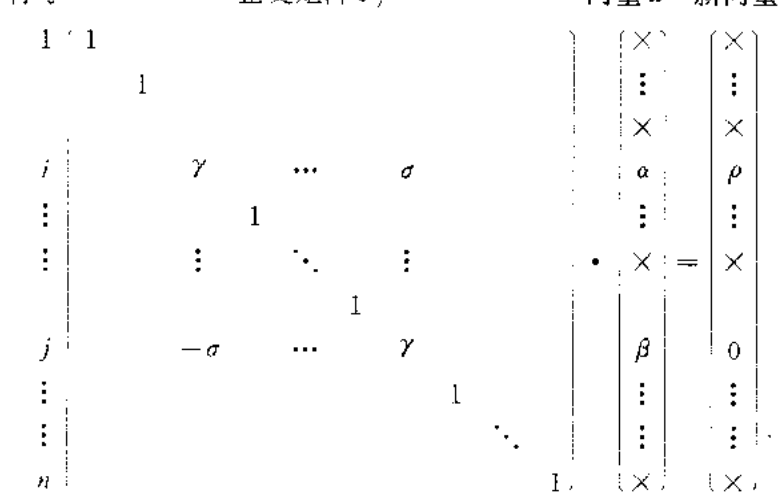
$$\rho=\gamma\alpha+\sigma\beta=(\alpha^2+\beta^2)^{1/2}=\|\boldsymbol{x}\|_2$$

因此，对给定的向量 $\boldsymbol{x}=(\alpha,\beta)^{\mathrm{T}}$，计算

$$\begin{aligned}\rho&=(\alpha^2+\beta^2)^{1/2},\\ \sigma&=\sin\theta=\beta/\rho,\\ \gamma&=\cos\theta=\alpha/\rho,\end{aligned} \tag{2.41}$$

我们得到一个新向量 $\boldsymbol{P}\boldsymbol{x}=(\rho,0)^{\mathrm{T}}$，它的模与 $\boldsymbol{x}$ 的模相同，而其中一个元素为零。在几何上新向量是把 $\boldsymbol{x}$ 顺时针旋转了 θ 角，使旋转后向量 $\boldsymbol{P}\boldsymbol{x}$ 与一个坐标轴重合，即它在另一个坐标轴上的投影为零。因此，矩阵 $\boldsymbol{P}$ 又叫旋转矩阵。

上述的 2×2 矩阵的旋转可以推广到 $n\times n$ 旋转矩阵的情况。设 $\boldsymbol{x}$ 为一个 n 维向量，现在要通过旋转把它的第 j 个元素化为零，那么我们可以将 $\boldsymbol{x}$ 乘以下图所示的正交矩阵 $\boldsymbol{P}_j$：



上图中的×号表示任意非零元素，图中的参数 ρ、σ 和 γ 仍由(2.41)式计算。

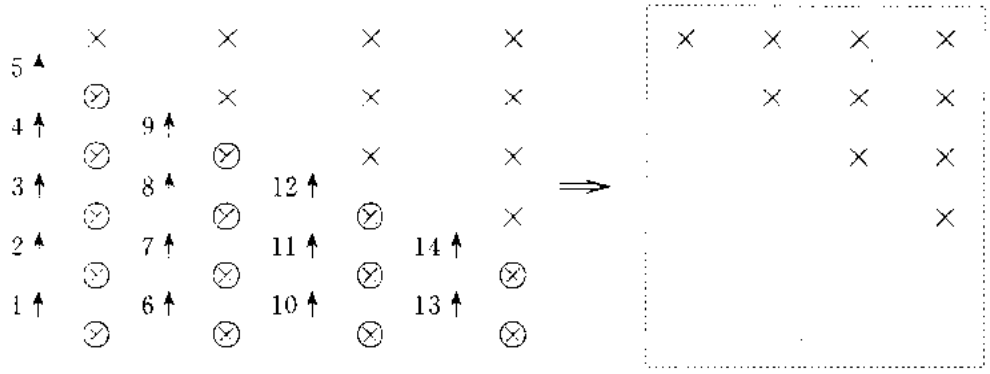
既然通过一次旋转就可以把一个 n 维向量 $\boldsymbol{x}$ 的一个元素化为零，那么通过 $(n-k)$ 次旋转便可以把 $\boldsymbol{x}$ 化为只含 k 个非零元素的 n 维向量。对于一个矩阵来说，如果我们对它的列向量依次进行旋转，便可以把它化为一个上三角阵。设 $\boldsymbol{A}$ 为秩等于 $m\leqslant n$ 的 $n\times m$ 矩阵，它的 QR 分解就是找到一个 $n\times n$ 正交矩阵 $\boldsymbol{Q}^{\mathrm{T}}$ 使得

$$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A}=\begin{pmatrix}\boldsymbol{R}\\ \boldsymbol{0}\end{pmatrix}\text{ 或 }\boldsymbol{A}=\boldsymbol{Q}\begin{pmatrix}\boldsymbol{R}\\ \boldsymbol{0}\end{pmatrix}. \tag{2.42}$$

其中 $\boldsymbol{R}$ 为 $m\times m$ 上三角阵，$\boldsymbol{0}$ 为 $(n-m)\times m$ 的零矩阵。

用下面的图示可以看出经过 14 次旋转把一个 6×4 矩阵 $\boldsymbol{A}$ 化为一个正交矩阵 $\boldsymbol{Q}$ 和一个三角阵 $\boldsymbol{R}$ 的过程由图可见，头五次旋转把 $\boldsymbol{A}$ 的第一列化为仅含有一个非零元素的向量，把这五次旋转对应的旋转矩阵记为 $\boldsymbol{P}_1\cdots\boldsymbol{P}_5$，第二列的旋转序号为 6—9，第三列为 10—12，第四列为 13—14。整个旋转过程，即 QR 分解的过程可用矩阵表示为

$$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{P}_{14}\cdots\boldsymbol{P}_1\boldsymbol{A}=\begin{pmatrix}\boldsymbol{R}\\ \boldsymbol{0}\end{pmatrix} \tag{2.43}$$



由于正交矩阵的乘积仍为正交矩阵，可以证明 $\boldsymbol{Q}$ 为正交矩阵：

$$\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}}=\boldsymbol{P}_1^{\mathrm{T}}\boldsymbol{P}_2^{\mathrm{T}}\cdots\boldsymbol{P}_{14}^{\mathrm{T}}\boldsymbol{P}_{14}\cdots\boldsymbol{P}_2\boldsymbol{P}_1=\boldsymbol{I} \tag{2.44}$$

顺便提一句，用正交矩阵 $\boldsymbol{Q}^{\mathrm{T}}$ 乘以 $\boldsymbol{A}$ 所得到的上三角阵就是矩阵 $(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})$ 的 Cholesky 分解所得到的三角阵 $\boldsymbol{L}$ 的转置。实际上

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A}=(\boldsymbol{0}^{\mathrm{T}}\quad\boldsymbol{R}^{\mathrm{T}})\begin{pmatrix}\boldsymbol{R}\\ \boldsymbol{0}\end{pmatrix}=\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R} \tag{2.45}$$

由于正态方程的系数矩阵 $(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})$ 是对称正定的，它的 Cholesky 分解是 $(\boldsymbol{L}\boldsymbol{L}^{\mathrm{T}})$，因此我们说一个非奇异矩阵 $\boldsymbol{A}$ 的 QR 分解乃是对称正定矩阵的 Cholesky 分解的进一步发展。

在进行 QR 分解之后，含有 m 个未知数的 n 阶线性方程组可用以下方法求解。注意到 $\boldsymbol{A}$ 是 $n\times m$ 非奇异矩阵 $(n\geqslant m)$，将(2.42)代入 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ 得

$$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=\boldsymbol{R}\boldsymbol{x}=\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}=\boldsymbol{c} \tag{2.46}$$

其中 $\boldsymbol{c}$ 为 n 维向量，但只有 m 个非零元素。由于 $\boldsymbol{R}$ 为三角阵，求(2.46)的解的运算量为 $n^2/2$ 拍，求 $\boldsymbol{x}$ 是很快的。

(二)QR 分解的快速算法

以下 $n\times n$ 埃尔米特矩阵

$$\boldsymbol{P}=\boldsymbol{I}-\tau\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}} \tag{2.47}$$

通常被称为 Householder 变换，其中 τ 为非零标量，$\boldsymbol{u}=(\mu_1\cdots,\mu_n)^{\mathrm{T}}$ 为任意向量，且 $\boldsymbol{u}\neq\boldsymbol{0}$。

由于 $$\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}=\boldsymbol{I}-2\tau\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}+\tau^{2}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}=\boldsymbol{I}+(\tau^{2}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}-2\tau)\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$$

因此当 $\tau=0$ 或 $$\tau=2/\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} \tag{2.48}$$

时 $\boldsymbol{P}$ 为正交矩阵。将(2.48)代入(2.47)则又可将正交的 Householder 变换表为

$$\boldsymbol{P}=\boldsymbol{I}-2\frac{\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}}{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}} \tag{2.49}$$

其中 $$\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}=\mu_{1}^{2}+\mu_{2}^{2}+,\cdots,+\mu_{n}^{2}$$

为标量。

正如本节第一段所述，正交旋转是要找到一个正交矩阵 $\boldsymbol{P}$ 使得向量 $\boldsymbol{x}$ 变为

$$\boldsymbol{P}\boldsymbol{x}=\rho\boldsymbol{e}_{1} \tag{2.50}$$

其中 $$\boldsymbol{e}_{1}=(1,0,\cdots,0)^{\mathrm{T}}$$

既然 $\boldsymbol{P}$ 可表示为(2.49)的形式，那么如何选择 $\boldsymbol{u}$ 而构成正交矩阵 $\boldsymbol{P}$ 呢？由于正交变换不改变向量 $\boldsymbol{x}$ 的模，即 $$\|\boldsymbol{x}\|_{2}^{2}=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{x}=\rho^{2}\boldsymbol{e}_{1}^{\mathrm{T}}\boldsymbol{e}_{1}=\rho^{2}$$

因此有 $$\rho=\pm\|\boldsymbol{x}\|_{2} \tag{2.51}$$

将(2.49)代入(2.50)，注意 $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{x}$ 是个标量有

$$\boldsymbol{P}\boldsymbol{x}=\boldsymbol{x}-2\left\{\frac{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{x}}{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}}\right\}\boldsymbol{u}=\rho\boldsymbol{e}_{1}$$

即 $$2\left\{\frac{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{x}}{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}}\right\}\boldsymbol{u}=\boldsymbol{x}-\rho\boldsymbol{e}_{1} \tag{2.52}$$

其中括号内的分式仍然是个标量。如令 $\boldsymbol{x}^{\mathrm{T}}=(\xi_{1},\xi_{2},\cdots,\xi_{n})$ 并选取

$$\boldsymbol{u}=\boldsymbol{x}-\rho\boldsymbol{e}_{1}=\begin{pmatrix}\xi_{1}-\rho\\ \xi_{2}\\ \vdots\\ \xi_{n}\end{pmatrix}=\begin{pmatrix}\mu_{1}\\ \mu_{2}\\ \vdots\\ \mu_{n}\end{pmatrix} \tag{2.53}$$

再选取(2.51)中 ρ 的符号使得 $(\xi_{1}-\rho)$ 不为零，令 $\mu_{1}=\xi_{1}-\rho$ 我们有

$$\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}=(\boldsymbol{x}-\rho\boldsymbol{e}_{1})^{\mathrm{T}}(\boldsymbol{x}-\rho\boldsymbol{e}_{1})=\rho^{2}-2\rho\xi_{1}+\rho^{2}=-2\rho\mu_{1} \tag{2.54}$$

由(2.48)还有： $$\tau=2/\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}=-1/(\rho\mu_{1}) \tag{2.55}$$

上面得到的(2.51)，(2.53)，(2.55)和(2.47)给出了将一个任意向量 $\boldsymbol{x}$ 旋转为只含一个非零元素的新向量 $\rho\boldsymbol{e}_{1}$ 的 Householder 更换 $\boldsymbol{P}$。在计算机中这一过程可归纳为

$$\left.\begin{aligned}&\rho:=-\mathrm{sign}(\xi_{1})(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x})^{1/2}\\ &\boldsymbol{u}:=\boldsymbol{x}-\rho\boldsymbol{e}_{1}\\ &\tau:=-1(\rho\mu_{1})\\ &\boldsymbol{P}:=\boldsymbol{I}-\tau\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\end{aligned}\right\} \tag{2.56}$$

在第一段中已经讨论过了用正交变换把矩阵 $\boldsymbol{A}$ 的列向量依次旋转为上三角阵的方法

$$\boldsymbol{P}_{n}\cdots\boldsymbol{P}_{1}\boldsymbol{A}=\begin{pmatrix}\boldsymbol{R}\\ \boldsymbol{0}\end{pmatrix} \tag{2.57}$$

虽然这个方法在原则上是可行的，但并不是高效率的算法。采用下述的所谓 Given 算法可以使计算速度提高一倍。在把 $n\times m$ 矩阵 $\boldsymbol{A}$ 乘以 $\boldsymbol{P}_{1}$ 使它的第一列旋转之后，把 $\boldsymbol{P}_{1}\boldsymbol{A}$ 分块，我们得到一个 $(n-1)\times(m-1)$ 的新矩阵 $\boldsymbol{A}_{1}$：

$$P_1A=P_1\begin{pmatrix}\times & \times & \cdots & \times\\ \times & \times & \cdots & \times\\ \vdots & \vdots & & \vdots\\ \times & \times & \cdots & \times\end{pmatrix}-\left(\begin{array}{c|ccc}\times & \times & \cdots & \times\\ \hline & \times & \cdots & \times\\ & \vdots & \boldsymbol{A}_2 & \vdots\\ & \times & \cdots & \times\end{array}\right)\Longrightarrow A_1$$

因此，进一步将 P_1A 作旋转时我们没有必要将它乘以 $n\times n$ 矩阵 P_2，而只需将 A_2 乘以 $(n-1)\times(n-1)$ 正交矩阵 P'_2，即可将 A 的第二列化为只含两个非零元素的列向量。依次将 P'_i 减阶并乘以分块矩阵 A_{i-1}，当进行到 $i=m$ 时 A 便化为三角阵 R 了。由于旋转矩阵 P_i 是不断减阶的，旋转计算量随之逐渐减小，用此 Given 算法实际上比(2.57)的旋转方法快一倍。

在计算过程中还要注意到内存量的节省。由于 $P=I-\tau uu^{\mathrm{T}}$，τ 和 u 便给出了 $n\times n$ 矩阵 P 的全部信息，因此计算时可以不必储存整个矩阵 P。实际上，计算时只需存入向量 $(\rho,\xi_2,\cdots,\xi_n)^{\mathrm{T}}$ 和 $(\xi_1-\rho,\tau)^{\mathrm{T}}$ 便可，连 u 也可以不存。

(三)线性最小二乘问题的 QR 分解法

设 A 为 $n\times m$ 矩阵，$\mathrm{rank}(A)=m\leqslant n$，求解方程 $Ax=b$ 的问题称为线性最小二乘问题；当 $m>n$ 时唯一的精确解不存在。设预测误差向量 $e=b-Ax$，我们可以用

$$\min_x \|b-Ax\|_{1,2,\cdots,\infty}$$

为准则求预测误差的模的最小解估计，当取 $\|e\|_2$ 为最小时就是线性最小二乘法。用矩阵 A 的 QR 分解求解最小二乘问题具有独特的优点(见下节)。

由于 $A=QR$，其中 Q 为 $n\times n$ 正交矩阵，则有

$$\min_x \|b-Ax\|_2=\min_x \|Q^{\mathrm{T}}b-Q^{\mathrm{T}}Ax\|_2$$

注意到

$$Q^{\mathrm{T}}A=\begin{pmatrix}R\\ 0\end{pmatrix},R\text{ 为 }m\times m\text{ 上三角阵},$$

并把矩阵 Q 分块为

$$\underset{n\times n}{Q}=(\underset{n\times m}{Q_1},\ \underset{n\times(n-m)}{Q_2})$$

则

$$\|Q^{\mathrm{T}}b-Q^{\mathrm{T}}Ax\|_2^2=\left\|\begin{pmatrix}Q_1^{\mathrm{T}}b\\ Q_2^{\mathrm{T}}b\end{pmatrix}-\begin{pmatrix}Rx\\ 0\end{pmatrix}\right\|_2^2$$

$$=\|Q_1^{\mathrm{T}}b-Rx+Q_2^{\mathrm{T}}b\|_2^2\leqslant\|Q_1^{\mathrm{T}}b-Rx\|_2^2+\|Q_2^{\mathrm{T}}b\|_2^2 \tag{2.58}$$

由于上式右边的第二项与 x 无关，它对 x 的导数为零，此项在最小化时不起作用。容易看出，上式右边的第一项为零时 $\|b-Ax\|_2^2$ 取最小值，即预测误差的模的最小化等效于

$$\|Q_1^{\mathrm{T}}b-Rx\|_2^2=0 \tag{2.59}$$

因此，$Ax=b$ 的最小二乘解要满足的方程为

$$R\hat{x}=Q_1^{\mathrm{T}}b \tag{2.60}$$

式中 $\hat{x}$ 表示最小二乘解估计。

实际上，用矩阵运算可将最小化的过程表示如下：

$$\|Q_1^{\mathrm{T}}b-Rx\|_2^2=(Q_1^{\mathrm{T}}b-Rx)^{\mathrm{T}}(Q_1^{\mathrm{T}}b-Rx)$$

$$=(b^{\mathrm{T}}Q_1-x^{\mathrm{T}}R^{\mathrm{T}})(Q_1^{\mathrm{T}}b-Rx)$$
$$=b^{\mathrm{T}}Q_1Q_1^{\mathrm{T}}b-x^{\mathrm{T}}R^{\mathrm{T}}Q_1^{\mathrm{T}}b-b^{\mathrm{T}}Q_1Rx+x^{\mathrm{T}}R^{\mathrm{T}}Rx$$
$$=x^{\mathrm{T}}R^{\mathrm{T}}Rx-2x^{\mathrm{T}}R^{\mathrm{T}}Q_1^{\mathrm{T}}b+b^{\mathrm{T}}b$$

其中 $b^{\mathrm{T}}Q_1Rx=(x^{\mathrm{T}}R^{\mathrm{T}}Q_1^{\mathrm{T}}b)^{\mathrm{T}}$ 为一个数，它的转置为它本身。

令

$$\frac{\partial}{\partial x^{\mathrm{T}}}\| Q_1^{\mathrm{T}}b-Rx\|_2^2=2R^{\mathrm{T}}Rx-2R^{\mathrm{T}}Q_1^{\mathrm{T}}b=0$$

即得(2.60)式。由(2.58)式可知，预测误差的模为

$$\| e\|_2=\| b-A\hat{x}\|_2=\| Q_2^{\mathrm{T}}b\|_2$$

矩阵 A 的 QR 分解同时揭示了线性最小二乘问题的几何意义。把预测误差向量写为

$$e=b-A\hat{x}=b-Q_1R\hat{x}=b-Q_1Q_1{}^{\mathrm{T}}b=(I-Q_1Q_1{}^{\mathrm{T}})b$$

因为

$$I=QQ^{\mathrm{T}}=(Q_1\quad Q_2)\begin{pmatrix}Q_1{}^{\mathrm{T}}\\ Q_2{}^{\mathrm{T}}\end{pmatrix}=Q_1Q_1^{\mathrm{T}}+Q_2Q_2^{\mathrm{T}}$$

即

$$Q_2Q_2^{\mathrm{T}}=I-Q_1Q_1^{\mathrm{T}}$$

因此有

$$e=Q_2Q_2{}^{\mathrm{T}}b$$

由于 Q 是正交矩阵，它的列向量张成一个欧几里德空间 $\mathcal{R}(Q)$，而 Q_1 和 Q_2 相应为 $\mathcal{R}(Q)$的子空间，记为 $\mathcal{R}(Q_1)$和 $\mathcal{R}(Q_2)$。由于 $e=Q_2Q_2^{\mathrm{T}}b$，预测误差 e 即为数据向量 b 在 $\mathcal{R}(Q_2)$空间的正交投影，而由于 $A\hat{x}=Q_1R\hat{x}=QQ_1^{\mathrm{T}}b$，它为 b 在 $\mathcal{R}(Q_1)$空间的正交投影，而投影关系为

$$b=Q_1Q_1^{\mathrm{T}}b+Q_2Q_2^{\mathrm{T}}b \tag{2.61}$$

(四)求解最小二乘问题的算法比较

除了 QR 算法之外，线性最小二乘问题还可以用 Cholesky 方法和高斯消去法求解。根据使预测误差向量的模取极小的准则，因为

$$\| b-Ax\|_2^2=(b-Ax)^{\mathrm{T}}(b-Ax)=b^{\mathrm{T}}b-2x^{\mathrm{T}}A^{\mathrm{T}}b-x^{\mathrm{T}}A^{\mathrm{T}}Ax,$$

令 $\partial\| b-Ax\|_2^2/\partial x^{\mathrm{T}}=0$，可得

$$(A^{\mathrm{T}}A)\hat{x}=A^{\mathrm{T}}b \tag{2.62}$$

这就是所谓的法方程，如果 $A^{\mathrm{T}}A$ 不是奇异矩阵，那么用高斯半主元素消去法及 Cholesky 分解都可用来求解(2.62)。

(1)Cholesky 分解法

由于$(A^{\mathrm{T}}A)$是对称正定矩阵，Cholesky 分解用于解法方程(2.62)的速度是最快的，其计算量可作如下估计

计算 $B=(A^{\mathrm{T}}A)$	$mn^2/2$ 拍
Cholesky 分解 $B=LL^{\mathrm{T}}$	$n^3/6$ 拍
回代 $L\hat{L}\hat{x}=A^{\mathrm{T}}b$	n^2 拍

当 $m=n$ 时总计算量$\left(\frac{2}{3}n^3+n^2\right)$拍

(2)高斯消去法

计算　$B=A^{\mathrm{T}}A$　　$mn^2/2$ 拍

LU 分解　$B=LU$　　$n^3/3$ 拍

回代　$LU\hat{x}=A^{\mathrm{T}}b$　　n^2 拍

$m=n$ 时的总计算量$\left(\frac{5}{6}n^3+n^2\right)$拍

(3)QR 分解法

由于 $A=QR$，则法方程 $A^{\mathrm{T}}Ax=A^{\mathrm{T}}b$ 化为

$$R^{\mathrm{T}}Q^{\mathrm{T}}QRx=R^{\mathrm{T}}Q^{\mathrm{T}}b$$

即

$$Rx=Q^{\mathrm{T}}b$$

其中 R 为上三角阵。解法分两步

QR 分解　$A=QR$　　$2n^2\left(m-\frac{n}{3}\right)$拍

回代　$R\hat{x}=Q^{\mathrm{T}}b$　　n^2 拍

$m=n$ 时总计算量　　$\left(\frac{4}{3}n^3+n^2\right)$拍

比较以上三种方法可见，QR 分解是线性最小二乘问题求解算法中成本最高的，它的计算量约为 Cholesky 分解的一倍。然而，由于 QR 算法最为可靠，它已成为对付最小二乘问题的最受欢迎的方法，这一点在下面还要证明。在这里我们先来看一下直接解法方程的前两种算法为什么不可靠。Cholesky 方法和高斯消去法都要计算矩阵乘法 $A^{\mathrm{T}}A$，对于非良态的矩阵 A，做这样的乘积有可能出问题。例如，下列 3×2 矩阵 A 的秩显然是 2：

$$A=\begin{pmatrix}1 & 1\\ 10^{-4} & 0\\ 0 & 10^{-4}\end{pmatrix}$$

但是，在一个十进制 7 位的计算机中，计算 $A^{\mathrm{T}}A$ 后矩阵的秩降为 1：

$$A^{\mathrm{T}}A=\begin{pmatrix}1+10^{-8} & 1\\ 1 & 1+10^{-8}\end{pmatrix}\approx\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}$$

即 $\mathrm{rank}(A)=2$，$\mathrm{rank}(A^{\mathrm{T}}A)=1$。换句话说，法方程 $(A^{\mathrm{T}}A)x=A^{\mathrm{T}}b$ 有无穷多个解，这是由于我们在作矩阵相乘时把 A 中最小特征值的信息丢掉了。在一般情况下，在机器中作矩阵相乘都可能丢失有用信息，造成求解的困难。由于 QR 分解算法的“安全性”，现在线性最小二乘问题多用它求解。即使是对 $n\times n$ 线性方程组 $Ax=b$，许多人都喜欢用 QR 分解而不用高斯消去法求解，因为此时虽然增加了一些计算工作量，但是，却可以取得半主元素高斯消去法双精度计算时所具有的结果。

(五)扰动分析和广义逆

现在来进行线性最小二乘问题的扰动分析。设 δA 为矩阵 A 的微扰矩阵，且 A 与 $A+\delta A$ 是满秩的。再设 δb 为观测向量 b 的微扰向量，相应的解 x 的误差向量为 δx。由于 A 和 $A+\delta A$ 都为满秩矩阵，最小二乘解估计 $\hat{x}$ 和 $x+\delta x$ 都是唯一存在的。代入法方程(2.62)有

$$(A+\delta A)^{\mathrm{T}}(A+\delta A)(x+\delta x)=(A+\delta A)^{\mathrm{T}}(b+\delta b)$$

因此

$$\delta x=(A^{\mathrm{T}}A)^{-1}\{\delta A^{\mathrm{T}}(e+\delta e)+A^{\mathrm{T}}[\delta b-\delta A(x+\delta x)]\}\tag{2.63}$$

其中 e 为最小二乘法的预测误差向量

$$e=b-Ax \tag{2.64}$$

而
$$e+\delta e=(b-\delta b)-(A+\delta A)(x+\delta x) \tag{2.65}$$

由矩阵范数的性质和(2.63)式有

$$\|\delta x\|_2 \leqslant \|(A^TA)^{-1}\|_2 \|\delta A^T\|_2 \|e+\delta e\|_2$$
$$+\|(A^TA)^{-1}A^T\|_2\{\|\delta b\|_2+\|\delta A\|_2\|x+\delta x\|_2\} \tag{2.66}$$

当用 QR 分解而不用直接解法方程(2.62)时，由于

$$(A^TA)^{-1}=(R^TQ^TQR^T)^{-1}=(R^TR)^{-1} \tag{2.67}$$

$$\|(A^TA)^{-1}\|_2=\|(R^TR)^{-1}\|_2=\|R^{-1}\|_2^2 \tag{2.68}$$

将(2.67)和(2.68)代入(2.66)有

$$\|\delta x\|_2 \leqslant \|R^{-1}\|_2^2\|\delta A\|_2\|e+\delta e\|_2+\|R^{-1}\|_2\{\|\delta b\|_2+\|\delta A\|_2\|x+\delta x\|_2\} \tag{2.69}$$

上式右边的第一项表示由于方程组个数多于未知数时由残差 e 引起的扰动，而第二项为求解方程组 $Ax=b$ 本身带来的误差。如果残差 $\|e\|_2$ 很小，(2.69)退化为

$$\|\delta x\|_2 \leqslant \|R^{-1}\|_2(\|\delta b\|_2+\|\delta A\|_2\|x+\delta x\|_2) \tag{2.70}$$

当 $\|\delta A\|_2=0$ 时

$$\|\delta x\|_2 \leqslant \|R^{-1}\|_2\|\delta b\|_2$$

而当 $\|\delta b\|_2=0$ 时

$$\|\delta x\|_2/\|x+\delta x\|_2 \leqslant \|R^{-1}\|_2\|\delta A\|_2$$

可见，在 e 很小情况下用 QR 分解算法解线性最小二乘问题时结果是精确的，因为误差的模仅与 $\|R^{-1}\|$ 成比例。

然而，如果不用 QR 算法而直接解法方程(2.62)，则做乘法 A^TA 的运算就会产生较大的误差。设 H 为矩阵相乘时产生的误差矩阵，即做 A^TA 计算的结果为 A^TA+H，这时相应传递给最小二乘解估计 $\hat{x}$ 的误差为 δx，由于

$$(A^TA+H)(x+\delta x)=A^Tb \tag{2.71}$$

即
$$A^TA\delta x=-H(x+\delta x)$$

由(2.68)有

$$\|\delta x\|_2 \leqslant \|R^{-1}\|_2^2\|H\|_2\|x+\delta x\|_2 \tag{2.72}$$

说明误差与 $\|R^{-1}\|_2$ 的平方成正比。误差矩阵 H 通常可表示为

$$\|H\|_2=o(\varepsilon)\|A\|_2^2$$

其中 ε 为机器的精度，o 表示趋近于无穷小。(2.72)式说明，即使最小二乘问题的残差 e 很小，由矩阵相乘引起的误差也可能是很大的，因为它与 $\|R^{-1}\|$ 的平方成正比，并与 e 无关。这就是为什么 QR 分解法广泛用于最小二乘问题求解的原因。

从上面分析可以看到，范数 $\|R^{-1}\|$ 决定了最小二乘问题解估计的误差，那么它必然与条件数有关。为了导出非方阵的条件数的概念，即 $n\times m$ 矩阵 $A(n\neq m)$ 的条件数 cond(A) 的表达式，必须定义所谓的广义逆，因为非方阵的逆矩阵没有逆。由法方程(2.62)把最小二乘解写为

$$\hat{x}=(A^TA)^{-1}A^Tb=R^{-1}Q_1^Tb$$

我们可以定义 $n\times m$ 矩阵 A 的广义逆为

$$A^+=(A^TA)^{-1}A^T=R^{-1}Q_1^T \tag{2.73}$$

因此最小二乘问题的解估计可表为

$$\hat{x}=A^+b \tag{2.74}$$

当 A 为非奇异的方阵时，$A^+=A^{-1}$，变为在通常意义下的逆矩阵。

现在来考虑与 $n\times m$ 系数矩阵 A 对应的线性最小二乘问题的条件数 cond(A)。由于

$$\|A^+\|=\|R^{-1}Q_1^T\|_2$$

$$\|A^+\|_2^2=\|R^{-1}Q^TQR^{-1}\|=\|R^{-1}\|_2^2$$

则由(2.31)可定义

$$\mathrm{cond}(A)=\|A\|_2\|A^+\|_2=\|R\|_2\|R^{-1}\|_2=\mathrm{cond}(R) \tag{2.75}$$

由此可见，线性最小二乘问题的条件数与 $\|R^{-1}\|_2$ 成正比。

与广义逆有关的一个概念是伪逆(Pseudoinverse)，它也叫做 Moore-Penrose 广义逆矩阵，记为 G。给定一个 $n\times m$ 矩阵 A，可以唯一地定义一个矩阵 G，使它具有以下四点性质：

$$\left.\begin{aligned}&(\mathrm{i})AGA=A\\&(\mathrm{ii})GAG=G\\&(\mathrm{iii})AG=(AG)^T,\text{即 }AG\text{ 为对称矩阵}\\&(\mathrm{iv})GA=(GA)^H,\text{即 }GA\text{ 为厄米特矩阵}\end{aligned}\right\} \tag{2.76}$$

由(2.76)的定义可知，对于非奇异的方阵，$G=A^{-1}$。对于秩为 m 的 $n\times m$ 实矩阵 A，$G=A^+=R^{-1}Q_1^T$。把 A 做 QR 分解之后，有

$$AG=Q_1RR^{-1}Q_1^T=Q_1Q_1^T$$

可见 AG 即为(2.61)中讲到的 b 在 $\mathscr{R}(Q_1)$ 空间的投影矩阵，而 $GA=Q_2Q_2^T$ 为 b 在 $\mathscr{R}(Q_2)$ 空间的投影矩阵。从下节的讨论可知，伪逆可以通过奇异值分解计算。

四、奇异值分解(SVD)[1]

上面已经讲到，如果线性方程组的系数矩阵 A 为对称正定的非稀疏矩阵，则用 Cholesky 分解求解是最快和最精确的算法；如果 A 为非奇异矩阵而且不是方阵，用 QR 分解目前是最安全的。那么，对于实际的科学计算问题，如何定量地估计系数矩阵 A 是否接近奇异？换句话说，怎么计算条件数 cond(A)？如果 A 接近奇异，能否而且如何求线性方程组或法方程的解？本章讨论的奇异值分解的方法，就是目前可以用来对付上述问题的主要工具。当然，我们希望在地球物理问题中 A 最好不接近奇异，这样以前讲过的方法就够用了。不幸的是，在地球物理反演问题中涉及的系数矩阵 A 多半是接近奇异的，因为这时我们要用观测的信号去推测场源或传播介质的参数，而从信号向源推移的过程具有解析上的不稳定性。作为谱展开的一种形式，奇异值分解是这时最重要的工具。

奇异值分解的萌芽思想至少可追溯到上个世纪，如 Sylvester 于 1899 年(Messeger of Math，1899，Vol. 42)发表的文章。但作为数值计算的工具被人们所重视还是 1965 年以后的事。Golub 和 Kahan(1965)首次指出了直接由法方程系数矩阵计算奇异值存在的问题[28]，1970 年 Golub 和 Reinsch 提出了下面要讨论的奇异值分解算法[29]。算法的舍入误差分析和

[1] 英文奇异值分解 Singular Value Decomposition 的词头。

扰动分析的研究还在进行。

(一)奇异值分解定理

定理:对于任意 $n\times m$ 复矩阵 $\boldsymbol{A}$,存在着两个酉阵,即 $n\times n$ 酉阵 $\boldsymbol{U}$ 和 $m\times m$ 酉阵 $\boldsymbol{V}$,使得

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}=\boldsymbol{S}=\mathrm{Diag}(\alpha_1,\alpha_2,\cdots,\alpha_m,0\cdots0) \tag{2.77}$$

其中 Diag 是对角矩阵的表示,它的对角线元即为非负实数 $\alpha_i(i=1,\cdots,m)$,叫做矩阵 $\boldsymbol{A}$ 的奇异值,并满足

$$\alpha_1\geqslant\alpha_2\geqslant\cdots\geqslant\alpha_m\geqslant0 \tag{2.78}$$

我们下面仅就 $\boldsymbol{A}$ 为实矩阵的情况证明这个定理。首先,证明对于任意实矩阵 $\boldsymbol{A}$,存在着两个范数为单位的向量 $\boldsymbol{u}$ 和 $\boldsymbol{v}$,$\|\boldsymbol{u}\|_2=\|\boldsymbol{v}\|_2=1$,使得矩阵 $\boldsymbol{A}$ 对它们的变换满足以下关系

$$\boldsymbol{A}\boldsymbol{v}=\alpha\boldsymbol{u}\qquad\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}=\alpha\boldsymbol{v} \tag{2.79}$$

其中

$$\alpha=\|\boldsymbol{A}\|_2 \tag{2.80}$$

第一步,根据从属范数的定义,有

$$\|\boldsymbol{A}\|_2=\max_{\|\boldsymbol{x}\|_2=1}\|\boldsymbol{A}\boldsymbol{x}\|_2$$

记上式中取极大值的向量 $\boldsymbol{x}$ 为 $\boldsymbol{v}(m\times1)$,此时

$$\left.\begin{aligned}\|\boldsymbol{x}\|_2=\|\boldsymbol{v}\|_2=1,\\ \|\boldsymbol{A}\boldsymbol{v}\|_2=\|\boldsymbol{A}\|_2=\alpha.\end{aligned}\right\} \tag{2.81}$$

第二步,令 $\boldsymbol{y}$ 为任意 $n\times1$ 向量。由于

$$\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}\leqslant|\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}|\leqslant\|\boldsymbol{y}\|_2\|\boldsymbol{A}\|_2\|\boldsymbol{x}\|_2 \tag{2.82}$$

当 $\|\boldsymbol{y}\|_2=\|\boldsymbol{x}\|_2=1$ 和 $\boldsymbol{A}$ 不为零矩阵(即 $\alpha=\|\boldsymbol{A}\|_2\neq0$)时,上式可表为

$$\|\boldsymbol{A}\|_2=\max_{\|\boldsymbol{x}\|_2=\|\boldsymbol{y}\|_2=1}(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}) \tag{2.83}$$

记上式取极大值时的向量 $\boldsymbol{y}$ 为 $\boldsymbol{u}$,则应有 $\|\boldsymbol{y}\|_2=\|\boldsymbol{u}\|_2=1$。可以看出,选取 $\boldsymbol{y}=\boldsymbol{u}=\boldsymbol{A}\boldsymbol{v}/\alpha=\boldsymbol{A}\boldsymbol{v}/\|\boldsymbol{A}\boldsymbol{v}\|_2$ 便可满足(2.83)式,因为此时由(2.81)有标量

$$\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=\frac{(\boldsymbol{A}\boldsymbol{v})^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}}{\|\boldsymbol{A}\boldsymbol{v}\|_2}=\|\boldsymbol{A}\boldsymbol{v}\|_2=\|\boldsymbol{A}\|_2,$$

代入(2.81)有

$$\boldsymbol{u}=\boldsymbol{A}\boldsymbol{v}/\alpha \tag{2.84}$$

或

$$\boldsymbol{A}\boldsymbol{v}=\alpha\boldsymbol{u}$$

这样,我们便得到了(2.79)的第一式。与此对偶的是(2.79)的第二式,由(2.84)式作转置

$$\alpha\boldsymbol{u}^{\mathrm{T}}=\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}$$

因为 $\|\boldsymbol{u}\|_2=1$,所以

$$\alpha\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}=\alpha\|\boldsymbol{u}\|_2^2=\alpha=\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u} \tag{2.85}$$

但是由范数的 Cauchy-Schwarz 定理(2.7)可知,标量

$$\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}\leqslant\|\boldsymbol{v}\|_2\cdot\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}\|_2\leqslant\alpha=\|\boldsymbol{A}\|_2 \tag{2.86}$$

这时因为 $\|\boldsymbol{v}\|_2=\|\boldsymbol{u}\|_2=1$。要使 $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}$ 同时满足不等式(2.86)和等式(2.85)只有在 $\boldsymbol{v}$ 和 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}$ 线性相关才有可能(见本章第一节(2.7)的证明),因此由线性相关的定义有

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}=\gamma\boldsymbol{v},$$

其中，常数 γ 可由(2.85)式求得，即

$$\gamma=\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}-\alpha,$$

所以有

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}=\alpha\boldsymbol{v},$$

这就是(2.79)中的第二式。

现在我们已经证明了对于任意实数矩阵 $\boldsymbol{A}$，存在着向量 $\boldsymbol{u}$ 和 $\boldsymbol{v}$，满足

$$\|\boldsymbol{u}\|_2=\|\boldsymbol{v}\|_2=1,\boldsymbol{A}\boldsymbol{v}=\alpha\boldsymbol{u},\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}=\alpha\boldsymbol{v}$$

其中 $\alpha=\|\boldsymbol{A}\|_2$，下面，就可以来证明奇异值分解定理的成立。当 $\boldsymbol{A}$ 为零矩阵时，取 $\boldsymbol{S}$ 也为零矩阵，定理(2.77)即成立。当 $\boldsymbol{A}\neq\boldsymbol{0}$ 时，设 $\boldsymbol{u}_1$ 和 $\boldsymbol{v}_1$ 为满足(2.79)式的两个向量，而且 $\|\boldsymbol{u}_1\|_2=\|\boldsymbol{v}_1\|_2=1$，$\boldsymbol{u}_1$ 为 $n\times1$，而 $\boldsymbol{v}_1$ 为 $m\times1$，则有

$$\boldsymbol{A}\boldsymbol{v}_1=\alpha_1\boldsymbol{u}_1,\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}_1=\alpha_1\boldsymbol{v}_1 \tag{2.87}$$

现在以 $\boldsymbol{u}_1$ 和 $\boldsymbol{v}_1$ 构成两个正交矩阵，并以分块矩阵表示

$$\boldsymbol{V}_1=(\boldsymbol{v}_1,\boldsymbol{F})\qquad\boldsymbol{U}_1=(\boldsymbol{u}_1,\boldsymbol{G}) \tag{2.88}$$

其中 $\boldsymbol{F}$ 为 $m\times(m-1)$分块矩阵，$\boldsymbol{G}$ 为 $n\times(n-1)$分块矩阵，于是有

$$\boldsymbol{A}\boldsymbol{V}_1=(\boldsymbol{A}\boldsymbol{v}_1,\boldsymbol{A}\boldsymbol{F})=(\alpha_1\boldsymbol{u}_1,\boldsymbol{A}\boldsymbol{F})$$

$$\begin{aligned}\boldsymbol{U}_1^{\mathrm{T}}\boldsymbol{A}\boldsymbol{V}_1&=\begin{pmatrix}\boldsymbol{u}_1^{\mathrm{T}}\\\boldsymbol{G}^{\mathrm{T}}\end{pmatrix}(\alpha_1\boldsymbol{u}_1,\boldsymbol{A}\boldsymbol{F})\\&=\begin{pmatrix}\alpha_1&\boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{A}\boldsymbol{F}\\\alpha_1\boldsymbol{G}^{\mathrm{T}}\boldsymbol{u}_1&\boldsymbol{G}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{F}\end{pmatrix}\\&=\begin{pmatrix}\alpha_1&0\\0&\boldsymbol{A}_2\end{pmatrix}\end{aligned} \tag{2.89}$$

注意其中

$$\boldsymbol{A}_2=\boldsymbol{G}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{F}$$

代入(2.87)有

$$\boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{A}\boldsymbol{F}=(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}_1)^{\mathrm{T}}\boldsymbol{F}=(\alpha_1\boldsymbol{v}_1)^{\mathrm{T}}\boldsymbol{F}=\boldsymbol{0}$$

其中根据(2.88)式，$\boldsymbol{V}_1$ 为正交矩阵，它的列向量为 m 维欧几里德空间的正交基。也就是说，$\boldsymbol{V}_1$ 和子空间 $\mathscr{R}(\boldsymbol{F})$是垂直的，投影为零。对于(2.88)式中的 $\boldsymbol{u}_1$ 和 $\boldsymbol{G}$，同样有

$$\boldsymbol{G}^{\mathrm{T}}\boldsymbol{u}_1=\boldsymbol{0}$$

现在我们已用左乘 $\boldsymbol{U}_1^{\mathrm{T}}$ 和右乘 $\boldsymbol{V}_1$ 的方法证明可把 $\boldsymbol{A}$ 的第一行和第一列除对角线元素以外都化为零，仅保留的对角线元素 α_1 为非负标量，因为它是(2.80)式中表示的范数。用同样的办法还可以进一步把矩阵 $\boldsymbol{A}_2=\boldsymbol{G}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{F}$ 的首行和首列的非对角线元素化为零。用 $\boldsymbol{U}_2^{\mathrm{T}}$ 左乘 $\boldsymbol{A}$ 和用 $\boldsymbol{V}_2$ 右乘 $\boldsymbol{A}$，即

$$\boldsymbol{U}_2^{\mathrm{T}}\boldsymbol{A}\boldsymbol{V}_2=\begin{pmatrix}1&0\\0&\boldsymbol{U'}_2^{\mathrm{T}}\end{pmatrix}\begin{pmatrix}\alpha_1&0\\0&\boldsymbol{A}_2\end{pmatrix}\begin{pmatrix}1&0\\0&\boldsymbol{V'}_2\end{pmatrix}=\begin{pmatrix}\alpha_1&0&0\\0&\alpha_2&0\\0&0&\boldsymbol{A}_3\end{pmatrix} \tag{2.90}$$

其中 $\boldsymbol{U'}_2$ 和 $\boldsymbol{V'}_2$ 为正交矩阵。最后，用数学归纳法可以证明，当这个左乘和右乘的过程进行到第 m 次时，则有

$$U_m^{\mathrm{T}}\cdots U_2^{\mathrm{T}}U_1^{\mathrm{T}}AV_1V_2\cdots V_m=S=\begin{bmatrix}\alpha_1 & & & \\ & \alpha_2 & & \\ & & \ddots & \\ & & & \alpha_m \\ \cdots & \cdots & \cdots & \cdots \\ & & \mathbf{0} & \end{bmatrix}$$

令

$$U^{\mathrm{T}}=U_m^{\mathrm{T}}\cdots U_2^{\mathrm{T}}U_1^{\mathrm{T}}$$

$$V=V_1V_2\cdots V_m,$$

则得到定理(2.77)。

(二)SVD 的性质

矩阵的奇异值分解形式具有许多有用的性质,正是这些性质使 SVD 变成线性计算中最强有力的工具。

性质一:矩阵 A 在奇异值分解之后取得的奇异值的平方即为矩阵 $A^{\mathrm{T}}A$ 的特征值。由(2.77)有

$$A^{\mathrm{T}}A=VS^{\mathrm{T}}SV^{\mathrm{T}} \tag{2.91}$$

因此

$$A^{\mathrm{T}}AV=V\begin{bmatrix}\alpha_1^2 & & & \\ & \alpha_2^2 & & \\ & & \ddots & \\ & & & \alpha_m^2\end{bmatrix}$$

记 v_j 为矩阵 V 的列向量,有

$$A^{\mathrm{T}}Av_j=\alpha_j^2v_j \qquad j=1,\cdots m$$

因此 α_j^2 为 $A^{\mathrm{T}}A$ 的特征值,V 的列向量为 $A^{\mathrm{T}}A$ 的特征向量。

类似地,因为

$$AA^{\mathrm{T}}=USV^{\mathrm{T}}VS^{\mathrm{T}}U^{\mathrm{T}}=USS^{\mathrm{T}}U^{\mathrm{T}} \tag{2.92}$$

$$AA^{\mathrm{T}}U=U\begin{bmatrix}\alpha_1^2 & & & & \\ & \ddots & & & \\ & & \alpha_m^2 & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0\end{bmatrix}$$

因此 α_j^2 也是矩阵 AA^{T} 的特征值,同时 U 的列向量是 AA^{T} 的特征向量。

显然,当 A 为 $n\times n$ 的对称正定矩阵时,A 的奇异值分解退化为特征值分解,而且因为此时

$$A^{\mathrm{T}}A=AA^{\mathrm{T}}$$

由(2.91)和(2.92)有

$$VS^{\mathrm{T}}SV^{\mathrm{T}}=USS^{\mathrm{T}}U^{\mathrm{T}}$$

即

$$\boldsymbol{U}=\boldsymbol{V}$$

性质二：正交变换不改变矩阵的奇异值。设 $\boldsymbol{P}$ 为 $n\times n$ 正交矩阵，$\boldsymbol{Q}$ 为 $m\times m$ 正交矩阵，则

$$\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Q}=\boldsymbol{P}^{\mathrm{T}}\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{Q}=\boldsymbol{U}'\boldsymbol{S}\boldsymbol{V}'^{\mathrm{T}}$$

由于正交阵相乘仍为正交阵，$\boldsymbol{A}$ 的奇异值被保存下来。

性质三：矩阵 $\boldsymbol{A}$ 的奇异值与它的范数的极值有关。由于正交变换不改变矩阵的模，利用从属范数的定义有

$$\|\boldsymbol{A}\|_2=\|\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}\|_2=\|\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\|_2=\|\boldsymbol{S}\|_2=\max_{x\neq 0}\frac{\|\boldsymbol{S}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2}=\alpha_1$$

$$\|\boldsymbol{A}\|_F=\|\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}\|_F=\|\boldsymbol{S}\|_F=(\alpha_1^2+\cdots+\alpha_m^2)^{1/2}$$

由于

$$\|\boldsymbol{A}\boldsymbol{x}\|_2^2=\|\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}\|_2^2=\|\boldsymbol{S}\boldsymbol{y}\|_2^2$$

其中

$$\|\boldsymbol{y}\|_2=\|\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}\|_2=\|\boldsymbol{x}\|_2,\boldsymbol{y}^{\mathrm{T}}=(\eta_1,\cdots,\eta_m)$$

由从属范数性质有

$$\|\boldsymbol{A}\boldsymbol{x}\|_2\leqslant\|\boldsymbol{A}\|_2\|\boldsymbol{x}\|_2=\alpha_1\|\boldsymbol{x}\|_2 \tag{2.93}$$

同时用展开式表示，即

$$\|\boldsymbol{A}\boldsymbol{x}\|_2^2=\sum_{i=1}^{m}\alpha_i^2\eta_i^2\geqslant\alpha_m^2\sum_{i=1}^{m}\eta_i^2=\alpha_m^2\|\boldsymbol{x}\|_2^2 \tag{2.94}$$

其中 α_m 为最小的奇异值。把(2.93)和(2.94)式两边除以非负实数 $\|\boldsymbol{x}\|_2$，得不等式

$$\alpha_1\geqslant\frac{\|\boldsymbol{A}\boldsymbol{x}\|}{\|\boldsymbol{x}\|_2}\geqslant\alpha_m \tag{2.95}$$

于是，奇异值的变化范围即由从属范数和 $\boldsymbol{x}$ 的范数的比值决定：

$$\left.\begin{aligned}\alpha_1&=\max_{x\neq 0}\frac{\|\boldsymbol{A}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2}=\|\boldsymbol{A}\|_2\\ \alpha_m&=\min_{x\neq 0}\frac{\|\boldsymbol{A}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2}\end{aligned}\right\} \tag{2.96}$$

性质四：在解线性方程组 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ 时方阵 $\boldsymbol{A}$ 的条件数等于最大奇异值 α_1 和最小奇异值 α_m 之比值。当 $\boldsymbol{A}$ 为 $m\times m$ 方阵时将方程组写为

$$\begin{aligned}&\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}=\boldsymbol{b}\\ &\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}=\boldsymbol{U}^{\mathrm{T}}\boldsymbol{b}\\ &\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}=\boldsymbol{S}^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{b}\\ &\boldsymbol{x}=\boldsymbol{V}\boldsymbol{S}^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{b}\end{aligned} \tag{2.97}$$

则有

$$\boldsymbol{A}^{-1}=\boldsymbol{V}\boldsymbol{S}^{-1}\boldsymbol{U}^{\mathrm{T}}$$

根据条件数的定义和(2.96)式有

$$\mathrm{cond}(\boldsymbol{A})=\|\boldsymbol{A}\|_2\|\boldsymbol{A}^{-1}\|_2=\|\boldsymbol{S}\|_2\|\boldsymbol{S}^{-1}\|_2=\alpha_1\cdot\alpha_m^{-1}=\alpha_1/\alpha_m \tag{2.98}$$

性质五：矩阵 $\boldsymbol{A}$ 接近奇异的标志是最小的奇异值接近于零(条件数趋向无穷大)，此时与它最接近的矩阵为去掉 $\boldsymbol{A}$ 的最小奇异值 α_m 的矩阵。

当 $n\times m$ 矩阵 A 奇异时，它的秩 $\mathrm{rank}(A)=p<m$，因此只有 p 个非零的奇异值

$$S=\mathrm{diag}(\alpha_1,\alpha_2,\cdots,\alpha_p,0,\cdots,0) \tag{2.99}$$

对于一个 10 进制 9 位的计算机来说，如果 $p=m$ 而 $\alpha_m\leqslant 10^{-10}$，则矩阵 A 是在计算意义上奇异的。因此，SVD 可以用来确定矩阵有效的秩。

如果考虑计算过程中的舍入误差的传播和系数矩阵本身的微扰，那么在上述情况下即使 $\alpha_m>10^{-10}$，这个最小奇异值也不一定有意义。在一般情况下，设 SVD 的运算精度为 ε，则小于 ε 的奇异值 $\alpha_i<\varepsilon$ 都是不确定的，这种不确定性造成了计算的不稳定性，即数据的微小变化导致方程组解的相当大变化。从(2.97)式可以看出，线性方程组的解是与矩阵 S 的逆(奇异值的倒数矩阵)成正比的，奇异值越小数据中的误差就被放得越大，这就是反演常见的接近奇异的系数矩阵的情况。在这种情况下，为了找到稳定的(方差小的)解估计我们可以用 A 的最接近矩阵代替 A 求解。

矩阵 A 的最接近矩阵 B 定义为使下列范数取最小值的矩阵：

$$\min\|A-B\|_2 \qquad \text{或 } \min\|A-B\|_F \tag{2.100}$$

如果 $\mathrm{rank}(A)=p$，$p<m$，则容易看出，B 可表为

$$B=U\begin{pmatrix}\alpha_1 & & & & & \\ & \alpha_2 & & & & \\ & & \ddots & & & \\ & & & \alpha_q & & \\ & & & & 0 & \\ & & & & & \ddots \\ & & & & & & 0\end{pmatrix}V^{\mathrm{T}}=US_1V^{\mathrm{T}} \tag{2.101}$$

其中 $q<p$，实际上

$$A-B=U(S-S_1)V^{\mathrm{T}}=\alpha_{q+1}u_{q+1}v_{q+1}^{\mathrm{T}}+\cdots+\alpha_p u_p v_p^{\mathrm{T}} \tag{2.102}$$

$$\|A-B\|_{2\cdot F}=\sum_{i=q+1}^{p}\alpha_i\|u_iv_i^{\mathrm{T}}\|_{2\cdot F}$$

由于 $(\alpha_p,\cdots,\alpha_{q+1})$ 是最小的几个奇异值，B 和 A 最接近。

于是，用最接近矩阵 B 代替 A 去求方程组(2.97)的解估计便构成了对付不能稳定地求解的线性问题的一种战略，这样求出的解既不可能是精确解，也不是“最优”的解，而只能是一种“广义”的近似解。

性质六：矩阵 A 的行列式 $\det(A)$ 等于所有奇异值的乘积。由于

$$\det(A)=\det(U)\cdot\det(S)\cdot\det(V^{\mathrm{T}})$$

而正交矩阵 U 和 V 的行列式等于 ±1，则有

$$\det(A)=\pm\det(S)=\pm\alpha_1\alpha_2\cdots\alpha_n$$

由此易见当 $\alpha_n\to0$ 时，有 $\det(A)\to0$，标志矩阵 A 接近奇异。

(三)奇异值分解算法

通过本节第一段中 SVD 定理的证明可知，SVD 可以通过一系列的正交变换来实现，下面来讨论一种分两步走的 SVD 算法，它首先把矩阵 A 化为双对角阵，这种算法是 Golub 和 Reinsch 在 1970 年提出的[25]。

计算的第一步是用第三节讲过的 Householder 变换把矩阵 A 逐步化为双对角阵。设 U_1

为一个 Householder 变换，将 U_1 乘以 A 便把 A 的第一列化为除对角之外全为零的向量。以一个 6×5 的矩阵 A 为例，这个过程记为

$$U_1A=\begin{bmatrix}\times & \times & \otimes & \otimes & \otimes\\ 0 & \times & \times & \times & \times\\ 0 & \times & \times & \times & \times\\ 0 & \times & \times & \times & \times\\ 0 & \times & \times & \times & \times\\ 0 & \times & \times & \times & \times\end{bmatrix}$$

再设 V_1 也是一个 Householder 变换，用它右乘矩阵 U_1A，则可把上面用圆圈表示的第一行中前两个以外的元素化为零，得

$$U_1AV_1=\begin{bmatrix}\times & \times & 0 & 0 & 0\\ 0 & \times & \times & \otimes & \otimes\\ 0 & \otimes & \times & \times & \times\\ 0 & \otimes & \times & \times & \times\\ 0 & \otimes & \times & \times & \times\\ 0 & \otimes & \times & \times & \times\end{bmatrix}$$

下一步，类似于上述方法选择 Householder 变换 U_2 和 V_2，把上面用圆圈标出的元素化为零。当 A 被左乘五次和右乘三次时，它即被化为双对角阵：

$$U_5\cdots U_1AV_1\cdots V_3=\begin{bmatrix}\times & \times & 0 & 0 & 0\\ 0 & \times & \times & 0 & 0\\ 0 & 0 & \times & \times & 0\\ 0 & 0 & 0 & \times & \times\\ 0 & 0 & 0 & 0 & \times\\ 0 & 0 & 0 & 0 & 0\end{bmatrix}$$

对于一般的 $n\times m$ 矩阵来说，用 Householder 变换可逐步把它化为以下双对角阵

$$B=U_k\cdots U_1AV_1\cdots V_l=\tilde{U}A\tilde{V} \tag{2.103}$$

其中

$$\left.\begin{aligned}k&=\min(m,n-1)\\ l&=\min(n,m-2)\end{aligned}\right\} \tag{2.104}$$

也许有人会问，为什么不用 Householder 变换把矩阵 A 直接化为对角阵呢？在一般情况下，这是不可能的。因为，这样做左乘和右乘会相互干扰，右乘 V_i 后会把原来左下角已经化为零的元素变为非零元素。

奇异值分解的第二步是把双对角线矩阵 B 化为对角阵 S，其中 S 为奇异值矩阵。去掉全

为零的行或列，矩阵 B 可表为

$$B=\begin{bmatrix}\beta_1 & \gamma_1 & & & & & \\ & \beta_2 & \gamma_2 & & & & \\ & & \ddots & \ddots & & & \\ & & & \ddots & \ddots & & \\ & & & & \ddots & \ddots & \\ & & & & & \beta_{m-1} & \gamma_{m-1} \\ & & & & & & \beta_m \end{bmatrix}$$

如上节所述，矩阵 B 的奇异值的平方是三对角线矩阵 $C=B^TB$ 的特征值，因此从原理上说可以用求 C 的特征值的算法，找出矩阵 B 的奇异值，它即是矩阵 A 的奇异值，因为正交变换不改变矩阵的奇异值。由于矩阵 $C=B^TB$ 是对称正定的三对角阵，不难进行特征值分解求出它的特征值及特征向量。由 C 的特征值的开方可以得到 A 的奇异值矩阵 S，由 C 的特征向量和(2.103)式中的 $\tilde{U}$ 及 $\tilde{V}$ 矩阵便不难计算 U 和 V^T。

SVD 的计算过程是一个冗长和繁复的过程，目前，虽然已提出数种不同的算法，还没有一种是既精确又快的。由于 SVD 的计算成本太高，它的实际应用常常受到限制。

在大型计算机系统中一般都配有包含 SVD 子程序的计算程序包。较好的有在 LINPACK(线性程序包 1979 年版)中的双精度奇异值分解子程序 DSVDC 或单精度子程序 SSVDC[36]，以及在 IMSL(国际数学统计程序库 1976 年版)中的 LSVALR 子程序。计算速度最快的是特征值问题程序包 EISPACK(1976 年版)中的 SVD 子程序，但此子程序输出的奇异值不是按照上面讲的从大到小排列的，因此要注意重排奇异值和矩阵 U 的列及 V 的行。

奇异值分解在数学上有许多应用，它也是地球物理广义反演中最重要的工具之一，在下一章我们还要进一步讨论利用 SVD 的广义反演方法。

从矩阵特征值问题的角度作进一步研究，可以把矩 A 的奇异值分解看成是矩阵(A^TA)的特征分解，即解特征值问题：

$$A^TAx=\alpha^2x, \qquad A\in\mathscr{R}^{m\times n}$$

$$m\geqslant n \qquad 0\neq x\in\mathscr{R}^n, \qquad \alpha\geqslant 0$$

把奇异值分解进一步推广，可以导致所谓的广义奇异值问题，即求解以下特征问题：

$$A^TAx=\mu^2B^TBx, A,B\in\mathscr{R}^{m\times n}$$

$$m\geqslant n, \quad 0\neq x\in\mathscr{R}^n, \qquad \mu\geqslant 0,$$

其中 μ 称为广义奇异值。相应的求广义奇异值的算法称为 VZ 算法。显然，奇异值算法只是在 B 为正交矩阵时的 VZ 算法的特殊情况。关于 VZ 算法可以参考“应用数学与计算数学”1979 年第 5 期 C. F. Van Coan 的文章。

五、奇异值分解 Fortran 子程序

以下程序引自 W. H. Press, B. P. Flannery, S. A. Teukolsky 和 W. T. Vetterling 编写的“Numerical Recipes”一书 2.9 节，由剑桥大学出版社 1993 年出版(第二版)。

输入矩阵 A，它的实际维数为 $MP\times NP$，逻辑维数为 $M\times N$。例如，它有 200 行及 150 列，则实际维数为 200×150，但是只用其中 100 行及 100 列做奇异值分解，则 $M=N=100$。

它分解为 $A=UWV^T$。

输出的矩阵 U 存回矩阵 A，对角线矩阵中的奇异值存在向量 W，而矩阵 V 是未经转置的。

当 $M\geqslant N$ 时为正常输出。如果 $M<N$，则输出的矩阵 U 中只含有 M 行 M 列非零元素。

第二条语句对 N 的最大值进行限制，可根据机器情况进行修改。

```
      SUBROUTINE svdcmp(a,m,n,mp,np,w,v)
      INTEGER m,mp,n,np,NMAX
      REAL a(mp,np),v(np,np),w(np)
      PARAMETER (NMAX=500)
CU    USES pythag
      INTEGER i,its,j,jj,k,l,nm
      REAL anorm,c,f,g,h,s,scale,x,y,z,rv1(NMAX),pythag
      g=0.0
      scale=0.0
      anorm=0.0
      do 25 i=1,n
       l=i+1
       rv1(i)=scale*g
       g=0.0
       s=0.0
       scale=0.0
       if(i.le.m)then
        do 11 k=i,m
         scale=scale+abs(a(k,i))
11      continue
        if(scale.ne.0.0)then
         do 12 k=i,m
          a(k,i)=a(k,i)/scale
          s=s+a(k,i)*a(k,i)
12       continue
         f=a(i,i)
         g=-sign(sqrt(s),f)
         h=f*g-s
         a(i,i)=f-g
         do 15 j=l,n
          s=0.0
          do 13 k=i,m
           s=s+a(k,i)*a(k,j)
13        continue
          f=s/h
          do 14 k=i,m
           a(k,j)=a(k,j)+f*a(k,i)
14        continue
15       continue
        do 16 k=i,m
         a(k,i)=scale*a(k,i)
```

```
16        continue
        endif
      endif
      w(i)=scale *g
      g=0.0
      s=0.0
      scale=0.0
      if((i.le.m).and.(i.ne.n))then
        do 17 k=l,n
          scale=scale+abs(a(i,k))
17        continue
        if(scale.ne.0.0)then
          do 18 k=l,n
            a(i,k)=a(i,k)/scale

            s=s+a(i,k)*a(i,k)
18        continue
          f=a(i,l)
          g=-sign(sqrt(s),f)
          h=f*g-s
          a(i,l)=f-g
          do 19 k=l,n
            rv1(k)=a(i,k)/h
19        continue
          do 23 j=l,m
            s=0.0
            do 21 k=l,n
              s=s+a(j,k)*a(i,k)
21          continue
            do 22 k=l,n
              a(j,k)=a(j,k)+s*rv1(k)
22          continue
23        continue
          do 24 k=l,n
            a(i,k)=scale*a(i,k)
24        continue
        endif
      endif
      anorm=max(anorm,(abs(w(i))+abs(rv1(i))))
25    continue
    do 32 i=n,l,-1
      if(i.lt.n)then
        if(g.ne.0.0)then
          do 26 j=l,n
            v(j,i)=(a(i,j)/a(i,l))/g
26        continue
          do 29 j=l,n
            s=0.0
            do 27 k=l,n
```

```
            s=s+a(i,k)*v(k,j)
27          continue
            do 28 k=l,n
              v(k,j)=v(k,j)+s*v(k,i)
28          continue
29        continue
        endif
        do 31 j=l,n
          v(i,j)=0.0
          v(j,i)=0.0
31      continue
      endif
      v(i,i)=1.0
      g=rv1(i)
      l=i
32    continue
    do 39 i=min(m,n),1,-1
      l=i+1
      g=w(i)
      do 33 j=l,n
        a(i,j)=0.0

33      continue
      if(g.ne.0.0)then
        g=1.0/g
        do 36 j=l,n
          s=0.0
          do 34 k=l,m
            s=s+a(k,i)*a(k,j)
34          continue
          f=(s/a(i,i))*g
          do 35 k=i,m
            a(k,j)=a(k,j)+f*a(k,i)
35          continue
36        continue
          do 37 j=i,m
            a(j,i)=a(j,i)*g
37        continue
      else
        do 38 j= i,m
          a(j,i)=0.0
38      continue
      endif
      a(i,i)=a(i,i)+1.0
39    continue
    do 49 k=n,1,-1
      do 48 its=1,30
        do 41 l=k,1,-1
          nm=l-1
          if((abs(rv1(l))+anorm).eq.anorm)  goto 2
          if((abs(w(nm))+anorm).eq.anorm)  goto 1
```

```
41      continue
1      c=0.0
       s=1.0
       do 43 i=l,k
        f=s*rv1(i)
        rv1(i)=c*rv1(i)
        if((abs(f)+anorm).eq.anorm) goto 2
        g=w(i)
        h=pythag(f,g)
        w(i)=h
        h=1.0/h
        c= (g*h)
        s=-(f*h)
        do 42 j=1,m
         y=a(j,nm)
         z=a(j,i)
         a(j,nm)=(y*c)+(z*s)
         a(j,i)=-(y*s)+(z*c)
42      continue
43     continue
2      z=w(k)
       if(l.eq.k)then
        if(z.lt.0.0)then
         w(k)=-z
         do 44 j=1,n
          v(j,k)=-v(j,k)
--
44         continue
        endif
        goto 3
       endif
       if(its.eq.30) pause 'no convergence in svdcmp'
       x=w(l)
       nm=k-1
       y=w(nm)
       g=rv1(nm)
       h=rv1(k)
       f=((y-z)*(y+z)+(g-h)*(g+h))/(2.0*h*y)
       g=pythag(f,1.0)
       f=((x-z)*(x+z)+h*((y/(f+sign(g,f)))-h))/x
       c=1.0
       s=1.0
       do 47 j=l,nm
        i=j+1
        g=rv1(i)
        y=w(i)
        h=s*g
        g=c*g
        z=pythag(f,h)
        rv1(j)=z
```

```
        c=f/z
        s=h/z
        f= (x*c)+(g*s)
        g=-(x*s)+(g*c)
        h=y*s
        y=y*c
        do 45 jj=1,n
          x=v(jj,j)
          z=v(jj,i)
          v(jj,j)= (x*c)+(z*s)
          v(jj,i)=-(x*s)+(z*c)
45        continue
        z=pythag(f,h)
        w(j)=z
        if(z.ne.0.0)then
          z=1.0/z
          c=f*z
          s=h*z
        endif
        f= (c*g)+(s*y)
        x=-(s*g)+(c*y)
        do 46 jj=1,m
          y=a(jj,j)
          z=a(jj,i)
          a(jj,j)= (y*c)+(z*s)
          a(jj,i)=-(y*s)+(z*c)
46        continue
47      continue
        rv1(l)=0.0
        rv1(k)=f
        w(k)=x
48      continue
3     continue
49    continue
      return
      END
C  (C) Copr. 1986-92 Numerical Recipes Software .
```

第三章　广义反演方法

在本章中我们来讨论目前地球物理反演中应用最广的一大类反演方法，即广义反演。假定地球物理反演问题是以泛函方程组的形式给出的，离散化后我们便得到某种类型的非线性方程组。如果这种方程的唯一解存在，则利用第二章的方法便可以获得反演问题的解答。然而，由于地球物理数据的有限性和不精确性，涉及的方程组常常不存在精确的唯一解答，或者不能够用稳定收敛的数值方法求得其精确解，这时我们只能根据某种准则求出反演问题的一个可以接受的解估计。我们把这种情况下求反演问题的解估计的方法统称为广义反演方法，而把凡是利用泰勒级数展开把非线性反问题线性化的广义反演方法称为广义线性反演。在过去一些文献中把利用广义逆的方法称为广义反演，这是对广义反演的一种狭义的理解，实际上广义逆方法只是广义线性反演中比较典型的一种。

地球物理中的广义反演方法是现代数值分析方法和 Backus-Gilbert（以下简称为 BG）反演理论结合的产物，它们实际上是在离散地球模型的情况下 BG 反演理论的简化和应用。这一点在看完第四章 BG 理论之后就会十分明了。当然，从 BG 反演理论中的评价准则也可以导出所谓的 BG 广义反演方法，但是根据循序渐进的原则我们把 BG 广义反演放到下一章去讨论。在本章中我们只研究可以直接从现代数值分析方法中延伸出来的广义反演，并介绍其在地球物理问题中的典型应用。和数值分析不同的是，由于我们所研究的地球物理反演问题的解不唯一，我们不仅要依照某种准则来构造出一种可以接受的解估计，而且还要尽量评价得出的解估计与其它可能的解估计相比有什么特点，这里我们就应用了 BG 理论的基本思想。尤其要注意的是，我们不能单纯追求与数据拟合得最好的解估计，因为可以与数据拟合的解估计可能有无穷多，数据的拟合误差最小并不说明与真实地球模型拟合的误差最小。重要的问题是得出解估计的分辨率和方差，只有它们才能反映出我们求出的地球模型的质量。

广义线性反演首先要从地球模型的参数化（Parametrization）和非线性问题的线性化开始，现在就首先来讨论参数化及线性化的问题。

一、参数化、线性化和归一化

（一）地球模型的参数化

在本章中我们假定地球模型可以用有限个离散参数来描述，因此可以用一个向量来表示。由于实际地球的物理性质一般是一个三维空间的分块连续函数（见第五章），把这样的函数离散化为有限个离散参数的过程叫参数化。我们同时假定精确的地球物理数据是一个可以由模型通过计算得到的实数，而实际地球物理数据虽然包含着干扰，但它的来源与地球模型无关，因此实际数据可以表示为精确数据和干扰之和。在大多数情况下，我们都认为干扰可以用正态分布的高斯噪音来描述。对地球模型和数据作以上假定给广义反演带来的好处

是便于在计算机上实现，但是也给广义反演的应用带来一定的限制。在第五章中我们将会发现，这样的地球模型是比较粗糙的，而且实际数据并不是都可以用计算数据加上随机噪音来描述的。然而，鉴于广义反演技术充分利用了计算机的特点并具有相当的灵活性，仍不失为现时应用最广泛的反演工具。

记 m 为参数化以前的地球模型，它是空间点 $x=(x_1,x_2,x_3)$ 的分块连续函数。记 d_i 为第 i 次观测的地球物理数据，δd_i 为其中的干扰，则由上述假定常常可以把数据和模型的关系表为

$$d_i=A_i(m)+\delta d_i,\quad i=1,2,\cdots,n \tag{3.1}$$

其中 A_i 表示对第 i 个数据由模型 m 到 d_i 的规定的正演计算，即 m 的泛函。回忆泛函的定义可知，d_i 是取决于函数 m 的变量，因此(3.1)表示了一个泛函方程式。如果我们有 n 个观测数据则可由(3.1)式写出包含有 n 个方程的泛函方程组。这样，我们要研究的反演问题就变为从形如(3.1)的泛函方程组解出地球模型 m。

如果把地球模型写为空间点 x 的函数 $m(x)$，我们要反演的是对于任意点 x_0 的函数 $m(x_0)$ 的值，这种连续地球模型的反演问题将放到下一章讨论。不过，由于地球物理数据的有限性，可以推测由有限个 d_i，$(i=1,2,\cdots,n)$ 反演 $m(x)$ 时唯一解一般不存在。此外，由于地球模型 $m(x)$ 要用分块连续的函数表示，它的微分和积分在间断面上只能用广义函数来解释，而不是古典意义下的微积分。因此，对于本章和下一章研究的反演方法都致力于寻找和评价泛函方程组(3.1)的广义解，而不是古典解。

在广义反演中先把 $m(x)$ 离散化为包含有限个模型参数的向量 $\boldsymbol{m}$，同时把泛函方程组(3.1)化为代数方程组。当然多数情况下这样得到的是非线性方程组。

例如，对地震法而言，静止的各向同性完全弹性球状地球可以用密度 $\rho(r)$，体积模量 $k(r)$ 和切变模量 $\mu(r)$ 来表示，它们都是从球心出发的径向距离 r 的函数($0\leqslant r\leqslant a$，a 为地球半径)。这样的地球模型可写为

$$m=(\rho,k,\mu) \tag{3.2}$$

其中 m 既不是泛函，也不是向量(ρ,k,μ 等不是数，而是函数)，而是函数集。在这个集中函数随自变量 r 的变化规律有相同之处(例如，层状地球模型中有共同的间断点)。对于重力等单一方法，模型 m 本身就只是一个函数 $\rho(r)$。

利用我们关于地球内部构造的先验知识，把地球划分成地壳、地幔、外核和内核四层，设每一层内的地球物质是均匀各向同性的，于是函数 $\rho(r)$，$k(r)$ 和 $\mu(r)$ 可离散化为

内核：　半径 $0-r_1$，参数 ρ_1,k_1,μ_1

外核：　r_1-r_2，　ρ_2,k_2,μ_2

地幔：　r_2-r_3，　ρ_3,k_3,μ_3

地壳：　r_3-a，　ρ_4,k_4,μ_4

这样，模型便可以表示为由以下十五个元组成的向量：

$$\boldsymbol{m}^{\mathrm{T}}=(\rho_1,\cdots,\rho_4,k_1,\cdots,k_4,\mu_1,\cdots,\mu_4,r_1,r_2,r_3)$$

如果我们知道了震源的位置，并假设震源信号为尖脉冲，把地面上某一个观测台记录到的某一组体波的到达时间记为 t_i，利用射线理论和界面上弹性波反射和折射公式不难写出(详见第五和第六章)。

$$t_i=f_i(m),i=1,\cdots,n \tag{3.3}$$

式中 n 表示台站的总数。可见在参数离散化之后(3.1)式中的泛函可退化为函数，模型 m 退化为向量 $\boldsymbol{m}$，数据 $\boldsymbol{d}=(t_1,\cdots,t_n)^{\mathrm{T}}$ 也是向量。

在上例中要注意到模型向量的元可能是具有不同量纲的物理量，取决于所用的单位，它们的绝对值可能相差很大，因此在计算时要做归一处理。当然，把地球简单地分成四层的做法早已过时了，根据(3.3)式进行反演是没有现实意义的，我们只不过用这个例子做一个参数化的示范。

很容易看出，反问题的解答将取决于模型参数化的方式，因此参数化是广义反演中的一个非常重要的问题。遗憾的是，现在还没有模型参数化的理论，来说明某种参数化的方式优于其它方式，而只能根据反演研究的目的和对象以及我们已有的先验知识来进行参数化。选择地球模型参数的原则是尽可能利用先验知识减少模型参数的数目，并满足反演结果的分辨率和精度要求。关于模型参数化的艺术可参考本章后几节的应用实例。

在模型参数化时，还要考虑不同的参数化方法能否导致反问题的解答稳定地收敛。以上述层状地球模型为例，如果我们把层数加大，反演获得的解答是不是会远远偏离于原先四层模型的解答？由于反问题的解答还与数据的个数和性质有关，也与算子的性质有关，在模型参数化时还必须综合考虑数据和算子的性质，来估计反问题解答可能达到的分辨率和方差。必要时还要通过实际计算了解模型参数化对反问题解答的影响。

(二)反问题的线性化

在模型参数化之后，便可以把精确数据 d_i 和模型参数 m_j 的关系(3.1)表示成以下非线性方程组：

$$d_i=A_i(m_1,\cdots,m_J) \tag{3.4}$$

其中 $i=1,\cdots,n;j=1,\cdots,J$，用向量形式记

$$\boldsymbol{m}^{\mathrm{T}}=(m_1,\cdots,m_J)$$

$$\boldsymbol{d}^{\mathrm{T}}=(d_1,\cdots,d_n)$$

如果我们可以根据先验知识给出模型参数的初始猜测

$$\boldsymbol{m}_0^{\mathrm{T}}=(m_{10},\cdots,m_{J0})$$

则有可能用牛顿法解非线性方程组的迭代思想，用泰勒级数把(3.4)式右边展开为

$$d_i=A_i(\boldsymbol{m}_0)+\sum_{j=1}^{J}\left[\frac{\partial A_i}{\partial m_j}\right]_{m_{j0}}(m_j-m_{j0})+\sum_{j=1}^{J}\left[\frac{\partial^2 A_i}{\partial m_j^2}\right]_{m_{j0}}\frac{(m_j-m_{j0})^2}{2!}+\cdots \tag{3.5}$$

把(3.5)式中的2次以上的项忽略掉就得到了线性化的系统，即

$$\boldsymbol{y}=\boldsymbol{J}\boldsymbol{x} \tag{3.6}$$

其中 $\boldsymbol{y}$ 的元

$$\begin{aligned}y_i&=\Delta d_{i0}\\&=d_i-A_i(m_{10},\cdots,m_{J0}),i=1,\cdots,n\end{aligned} \tag{3.7}$$

为实测数据和用初始猜测取得的预测数据之差。向量 $\boldsymbol{x}$ 的元为

$$x_j=\Delta m_{j0}=m_j-m_{j0},j=1,\cdots,J \tag{3.8}$$

为模型参数的修改增量，而矩阵 $\boldsymbol{J}$ 为 Jacobian 矩阵，它的元为

$$g_{ij}=(\partial A_i/\partial m_j)_{m_{j0}} \tag{3.9}$$

即

$$\boldsymbol{J}=\begin{bmatrix}\partial A_1/\partial m_1 & \partial A_1/\partial m_2\cdots\partial A_1/\partial m_j\\ \partial A_2/\partial m_1 & \partial A_2/\partial m_2\cdots\partial A_2/\partial m_j\\ \vdots & \vdots\\ \partial A_n/\partial m_1 & \partial A_n/\partial m_2\cdots\partial A_n/\partial m_j\end{bmatrix}_{m_{j0}}$$

矩阵方程(3.6)给出了一个迭代格式。把 m_{j0} 的上角标改为 m_{jk}，$k=0,1,\cdots,K$，K 为迭代次数，则由每一次解方程(3.6)都可以修改我们对模型参数的猜测，经过 K 次迭代之后，就可以得到模型参数的逼近值。

迭代计算的流程表示于图 3.1。注意在广义反演中迭代停止的准则是模型修改增量的模小到可以忽略，而不是数据拟合差的模很小，这是与最小二乘法不同的地方。

地球物理反演问题也有可能用隐式的一般非线性方程给出，即

$$f(\boldsymbol{d},\boldsymbol{m})=0$$

在模型参数化后，由于数据 $\boldsymbol{d}$ 为已知，代入后可得方程组

$$f_i(\boldsymbol{m})=0,i=1,2,\cdots,n \tag{3.10}$$

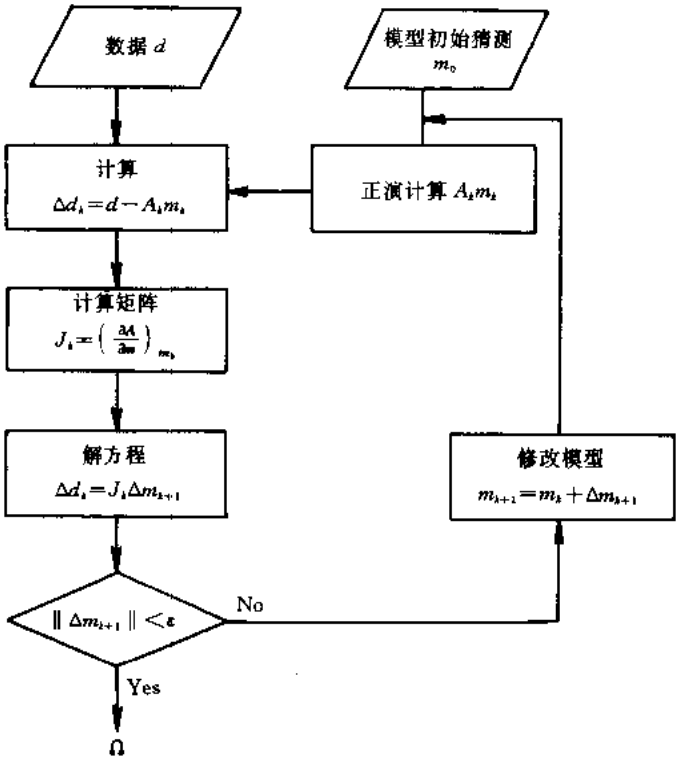


图 3.1　广义反演的一种迭代流程

用牛顿法的思想来解非线性方程组(3.10)时可在模型的初始猜测 $\boldsymbol{m}^0$ 处把 f_i 展成泰勒级数。令 $\boldsymbol{f}^{\mathrm{T}}=(f_1,\cdots,f_n)$，得

$$\boldsymbol{f}(\boldsymbol{m}_0)+\boldsymbol{J}_0(\boldsymbol{m}-\boldsymbol{m}_0)+\cdots=0$$

其中 $\boldsymbol{J}_0=\boldsymbol{J}(\boldsymbol{m}_0)=(\partial f_i/\partial m_j)m_{j0}$ 为 Jacobian 矩阵。取上式中的 0 阶及 1 阶项作线性近似，则此

方程组的解的一级近似变为以下线性方程组的解

$$\boldsymbol{f}(\boldsymbol{m}_0)+\boldsymbol{j}_0(\boldsymbol{m}_1-\boldsymbol{m}_0)=0$$

它可写为

$$\boldsymbol{m}_1=\boldsymbol{m}_0-(\boldsymbol{j}_0)^{-1}\boldsymbol{f}(\boldsymbol{m}_0)$$

对于迭代过程的第 k 步，模型 $\boldsymbol{m}$ 的 $(k+1)$ 次近似可由求解以下线性方程组获得

$$\boldsymbol{j}_k(\boldsymbol{m}_{k+1}-\boldsymbol{m}_k)=-\boldsymbol{f}(\boldsymbol{m}_k)。$$

上述线性化的过程能否收敛是我们首先要考虑的问题。大致地说，如果 $\boldsymbol{r}$ 是方程组 $\boldsymbol{f}(\boldsymbol{m})=\boldsymbol{0}$ 的一个解而且 Jacobian 矩阵不奇异，则 $\boldsymbol{f}$ 对 $\boldsymbol{m}$ 的元的二阶偏导数在 $\boldsymbol{r}$ 附近是连续的。这时在 $\boldsymbol{m}_0$ 足够接近 $\boldsymbol{r}$ 的条件下牛顿迭代过程是收敛的。而且，如果记误差向量为 $\boldsymbol{e}_k=\boldsymbol{m}_k-\boldsymbol{r}$，则 $\|\boldsymbol{e}_{k+1}\|/\|\boldsymbol{e}_k\|^2$ 在 $k\to\infty$ 时是有界的，说明收敛的阶次为 2[15]。

其次，每一次迭代的计算速度也必须考虑。牛顿法迭代的主要计算量是计算 Jacobian 矩阵和解线性方程组(3.6)。解线性方程组的计算量已经在第二章中详细讨论过，而直接计算 Jacobian 矩阵常常带来许多困难。用有限差分来代替偏导数计算的所谓正割法(Secant method)虽然可以减少计算的工作量，但是这样做有导致迭代发散的危险，因为在高维情况下推广正割法而取得的关于 $\boldsymbol{J}(\boldsymbol{m})$ 的近似常常是奇异的。在多维最优化方法中常用的拟牛顿法这时常可使用[30]，它的思想是在开始迭代时用比较粗糙的 $\boldsymbol{J}(\boldsymbol{m})$ 近似，而在迭代过程中不断提高它的精度。

关于非线性方程组的求解方法在最优化算法的书籍中都可以找到，但是地球物理反演中直接应用这些数值计算方法常常得不到满意的结果。因为，地球物理数据中包含的不仅仅是舍入误差，它与参数化后的模型计算出来的预测数据可能有很大差别。此外，Jacobian 矩阵常可能接近奇异，等等。因此，还要结合地球物理数据的特点来研究非线性反演问题，这就是以下各节要讨论的广义线性反演方法。如果在泰勒展开式中取了高次项，则可以导出广义非线性反演方法，它们在地球物理反演中还没有得到广泛应用。

近年来的研究表明，非线性反问题遵循的规律要比线性反问题复杂得多，决不像牛顿迭代法那样，用迭代加解线性方程组就可以解决。因此，类似于图列的反演流程只给出了处理非线性反问题的一个出发点，对非线性反问题的深入讨论涉及许多新概念和新思路，将在第几章讨论。

(三)归一化

在本节上面所举的例子(3.3)中，参数化后的地球模型为

$$\boldsymbol{m}^{\mathrm{T}}=(\rho_1,\cdots\rho_4,k_1,\cdots k_4,\mu_1,\cdots,\mu_4,r_1,r_2,r_3)$$

它的元为不同量纲的物理量。取决于量纲，它们的数值可以相差几个以上的数量级，此时 Jacobian 矩阵的列向量的模也差别极大，以至性状很坏。因此，在进行数值反演计算之前，应该把所有模型参数化为无量纲的量，这个过程称为归一化。对于数据也有类似的情况，例如，在用地震波走时和振幅作联合反演时，用 SI 单位制则走时以 s 为单位，位移以 m 为单位，因此位移常小于 10^{-8} 数量级。注意到不同的数据方差也不同，因此用方差或其它统计参数使数据和模型归一化常常是必不可少的。

以数据为例，假定各个数据在统计上是独立的，其标准差为 $\sigma_i, i=1,\cdots,n$，则 d_i/σ_i 是一个无量纲的量。记对角阵

$$\boldsymbol{E}=\begin{bmatrix} 1/\sigma_1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & 1/\sigma_n \end{bmatrix}$$

则可称 $\boldsymbol{d}'=\boldsymbol{Ed}$ 为归一化的数据。对于线性方程组 $\boldsymbol{d}=\boldsymbol{Am}$，数据规格化后变为

$$\boldsymbol{EAm}=\boldsymbol{Ed} \tag{3.11}$$

或写为

$$\boldsymbol{A}'\boldsymbol{m}=\boldsymbol{d}',\ \boldsymbol{A}'=\boldsymbol{EA},\ \boldsymbol{d}'=\boldsymbol{Ed}$$

如果要用最小二乘法求解方程组，则此时我们要把预测误差 $\boldsymbol{r}$ 与 $\boldsymbol{E}$ 的积的 L_2 模最小化。即使

$$\| \boldsymbol{Er} \|_2=\boldsymbol{r}^{\mathrm{T}}\boldsymbol{E}^{\mathrm{T}}\boldsymbol{Er}=\boldsymbol{r}^{\mathrm{T}}\boldsymbol{C}_d^{-1}\boldsymbol{r} \tag{3.12}$$

最小化，其中

$$\boldsymbol{r}=\boldsymbol{Am}-\boldsymbol{d},$$

而 $\boldsymbol{C}_d$ 为数据的协方差矩阵。用标准差组成的矩阵去对数据归一化相当于给数据加权，即给误差大的数据加上较小的权，这也有利于提高反演计算的精度。

如果数据在统计上是相关的，则归一化时必须考虑数据的协方差矩阵。我们知道，把一个有误差数据 $\tilde{d}_i$ 当作随机变量处理时它的方差定义为

$$\mathrm{Var}(\tilde{d}_i)=\lim_{L\to\infty}\frac{1}{L}\sum_{l=1}^{L}(\tilde{d}_{il}-d_i)^2 \tag{3.13}$$

式中 L 为测量次数，d_i 为随机变量 $\tilde{d}_i$ 的数学期望 $E(d_i)$，也就是精确的数据，而 $\tilde{d}_{il}$ 为第 l 次的测量值。两个带误差的数据 $\tilde{d}_i$ 和 $\tilde{d}_j$ 在统计上相关时它们的互协方差定义为以下乘积的数学期望

$$\begin{aligned}\mathrm{Cov}(\tilde{d}_i,\tilde{d}_j)&=E\{(\tilde{d}_i-d_i)(\tilde{d}_j-d_j)\}\\&=\lim_{L\to\infty}\frac{1}{L}\sum_{l=1}^{L}(\tilde{d}_{il}-d_i)(\tilde{d}_j-d_j)\end{aligned} \tag{3.14}$$

对于整个数据集，$\tilde{\boldsymbol{d}}$ 乃是随机向量，它的协方差矩阵定义为

$$\boldsymbol{C}_d=\begin{bmatrix} \mathrm{Var}(\tilde{d}_1) & \mathrm{Cov}(\tilde{d}_1,\tilde{d}_2) & \cdots & \mathrm{Cov}(\tilde{d}_1,\tilde{d}_n)\\ \mathrm{Cov}(\tilde{d}_2,\tilde{d}_1) & \mathrm{Var}(\tilde{d}_2) & \cdots & \mathrm{Cov}(\tilde{d}_2,\tilde{d}_n)\\ \vdots & \vdots & & \vdots\\ \mathrm{Cov}(\tilde{d}_n,\tilde{d}_1) & \mathrm{Cov}(\tilde{d}_n,\tilde{d}_2) & \cdots & \mathrm{Var}(\tilde{d}_n) \end{bmatrix} \tag{3.15}$$

因此 $\boldsymbol{C}_d$ 为 $n\times n$ 对称正定矩阵。利用矩阵特征值分解的方法可以把协方差矩阵分解为

$$\boldsymbol{C}_d=\boldsymbol{V}\Lambda\boldsymbol{V}^{\mathrm{T}}, \tag{3.16}$$

其中 $\boldsymbol{V}$ 为正交矩阵，它的列向量为特征向量；Λ 为对角阵，对角线元素为特征值。取归一化矩阵

$$\boldsymbol{E}=\Lambda^{-1/2}\boldsymbol{V}^{\mathrm{T}}, \tag{3.17}$$

则满足(3.12)中 $\boldsymbol{E}^{\mathrm{T}}\boldsymbol{E}=\boldsymbol{C}_d^{-1}$ 的要求，可把线性方程组按(3.11)式归一化。

在进行实际地球物理数据的反演时，按(3.17)式作归一化处理要求给定数据的协方差矩阵 $\boldsymbol{C}_d$，这要求多次重复观测，增加数据采集的成本。在数据统计上不相关而且是由多台不

同仪器观测取得时，可以利用不同仪器在同一基点处重复观测的数据确定每台仪器的方差，并把此方差视为某台仪器在某天观测的数据的方差，用来对不同仪器观测的数据作归一化。

二、广义最小二乘法

在参数化、线性化和归一化之后，我们将结合地球物理反演的情况讨论形如(3.6)的线性方程组的各种解法。首先来看看基于以数据拟合差的 L_2 模取极小为准则的一类算法，即广义最小二乘法。由于算法成熟，最小二乘法的各种变种在地球物理反演时应用非常广泛。由于在上一章中我们已讨论过常规的最小二乘法及其扰动分析，本节讨论应用最小二乘及最小长度等准则求反问题解估计的一门技术，统称为广义最小二乘法。顾名思义，解估计指的是针对具体的问题和数据按照设计的算法取得的计算结果。鉴于实际地球物理数据的反演问题非唯一性和算法的多样性，解估计关联问题、数据和算法三个方面，千万不要把它和数学上由逻辑推理导出的古典解等同起来。换句话说，解估计是我们实际算出来的东西，即使反问题的解(即真解或精确的地球模型)真的存在和唯一，我们也只能希望它尽量接近真解，但实际上它可能与真解相差很远。由于精确的地球模型并不知道，因此我们无法用解估计与真解的差别(如解估计与真解之差的模)来衡量解估计的质量，而只能按 BG 理论用分辨率和方差来衡量它的质量。下面，我们先来看看最小二乘法对数据和模型隐含的要求。

(一)最小二乘法隐含的要求

对于方程式 $\boldsymbol{d}=\boldsymbol{Gm}$，$\boldsymbol{G}$ 为 $n\times J$，$J\leqslant n$，最小二乘法的预测误差是

$$\|\boldsymbol{e}\|_2^2=(\boldsymbol{d}-\boldsymbol{Gm})^{\mathrm{T}}(\boldsymbol{d}-\boldsymbol{Gm}) \tag{3.18}$$

由(2.62)式可知，对(3.18)取极小便得到最小二乘解估计

$$\hat{\boldsymbol{m}}=[\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}]^{-1}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{d} \tag{3.19}$$

由(3.18)式可知，最小二乘准则中预测误差是向量 $\boldsymbol{e}$ 的欧几里德长度的平方，即 L_2 范数。

回忆第二章中向量的范数定义可知，一个向量的长度可以用任意范数来衡量，而不一定要用 L_2。因此，我们有理由用任意范数

$$\|\boldsymbol{e}\|_\rho=\left[\sum_i |e_i|^\rho\right]^{1/\rho} \tag{3.20}$$

的最小化来求方程组 $\boldsymbol{Gm}=\boldsymbol{d}$ 的解估计。那么用什么范数的最小化最好呢？这要看数据 $\boldsymbol{d}$ 的统计性质。假定如图 3.2 所示，在直线拟合中第 k 个数据有很大的偏差，由(3.20)式可知，当数据很多时 $|e_k|$ 对低 ρ 的范数 $\|\boldsymbol{e}\|_\rho$ 的值影响不大，而对大 ρ 值的范数影响甚大。换句话说，低阶范数对不同尺度的误差给以的权相差不大，而高阶范数对大误差给予优惠。这样看来，在最小化时选取什么范数必须考虑数据中误差的特性。当数据中包含有许多误差很大的"坏"数据时，它的概率分布曲线很宽，属于"长尾巴型"，这时不宜使用低阶范数。实际上用 L_2 范数隐含着数据的统计特性服从"短尾

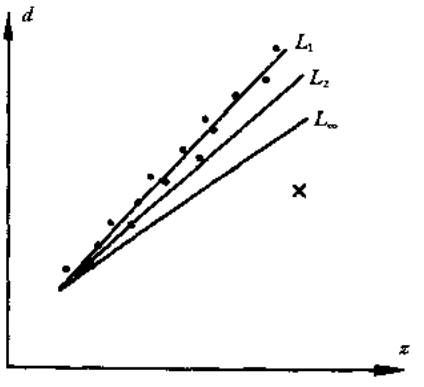


图 3.2　对不同范数作最小化时直线拟合的示意图

×表示一个偏差很大的数据，圆点为偏差小的测量数据，z 为拟合直线的坐标

巴"的高斯分布。

设 $\tilde{d}_i$ 为服从高斯分布的某个观测数据，$\bar{d}$ 和 σ^2 为它的均值和方差，则此数据的概率函数表为

$$P(\tilde{d}_i)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(\tilde{d}_i-\bar{d})^2}{2\sigma^2}\right] \tag{3.21}$$

如果假定数据集中所有数据的均值和方差都分别为 $\bar{d}$ 和 σ^2，由于 n 个互不相关的数据的联合概率分布等于 n 个独立的高斯分布的概率的乘积，则数据向量 $\tilde{\boldsymbol{d}}$ 的联合概率分布为

$$\begin{aligned}P(\tilde{\boldsymbol{d}})&=\prod_{i=1}^{n}(2\pi)^{-1/2}\sigma^{-1}\exp\left[-\frac{(\tilde{d}_i-\bar{d})^2}{2\sigma^2}\right]\\&=(2\pi)^{-n/2}\sigma^{-n}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\tilde{d}_i-\bar{d})^2\right]\end{aligned} \tag{3.22}$$

如果应用下节第四段将讨论的最大似然的思想来对 $\lg P(\tilde{\boldsymbol{d}})$ 求极大，代入(3.22)便等效于对

$$\sum_{i=1}^{n}(\tilde{d}_i-\bar{d})^2=\|\boldsymbol{e}\|_2^2$$

求极小，这就是最小二乘准则(3.18)。因此，最小二乘法的应用条件是数据服从高斯分布。

虽然高斯分布的数据在地球物理测量中是很常见的，但是大多数据集中总有那么几个"坏"数据，它们的存在说明数据并不完全服从高斯分布的统计性质。如果在反演之前不曾把这些坏数据去掉的话，那么最好用使 L_1 范数最小化的反演方法，如线性规划法等，因为这时最小二乘法给予坏数据的权仍然过大。这个问题我们将留到本章第五节讨论。

(二)混定方程组和欠定方程组

在第二章第三节中已讨论过，对于数据个数 n 大于未知参数个数 J 的超定方程组的情况如果矩阵 $(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})$ 不奇异，则由(3.19)式给出的最小二乘解估计在理论上是唯一存在的。如果此时 $n=J$，则存在精确解 $\boldsymbol{m}$ 使得预测误差为零。但是在地球物理反演中常遇到矩阵 $(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})$ 奇异或接近奇异的情况，这时虽然 $n\geqslant J$，线性反问题的最小二乘解估计仍然不能由(3.19)式算出。由于这类反问题不能用对付超定方程组的常规最小二乘法求解估计，我们把它们称为混定方程组求解问题(mixed-determined problems)，在地球物理反演中经常遇到。

例如，假设要利用全国地震台网观测的天然地震走时来反演中国境内岩石圈的平均速度。把中国岩石圈按经纬度分划成 J 块，使 J 小于走时数据的总数 n，于是可得到一个超定方程组。然而，由于地震台站分布极不均匀，在内地台网密集，而在西部边远地区极其稀疏，相应内地岩石圈的波速反演才是真正超定的，而在边远地段波速参数的反演却是欠定。对整个中国岩石圈波速结构的反演问题因其部分模型参数超定而另一部分欠定，属于混定问题。再如对有限大小的磁性体几何参数的反演，在磁测剖面很短时顶界面参数(或埋深)的反演为超定，而底界面参数(或延深)的反演却是欠定的，也属于混定问题。

如果可能，我们可以希望通过把模型参数分为超定和欠定两类的办法来对付混定问题。这就是说，我们要把方程式 $\boldsymbol{G}\boldsymbol{m}=\boldsymbol{d}$ 写成分块形式 $\boldsymbol{G}'\boldsymbol{m}'=\boldsymbol{d}'$，其中 $\boldsymbol{m}'$ 的上部为超定的模型参数 $\boldsymbol{m}'_1$，而下部包含欠定的参数 $\boldsymbol{m}'_2$：

$$\begin{pmatrix}\boldsymbol{G}'_1 & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{G}'_2\end{pmatrix}\begin{pmatrix}\boldsymbol{m}'_1\\ \boldsymbol{m}'_2\end{pmatrix}=\begin{pmatrix}\boldsymbol{d}'_1\\ \boldsymbol{d}'_2\end{pmatrix}$$

如果这样的分块可以实现，则对 $\boldsymbol{m}'_1$ 采用最小二乘法反演，而用下面要讨论的对付欠定方程

组的方法反演 m'_2。实际上，矩阵 G 的奇异值分解就是把 m'_1 和 m'_2 分开的一种途径，由于它已经超出最小二乘法的范围，我们将放到本章第四节讨论。

对于病态的方程组 $Gm=d$，即矩阵 (G^TG) 的条件数较大的情况，阻尼最小二乘法也是常用的。这时我们不仅要使预测误差最小化，而且也要使解的长度 $L=m^Tm$ 最小化。把以下目标函数取极小

$$\phi(m)=\|e\|_2+\varepsilon^2L=e^Te+\varepsilon^2m^Tm \tag{3.23}$$

其中权 ε^2 称为阻尼系数，当它足够大时极小化的过程将主要对(3.23)式右边第二项进行，因为此时对预测误差项的权相对减少，而使解估计的预测误差相应加大。将(3.23)写为

$$\begin{aligned}\phi&=(Gm-d)^T(Gm-d)+\varepsilon^2m^Tm\\&=m^TG^TGm-2m^TG^Td+d^Td+\varepsilon^2m^Tm\end{aligned}$$

极小化时令 $\partial\phi/\partial m^T=0$，得

$$2G^TGm-2G^Td+2\varepsilon^2m=0$$

因此，阻尼最小二乘法的解估计为

$$\hat{m}=[G^TG+\varepsilon^2I]^{-1}G^Td \tag{3.24}$$

其中阻尼系数 ε^2 的选择主要靠实验，它起到在解估计的分辨率和方差之间的调节作用。

对于欠定方程组即 $n<J$ 的情况，满足方程组的解有无穷多个，说明数据不足时模型参数不能唯一地确定。最常用的方法是用模型参数的欧几里德以度 $L=m^Tm=\Sigma m_j^2$ 最小为准则求解估计。实际上，L 表示一个物理系统的能量(离散化后为功率)，取 L 极小说明求系统处于能量最稳定状态下的参数。我们把预测误差为零作为约束，即令 $e=d-Gm=0$，用拉格朗日乘子法把 L 取极小的条件极值问题化为以下无条件极值问题

$$\phi=L+\sum_{i=1}^{n}\lambda_ie_i=\sum_{j=1}^{J}m_j^2+\sum_{i=1}^{n}\lambda_i\Big(d_i-\sum_{j=1}^{J}G_{ij}\cdot m_j\Big)=\min \tag{3.25}$$

其中 λ_i 为拉格朗日乘子。记 $\lambda=(\lambda_1,\cdots,\lambda_n)^T$，则上述目标函数 ϕ 可写为

$$\phi(m)=m^Tm+\lambda^Te=m^Tm+\lambda^T(d-Gm) \tag{3.26}$$

令

$$\partial\phi/\partial m^T=2m-G^T\lambda=0$$

得

$$m=G^T\lambda/2 \tag{3.27}$$

将(3.27)代回入约束方程

$$d=Gm=G[G^T\lambda/2]$$

便有

$$\lambda=2(GG^T)^{-1}d \tag{3.28}$$

把(3.28)代回(3.27)便得到欠定方程组的最小长度解估计

$$\hat{m}=G^T(GG^T)^{-1}d \tag{3.29}$$

(三)最平解估计和最光滑解估计

上面我们讨论了用模型参数的最小长度特征作为先验知识求一种解估计的方法。作为模型参数的总体特征，除了能量外还可以有平度，光滑度等，它们也可以用作求解估计的准则。例如，如果已知模型参数是变化缓慢的函数(如海水密度)的取样值，则我们可以用模型的一阶导数的模取极小为准则，求模型参数的所谓“最平”解估计[31]。用有限差分代替微分，模型向量 m 的“平度”可用下式计算

$$\boldsymbol{l}=\begin{bmatrix}-1 & 1 & & & & \\ & -1 & 1 & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & -1 & 1\end{bmatrix}\begin{bmatrix}m_1\\ m_2\\ \vdots\\ \vdots\\ \vdots\\ m_j\end{bmatrix}=\boldsymbol{Dm} \tag{3.30}$$

其中 $\boldsymbol{D}$ 称为平度矩阵。应用 $\boldsymbol{l}$ 的 L_2 模表示模型参数的总体平度，则最平解估计可以令下式取极小获得

$$L=\boldsymbol{l}^{\mathrm{T}}\boldsymbol{l}=\boldsymbol{m}^{\mathrm{T}}\boldsymbol{D}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{m}=\boldsymbol{m}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{m} \tag{3.31}$$

其中 $\boldsymbol{W}$ 等效于对模型参数 $\boldsymbol{m}$ 的权系数矩阵，表为

$$\boldsymbol{W}=\begin{bmatrix}1 & -1 & & & & & & & \\ -1 & 2 & -1 & & & & & & \\ & -1 & & 2 & & -1 & & & \\ & & \ddots & & \ddots & & \ddots & & \\ & & & \ddots & & \ddots & & \ddots & \\ & & & & \ddots & & \ddots & & \ddots \\ & & & & & -1 & & 2 & -1 \\ & & & & & & -1 & & 1\end{bmatrix} \tag{3.32}$$

说明最平解估计为对模型向量加权的一种解估计。

在求欠定方程组 $\boldsymbol{G}\boldsymbol{m}=\boldsymbol{d}$ 的最平解估计时我们把方程组作为约束，用拉格朗日乘子法(3.31)式取极小时的解估计，即令

$$\phi(\boldsymbol{m})=L-\boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{d}-\boldsymbol{G}\boldsymbol{m})=\boldsymbol{m}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{m}+\boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{d}-\boldsymbol{G}\boldsymbol{m})=\min$$

由

$$\partial\phi/\partial\boldsymbol{m}^{\mathrm{T}}=2\boldsymbol{W}\boldsymbol{m}-\boldsymbol{G}^{\mathrm{T}}\boldsymbol{\lambda}=0$$

得

$$\boldsymbol{m}=\boldsymbol{W}^{-1}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{\lambda}/2$$

代回入方程

$$\boldsymbol{d}=\boldsymbol{G}\boldsymbol{m}=\boldsymbol{G}\boldsymbol{W}^{-1}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{\lambda}/2$$

便可求出乘子向量 $\boldsymbol{\lambda}$，最后得最平解估计为

$$\hat{\boldsymbol{m}}=\boldsymbol{W}^{-1}\boldsymbol{G}^{-1}(\boldsymbol{G}\boldsymbol{W}^{-1}\boldsymbol{G}^{\mathrm{T}})^{-1}\boldsymbol{d} \tag{3.33}$$

在方程组病态的情况下，我们还可以用阻尼最小二乘法的思想求最平解估计。这时，把目标函数取为

$$\phi(\boldsymbol{m})=(\boldsymbol{G}\boldsymbol{m}-\boldsymbol{d})^{\mathrm{T}}(\boldsymbol{G}\boldsymbol{m}-\boldsymbol{d})+\varepsilon^2\boldsymbol{m}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{m}=\min$$

令

$$\partial\phi/\partial\boldsymbol{m}^{\mathrm{T}}=2\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{m}-2\boldsymbol{G}^{\mathrm{T}}\boldsymbol{d}+2\varepsilon^2\boldsymbol{W}\boldsymbol{m}=0$$

得

$$\hat{\boldsymbol{m}}=(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}+\varepsilon^2\boldsymbol{W})^{-1}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{d} \tag{3.34}$$

如果按(3.12)式同时对数据归一化，即把目标函数取为

$$\phi(\boldsymbol{m})=(\boldsymbol{E}\boldsymbol{G}\boldsymbol{m}-\boldsymbol{E}\boldsymbol{d})^{\mathrm{T}}(\boldsymbol{E}\boldsymbol{G}\boldsymbol{m}-\boldsymbol{E}\boldsymbol{d})+\varepsilon^2\boldsymbol{m}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{m}$$

则类似地可得数据归一化之后的最平阻尼二乘法解估计为

$$\hat{m}=(G^{\mathrm{T}}C_d^{-1}G+\varepsilon^2W)^{-1}G^{\mathrm{T}}C_d^{-1}d \tag{3.35}$$

其中 C_d^{-1} 为数据的协方差矩阵，$E^{\mathrm{T}}E=C_d^{-1}$。

至于所谓的最光滑解估计是以模型参数的二阶导数取极小为准则得到的解估计。用差分代替微分，所谓的光滑矩阵 D' 与(3.32)中的权矩阵 W 具有同样的形式。因此最光滑解估计只要用 D' 代替(3.30)式中矩阵 D 便可用上述方法导出，而不需赘述。

最小能量、最平和最光滑解估计是在信息不足时用先验知识加入所谓的“宽约束”的典型例子，与最小二乘法相比，它们导致解估计分辨率的降低，但同时也减少了方差。另一类先验知识是所谓的“紧约束”，下面就来看看其中一个典型例子。

(四)使用线性约束的最小二乘法

在某些情况下，例如在有钻井或测井资料的情况下，我们有可能知道地球模型参数所满足的线性约束条件，它可写为 $Fm=h$。例如，当我们已知 $m_1=h_1$ 及 $m_2=h_2$ 时，上述线性约束便有以下形式

$$Fm=\begin{bmatrix}1 & & & & & \\ & 1 & & & & \\ & & 0 & & & \\ & & & \ddots & & \\ & & & & \ddots & \\ & & & & & 0\end{bmatrix}\begin{bmatrix}m_1\\ m_2\\ \vdots\\ \vdots\\ \vdots\\ m_j\end{bmatrix}=\begin{bmatrix}h_1\\ h_2\\ 0\\ \vdots\\ \vdots\\ 0\end{bmatrix}=h$$

在知道模型参数的加权平均或其它形式的线性组合时，这种约束条件都可以用来提高最小二乘解估计的质量。

用拉格朗日乘子法把 $Fm=h$ 作为约束并求预测误差 $|e^{\mathrm{T}}e|$ 的极小，可归结为令

$$\phi(m)=(Gm-d)^{\mathrm{T}}(Gm-d)+2\lambda^{\mathrm{T}}(Fm-h)=\min \tag{3.36}$$

式中 $2\lambda_i$ 为拉格朗日乘子。令 $\partial\phi/\partial m^{\mathrm{T}}=0$ 得

$$G^{\mathrm{T}}Gm+F^{\mathrm{T}}\lambda=G^{\mathrm{T}}d \tag{3.37}$$

而在(3.36)中令 $\partial\phi/\partial\lambda^{\mathrm{T}}=0$ 有约束

$$Fm=h \tag{3.38}$$

合并(3.37)和(3.38)，我们得到以分块矩阵表示的方程

$$\begin{bmatrix}G^{\mathrm{T}}G & F^{\mathrm{T}}\\ F & 0\end{bmatrix}\begin{bmatrix}m\\ \lambda\end{bmatrix}=\begin{bmatrix}G^{\mathrm{T}}d\\ h\end{bmatrix} \tag{3.39}$$

此方程的系数矩阵为方阵，大小为 $(j+p)\times(j+p)$，p 为向量 h 中不等零的元的个数。对此方阵求逆并乘以(3.39)式的右端向量便可同时得到在线性约束条件下的最小二乘解估计 $\hat{m}$ 和乘子向量 λ。

三、广义线性反演

在上节中我们已经导出了各种推广形式的最小二乘法的解估计，其准则是根据预测误差的极小化或模型参数的长度或平度的简单性。归纳解估计的表达式(3.19)、(3.24)、

(3.29)、(3.33)、(3.34)及(3.35)，我们可以把线性反问题的解估计表为[31]

$$\hat{m}=Hd+v \tag{3.40}$$

其中 H 为某种矩阵，而 v 为某种向量，它们都与数据本身无关。推广到一般情况，线性反问题的解估计都可以用(3.40)表示。在广义反演中，习惯上把矩阵 H 称之为广义逆，并记为 G^{-g}，注意它与第二章定义的广义逆矩阵 G^{+} 的含义不同，即 G^{-g} 包含了 G^{+} 作为特殊情况，但并不限于 G^{+}。由(3.40)式可知，解估计的性质与 H 的性质密切相关。广义线性反演的基本点就在深入地研究广义逆的性质来预测解估计，从而使地球模型的解答符合地球物理解释的要求。

在上一节中，我们已经知道对超定方程组的最小二乘解估计，其广义逆为

$$G^{-g}=G^{+}=(G^{T}G)^{-1}G^{T} \tag{3.41}$$

对欠定方程组的最小长度解估计为

$$G^{-g}=G^{T}(GG^{T})^{-1} \tag{3.42}$$

等等，但是还没有讨论这些广义逆的性质。现在我们就从一般情况下广义反演中最重要的概念，即分辨率和方差入手，研究解估计对广义逆的依赖性。

(一)分辨率矩阵和方差

现在来研究线性方程组 $Gm=d$ 的解和解估计之间的关系，从而可以在不依赖于数据的情况下了解解估计的性质。先假定(3.40)式中向量 v 为零向量，将它代入此方程式有

$$\hat{m}=G^{-g}d=G^{-g}[Gm]=(G^{-g}G)m=Rm \tag{3.43}$$

式中 $R=G^{-g}G$ 为 $J\times J$ 矩阵。当 $R=I$ 时有 $\hat{m}=m$，即对任何数据解估计都是精确解。当 $R\neq I$ 时，模型参数的某个估计值 $\hat{m}_j$ 不仅与 m_j 有关，而且还与 m_j 邻近的参数 $m_{j\pm1}, m_{j\pm i}\cdots$ 有关，它实际上是真实模型参数 $\{m_j\}$，$j=1,\cdots,J$，的加权平均(见图3.3)，其中的权以矩阵 R 的行向量表示。由此可见，当 R 中各行的尖峰越窄，即越接近单位脉冲，而 R 越趋近对角线矩阵，则 $\hat{m}_j$ 越接近于 m_j 而与相邻参数的关系越不密切。换句话说，这时解估计的分辨率越高。由于 R 的各行是窥测解估计分辨率的窗口，因此 R 被称为模型的分辨率矩阵。对 $\hat{m}_k$ 的分辨率的定量估计可以用以下分辨核表示。

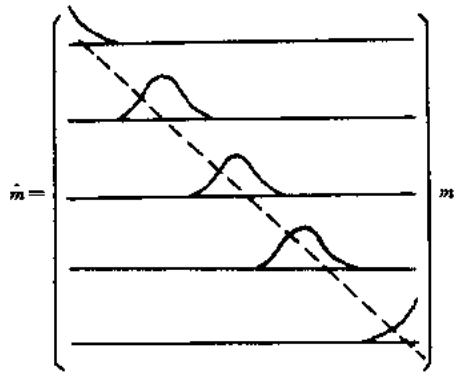


图 3.3 模型分辨率矩阵 R 的示意图，说明解估计与真解之间是加权平均关系，其中权函数为矩阵 R 的行

$$r_k=\sum_{j=1}^{J}(G_{ki}^{-g}G_{ij}-\delta_{kj})^2 \tag{3.44}$$

它是矩阵 R 中第 k 行与单位脉冲函数的差的 L_2 模。当 r_k 越大时模型估计值 $\hat{m}_k$ 偏离 m_R 越大，说明分辨率越差。因此，通过计算分辨率矩阵可使我们在取得数据之前就了解解估计 $\hat{m}$ 的性质。

假如我们已有了某种模型参数的解估计 $\hat{m}=G^{-g}d$，用它乘以 G 便得到了所谓的预测数据 $\hat{d}$

$$\hat{d}=Gm=G(G^{-g}d)=(GG^{-g})d=Pd \tag{3.45}$$

其中 $P=GG^{-g}$ 称为信息密度矩阵[20]，也有人称为数据分辨率矩阵[31]。信息密度矩阵 P 是实

际数据 $\boldsymbol{d}$ 和预测数据 $\hat{\boldsymbol{d}}$ 拟合程度的度量。例如，当 $\boldsymbol{P}=\boldsymbol{I}$ 时有 $\boldsymbol{d}=\hat{\boldsymbol{d}}$，说明预测误差等于零。反之，$\boldsymbol{P}$ 与单位矩阵 $\boldsymbol{I}$ 的偏离越大，说明预测误差越大。

当 $\boldsymbol{P}$ 不是单位矩阵时，它每一行与脉冲函数的差别反映了预测数据 $\hat{d}_i$ 在测值 d_i 中起的重要性，因此 $\boldsymbol{P}$ 的对角线元素构成的向量

$$\boldsymbol{p}=\mathrm{diag}(\boldsymbol{P})$$

又被称为数据的“重要性”向量。

由于信息密度矩阵 $\boldsymbol{P}$ 与数据无关，而只与系数矩阵（数据核）有关，因此研究它将有助于设计数据采集的参数如取样密度等。

除了分辨率之外，解估计的方差也是评价解估计质量的重要参数。与分辨率不同，$\hat{\boldsymbol{m}}$ 的方差不仅与广义逆有关，还与数据的方差有关。把数据向量 $\boldsymbol{d}$ 视为 n 维向量空间 $\mathscr{R}_d$ 中的一个点，而模型 $\boldsymbol{m}$ 视为另一个 n 维空间 $\mathscr{R}_m$ 中的一个点，则广义逆 $\boldsymbol{G}^{-g}$ 乃是从 $\mathscr{R}_d$ 空间向 $\mathscr{R}_m$ 空间的一个映射，它同时也把数据的误差放大为解估计的误差。回忆数据的协方差矩阵 $\boldsymbol{C}_d$ 为(3.15)式定义的对称正定矩阵，由于 $\boldsymbol{m}=\boldsymbol{G}^{-g}\boldsymbol{d}$ 是解估计 $\hat{\boldsymbol{m}}$ 的数学期望，则由定义可知它的协方差矩阵为

$$\begin{aligned}\boldsymbol{C}_m &=(\boldsymbol{G}^{-g}\tilde{\boldsymbol{d}}-\boldsymbol{G}^{-g}\boldsymbol{d})(\boldsymbol{G}^{-g}\tilde{\boldsymbol{d}}-\boldsymbol{G}^{-g}\boldsymbol{d})^{\mathrm{T}}\\ &=\boldsymbol{G}^{-g}(\tilde{\boldsymbol{d}}-\boldsymbol{d})(\tilde{\boldsymbol{d}}-\boldsymbol{d})^{\mathrm{T}}(\boldsymbol{G}^{-g})^{\mathrm{T}}\\ &=\boldsymbol{G}^{-g}\boldsymbol{C}_d(\boldsymbol{G}^{-g})^{\mathrm{T}}\end{aligned}\tag{3.46}$$

其中 $\tilde{\boldsymbol{d}}$ 为带误差的数据。当任两个数据 $\tilde{d}_i$ 和 $\tilde{d}_j$ 在统计上独立时，由于 $(\tilde{d}_i-d_i)$ 和 $(\tilde{d}_j-d_j)$ 在统计上不相关，$\boldsymbol{C}_d$ 变为以方差 $\sigma_i^2=\mathrm{Var}(\tilde{d}_i)$ 为元的对角阵，此时解估计的方差变为

$$\mathrm{Var}(\hat{m}_k)=\sum_{i=1}^{n}(G_{ki}^{-g})^2\mathrm{Var}(\tilde{d}_i)\tag{3.47}$$

在令数据的方差为单位时，我们可以考察解估计协方差与广义逆的关系。由(3.46)此时有

$$\boldsymbol{C}_{m.u}=\boldsymbol{G}^{-g}(\boldsymbol{G}^{-g})^{\mathrm{T}}\tag{3.48}$$

$\boldsymbol{C}_{m.u}$ 称为解估计单位协方差矩阵，它可以用来衡量数据中的误差在映射到解空间之后的放大程度。

现在让我们用上节讨论过的广义最小二乘法的广义逆为例，评价它们的解估计的分辨率。由(3.41)式可知，超定方程组的最小二乘解估计的模型分辨率矩阵为

$$\boldsymbol{R}=\boldsymbol{G}^{-g}\boldsymbol{G}=[\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}]^{-1}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}=\boldsymbol{I}$$

说明这时解估计的分辨率最高。信息密度矩阵

$$\boldsymbol{P}=\boldsymbol{G}\boldsymbol{G}^{-g}=\boldsymbol{G}[\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}]^{-1}\boldsymbol{G}^{\mathrm{T}}$$

不等于单位矩阵，说明此时数据多余。解估计的单位协方差矩阵由(3.41)代入(3.48)有

$$\boldsymbol{C}_{m.u}=[\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}]^{-1}$$

它取决于矩阵 $(\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}})$ 的逆的性质。对于欠定方程组的最小长度解估计(3.42)，相应地有模型分辨率矩阵

$$\boldsymbol{R}=\boldsymbol{G}^{\mathrm{T}}(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})^{-1}\boldsymbol{G}$$

它远离单位矩阵说明此时解估计的分辨率差。信息密度矩阵

$$\boldsymbol{P}=\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}(\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}})^{-1}=\boldsymbol{I}$$

说明欠定方程组中数据已被充分利用而没有多余。把(3.42)代入(3.48)得单位协方差矩阵为

$$C_{m,u}=G^{\mathrm{T}}(GG^{\mathrm{T}})^{-2}G$$

说明由于数据不足解估计的方差仍然较大。类似地，由(3.29)至(3.35)各式我们也不难用分辨率和方差对阻尼最小二乘及最平解估计进行评价。

(二)展布和最小展布解估计

对于线性反问题，我们已经知道可以由广义逆导出解估计的分辨率和单位方差。那么能否反过来根据我们对分辨率或方差的要求为准则来选取广义逆呢？在广义线性反演中所关心的正是这个问题。

首先来考虑找一个最佳分辨率的解估计的问题。由于在最理想的情况下分辨率矩阵 $\boldsymbol{R}$ 为单位矩阵 $\boldsymbol{I}$，我们还可以用矩阵$(\boldsymbol{R}-\boldsymbol{I})$的模来衡量分辨率。当取$(\boldsymbol{R}-\boldsymbol{I})$的 L_2 模作为分辨率的量度时，它被称为矩阵 $\boldsymbol{R}$ 的非对角线元素的展布，记为

$$\mathrm{spread}(\boldsymbol{R})=\|\boldsymbol{R}-\boldsymbol{I}\|_2^2=\sum_{i=1}^{J}\sum_{j=1}^{J}[R_{ij}-I_{ij}]^2 \tag{3.49}$$

类似的，对数据分辨率矩阵 $\boldsymbol{P}$，其展布定义为

$$\mathrm{spread}(\boldsymbol{P})=\|\boldsymbol{P}-\boldsymbol{I}\|_2^2=\sum_{i=1}^{n}\sum_{j=1}^{n}[P_{ij}-I_{ij}]^2 \tag{3.50}$$

注意在第四章中我们还将讨论另一种形式的展布，叫 BG 展布。为了区别这两种不同的情况，由(3.49)和(3.50)定义的展布又叫 Dirichlet 展布。

定义了展布之后，我们可以取展布最小为准则寻求广义逆，以便得到高分辨率的解估计。例如为了最充分地利用数据，要求使 spread$(\boldsymbol{P})$取极小的广义逆 $\boldsymbol{G}^{-g}$。这时把 spread$(\boldsymbol{P})$的第 k 行作为目标函数

$$\begin{aligned}\phi_k&=\sum_{i=1}^{n}(P_{ki}-I_{ki})^2=\sum_{i=1}^{n}P_{ki}^2-2\sum_{i=1}^{n}P_{ki}I_{ki}+\sum_{i=1}^{n}I_{ki}^2\\&=\min\end{aligned} \tag{3.51}$$

把 $\boldsymbol{P}=\boldsymbol{G}\boldsymbol{G}^{-g}$ 代入(3.51)便有

$$\begin{aligned}\phi_k=&\sum_{i=1}^{n}\Big[\sum_{j=1}^{J}G_{kj}G_{ji}^{-g}\Big]\Big[\sum_{l=1}^{J}G_{kl}G_{li}^{-g}\Big]-2\sum_{i=1}^{n}\Big[\sum_{j=1}^{J}G_{kj}G_{ji}^{-g}\delta_{ki}\Big]+\sum_{i=1}^{n}\delta_{ki}^2\\=&\sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{l=1}^{J}G_{ji}^{-g}G_{li}^{-g}G_{kj}G_{kl}-2\sum_{j=1}^{n}G_{jk}^{-g}G_{kj}+1\end{aligned}$$

为求使 ϕ_k 极小的广义逆，令

$$\begin{aligned}\frac{\partial\phi_k}{\partial G_{qr}^{-g}}&=2\sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{l=1}^{J}\delta_{jq}\delta_{ir}G_{li}^{-g}G_{kj}G_{kl}-2\sum_{j=1}^{n}\delta_{jq}G_{kj}\\&=2\sum_{l=1}^{J}G_{lr}^{-g}G_{kq}G_{kl}-2G_{kq}=0\end{aligned}$$

上式写成矩阵形式为

$$\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{G}^{-g}-\boldsymbol{G}^{\mathrm{T}}=0$$

或

$$\boldsymbol{G}^{-g}=[\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}]^{-1}\boldsymbol{G}^{\mathrm{T}}$$

这就是(3.41)式。上述推导说明取数据分辨率的展布最小导致超定方程组的最小二乘解估计。由此可知，取矩阵 $\boldsymbol{P}$ 的展布极小和取预测误差极小是等价的。

对于欠定方程组，我们希望取模型的分辨率矩阵 $\boldsymbol{R}$ 的展布极小来找广义逆。将 $\boldsymbol{R}=$

$G^{-g}G$代入(3.49)并取

$$\phi_k=\sum_{j=1}^{J}(R_{kj}-I_{kj})^2=\min$$

可用类似的方法导出形如(3.42)式的广义逆，说明使 spread($\boldsymbol{R}$)的 L_2 模极小的解估计就是最小长度解估计。

(三)随机逆和最小方差解估计

当数据带有误差时我们不仅要考虑解估计的分辨率，而且还必须使解估计的方差处在允许范围之内。由于解估计的方差不仅与广义逆有关，而且也和数据方差有关，因此在把方差作为准则求广义逆时要排除与数据有关的因素，只用单位协方差矩阵 $\boldsymbol{C}_{m,u}$作推导。类似于展布，我们也可以用 $\boldsymbol{C}_{m,u}$的 L_2 模来衡量单位协方差矩阵的大小。对于统计上独立的模型参数，由于协方差矩阵的非对角元素为零，它的大小可写为

$$\text{size}(\boldsymbol{C}_{m,u})=\|(\text{Var}_u\hat{\boldsymbol{m}})^{1/2}\|_2^2 \tag{3.52}$$

其中单位方差可由(3.47)式表为

$$\text{Var}_u(\hat{m}_k)=\sum_{i=1}^{n}(G_{ki}^{-g})^2 \tag{3.53}$$

有了(3.52)的定义后我们便可以求解估计方差的 L_2 模的最小解估计。

更全面的考虑是不仅要顾及数据误差的传播，同时还要顾及解估计的分辨率不足引起的误差[32]。设带误差的数据为 $\tilde{\boldsymbol{d}}$，则解估计可表为

$$\begin{aligned}\hat{\boldsymbol{m}}&=\boldsymbol{G}^{-g}\tilde{\boldsymbol{d}}=\boldsymbol{G}^{-g}\tilde{\boldsymbol{d}}-\boldsymbol{G}^{-g}\boldsymbol{d}+\boldsymbol{G}^{-g}\boldsymbol{d}\\&=\boldsymbol{G}^{-g}\boldsymbol{G}\boldsymbol{m}+\boldsymbol{G}^{-g}(\tilde{\boldsymbol{d}}-\boldsymbol{d})\\&=\boldsymbol{R}\boldsymbol{m}+\boldsymbol{G}^{-g}(\tilde{\boldsymbol{d}}-\boldsymbol{d})\end{aligned} \tag{3.54}$$

因此解估计的误差可表为

$$\hat{\boldsymbol{m}}-\boldsymbol{m}=(\boldsymbol{R}-\boldsymbol{I})\boldsymbol{m}+\boldsymbol{G}^{-g}(\tilde{\boldsymbol{d}}-\boldsymbol{d}) \tag{3.55}$$

从而误差($\hat{\boldsymbol{m}}-\boldsymbol{m}$)的协方差矩阵为

$$\boldsymbol{C}=(\boldsymbol{R}-\boldsymbol{I})\boldsymbol{C}_m(\boldsymbol{R}-\boldsymbol{I})^{\mathrm{T}}+\boldsymbol{G}^{-g}\boldsymbol{C}_d(\boldsymbol{G}^{-g})^{\mathrm{T}} \tag{3.56}$$

注意上式中有两项，第一项与分辨率矩阵 $\boldsymbol{R}$ 有关，是由解估计的分辨率不足引起的误差，而第二项为由数据误差的传播引起的解估计的误差项。由于($\hat{\boldsymbol{m}}-\boldsymbol{m}$)的方差为矩阵 $\boldsymbol{C}$ 的对角线元素，即 $C_{kk}=\text{Var}(\hat{m}_k-m_k)$，我们可以用($\hat{\boldsymbol{m}}-\boldsymbol{m}$)的方差的 L_2 模取极小为准则选取广义逆 $\boldsymbol{G}^{-g}$。也就是说，令 $\phi=\sum_k C_{kk}^2$，代入(3.56)及 $\boldsymbol{R}=\boldsymbol{G}^{-g}\boldsymbol{G}$ 并令

$$\partial\phi/\partial G_{qr}^{-g}=0$$

最后可以导出被称为狭义的广义线性反演广义逆，它表为

$$\begin{aligned}\boldsymbol{G}^{-g}&=\boldsymbol{C}_m\boldsymbol{G}^{\mathrm{T}}(\boldsymbol{G}\boldsymbol{C}_m\boldsymbol{G}^{\mathrm{T}}+\boldsymbol{C}_d)^{-1}\\&=(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{C}_d^{-1}\boldsymbol{G}+\boldsymbol{C}_m^{-1})^{-1}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{C}_d^{-1}\end{aligned} \tag{3.57}$$

因此，相应的最小方差解估计写为

$$\begin{aligned}\hat{\boldsymbol{m}}&=\boldsymbol{C}_m\boldsymbol{G}^{\mathrm{T}}(\boldsymbol{G}\boldsymbol{C}_m\boldsymbol{G}^{\mathrm{T}}-\boldsymbol{C}_d)^{-1}\boldsymbol{d}\\&=(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{C}_d^{-1}\boldsymbol{G}+\boldsymbol{C}_m^{-1})^{-1}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{C}_d^{-1}\boldsymbol{d}\end{aligned} \tag{3.58}$$

将(3.57)式代回 $\boldsymbol{C}$ 的表达式(3.56)有协方差矩阵

$$\boldsymbol{C}=(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{C}_d^{-1}\boldsymbol{G}+\boldsymbol{C}_m^{-1})^{-1} \tag{3.59}$$

而分辨率矩阵为

$$R=CG^{\mathrm{T}}C_d^{-1}G \tag{3.60}$$

在(3.57)中导出的最小方差解估计最早是由 Franklin(1970)在假定模型 m 为二阶高斯随机过程的一次实现的条件下推导出来的，它称之为随机逆[33]。以后，Jackson(1979)等人又证明关于高斯随机过程的假设是不必要的[32]。

作为随机逆(3.57)的应用的一个例子，看看广泛用于病态的超定方程组求解的所谓"沿脊回归"技术[34]，它实际上是取协方差矩阵 $C_m=\tau^2 I$ 的特殊情况，其中 τ^2 为待定参数。此时由(3.56)可知解估计误差的协方差矩阵为

$$C=\tau^2(R-I)(R-I)^{\mathrm{T}}+G^{-g}C_d(G^{-g})^{\mathrm{T}}$$

当 τ^2 很小时上式第一项减小，说明对分辨率不足引起误差加较小的权，而对第二项数据传播引起的误差加较大的权。因此，τ^2 实际上是在解估计的分辨率和方差之间起调节作用的参数，而所谓的沿脊回归技术只是随机逆的一个特例。

(四)最大的似然解估计

最大似然解估计也是随机逆应用的一个特例。如果不对解估计的方差作任何限制，即令 $C_m^{-1}=0$，则由(3.58)式可以得到一种取分辨率最高的解估计

$$\hat{m}=(G^{\mathrm{T}}C_d^{-1}G)^{-1}G^{\mathrm{T}}C_d^{-1}d \tag{3.61}$$

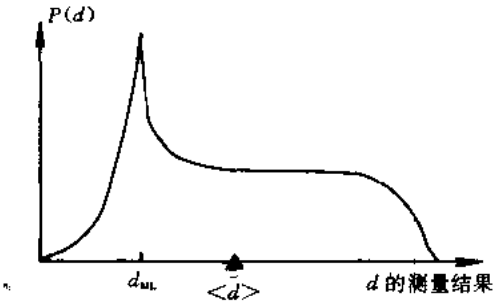


图 3.4 对数据 d 的概率分布 $P(d)$ 的最大似然点 d_{ML}，它是出现可能性最大的测量值，但不一定为均值 $\bar{d}$

其模型分辨率矩阵 $R=I$。式(3.61)只在矩阵 G 为非奇异时成立。实际上，(3.61)式给出的解估计就是用数据的协方差矩阵作归一化后由最小二乘法得到的解估计，即对

$$(d-Gm)^{\mathrm{T}}C_d^{-1}(d-Gm)$$

取极小的解估计。

把(3.61)式称为最大似然解估计的原因是它和用概率密度的条件极值导出的解相同，而且最大似然法是先出现的经典方法。回忆一下最大似然的思想，让我们来看看图3.4所示的随机变量 d 的概率分布 $P(d)$。如果 $P(d)$ 是一个较复杂的曲线，它的尖峰处称为最大似然点 d_{ML}，与处于重心平衡位置的均值 $\bar{d}$ 可能并不重合。

如果一个向量 m 的估计 $\hat{m}$ 是由数据 d 导出的，而 d 是带有误差的随机向量，而估计误差 $(m-\hat{m})$ 也是一个取决于 d 的随机向量。所谓的似然函数乃是给定 m 的条件下 d 出现的条件概率 $P(d/m)$，它取极大时对 m 的估计 $\hat{m}_{ML}$ 被称为最大似然估计。由于概率 $P>0$ 而且 $\lg P(d/m)$ 和 $P(d/m)$ 对相同的 m 取极大值，所以极大似然估计应满足以下方程

$$\left.\frac{\partial \lg P(d/m)}{\partial m^{\mathrm{T}}}\right|_{m=\hat{m}_{ML}}=0 \tag{3.62}$$

由于数据与模型的关系为 $Gm=d$，Gm 可视为随机向量 d 的数学期望，假定数据 d 符合多变量高斯分布，它的概率密度可写为

$$P(d)\propto\exp\left[-\frac{1}{2}(d-Gm)^{\mathrm{T}}C_d^{-1}(d-Gm)\right] \tag{3.63}$$

因此，在给定模型 m 的条件求 d 出现概率最大的解估计可由(3.63)式代入(3.62)式导出。

注意到(3.63)式指数项中的负号，代入(3.62)式后变为对

$$(\boldsymbol{d}-\boldsymbol{Gm})^{\mathrm{T}}\boldsymbol{C}_d^{-1}(\boldsymbol{d}-\boldsymbol{Gm})$$

取极小，从而导出(3.61)式。这就说明了最大似然解估计等效于用 $\boldsymbol{C}_d^{-1}$ 加权的最小二乘解估计。

(五)模型先验信息的利用[32]

在某些情况下我们可能对模型有了相当的了解。例如，在地震台站密布的地区，我们有可能已经用以前发生的地震对该区的地下构造进行过反演，得到了模型参数 $\boldsymbol{m}_0$，在积累了新的数据 $\boldsymbol{d}$ 以后，我们想用新的数据重新研究地下构造，以提高模型参数确定的精度和分辨率。把关于模型 $\boldsymbol{m}$ 的先验信息 $\boldsymbol{m}_0$ 表为

$$\boldsymbol{m}_0=\boldsymbol{I}\boldsymbol{m}+\Delta\boldsymbol{m} \tag{3.64}$$

其中 $\boldsymbol{I}$ 为单位矩阵、$\Delta\boldsymbol{m}$ 为猜测误差。对于新获得的数据，假设它满足以下矩阵方程式

$$\boldsymbol{Gm}=\boldsymbol{d}+\delta\boldsymbol{d} \tag{3.65}$$

其中 $\delta\boldsymbol{d}$ 表示观测误差向量。按照以下组合和分块矩阵的记法构造新的矩阵和向量

$$\boldsymbol{y}=\begin{bmatrix}\boldsymbol{d}\\ \cdots\\ \boldsymbol{m}_0\end{bmatrix}$$

$$\boldsymbol{A}=\begin{bmatrix}\boldsymbol{G}\\ --\\ \boldsymbol{I}\end{bmatrix}$$

和

$$\boldsymbol{e}=\begin{bmatrix}\delta\boldsymbol{d}\\ \cdots\\ \Delta\boldsymbol{m}\end{bmatrix}$$

则可以把以上两个方程(3.64)和(3.65)合并为一个新的方程

$$\boldsymbol{y}=\boldsymbol{Am}+\boldsymbol{e} \tag{3.66}$$

如果我们把数据和模型的协方差矩阵也按以下方式组合成一个新的协方差矩阵

$$\boldsymbol{C}=\begin{bmatrix}\boldsymbol{C}_d & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{C}_m\end{bmatrix}$$

则利用最大似然解估计的公式(3.61)可得

$$\hat{\boldsymbol{m}}_{ML}=(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{y}$$

现在把各个分块矩阵代入上式，有

$$\begin{aligned}\hat{\boldsymbol{m}}_{ML}&=\left\{\begin{bmatrix}\boldsymbol{G}^{\mathrm{T}} & \boldsymbol{I}\end{bmatrix}\begin{bmatrix}\boldsymbol{C}_d^{-1} & 0\\ \cdots & \cdots\\ 0 & \boldsymbol{C}_m^{-1}\end{bmatrix}\begin{bmatrix}\boldsymbol{G}\\ --\\ \boldsymbol{I}\end{bmatrix}\right\}^{-1}\begin{bmatrix}\boldsymbol{G}^{\mathrm{T}} & \boldsymbol{I}\end{bmatrix}\begin{bmatrix}\boldsymbol{C}_d^{-1} & 0\\ \cdots & \cdots\\ 0 & \boldsymbol{C}_m^{-1}\end{bmatrix}\begin{bmatrix}\boldsymbol{d}\\ \boldsymbol{m}_0\end{bmatrix}\\ &=\left\{\begin{bmatrix}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{C}_d^{-1} & \boldsymbol{C}_m^{-1}\end{bmatrix}\begin{bmatrix}\boldsymbol{G}\\ --\\ \boldsymbol{I}\end{bmatrix}\right\}^{-1}\begin{bmatrix}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{C}_d^{-1} & \boldsymbol{I}\boldsymbol{C}_m^{-1}\end{bmatrix}\begin{bmatrix}\boldsymbol{d}\\ \boldsymbol{m}_0\end{bmatrix}\\ &=(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{C}_d^{-1}\boldsymbol{G}+\boldsymbol{C}_m^{-1})^{-1}(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{C}_d^{-1}\boldsymbol{d}+\boldsymbol{I}\boldsymbol{C}_m^{-1}\boldsymbol{m}_0)\end{aligned} \tag{3.67}$$

于是，便得到了应用模型的先验估计来确定线性反问题的解估计的反演公式。

值得注意的是，在使用上述公式时要求关于模型的先验知识 $\boldsymbol{m}_0$ 与数据 $\boldsymbol{d}$ 线性无关，即它不能是由 $\boldsymbol{d}$ 直接或间接获得的。在用本章第一节讨论的非线性反问题线性化的方法用迭代过程求非线性反问题的近似解时，第$(k-1)$次迭代和第 k 次迭代所得到的近似解都是来自于同一组数据 $\boldsymbol{d}$，因此，不能把第$(k-1)$次迭代的近似解 $\boldsymbol{m}^{k-1}$ 作为先验知识 $\boldsymbol{m}_0$ 用上式求第

k 次的近似解。

四、利用自然逆的广义反演方法

上面我们虽然导出了许多种广义逆，但是它们在矩阵 G 奇异或者接近奇异时都不能直接在计算机上实现。因此在进行线性反演之前，常常有必要先研究矩阵 G 的秩及条件数，看看要求解的线性方程组 $Gm=d$ 是否属于混定的或欠定的方程组。在地球物理反演中，混定方程组是经常出现的。上一章曾讲过，对于矩阵的秩和条件数，我们还只能用奇异值分解这个工具，因此本节来讨论奇异值分解用于广义逆的情况。

（一）自然逆

由第二章第四节奇异值分解定理，$n\times J$ 实矩阵 G 可分解为 $n\times n$ 正交阵 U 和 $J\times J$ 正交阵 V 和一个对角阵 S：

$$G=USV^{\mathrm{T}}$$

取广义逆为

$$H_L=G^{-g}=(USV^{\mathrm{T}})^{-1}=VS^{-1}U^{\mathrm{T}} \tag{3.68}$$

这样定义的广义逆被称为自然逆，或 Lanczos 逆。相应的解估计为

$$\hat{m}=VS^{-1}U^{\mathrm{T}}d \tag{3.69}$$

显然，当 $n=J$ 而且 G 为非奇异时 U 和 V 都为 n 阶对称正交矩阵，其分辨率矩阵

$$P=GH_L=USV^{\mathrm{T}}VS^{-1}U^{\mathrm{T}}=I$$

$$R=H_LG=VS^{-1}U^{\mathrm{T}}USV^{\mathrm{T}}=I$$

由此可知方程 $Gm=d$ 有唯一解。

对于 $n>J$ 而且 $\mathrm{rank}(G)=J$ 时，广义逆矩阵

$$G^{+}=(G^{\mathrm{T}}G)^{-1}G^{\mathrm{T}}=(VS^{2}V^{\mathrm{T}})^{-1}VSU^{\mathrm{T}}$$

$$=VS^{-1}U^{\mathrm{T}}=H_L$$

因此 Lanczos 逆蜕化为由最小二乘定义的广义逆，相应的估计向量 $\hat{m}$ 即为最小二乘解。注意正定对称矩阵 $(G^{\mathrm{T}}G)$ 在奇异值分解之后得到的两个正交矩阵（左矩阵和右矩阵）是相同的。在一般情况下对应 Lanczos 逆的分辨率矩阵为

$$R=H_LG=VV^{\mathrm{T}} \tag{3.70}$$

由(3.44)式可知其分辨核为

$$r_k=\sum_{j=1}^{J}(V_{kj}V_{jk}-\delta_{kj})^2 \tag{3.71}$$

类似地由定义不难得到对应 Lanczos 逆的信息密度矩阵为

$$P=GH_L=UU^{\mathrm{T}} \tag{3.72}$$

现在，我们要来讨论以下两种情况下自然逆的应用：(a)当系数矩阵 G 为奇异时 Lanczos 逆不存在和(b)当数据带误差时解的估定向量 $\hat{m}$ 有方差。

当 G 奇异时，设 $\mathrm{rank}(G)=r<J$。根据奇异值的性质可知，此时最小的 $(n-r)$ 个奇异值等于零：

$$G=U\begin{bmatrix}\alpha_1 & & & & & \\ & \alpha_2 & & & & \\ & & \ddots & & & \\ & & & \alpha_r & & \\ & & & & 0 & \\ & & & & & \ddots \\ & & & & & & 0 \\ & & & & & & & 0 \\ & & & & & & & & \ddots \\ & & & & & & & & & 0\end{bmatrix}V^{\mathrm{T}}$$

由于 $\alpha_J=0$，因此条件数 $\mathrm{Cond}(G)=\infty$，因而无法直接由自然逆求得解估计，而要用到本节后面将要讨论的修正方法。

对于欠定方程组($n<J$)如果我们利用模型 m 的 L_2模取极小的准则求它的广义解，则有最小长度解估计

$$\hat{m}=G^{\mathrm{T}}[GG^{\mathrm{T}}]^{-1}d \tag{3.29}$$

代入奇异值分解公式有

$$\hat{m}=VSU^{\mathrm{T}}[US^2U^{\mathrm{T}}]^{-1}d=VS^{-1}U^{\mathrm{T}}d$$

说明自然逆也包括了最小长度解估计的广义逆。要注意的是在 $n<J$ 的情况下此时矩阵 V 第 n 列以后的元素在求 $\hat{m}$ 时实际并不起作用，即有

$$VS^{-1}=\begin{bmatrix}V_1 & V_2\end{bmatrix}\begin{bmatrix}\alpha_1^{-1} & & \\ & \ddots & \\ & & \alpha_n^{-1} \\ & O & \end{bmatrix}=\begin{bmatrix}V_1 & O\end{bmatrix}S^{-1}=V_nS^{-1}$$

最小长度解估计实际上为

$$\hat{m}=V_nS^{-1}U^{\mathrm{T}}d \tag{3.73}$$

(二)Wiggins 方法[20]

现在来研究混定方程组的情况，这时虽然有 $n\geqslant J$ 和 $\mathrm{rank}(G)=J$，但一些奇异值非常小，即

$$S=\mathrm{diag}(\alpha_1,\cdots,\alpha_q,\alpha_{q+1},\cdots\alpha_J)$$

其中 $\alpha_{q+1},\cdots,\alpha_J$ 很小。因此，在 S^{-1} 中 $\alpha_{q+1}^{-1},\cdots,\alpha_J^{-1}$ 很大，由(3.46)式可知，模型的协方差矩阵可写为

$$C_m=H_LC_dH_L^{\mathrm{T}}=VS^{-1}U^{\mathrm{T}}C_dUS^{-1}V^{\mathrm{T}}$$

对于高斯分布的数据 $C_d=\sigma_d^2I$，C_m 的对角线元素即为估计模型参数的方差

$$\mathrm{Var}(\hat{m}_k)=\sum_{i=1}^{J}\left(\frac{V_{ki}}{\alpha_i}\right)^2\sigma_d^2 \tag{3.74}$$

由上式可见，当 $\alpha_{q+1},\cdots\cdots,\alpha_J\to 0$时方差 $\mathrm{Var}(\hat{m}_k)\to\infty$。为了抑制方差，必须采取其它措施。

在第二章的第四节讨论 SVD 的性质时我们给出了最接近矩阵的定义，即如果 B 是 G

的最接近矩阵，则有

$$\|G-B\|_{2,F}=\min$$

其中 $B=US_eV^T$，而

$$S_e^{-1}=\mathrm{diag}(\alpha_1^{-1},\cdots,\alpha_q^{-1},0,\cdots,0)$$

既然 B 是 G 的最接近矩阵，取近似

$$d=Gm\simeq Bm=US_eV^Tm \tag{3.75}$$

注意到 m 的维数为 J，而 (US_eV^T) 的秩为 q，(3.75)变为欠定方程组。在条件(3.75)之下求的最小长度 $\|m\|_2^2$ 的解估计，可得

$$\hat{m}=VS_e^{-1}U^Td \tag{3.76}$$

其中

$$S_e^{-1}=\mathrm{diag}(\alpha_1^{-1},\cdots,\alpha_q^{-1},0,\cdots,0)$$

对于混定方程组的上述解估计最早是由 Wiggins(1972)年用于地球物理反演的[20]，因此被称为 Wiggins 法。他建议在反问题求解之前设定一个解估计方差的门限值 t_k，由满足

$$\sum_{i=1}^{q}\left(\frac{V_{ki}}{\alpha_i}\right)^2<t_k$$

的方法选取保留的奇异值的个数 q，q 称之为数据的有效自由度(或矩阵 G 的有效秩)。由(3.74)知 Wiggins 法求得的解估计的方差减少为

$$\mathrm{Var}(\hat{m}_k)=\sum_{i=1}^{q}(V_{ki}/\alpha_i)^2\mathrm{Var}(d_i) \tag{3.77}$$

但是，这时模型的分辨率也随之降低，因为(3.75)式已退变为欠定情况。利用(3.73)式中 V_e 的表示及(3.70)式可知，Wiggins 法的数据分辨率矩阵为

$$R_e=V_eV_e^T=\begin{bmatrix}V_1V_1^T & O\\ O & O\end{bmatrix}$$

与不作修正的自然逆解估计(3.69)相比，显然有

$$\mathrm{spread}(R_e)=\|R_e-I\|_2^2>\|R-I\|_2^2 \tag{3.78}$$

说明 R_e 的展布增大，即 Wiggins 法是以牺牲分辨率改善方差的一种折衷方法。实际上，选取 q 越小 R_e 的展布越大，同时方差减小越多(见图3.5)。实际上，令 q 从小到大逐渐变化，然后用(3.78)和(3.77)计算展布 $\mathrm{spread}(R_e)$ 和 $\mathrm{Var}(\hat{m}_k)=\sigma_m^2$，就可以作出如图3.6的最佳折衷曲线。最佳折衷的 q 值应位于最接近原点的左下方，因为原点处展布和方差都为零。

(三)应用 SVD 的随机逆

在第三节我们已经讨论了最小方差解估计和随机逆，即

$$G^{-g}=(G^TC_d^{-1}G+C_m^{-1})^{-1}G^TC_d^{-1} \tag{3.57}$$

同时，我们也讨论了取模型的协方差矩阵 $C_m=\tau^2I$ 的沿脊回归方法，其中 τ^2 是一种在解估计的方差的分辨率之间取折衷的参数。对于高斯分布的情况如果取

$$\tau=\sigma_m/\sigma_d \tag{3.79}$$

并令 $C_m=\sigma_m^2I$ 和 $C_d=\sigma_d^2I$，代入(3.57)式

有

$$\begin{aligned}G^{-g}&=[\sigma_d^{-2}G^TG+\sigma_m^{-2}I]^{-1}G^T\sigma_d^{-2}\\&=[G^TG+\tau^2I]^{-1}G^T\end{aligned} \tag{3.80}$$

于是，对超定的方程组，广义逆实际上和使预测误差和模型长度取极小的阻尼最小二乘解估

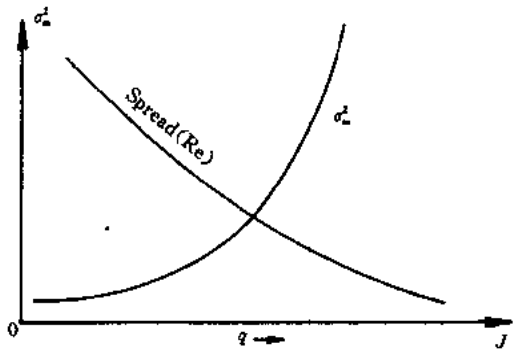


图 3.5　分辨率和方差随 q 变化的示意图

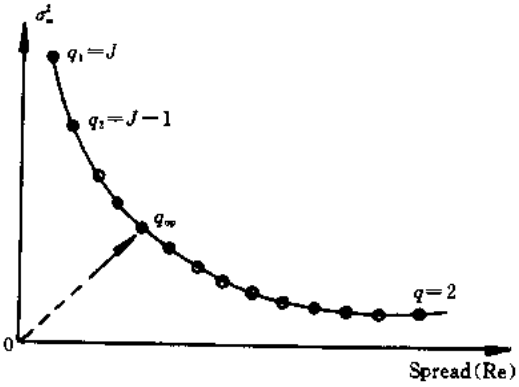


图3.6　在分辨率和方差之间取折衷曲线，最佳折衷参数 q_{op} 位于接近原点的左下方

计(3.24)相当。换句话说，折衷参数 τ^2 相当于阻尼最小二乘法中的阻尼系数 ε^2，只不过在阻尼最小二乘法中 ε^2 没有说明它的数学含义，在随机逆中 ε 是比值 σ_d/σ_m。

对于混定及欠定方程组 $\boldsymbol{Gm}=\boldsymbol{d}$ 的情况，由于矩阵 $\boldsymbol{G}$ 奇异或接近奇异，我们要用奇异值分解求这时的广义逆。应用第二章第四节讨论过的奇异值分解定理，这时由(3.80)可把广义逆写为

$$\begin{aligned}\boldsymbol{H}_L&=[\boldsymbol{VS}^2\boldsymbol{V}^{\mathrm{T}}+\tau^2\boldsymbol{I}]^{-1}\boldsymbol{VSU}^{\mathrm{T}}\\&=[\boldsymbol{V}(\boldsymbol{S}^2+\tau^2\boldsymbol{I})\boldsymbol{V}^{\mathrm{T}}]^{-1}\boldsymbol{VSU}^{\mathrm{T}}\\&=\boldsymbol{V}(\boldsymbol{S}^2+\tau^2\boldsymbol{I})^{-1}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{VSU}^{\mathrm{T}}\\&=\boldsymbol{V}(\boldsymbol{S}^2+\tau^2\boldsymbol{I})^{-1}(\boldsymbol{S}^{-1})^{-1}\boldsymbol{U}^{\mathrm{T}}\\&=\boldsymbol{V}[\boldsymbol{S}^{-1}(\boldsymbol{S}^2+\tau^2\boldsymbol{I})]^{-1}\boldsymbol{U}^{\mathrm{T}}\\&=\boldsymbol{V}[\boldsymbol{S}+\tau^2\boldsymbol{S}^{-1}]^{-1}\boldsymbol{U}^{\mathrm{T}}\\&=\boldsymbol{VS}_\tau^{-1}\boldsymbol{U}^{\mathrm{T}}\end{aligned}\tag{3.81}$$

其中

$$\boldsymbol{S}_\tau=\boldsymbol{S}+\tau^2\boldsymbol{S}^{-1}=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_J,0,\cdots,0)$$

$$\lambda_j=\alpha_j+\tau^2/\alpha_j,\ j=1,2,\cdots,J\leqslant n$$

表示对奇异值的修改。因此，用 SVD 表示的随机逆的解估计为

$$\hat{\boldsymbol{m}}=\boldsymbol{VS}_\tau^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{d},\boldsymbol{S}_\tau^{-1}=\boldsymbol{S}(\boldsymbol{S}^2+\tau^2\boldsymbol{I})^{-1}\tag{3.82}$$

分辨率矩阵为

$$\boldsymbol{R}_\tau=\boldsymbol{H}_L\boldsymbol{H}=\boldsymbol{VS}_\tau^{-1}\boldsymbol{SV}^{\mathrm{T}}=\boldsymbol{VS}^2(\boldsymbol{S}^2+\tau^2\boldsymbol{I})^{-1}\boldsymbol{V}^{\mathrm{T}}\tag{3.83}$$

注意到矩阵 $[\boldsymbol{S}_\tau^{-1}\boldsymbol{S}]$ 的对角元为

$$\left(\frac{\alpha_j}{\lambda_j}\right)=\left(\frac{\alpha_j^2}{\alpha_j^2+\tau^2}\right)\leqslant 1$$

可见与 Lanczos 自然逆对应的分辨率矩阵(3.70)相比，随机逆的分辨率矩阵(3.83)更加偏离于单位矩阵。因此在 $\tau^2>0$ 的情况下，τ^2 越大解估计 $\hat{\boldsymbol{m}}$ 的分辨率越低。实际上，这时 Direchlet 展布为

$$\text{spread}(\boldsymbol{R}_r)=\sum_{i=1}^{J}\sum_{j=1}^{J}\left(\frac{\alpha_j^2 V_{ij}V_{ji}}{\alpha_j^2+\tau^2}-\delta_{ij}\right)^2 \tag{3.84}$$

对于高斯分布的数据，应用 SVD 之后随机逆解估计的协方差矩阵为

$$\boldsymbol{C}_m=\boldsymbol{H}_L\boldsymbol{C}_d\boldsymbol{H}_L^{\mathrm{T}}=\boldsymbol{V}\boldsymbol{S}_\tau^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{C}_d\boldsymbol{U}\boldsymbol{S}_\tau^{-1}\boldsymbol{V}^{\mathrm{T}}=\sigma_d^2\boldsymbol{V}\boldsymbol{S}_\tau^{-2}\boldsymbol{V}^{\mathrm{T}}$$

因此解估计的方差为

$$\text{Var}(\tilde{m}_k)=\sum_{j=1}^{J}\left(\frac{\alpha_j V_{kj}}{\alpha_j^2+\tau^2}\right)^2\sigma_d^2 \tag{3.85}$$

为了选取在分辨率和方差之间取得最佳折衷的解估计，取参数 τ^2 为$10^{-8}\sim10^{-1}$之间的一系列数值，便可以由(3.84)和(3.85)式作出类似于图3.6的折衷曲线。此时 τ^2 起到图3.6中的 q 的倒数的作用。然后，可按离原点最近的原则选取最佳折衷参数 τ^2。通过以上作折衷曲线的过程，我们实际上评价了一系列解估计，相当于作算子谱展开。算子谱展开的作用是以直观的形式把模型参数中可以由数据准确确定的成分和其它统计上不可靠的成分区分开。具体地说，在进行奇异值分解之后，相对大的奇异值及相应的左右特征向量表示模型参数中比较肯定的和可靠的成分，而很小的奇异值和相应的特征向量则表示统计上不可靠的成分，它们在有限位的计算中会造成数值运算过程中严重的不稳定性。Wiggins 方法在处理奇异值时把小的奇异值及对应的特征向量删除，实际上是甩掉模型参数中的不可靠成分，以此减少解估计的方差。但是这样做同时也可能严重地损害了解估计的分辨率。当奇异值曲线呈阶梯形时(见图3.7a)，Wiggins 方法是比较适用的，因为这时确定成分和不可靠成分差别很大，甩掉不确定成分对解估计分辨率的影响不大。当奇异值缓慢下降时不好确定哪些是不确定成分，用随机逆求解混定方程组便不存在 Wiggins 法的困难(图3.7b 的情况)。此外，由折衷曲线的渐近线还能了解对于给定问题能达到的最高分辨率及最小方差是多少。这样地球物理反演就不仅仅是考虑某一个解估计，而是对所有可能的解估计作评价。这一套技术乃是 Backus-Gilbert 方法在离散模型的情况下的变种。

(四)奇异值的修改

从上面两节可以看出，无论是 Wiggins 方法还是随机逆，它们与 Lanczos 自然逆的区别都在于对奇异值的修改。由于把矩阵 $\boldsymbol{G}$ 作奇异值分解的过程也就是把线性算子对角化的过程，我们在上一章已经知道，矩阵 $\boldsymbol{G}$ 的奇异值的平方就是矩阵($\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}$)的本征值，而本征值代表了线性算子($\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}$)的谱。因此，奇异值分解是有限维线性空间算子的谱展开的一种形式(详见第四章)。用 SVD 求解混定方程组时，就常常出现阶梯状的奇异值曲线。

在地球物理中还有一种常见情况，此时奇异值平缓单调地减少(图3.7b)，在算子 $\boldsymbol{G}$ 接近奇异时常常出现这种情况，由此引起的不稳定性叫解析上的不稳定性。对于这种情形，用 Wiggins 方法把小于方差门限 t_k 的奇异值甩掉就会引起分辨率明显地变坏，因为保留的和甩掉的成分差别不明显。由于随机逆法并不甩掉任何奇异值，而是对所有奇异值进行连续的修改(修改的幅度取决于折衷参数 τ^2的选择)，因此，可以预期这时用随机逆法将优于 Wiggins 法。详见本章第六节例子。

随机逆法虽然保留了全部奇异值而仅仅对统计上不可靠的成分进行抑制，但它不仅修改了较小的奇异值，对相对较大的奇异值也进行了修改，这就使得模型中的确定成分也受到了畸变，同时也降低了分辨率。注意到在 Wiggins 方法相对大的奇异值是不被修改的，可知在处理大的奇异值方面，Wiggins 方法是具有优点的。作者曾注意到了随机逆和 Wiggins 方

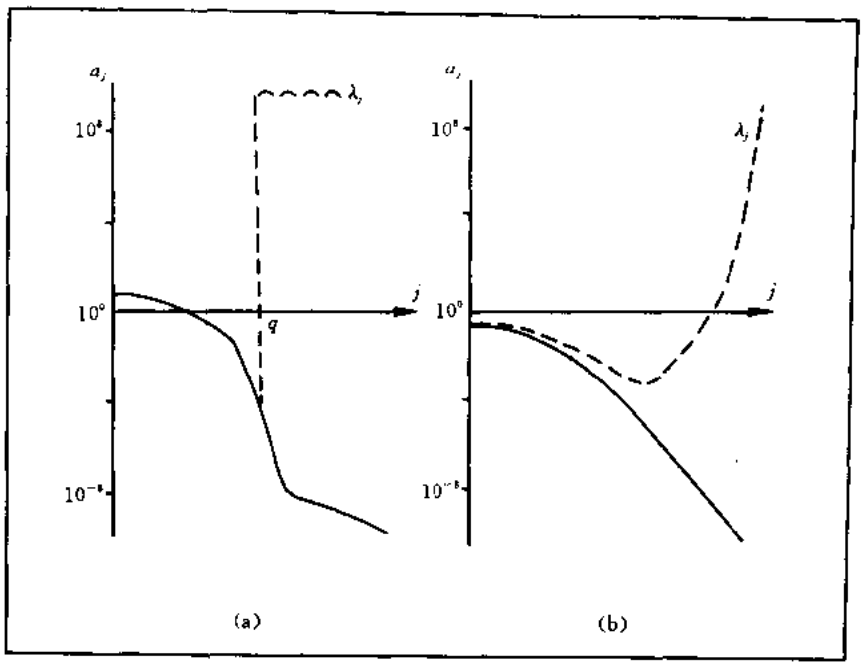


图 3.7　两种不同类型的奇异值曲线及其修改
(a)阶梯型；(b)缓慢下降型

法各具有的优缺点，并找到了吸取两者之长的奇异值修改方法，即对大的奇异值尽量趋近 Wiggins 的准则，而对小的奇异值则趋近于随机逆对奇异值的修改。具体地说，设 t 为 Wiggins 方法中的保留的最小奇异值的门限，q 为对应小于 t 的奇异值的个数(有效自由度)，用以下公式修改奇异值：

$$\left.\begin{aligned} \alpha'_k &= t^2/\alpha_k && k \geqslant q \\ \alpha'_k &= \alpha_k + \frac{k}{q}\left(\frac{t^2}{\alpha_{q+1}} - \alpha_q\right) && k < q \end{aligned}\right\} \tag{3.86}$$

相应地修改后的奇异值矩阵为

$$\boldsymbol{S}_r = \mathrm{diag}(\alpha'_1, \alpha'_2, \cdots, \alpha'_j, 0, \cdots, 0)$$

解估计仍用(3.82)式计算。利用这种奇异值分解取得解估计，其分辨率矩阵仍然为

$$\boldsymbol{R} = \boldsymbol{V}\boldsymbol{S}_r^{-1}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}$$

单位方差为

$$\mathrm{Var}(\hat{m}_k) = \sum_{j=1}^{J}\left(\frac{V_{kj}}{\alpha'_j}\right)^2$$

由于此法兼取了 Wiggins 方法和随机逆的优点，解估计的质量可以提高(见本章第六节)。在理论上这种改进的奇异值修改方法相应于模型 $\boldsymbol{m}$ 的协方差矩阵 $\boldsymbol{C}_m$ 为对角阵，但对角元素可以不相符，即不要求 $\boldsymbol{C}_m = \sigma^2\boldsymbol{I}$ 的高斯分布情况。

五、线性规划和联合线性反演

在以上三节中我们讨论了以各种取 L_2 模为极小的线性反问题的解估计，如同第二节所

述，其中暗示数据的统计特征服从高斯分布。本节将讨论 L_2模极小化以外的其它线性反演方法，以及不同类型的地球物理数据的联合线性反演问题，以便在选择反演方法时互相对比和借鉴。

(一)L_1模的极小化

假定一个实际地球物理数据 d 在统计上是属于双边指数分布的随机变量，其数学期望为 $\bar{d}$，则它的概率分布可写为

$$P(d)=\frac{1}{2\lambda}\exp\left(-\frac{|d-\bar{d}|}{\lambda}\right) \tag{3.87}$$

为了便于与高斯分布的概率作比较，指数分布中的偏差参数可设为 $\lambda=\sigma/\sqrt{2}$，其中 σ 为高斯分布的标准差。于是，(3.87)式可写为

$$P(d)=\frac{1}{\sqrt{2}\sigma}\exp\left(-\frac{\sqrt{2}|d-\bar{d}|}{\sigma}\right) \tag{3.88}$$

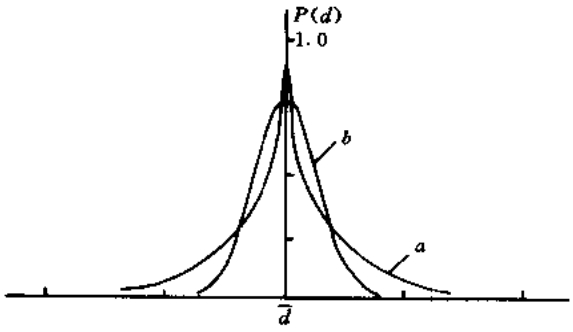


图 3.8 指数分布(a)和高斯分布(b)概率曲线的比较，说明指数分布具有较长的尾巴

比较(3.88)和(3.21)式在相同的 $\bar{d}$ 和 σ 的条件下得到如图3.8所示的指数分布曲线和高斯分布曲线。由图3.8可见，指数分布的概率曲线比较尖，但在远离均值 $\bar{d}$ 的情况下数据出现的概率比高斯分布大，即具有较长的“尾巴”。说明指数分布的数据容易出现如图3.2所示的个别数据偏差很大的情况。这时，我们希望反演算法要能对付数据集中包含的极少数方差很大的“坏”数据，即具有较大的韧性。

正如 L_2模与高斯分布有关一样，指数分布与 L_1模有类似的关系。为了导出这种关系，考虑 n 个不相关的数据的联合概率分布，它是 n 个独立的指数分布(3.88)的乘积

$$P(\boldsymbol{d})=2^{-\frac{n}{2}}\sigma^{-n}\exp\left(-\frac{\sqrt{2}}{\sigma}\sum_{i=1}^{n}|d_i-\bar{d}|\right) \tag{3.89}$$

其中假定每一个数据的均值都为 $\bar{d}$，方差都为 σ^2。采用最大似然的思想对 $P(\boldsymbol{d})$取极大，就要求对指数中的求和项取极小，即令

$$E=\sum_{i=1}^{n}|d_i-\bar{d}|=\|\boldsymbol{e}\|_1$$

显然，E 是预测误差向量 $\boldsymbol{e}$ 的 L_1模。

现在来考虑线性反问题 $\boldsymbol{Gm}=\boldsymbol{d}$。假定数据和模型参数在统计上不相关，而且我们知道了它们的方差 σ_d^2 和 σ_m^2，则由(3.88)式在给定 $\boldsymbol{m}$ 时 $\boldsymbol{d}$ 出现的指数型联合条件概率可写为[16]

$$P(\boldsymbol{d}/\boldsymbol{m})=2^{-(n+J)/2}\prod_{i=1}^{n}\sigma_{di}^{-1}\prod_{j=1}^{J}\sigma_{mj}^{-1}$$
$$\times\exp\left[-\frac{1}{\sqrt{2}}\sum_{i=1}^{n}|e_i|\sigma_{di}^{-1}-\frac{1}{\sqrt{2}}\sum_{j=1}^{J}|l_j|\sigma_{mj}^{-1}\right] \tag{3.90}$$

式中 $e=d-Gm$ 为预测误差向量，而 $l=m-\bar{m}$ 为解误差的长度，在这里 $\bar{m}$ 表示解的数学期望。用最大似然的原理确定模型参数时要将上式中的指数项极小化，即令

$$E+L=\sum_{i=1}^{n}|e_i|\sigma_{di}^{-1}+\sum_{j=1}^{J}|l_i|\sigma_{mj}^{-1}=\min \tag{3.91}$$

上式中第一项为用 σ_{di}^{-1} 加权的预测误差的 L_1 模，而第二项是用 σ_{mj}^{-1} 加权的模型解误差的长度的 L_1 模，权系数都是标准差的倒数，而不是像高斯分布时方差的倒数。因此，L_1 模的线性问题的解法并不像用高阶模的解法那样，对方差大的坏数据和模型参数加较大的权。

(二)L_1模反问题和线性规划

我们来讨论 L_1 模反问题如何转化为线性规划(Linear programming)问题[1]。线性规划问题是首先在经济和企业管理中发展起来并已经被深入地研究过的问题，有种 L_1 成熟的解法[35,1]。因此，只要实现了这种转化，L_1 模线性反问题便易于借用机器中常配有的线性规划的软件求出反问题的解估计。

线性规划问题指的是，在约束

$$\boldsymbol{A}\boldsymbol{x}\begin{Bmatrix}\geqslant\\=\\\leqslant\end{Bmatrix}\boldsymbol{b}\text{ 及 }\boldsymbol{x}\geqslant\boldsymbol{0} \tag{3.92}$$

的条件下求使得 $\phi=\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}$ 取极小的解估计 $\boldsymbol{x}$。在管理学中，ϕ 表示一个生产多种产品的工厂可实现的总利润，而每一种产品的产量用 $\boldsymbol{x}$ 表示，而它的利润用 $\boldsymbol{c}$ 表示。约束条件表示劳动力及机械设备能力的限制，以及产量只能是正值。

在地球物理反问题中，如果它被归纳为欠定的线性方程组 $\boldsymbol{Gm}=\boldsymbol{d}$，而模型本身是一个变量 u 的函数 $m(u)$，则在连续情况模型的 L_1 模可表为

$$L=\|m\|_1=\int_a^b|m(u)|\mathrm{d}u$$

例如 m 代表层速度或电导率，而 u 表示深度 z 等等，(a,b) 表示 z 的变化区间。把区间 (a,b) 离散化(即分层)，分成的层面深度为 $u_0,u_1,\cdots,u_J$，其中 $a=u_0,b=u_J$，则模型参数 $m_j,j=1,\cdots,J$，表示在层面 u_{j-1} 和 u_j 之间的层速度(或电导率)，因此模型的离散 L_1 模表为

$$L=\sum_{j=1}^{J}|m_j|r_j \tag{3.93}$$

其中 $r_j=u_j-u_{j-1}$。在约束条件

$$\sum_{j=1}^{J}G_{ij}m_j=d_i,i=1,2,\cdots,n \tag{3.94}$$

之下求使 L 极小或负 L 极大的模型参数 $\boldsymbol{m}^{\mathrm{T}}=(m_1,\cdots,m_J)$ 的问题便化为线性规划问题(3.92)。

如果观测数据是带误差的，而且误差 δd_i 可以估计，则约束条件(3.94)又可以化为不等式的形式：

$$\left.\begin{aligned}\sum_{j=1}^{J}G_{ij}m_j&\leqslant d_i+\delta d_i\\-\sum_{j=1}^{J}G_{ij}m_j&\leqslant\delta d_i-d_i\end{aligned}\right\} \tag{3.95}$$

在用线性规划方法求 L_1 模及问题的解估计时还必须注意以下技术问题。

(c)权函数。在(3.93)中，不同的$|m_j|$可能相差很大，绝对值小的$|m_j|$或γ_j对目标函数的贡献很小，对它们的估定值的分辨率相应较低，因此可以给它加较大的权，使它在目标函数中占有一定的位置。例如，地震道反演的目标往往是某些与油气有关的小薄层，薄层模型参数(如反射系数或波阻抗差异)往往较小，厚度γ_j也很小。作线性规划时必须加较大的权，使它们在目标函数最小化的过程中起到一定分量的作用。一般来说，设权函数离散化为序列$\{W_j\}$，$j=1,\cdots,J$，则我们可用下式代替(3.93)作最小化：

$$L=\sum_{j=1}^{J} W_j\gamma_j|m_j| \tag{3.96}$$

如参照(3.90)式用标准差加权时，可取$W_j=\sigma_{m_j}^{-1}$。

(b)在用单纯形法解线性规划问题时要求m_j全取正值。当模型参数m_j在物理上可正可负时，可以把它设为两个正变量之差：

$$m_j=p_j-q_j，其中\ p_j\geqslant 0,q_j\geqslant 0 \tag{3.97}$$

相应地，目标函数变为

$$L=\sum_{j-1}^{J} W_j\gamma_j(p_j+q_j) \tag{3.98}$$

相应约束条件也由(3.97)代入(3.94)得到，用线性规划法求出两个向量$\boldsymbol{p}$和$\boldsymbol{q}$之后再由(3.97)得到模型$\boldsymbol{m}$。当然，这时计算工作量要增加数倍。

(c)除了观测数据之外，在用线性规划法求解时加入新的约束条件是很方便的。例如，我们知道地球物质密度一般在0～10g/cm³之间，把类似的模型的上下界作为新的约束，常常可以提高解估计的质量。这种约束的形式为

$$\left.\begin{aligned} m_j&\leqslant m_j^{+}\\ -m_j&\leqslant -m_j^{-}\end{aligned}\right\}j=1,2,\cdots,J \tag{3.99}$$

式中m^{-}和m^{-}表示上界和下界，把(3.99)加到(3.94)或(3.95)的约束之中，可以增加$2J$个约束不等式。除了(3.99)形式的约束之外，还可以有以下线性组合形式的约束(如用平均值作约束)。

$$\xi_0\leqslant\sum_{j-1}^{J}\beta_j m_j\leqslant\xi_1 \tag{3.100}$$

线性规划算法的优点之一就是约束条件的加入十分方便。把人们对地球或地下构造的先验知识以各种约束条件表示出来，然后应用线性规划反演可以提高解估计的质量。如前所述，L_1模反演的另一个优点是具有较大的韧性，即少数坏数据的存在对解估计质量的影响较小。

现在，来简单地讨论一下线性规划的算法问题。从几何的角度看，(3.94)式中的每一个等约束条件都表示了n维空间中的一个超平面，因此(3.95)中$2n$个不等式约束表示了n维空间中的一个凸的超多边形，而满足(3.95)约束条件的解$\boldsymbol{m}$作为n维空间的点必定位于此凸多边形上或体内。然而，在(3.93)中规定$\boldsymbol{m}$的目标函数L必须取极小(或负L取极大)，说明$\boldsymbol{m}$必定位于此凸多边形的顶点上。用(3.93)式计算每一个顶点上相应$\boldsymbol{m}$的目标函数，并挑出其中L最小的一个顶点，它就是线性规划问题的最佳可行解估计。由此可见，线性规划算法的步骤可分为三步。首先由约束条件解出凸多边形的顶点，然后，计算每个顶点上的目标函数，最后找出目标函数取极大或极小顶点。如果目标函数取极值的顶点不止一个，则说

明线性规划问题没有唯一解，相连这些顶点之间边上的所有点都是问题的最佳可行解。但是，如果约束条件相互矛盾时，它在 n 维空间对应的多边形便不是凸的或者不封闭，此时用线性规划算法得不到最佳可行解。

上述分三步的算法是可以用来求解的，然而当约束很多时计算极其缓慢，因为顶点数目很多，因此要避免计算所有的顶点。单纯形法(Simplex Method)就是其中最常用的一种。

单纯形法的计算可分为以下四步：

1. 任选一个角点作始点，每个角点都与 n 条边相连。

2. 在这个角点找一条边，沿这条边上目标函数 $L(\boldsymbol{m})$ 应该逐渐减小(对于求极大的问题目标函数沿此边增大)。如果找不到这样的一条边，则在这个角点目标函数已经达到最小，此角点即为最优解。反之则做第三步。

3. 沿着找出的边找下一角点，然后重复第二步中的找边过程。此时，我们不可能返回原先的角点上去，因为目标函数在每一步都是减小的。

4. 在重复2～3步时，如果找不到一个角点沿其连接的边目标函数还可减少，则此角点 $\boldsymbol{m}$ 即为线性规划问题的解。

在具体计算时，单纯形法对(3.95)式中的每一个约束定义一个松驰变量 $\omega_j \geqslant 0$，$j=1$，…，$2J$，使这个不等式变为等式。例如，松驰变量 ω_1 把

$$G_{11}m_1+G_{12}m_2+\cdots+G_{1n}m_n \leqslant b_1$$

变为

$$a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n+\omega_1=b, \omega_1 \geqslant 0$$

而松驰变量本身也是未知数。于是，由(3.93)和(3.95)的上述转化就组成了求解 $(n+2J)$ 个方程组成的新方程

$$\boldsymbol{Ax}=\boldsymbol{b}$$

其中 $\boldsymbol{A}$ 为 $2J\times(n+2J)$，$\boldsymbol{x}$ 为 $n+2J$，包含了 m_j 和 ω_j。线性规划问题就变成求解这个欠定方程组。

(三)线性规划方法的应用

由于线性规划方法韧性好而且便于约束条件的加入，在地球物理反演中有许多应用。理论方面的探讨可参考法国科学家 Sabatier 的一篇颇有影响的文章[36]在重磁解释方面的应用可参考文献[37]。下面来介绍两个关于大地电磁及地震方面应用的例子。

第一个例子是王家映提出的对大地电磁测深结果作类似于反射地震的解释，即求出水平层状介质的电磁反射系数[3]。在大地电磁法中观测地面不同频率和方向的电场和磁场，它的比值

$$Z(\omega)=\boldsymbol{E}(\omega)/\boldsymbol{H}(\omega)$$

称为大地电磁阻抗，它可展开为

$$Z(\omega)=Z_0(\omega)\left[1+2\sum_{n=1}^{\infty} q_n \mathrm{e}^{-2\mathrm{i}k_n h_n}\right] \tag{3.101}$$

其中左边方括号内为传输函数，而 $Z_0(\omega)=\sqrt{\mathrm{i}\omega\mu\rho_1}$，是整个空间为表层电阻率 ρ_1 时大地电磁之阻抗，ω 为任意频率；$k_n=\sqrt{-\mathrm{i}\omega\mu\sigma_n}$，是第 n 个微层之复波数。σ_n 为该层之电导，h_n 是第 n 个微层之厚度；q_n 为电磁反射函数在不考虑吸收时之离散采样值；μ 是导磁系数，取 $4\pi\times$

10^{-7}为常数。

设

$$R(\omega)=\left(\frac{Z(\omega)}{Z_0(\omega)}-1\right)\Big/2$$

$$=\sum_{n=1}^{\infty}q_n e^{-2nk_n h_n} \tag{3.102}$$

虽然 q_n 是一个无限收敛的时间序列，在取 N 较大时，可对(3.102)进行截断。把观测频带内所有频率离散化后有

$$R(\omega_j)=\sum_{n=1}^{N}q_n e^{-2nk_{jn}h_n} \tag{3.103}$$

上式写成矩阵方程为

$$\boldsymbol{r}=\boldsymbol{P}\boldsymbol{q} \tag{3.104}$$

其中，数据为对不同频率 ω_j 由阻抗计算来的 $R(\omega_j)$。地球模型为 N 层的电磁反射系数，而系数矩阵为

$$\boldsymbol{P}=\begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1N} \\ P_{21} & P_{22} & \cdots & P_{2N} \\ \vdots & & & \\ P_{M1} & P_{M2} & \cdots & P_{MN} \end{bmatrix}$$

其中 $j=1,2,\cdots M$，

$$P_{jn}=e^{-2nk_{jn}h_n}$$

而 $k_{jn}=\sqrt{-i\omega_j\mu\sigma_n}$，$\{\omega_j\}$为频率 ω 的离散集。显然，(3.104)是一个线性方程组。$M<N$ 时欠定。用线性规划方法时设目标函数为电磁反射系数模型的 L_1模

$$\phi=\sum_{n=1}^{N}|q_n|$$

由(3.103)式可写出约束条件

$$R(\omega_j)+\delta R_j\geqslant\sum_{n=1}^{N}q_n e^{-2nk_{jn}h_n}$$

$$R(\omega_j)-\delta R_j\leqslant\sum_{n=1}^{N}q_n e^{-2nk_{jn}h_n}$$

这里 δR_j 为 $R(\omega_j)$的误差。由于 q_n 可正可负，由(3.97)可知，还须设

$$q_n=X_n-Y_n$$

其中

$$X_n,Y_n\geqslant 0$$

相应地，电磁反射系数序列的反问题变为在约束条件

$$\left.\begin{aligned} R(\omega_j)+\delta R_j&\geqslant\sum_{n=1}^{N}X_nP_{jn}-\sum_{n=1}^{N}Y_nP_{jn} \\ R(\omega_j)-\delta R_j&\leqslant\sum_{n=1}^{N}X_nP_{jn}-\sum_{n=1}^{N}Y_nP_{jn} \end{aligned}\right\} \tag{3.105}$$

下求目标函数

$$\phi=\sum_{n=1}^{N}(X_n-Y_n) \tag{3.106}$$

的极小。在此例中线性规划方法的应用效果较好，详见文献[3]。

第二个例子为反射地震的线性规划反演[4]。对于水平层状完全弹性介质，叠后反射地震道 $S(t)$ 可写为零相位子波 $W(t)$ 和反射系列 $R(t)$ 的褶积

$$S(t)=W(t)*R(t)$$

上式离散化为

$$\sum_{j=1}^{J} W_j R_{i-j}=S_i, i=1,2,\cdots,N \tag{3.107}$$

其中 N 为地震道数据总数，J 为反射界面的数目。应用线性规划反演反射序列时把(3.107)式作为约束并求 $\boldsymbol{R}^{\mathrm{T}}=(R_1,\cdots,R_J)$ 的 L_1 模的极小

$$\phi=\|\boldsymbol{R}\|_1=\sum_{j=1}^{J}|R_j| \tag{3.108}$$

考虑到数据集 $\{S_i\}$ 本身含有误差，可以把约束条件改为类似于(3.95)式的不等式约束。在第七节中我们还将讨论用 L_2 模取极小的地震道反演方法并对比它们的优缺点。

(四)联合线性反演

在某一个地区内地球物理调查常常使用多种地球物理方法，取得了多种地球物理数据。当然，地球物理资料解释得出的关于地下构造的模型必须说明所有的地球物理测量结果，而不只与单一方法的数据保持一致。这就是在综合地球物理解释中的所谓一致性原则。于是，我们希望在地球物理反演时输入多种不同类型的观测数据，得到同一种地下构造的模型。所谓联合反演就是从这种思想延伸出来的。

联合线性反问题又分两种情况。在第一种情况下，地球模型完全相同，只不过观测的仪器不同而使所得数据具有不同的含义。以直流电测深、交流频率测深和大地电磁测深为例，这三种方法测量的数据可分别记为 $\boldsymbol{d}_1$，$\boldsymbol{d}_2$ 和 $\boldsymbol{d}_3$，然而，对于水平层状地球模型来说，三种方法要反演的参数都为各层的电阻率和层面深度，记为

$$\boldsymbol{m}^{\mathrm{T}}=[\rho_1,\rho_2,\cdots,\rho_J,h_1,h_2,\cdots h_J]$$

如果通过转换可对这三种不同方法列出相应的线性方程组

$$\left.\begin{aligned}\boldsymbol{d}_1&=\boldsymbol{A}_1\boldsymbol{m}-\delta\boldsymbol{d}_1\\ \boldsymbol{d}_2&=\boldsymbol{A}_2\boldsymbol{m}-\delta\boldsymbol{d}_2\\ \boldsymbol{d}_3&=\boldsymbol{A}_3\boldsymbol{m}-\delta\boldsymbol{d}_3\end{aligned}\right\} \tag{3.109}$$

用本章第一节讲过的规一化方法用标准差组成的矩阵 $\boldsymbol{E}$ 对数据加权，容易把以上三个方程合并为一个方程

$$\begin{bmatrix}\boldsymbol{E}_1^{-1}\boldsymbol{d}_1\\ \boldsymbol{E}_2^{-1}\boldsymbol{d}_2\\ \boldsymbol{E}_3^{-1}\boldsymbol{d}_3\end{bmatrix}=\begin{bmatrix}\boldsymbol{E}_1^{-1}\boldsymbol{A}_1\\ \boldsymbol{E}_2^{-1}\boldsymbol{A}_2\\ \boldsymbol{E}_3^{-1}\boldsymbol{A}_3\end{bmatrix}\{\boldsymbol{m}\} \tag{3.110}$$

联合反演问题就变成求解上述方程组。在这种情况下用联合反演的好处是解估计的精度和分辨率比任一种单一方法反演都高[38]。如果用广义反演的某种方法(如随机逆法)对解估计进行评价可以发现，联合反演的折衷曲线位于所有单一方法的折衷曲线的下方，说明联合反演提高了解估计的质量。

第二种联合反演的情况是模型参数中仅有一部分相同。例如，假定水平地球波阻抗的差异，而且分层相同，则对于上一节所述的大地电磁测深和地震道反演而言，其物性参数各不

相同，但是，层面深度是相同的。同时用纵波和横波地震道，反演层状地球参数也有类似情况。总之，我们可以记第一种方法的数据为$\{u_i\}$，$i=1,2,\cdots,l$，模型为

$$\boldsymbol{m}_1^{\mathrm{T}}=(x_1,\cdots,x_J,h_1,\cdots,h_J)$$

相应的线性方程式为$\boldsymbol{u}=\boldsymbol{G}_1\boldsymbol{m}_1$，其中$\boldsymbol{G}_1$为$(l\times 2J)$，$\{x_i\}$为第一种方法对应的物性参数。同时记第二种方法数据为$\{v_i\}$，$i=1,2,\cdots,n$，模型为

$$\boldsymbol{m}_2^{\mathrm{T}}=(h_1,\cdots,h_J,y_1,\cdots y_J)$$

相应的方程为$\boldsymbol{v}=\boldsymbol{G}_2\boldsymbol{m}_2$，其中$\boldsymbol{G}_2$为$(n\times 2J)$，$\{y_j\}$为第二种方法的地层物性参数。联合反演的地球模型可用以下向量表示

$$\boldsymbol{m}^{\mathrm{T}}=(x_1,\cdots,x_J,h_1,\cdots,h_J,y_1,\cdots y_J)$$

反问题归结为求解以下方程组$\boldsymbol{d}=\boldsymbol{G}\boldsymbol{m}$，即

$$\begin{bmatrix}u_1\\ \vdots\\ u_l\\ v_1\\ \vdots\\ v_n\end{bmatrix}=\left[\begin{array}{c|c}\boldsymbol{G}_1 & \boldsymbol{0}\\ \hline \boldsymbol{0} & \boldsymbol{G}_2\end{array}\right]\begin{bmatrix}x_1\\ \vdots\\ x_J\\ h_1\\ \vdots\\ h_J\\ y_1\\ \vdots\\ y_J\end{bmatrix}\tag{3.111}$$

其中$\boldsymbol{G}$的行数为$(l+n)$，列数为$3J$。注意在以上公式中$\boldsymbol{u}$和$\boldsymbol{v}$及$\boldsymbol{x}$和$\boldsymbol{y}$可能是不同量纲的物理量，因此，一定要对数据和模型参数归一化。此外，矩阵$\boldsymbol{G}_1$和$\boldsymbol{G}_2$的元的排法是不同的。如果它们是 Jacobian 矩阵，在$\boldsymbol{G}_1$中对物理参数x_i的偏导数排在前面，而在$\boldsymbol{G}_2$中对y_i的偏导数排在后面。

在上述联合反演的公式构成时，隐含着我们施加了如下先验知识，即具有不同的物理性质$\{x_i\}$和$\{y_i\}$的介质的分层是相同的。对于复杂的实际地球来说，这种假定并不是普遍成立的。例如，低速带并不都是高电导带，密度分界面和岩石磁性分界面吻合的情况则更为少见。因此，进行联合反演之前一定要先了解岩石的不同物理性质的相干性，才能决定联合反演方案是否可行。否则，联合反演有可能给出比单一方法反演更糟的解估计。

六、广义线性反演在位场处理解释中的应用

在上节中我们已经讨论了线性规划方法在大地电磁和反射地震法中应用的例子，本节中讨论用随机逆方法进行位场数据反演的例子。假设地球模型为单个密度或磁性界面，由位场观测数据反演界面的深度变化的广义反演方法已由 Pedersen 讨论过[39]。在位场反问题中几何参数反演属于非线性反问题，要首先按本章第一节的线性化方法化为线性问题迭代求解。由于线性化后反问题的系数矩阵接近奇异，而且奇异值曲线呈阶梯状，所以 Pederson 采用第四章中的 Wiggins 方法求界面的深度变化，并讨论了解估计的分辨率和方差。下面我们来讨论数学上的一个典型的不适定问题：在任意曲面上观测的位场向下延拓到源的反问题，并分析在用不同广义反演技术时如何改善解估计的质量。这个工作是编者在1983年进行的，

见参考文献[3]。

(一)用于位场向下延拓问题的随机逆方法

设 $\psi(x',y',z')$ 为某一曲面 S 上的调和函数，S 可表为单值函数 $Z'=h(x',y')$；再设 $\phi(x,y,-H)$ 为平面 $Z=-H$ 上的调和函数，而 $Z=0$ 平面通过 $h(x',y')$ 的最小点。因此，h 表示曲面 S 的相对高程，H 表示从 S 面最低点向下延拓的深度。如果在 $Z=-H$ 平面以上没有场源存在，则在 S 上的调和函数可表为[40]

$$\psi(x',y',h)=\frac{1}{2\pi}\iint_{-\infty}^{\infty}\frac{[h(x',y')+H]\phi(x,y,-H)}{[(x-x')^2+(y-y')^2+(h(x',y')+H)^2]^{3/2}}\mathrm{d}x\mathrm{d}y \qquad (3.112)$$

用一个有限大小的矩形网格对 ψ 和 ϕ 进行取样，则(3.112)式可被离散化为

$$\psi(m,n)=\frac{1}{2\pi}\sum_{j=1}^{J}\sum_{i=1}^{I}\frac{[h(m,n)+H]\Delta x\Delta y}{\{[(m-i)\Delta x]^2+[(n-j)\Delta y]^2+[h(m,n)+H]^2\}^{3/2}}\phi(i,j) \qquad (3.113)$$

其中 Δx 和 Δy 为取样间距，J 和 I 分别为在 y 和 x 方向的取样总数，数据在 x 和 y 方向的取样总数分别记为 M 和 N，即 $m=1,\cdots\cdots,M$；$n=1,\cdots\cdots,N$。为把(3.113)式写为线性方程组，可以把取样矩阵 $\psi(m,n)$ 和 $\phi(i,j)$ 按行依次重排为向量。数据向量

$$\Psi=\{\psi_k\},k=1,2,\cdots,K,K=MN$$

模型向量 $\boldsymbol{\phi}=\{\phi_l\}$，

$$l=1,2,\cdots,L,L=IJ$$

而且

$$k=(n-1)I+m$$

$$l=(j-1)I+i$$

这样，(3.113)式便可以用矩阵方程 $\Psi=G\boldsymbol{\phi}$ 表示，其中 G 为 $K\times L$ 矩阵，它的元素由下式计算：

$$g_{kl}=[h(m,n)+H]\Delta x\Delta y/2\pi R^3$$

其中

$$R=\{(m-i)^2\Delta x^2+(n-j)^2\Delta y^2+[h(m,n)+H]^2\}^{1/2}$$

于是，上几节提及的广义反演方法便可以用来找 $\boldsymbol{\phi}$ 的最佳折衷解。

图3.9c 表示在30°山脊下方存在一个垂直磁化直立板状体的情况。板宽度 $2b=\Delta x$，埋深 $H=2\Delta x$，延深为无限，山脊高度为 $3.75\Delta x$，Δx 为任意。设板的有效磁化强度 $M=100\mathrm{A/m}$，地形上的磁场垂直分量的数据为16个，观测误差为高斯白噪音，则数据可用以下公式计算

$$\Delta Z_i=2M(1+0.05\mathrm{e})\left\{\arctan\frac{i\Delta x-c+b}{h_i+H}-\arctan\frac{i\Delta x-c-b}{h_i+H}\right\}$$

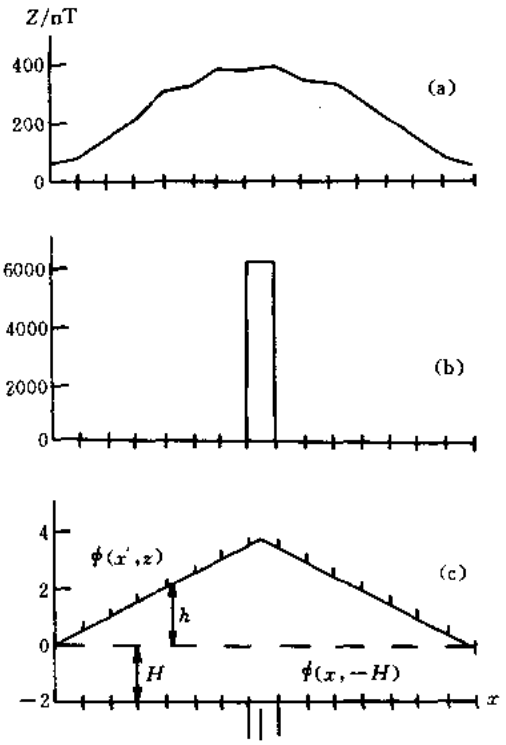


图 3.9 磁场向下延拓的例子

横轴为任意长度单位，以小竖道标明取样位置，磁场单位为 nT

(a)地形上的磁异常数据；(b)在直立板顶面的理论磁异常；(c)地形和场源的几何图形

式中 h_i 为第 i 个点的相对高程，$i=1,2,\cdots,16$；$c=8.5\Delta x$ 为板中心的横坐标；e 为高斯分布的随机误差。数据集$\{\Delta Z_i\}$表示于图3.9a，而在板顶面的理论磁异常示于图3.9b，为一个宽度等于 Δx 的有限幅度脉冲函数。对相应系数矩阵 G 作奇异值分解之后，求出 G 的奇异值如图3.10中的曲线 r 所示。由于最小的奇异值 $\lambda_{16}<10^{-7}$，矩阵 G 接近奇异，常规方法求出的 $\hat{\phi}$ 方差非常大。用如(3.82)式表示的随机逆作反演时，折衷参数可在$0.005\sim10^{-5}$之间选取，相应修改的奇异值示于图3.10的曲线 a—e，延拓结果如图3.11(曲线 a—e)所示。

如图3.9所示，理论上解为宽 Δx 的脉冲，用以下参数来评价所求的解估计 $\hat{\phi}$ 便使计算简化：延拓结果主峰的半值点宽度 W 反映了它的分辨本领，它越接近于 Δx 说明延拓结果分辨率越高。延拓结果侧瓣的极值和主峰值之比的绝对值$|A_s/A_m|$表示了最大误差。因此，W 越接近于 Δx，且$|A_s/A_m|$越小，表示估定的模型 $\hat{\phi}$ 质量越高。服从于 BG 理论，随 r^2增大，误差 $|A_s/A_m|$减小，但 W 增大，详见表3.1。由表可见，当

$$\alpha=1\times10^{-5}$$

时给出了最佳折衷的解估计，这时 $W=2.4\Delta x$，最大的相对误差为21.1%。

表 3.1　向下延拓结果的分辨宽度和最大误差

$r^2(\times10^{-3})$	$W(\Delta x)$	$\lvert A_s/A_m\rvert(\%)$
0.5	4.7	19.3
0.1	4.2	18.7
0.05	3.3	16.3
0.01	2.4	21.1
0.005	2.1	31.3
0.0025	1.9	35.8
0.001	1.7	54.0
0.0005	1.4	68.5

图3.12a 表示在走向为 y 方向的30°山脊下方一方柱体产生的重力异常，竖线表示地形等高线。方柱体剩余密度为$0.5g/cm^3$，体积为$400\times400\times200m^3$，顶面位于 $-200m$ 处。地形上重力异常呈长条状，垂直于山脊走向。用随机逆法将地形上的重力数据延拓到相对高程为零的水平面上之后(图3.12b)，异常变为等轴状，与方柱源吻合。

(二)Wiggins 方法的改善

为了找在分辨本领和方差之间取最佳折衷的近似解，还可以试用 Wiggins 提出的方法，即由(3.76)式求解估计，其中数据的有效自由度规定解估计的最大允许方差

$$\sum_{i=1}^{q}\left\{\frac{V_{ki}}{d_i}\right\}^2<t_k$$

式中 t_k 称为门限。不断改变有效自由度 q 并作折衷曲线，最后便可发现在分辨本领和方差之间取最佳折衷的 $\hat{\phi}$。

用 Wiggins 方法向下延拓时，结果不理想。图3.13表示了对图3.9的例子得出的相应折衷曲线(a)由 Wiggins 方法给出，曲线(d)由随机逆法给出。由于曲线(a)远在曲线(d)之上，说明 Wiggins 法对此例的延拓结果质量较随机逆法差，即在取最佳折衷的情况下，仍然有较大的误差和分辨宽度。

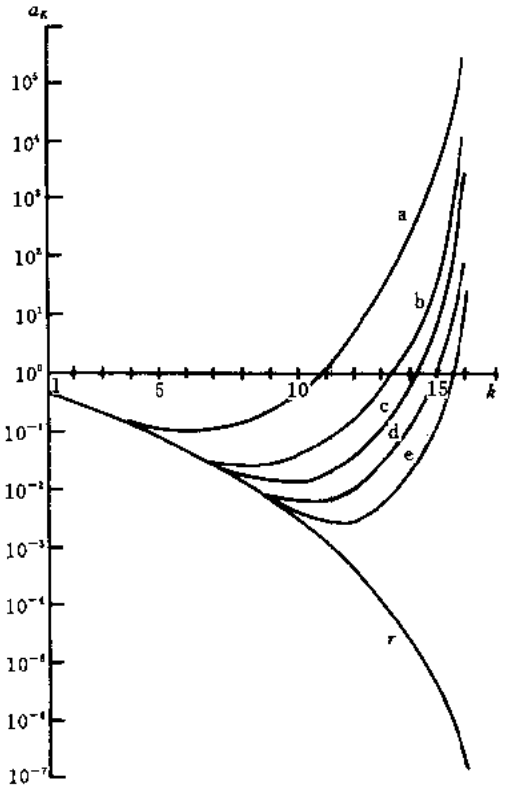


图 3.10　奇异值和奇异值的修改

r:奇异值曲线;a—e:修改后的奇异值,折衷参数τ^2分别为5,0.5,0.05,0.025和0.01乘以10^{-3}

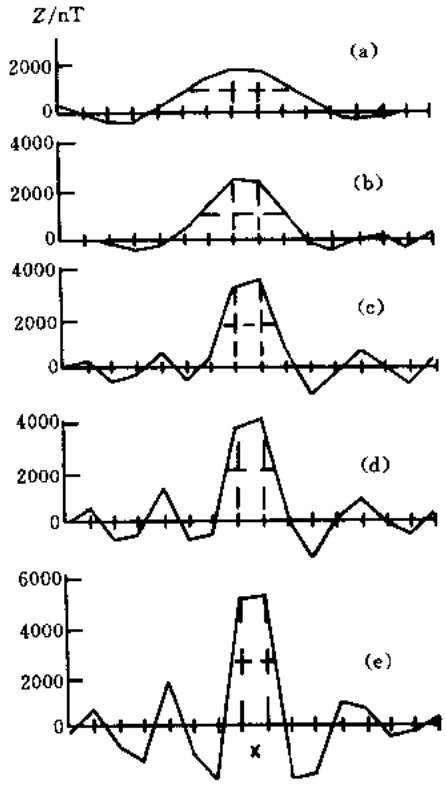


图 3.11　向下延拓的结果

水平坐标为任意,短竖道表示取样点,竖虚线标明源的宽度,横虚线表示主峰半宽度

(a)—(e)相应于图3.10中的奇异值修改曲线 a—e

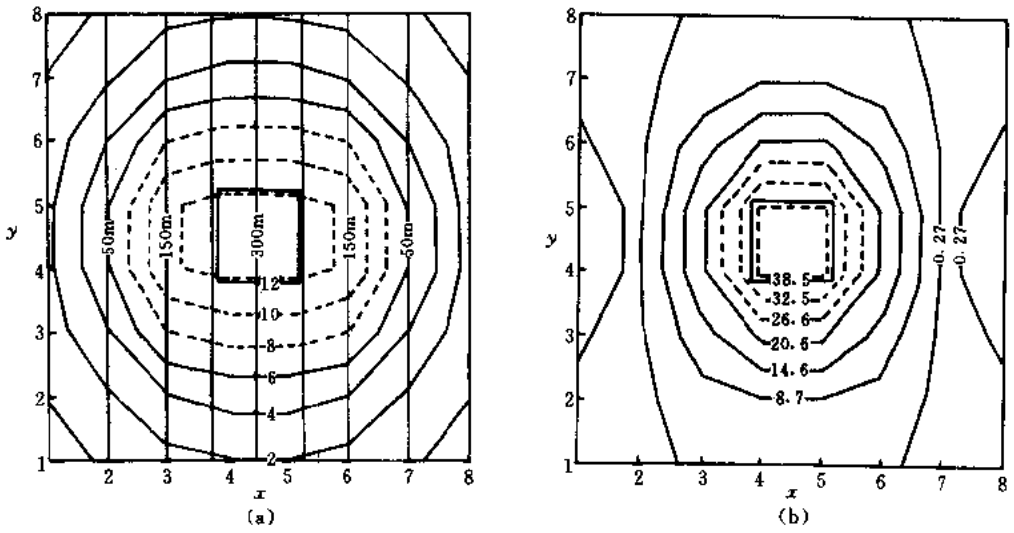


图 3.12　重力异常的延拓

重力等值线单位为 g.u.,x 和 y 轴的单位为 $\Delta x=\Delta y=200$m

(a)山脊上的重力异常,竖线为地形等高线,间距为50m;(b)向下延拓到地下相对高程为零的平面上的重力异常

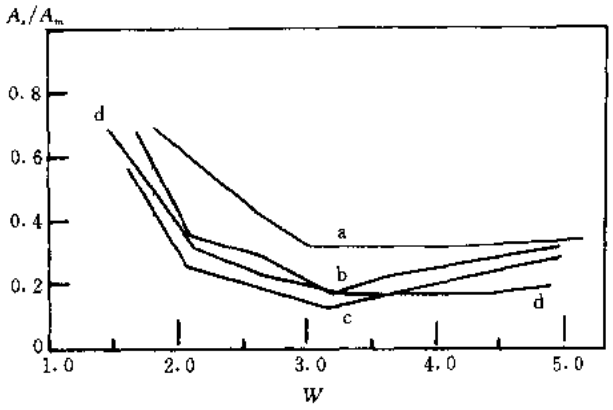


图 3.13 位场向下延拓问题的折衷曲线

(a)Wiggins 方法;(b)映像法;(c)按(3.86)式的改进方法;(d)随机逆法

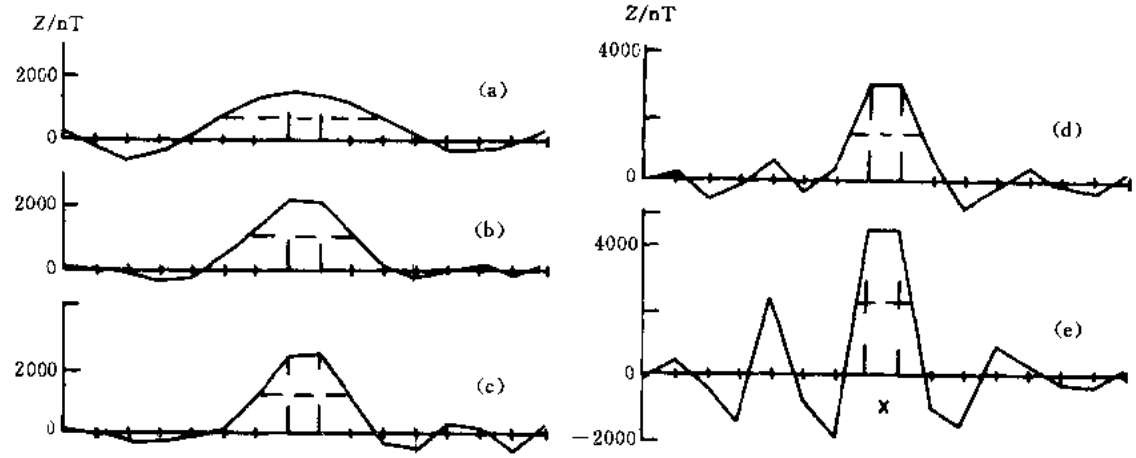


图 3.14 用改进方法取得的延拓结果

曲线 a—e 对应的门限值 t 相应为0.05,0.01,0.005,0.0025和0.001

为改善 Wiggins 法,我们可以下式来修改较小的奇异值:

$$\lambda'_k=t^2/\lambda_k,k>q$$

式中 t 为统一的门限值。由于修改后的奇异值是原奇异值的一种映像,可以把这种修改的 Lanczos 逆称为映像法,它对应的折衷曲线示于图3.13的曲线 b。映像法的折衷曲线位于曲线 a 和曲线 d 之间,表明映像法优于 Wiggins 法,但比随机逆差。

如果假定 $\hat{\phi}$ 的协方差矩阵为对角阵,则可用(3.86)式修改奇异值。相应的折衷曲线为图3.13中的曲线 c,向下延拓的结果示于图3.14。由图3.13可见,曲线 c 位于曲线 a、b 和 d 之下,说明按(3.86)式修改奇异值的广义反演方法对位场向下延拓问题优于其它方法。在图3.14中的曲线 d 对应于最佳折衷的延拓结果,主峰的半值点宽度

$$W=2.1\Delta x$$

最大相对误差为25%。

本节中讨论的位场向下延拓的例子表明,在应用随机逆方法时,(3.86)式修改奇异值可以进一步提高向下延拓问题最佳折衷解估计的估量。

用广义线性反演法对解估计进行评价时，只需要将系数矩阵作一次奇异值分解，因此对于大量数据的处理比用下章将讨论的 BG 方法快。但是，由于广义反演将模型离散化，模型的分辨本领便受抽样间距的限制。高分辨本领的广义反演一般要求减小取样间距，又导致欠定方程组的求解问题，需要加入先验数据或更多约束条件。

七、广义线性反演在地震处理解释中的应用

广义线性反演方法在地震学的各个领域中有重要的应用。在理论地震学中的应用可参考文献〔41〕。在地震勘探处理方面，广义反演和线性评价对最佳反褶积滤波器的设计具有重要意义[42,43]。对于叠后反射地震数据的反演，文献〔44〕给出了比第五节中讨论过的地震道线性规划反演更为复杂的地层模型，即考虑到层间多次波的情况。以后这种方法又进一步推广到叠前地震数据的反演之中[45]。本节要讨论的是邹振桓同志和编者进行叠后反射地震道广义反演的方法和技术[8]，重点在于了解与非线性反问题线性化有关的迭代收敛性及系数矩阵计算等问题。

(一)由反射地震道反演地层波阻抗的问题

叠后的零炮检距反射地震道数据可用泛函 $S(t,m)$ 表示，其中 $t\geqslant 0$ 为时间，m 为描述地球模型的函数，$S(t,m)$ 表示地球对 $t=0$ 时刻的地面震源的响应，在海水中为压力，在陆上为地面垂直振动的位移速度[4]。在水平层状均匀介质的情况下，m 为随深度 j 变化的波阻抗函数 $m(j)$。由于直接反演 $m(j)$ 要事先知道地层的层速度，我们转而反演 $m(t)$，因为在一维模型的情况下 t 和 z 是对应的。广义反演的第一步是把地层模型参数化。当然，我们可以按地震道取样的时间间隔 Δt 把 $m(t)$ 参数化为 $\{m(j\Delta t)\}$，其中 $j=1,\cdots,J$，但这时模型参数太多。例如 $\Delta t=4$ms 时4秒的地震记录有1000个地震数据 $\{S_i\}$，相应 $\{m_j\}$ 也有1000个参数，线性化后 Jacobian 矩阵为1000×1000，因此解方程组(3.6)时成本太高。另一种参数化的方法是按反射信号把地层划分为若干层，设每层的波阻抗为 ρ_jV_j，用反射双程走时表示的深度为 T_j，$j=1,\cdots,J+1$，$j=J+1$ 表示无反射的最底层。因此，地层模型在参数化后可表示为向量

$$\boldsymbol{m}^{\mathrm{T}}=(\rho_1V_1,\cdots,\rho_{J+1}V_{J+1},T_1,\cdots,T_J) \tag{3.114}$$

相应地震道表为 $S(t,m)$，离散采样之后记为向量 s。在一般情况下 s 和 m 不是线性关系。

实际上，在考虑层间多次波的情况下，反射地震道并不是如(3.107)式所示的子波 $W(t)$ 和界面反射系数的褶积，而是 $W(t)$ 和所谓的反射函数 $R(m)$ 的褶积[46,6]

$$s(m)=w*r(m) \tag{3.115}$$

其中 w 为子波 $W(t)$ 按 Δt 离散化后的向量，而 r 为 $R(m)$ 离散化后的向量，它的元可用以下递推公式计算

$$R_j=\frac{K_j+R_{j+1}e^{i\omega\Delta t}}{1+K_jR_{j+1}e^{i\omega\Delta t}},\ j=1,\cdots,J+1 \tag{3.116}$$

式中 $e^{i\omega\Delta t}$ 为单位延迟因子，ω 为频率，而 K_j 为界面反射系数

$$K_j=\frac{\rho_{j+1}V_{j+1}-\rho_jV_j}{\rho_{j+1}V_{j+1}+\rho_jV_j} \tag{3.117}$$

对于最底层因为无反射，有 $R_{J+1}=K_{J+1}$。因此合成地震道可以由下向上地由(3.115)和(3.116)式计算出来。注意(3.115)式中 s 和 r 呈线性关系，而不是与地层模型 m 呈线性关

系，因此由反射地震数据 s 估计 m 的问题为非线性反问题。按本章第一节(3.5)式的线性化方法求解估计时，涉及到如何计算 Jacobian 矩阵(3.9)，如何解方程组(3.6)以及如何确定(3.115)式中的子波及迭代收敛性等问题。解方程(3.6)的问题已由本章第二、三、四各节讨论，现在先来看看如何计算系数矩阵。

(二)系数矩阵的计算

按照(3.9)式及(3.114)式中对 m 的规定，必须分别求 $\partial S/\partial T_j$ 及 $\partial S/\partial(\rho_j V_j)$ 作为系数矩阵的元。然而在(3.115)式中看不出 S 与 T_j 及 $(\rho_j V_j)$ 的直接关系，这时不便于求出这些偏导数的计算显式，但忽略多次波时(3.115)变为

$$S(t)=W(t)*\sum_{j=1}^{N}K_j\delta(t-T_j) \tag{3.118}$$

式中假定了震源为单位振幅，K_j 为对应 T_j 时刻的初次反射波幅度，由(3.117)计算。对(3.118)求偏导数有

$$\partial S/\partial T_j=-K_jW(t)*\delta'(t-T_j) \tag{3.119}$$

式中 Dirac-δ 函数的偏导数要用数值逼近方法计算。用柯西函数逼近 δ 函数时，有

$$\delta(t)=\lim_{q\to\infty}\frac{q}{\pi(1+q^2t^2)}=\begin{cases}0, & t\neq 0\\ \infty & t=0\end{cases}$$

因此

$$\delta'(t-T_j)=\partial\delta(t-T_j)/\partial T_j=\lim_{q\to\infty}\frac{2q^3}{\pi}(T_j-t)e^{-q^2t^2}$$

由上式代入(3.119)式可计算出对 T_j 的偏导数矩阵的元。

用解析方法求偏导数 $\partial S/\partial(\rho_j V_j)$ 可用(3.118)和(3.117)式直接获得

$$S(t)=W(t)*\sum_{j=1}^{N}\frac{I_j-I_{j-1}}{I_j+I_{j-1}}\delta(t-T_j),\ I_j=\rho_jV_j \tag{3.120}$$

而 Jacobian 矩阵的元为

$$\frac{\partial S(i\Delta t)}{\partial I_j}=\frac{2I_{j-1}}{(I_j+I_{j-1})^2}W(i\Delta t)*\delta\{[i-(j-1)]\Delta t\}-\frac{2I_{j+1}}{(I_j+I_{j+1})^2}W(i\Delta t)*\delta[(i-j)\Delta t] \tag{3.121}$$

由于 T_j 和 (ρ_jV_j) 是具有不同量纲的物理量，如果不进行归一化则此时(3.6)式中的矩阵可能非常接近奇异，因此要把方程(3.6)式乘以权系数矩阵使之归一化。权系数的选择取决于所用的单位，以使归一化后的系数矩阵的对角线元素绝对值变化不大为宜。对地震道反演问题归一化之后得出的方程组常常是良态的，只在薄层很多时才病态，因此用各种广义的最小二乘法就可以取得解估计。

(三)子波及比例因子

按上述方法依(3.115)式反演地层参数时必须假定时不变子波 $W(t)$ 为已知。在有钻井及声测井资料的地震剖面上 $W(t)$ 可通过子波反褶积求取，而在没有钻孔资料时子波是未知的，也要作为模型参数来求。如果子波很复杂，则由(3.115)式可知这种反问题不可能有唯一解。如果实际子波比较简单，即可用少数参数来描述，则求反问题的解估计还是可能的。设子波 $W(t)$ 为限带恒相位 ϕ 的子波，则其振幅谱为梯形，可用最低截止频率 f_1 及最高截止频率 f_2 及幅度 ψ 来描述，而 $W(t)$ 可通过逆 Fourier 变换求得。这时，地震道反演的模型向量

(3.114)式变为

$$\boldsymbol{m}^{\mathrm{T}}=(\rho_1 V_1,\cdots\cdots,\rho_{J+1}V_{J+1},T_1,\cdots,T_J,\psi,f_1,f_2,\phi) \tag{3.122}$$

地震道 $S(t)$对子波参数的偏导数可用有限差分法求之。

在(3.118)式中假定了震源幅度为单位，这样用它计算出来的地震道与实际地震道将相差一个比例因子 α，即(3.118)式用于与地震道拟合时要改为

$$S(t)=\alpha W(t) * \sum_{j=1}^{N} K_j\delta(t-T_j)$$

由于这个比例因子也是未知的，因此也要当作模型参数来求。由于子波振幅谱的幅度 ψ 相当于对 $S(t)$乘以比例因子的作用，考虑 α 时可令 $\psi=1$，即用 α 代替(3.122)中的 ψ。由上式可知

$$\alpha^{(k)}=\sum_i S_i/\sum_i S_i^{(k)}$$

其中 S_i 为地震道数据，$S_i^{(k)}$ 为第 k 次迭代时由(3.115)式计算出的合成地震道数据。于是，偏导数为

$$\frac{\partial S^{(k)}}{\partial \alpha}=\frac{\alpha^{(k)}-\alpha^{(k-1)}}{\alpha^{(k)}}$$

(四)收敛性

应用广义线性反演方法求非线性反问题的解估计时，收敛性取决于初值的选择。图3.15和3.16给出地震道反演收敛性的例子。假定真实地层为七层波阻抗模型，其中第五层为低速薄层(图3.15a)。如果给定的初始猜测的趋势与模型相同，即使偏差较大(见图3.15a 中的虚线)，反演迭代过程也收敛，其解估计的最大相对误差为16%左右(图3.15b)。但是如果初始猜测的变化趋势给错，如把第五层猜成高速层(图3.16a 中的虚线)，则反演结果不能揭露此低速层的存在(见图3.16b)，说明此时即使迭代收敛其解估计的分辨率低而且方差很大。此

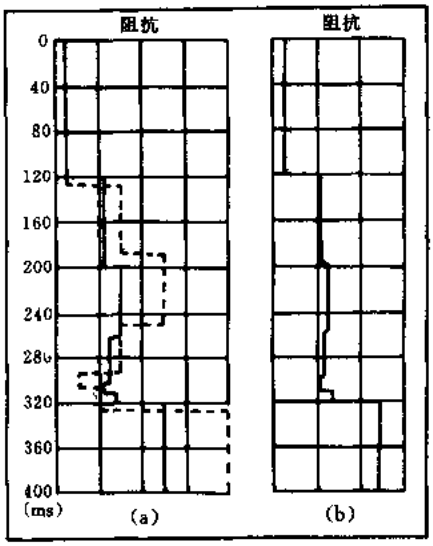


图 3.15 七层波阻抗模型(a)和广义线性反演结果(b)

(a)中的实线为波阻抗曲线，虚线为它的初始猜测

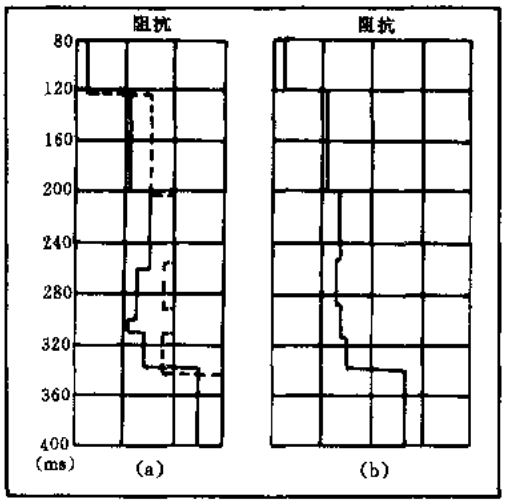


图 3.16 七层波阻抗模型(图(a)中实线)及其反演结果(b)

(a)中的虚线为初始猜测，它把第五层错猜成高速薄层

外当子波不符合(1.122)式的假设条件时迭代过程可能发散。

通过本节中讨论的例子可以看到，广义线性反演方法在用于解非线性反问题时要求事先对地球模型有一定的先验了解并把它在初始猜测中反映出来。对线性反问题我们总可以用本章2～5节所示的方法找到某种解估计，关键在于如何使解估计同时具有高的分辨率和小的方差。对于非线性广义线性反演则不同，只有在给定初始猜测接近真实地球模型时才可以求出有意义的解估计，这是广义线性反演的局限性，关于非线性反演的深入研究，请参见以下几章。

第四章　Backus-Gilbert 反演理论

在上一章讨论了离散地球模型的广义反演方法之后，本章集中讨论连续地球模型反问题的理论和方法。连续地球模型指的是可以用空间坐标的连续函数或函数集表示的地球内介质的物理性质。与其它模型相比，连续地球模型具有以下两个特点。首先，连续地球模型在研究的地球内部区域的任一点取值，在数学上属于无限维函数空间，而离散地球模型反映的是离散网格内地球物理性质的平均值，属于有限维向量空间。从理论上说，连续地球模型比离散地球模型具有更高的分辨率。然而，从有限维向无限维过渡并不总是可能的和无条件的，在数学上有一个收敛性的问题，因此，连续地球模型的反演比上一章讨论的离散地球模型的反演要困难。另外，连续地球模型是用经典意义上的函数或函数集表示的，不能描述地球内部存在函数取不定值的间断面的情况。在后一种情况下地球模型必须用广义函数描述；而广义函数不属于函数空间，而属于泛函空间，这是更困难的数学问题。关于用广义函数描述的地球模型将放到下一章讨论，本章只讨论连续地球模型的情况。

对于连续地球模型的反演问题我们必须从数学物理的角度研究解的存在性、非唯一性及算法稳定性等问题。正如第一章所述，Backus 和 Gilbert(BG)首先系统地研究了有限和不精确数据的反问题解的非唯一性，并导出了对解估计评价的一系列准则和方法[17,18,19]。根据 BG 评价准则也可以导出相应离散地球模型的广义反演方法，也将在本章进行讨论和对比。BG 反演理论和评价方法虽然具有理论上的严格性，但是要求繁复的计算，造成实际应用中的某些限制。在许多时候我们往往只要求构成某种稳定的算法，来计算误差不超过一定范围的解估计，这时合适的数学工具就是所谓的正则化(regularization)方法，于是，BG 理论和正则化方法就构成了本章研究连续地球模型反问题的主要内容。

一、连续地球物理模型的反问题

(一)模型空间和数据空间

现在让我们重新从地球物理模型和数据的特点出发，来研究对它们合适的数学描述。从自然哲学的观点来看，地球模型并不是客观存在的复杂地球本身，而是不同的人对于它的了解和认识。由于观测及分析技术的限制和主观认识的差异，地球模型可能有无穷多个。但是其中只有一个与真实地球吻合最好。从这个意义上说，我们在进行地球物理反演的时候，就不仅要去研究某一个可能的地球模型，而且要去研究所有可能的地球模型的总体，并对它们进行评价，才能找出其中一个最符合真实地球的模型作为反演问题的解答，那么，应该如何描述所有可能的地球模型的总体，又怎么去评价它们呢?这就是 Backus 和 Gilbert 于1967～1970年开创的地球物理反演理论的出发点。

Backus 和 Gilbert 认为，地球模型在数学上可以用有限个有序的函数的集合表示。对于不同的地球模型，函数集中各个函数的形式可以有所差别的。由于任意一个有限的函数集都

可以表示成为无穷维抽象空间的一个元，则所有可能的地球模型构成了一个无穷维的抽象空间，这个抽象空间称为模型空间，记为 $\mathscr{H}_m$。

为了了解各种模型的特点和差异，我们要求模型空间是可以进行运算的抽象空间，而且可以用一个非负标量来描述空间中元的大小。于是我们可以选用线性赋范空间作为模型空间。此外，我们还希望对于模型空间具有便于运算的形式，因此，Backus 和 Gilbert 便(注意正则化方法不同于 BG 理论，模型空间并不限于希尔伯特空间)选择希尔伯特空间作为模型空间在选择希尔伯特空间作为模型空间时，要首先根据所研究的地球物理反演问题定义空间中一对元的内积。

以静止的各向同性完全弹性层状地球为例，对于天然地震数据反演来说，如(3.2)所述，地球模型可以用密度 $\rho(r)$，体积模量 $k(r)$和切变模量 $\mu(r)$来描述：

$$m=[\rho(r),k(r),\mu(r)] \tag{4.1}$$

其中 r 为矢径，$r\in[0,a]$，a 为地球半径。注意在本章中函数也用小写罗马字母表示，不要把它与实数混淆起来。由此可见，这时地球模型为由三个函数组成的有序函数集。如果把这样的一个函数集视为一个线性赋范空间的一个元，而另一个可能的地球模型表示为

$$m'=[\rho'(r),k'(r),\mu'(r)],r\in[0,a] \tag{4.2}$$

在线性赋范空间内则 m 和 m'的线性组合

$$bm+b'm'=(b\rho+b'\rho,bk+b'k,b\mu+b'\mu) \tag{4.3}$$

也表示一个地球模型，其中 b 和 b'为两个实数。用以下方式来定义两个地球模型的内积

$$(m,m')=\int_0^a \mathrm{d}r\cdot r^2[\rho(r)\rho'(r)+k(r)k'(r)+\mu(r)\mu'(r)] \tag{4.4}$$

则可把模型空间定义为一个希尔伯特空间 $\mathscr{H}$，其中一个可能的地球模型 m 乃是 $\mathscr{H}$ 中的一个元。

在模型空间 $\mathscr{H}$ 中，必有一个元(模型)是最接近真实地球模型的。这个元称为实际地球模型，并记为 m_0。地球物理反演的目的是要求出对 m_0的合适的解估计 $\hat{m}$；我们知道希尔伯特空间是无限维赋范空间中最简单的类型，这种简单性是由于附加的内积条件带来的。内积的引入使希尔伯特空间赋有类似于欧几里得几何的某些性质[47]。例如，希尔伯特空间中两个元 m 和 m'之间的距离可以用范数定义为

$$\|m-m'\|=(m-m',m-m')^{1/2} \tag{4.5}$$

它们相对于原点0的夹角定义为

$$\cos\phi=\frac{|(m,m')|}{\|m\|\|m'\|}\leqslant 1 \tag{4.6}$$

另外，希尔伯特空间中一个无限维向量 m 在另一个向量 f 上的投影可表为

$$\mathrm{Proj}(m)=\frac{(f,m)}{(f,f)}f \tag{4.7}$$

投影的概念对求反问题的解估计至关重要。我们在许多情况下不能求得连续地球模型 m 本身，这时可以考虑求 m 在某个流形上的投影，因为投影是一种与 m 最接近的解估计。

由(4.6)式可以看出，在希尔伯特空间中两个元或两个向量正交(即 $\phi=90°$)的充分必要条件是

$$(m,m')=0$$

设 $\mathscr{M}$ 为希尔伯特空间中的一个完备子空间，即 $\mathscr{M}\subset\mathscr{H}$，为无限维向量子集，则 $\mathscr{H}$ 中所有

正交于 $\mathscr{M}$ 的向量组成的集叫做 $\mathscr{M}$ 的正交补，记为 $\mathscr{M}^{\perp}$。由于任意两个正交于 $\mathscr{M}$ 的向量的和仍然正交于 $\mathscr{M}$，因此 $\mathscr{M}^{\perp}$ 也是 $\mathscr{H}$ 的一个子空间。再由 $\mathscr{M}^{\perp\perp}=\mathscr{M}$ 可知 $\mathscr{M}^{\perp}$ 是一个完备的子空间。正交补的引入在分析反问题的多解性及正则化时非常重要。根据泛函分析中关于希尔伯特空间的定理，当 $\mathscr{M}$ 为模型空间 $\mathscr{H}_m$ 中的一个完备子空间时，有

$$\mathscr{H}_m=\mathscr{M}\oplus\mathscr{M}^{\perp} \tag{4.8}$$

其中⊕表示两个子空间的直和(*direct sum*)，其定义如下：记子空间 $\mathscr{A}$ 中的任一向量为 a，而子空间 $\mathscr{B}$ 中的任一向量为 b，如果空间 $\mathscr{C}$ 中的任一向量都可以唯一地用范数 $\|a+b\|$ 表示，则称 $\mathscr{C}$ 为子空间 $\mathscr{A}$ 和 $\mathscr{B}$ 的直和，并记为 $\mathscr{C}=\mathscr{A}\oplus\mathscr{B}$。在上述定义中的向量包括无限维向量，因此可以和函数一样用小写罗马字母表示。

现在假定实际地球模型 $m_0\in\mathscr{H}_m$ 可以分解为

$$m_0=m'+m''$$

其中 $m'\in\mathscr{M}$，$m''\in\mathscr{M}^{\perp}$，由(4.8)式可知，$m'$ 是在 $\mathscr{M}$ 中与 m_0 最接近的元，即连续地球模型的一种“最佳”解估计。由此可知把模型空间放入希尔伯特空间之后由内积带来的一系列性质，可以用于地球物理反演的计算和实现。

(二)数据空间

根据 BG 理论，可以把地球物理数据的共性抽象地表示为以下各点。

(1)一个精确的地球物理数据 d_i 是一个可以由模型 m 通过规定的运算取得的实数，即合成数据，

$$d_i=G_i(m) \tag{4.9}$$

其中 G_i 表示对于第 i 个数据函数 m 的泛函，即规定的运算。

(2)由于时间和空间的无限性和可分性，由一个模型 m 计算出来数据可以有无穷多个。在规定了内积之后，无限维的数据集也可以表示为一个希尔伯特空间中的一个元 d，d 称为理想数据，即无限和精确的地球物理数据，此时类似于(4.9)式 d 可表为

$$d=Gm \tag{4.10}$$

其中 G 为正演问题规定的算子。

(3)由于地球模型 m 可能有很多，由(4.10)对应的理想地球数据 d 也有很多，它们组成了一个函数空间的一个子集或者全部，称为数据空间。与模型空间对应，我们也可以把数据空间规定为一个希尔伯特空间，并记为 $\mathscr{H}_d$。算子 G 即为从模型空间到数据空间的映射，$G:\mathscr{H}_m\to\mathscr{H}_d$。

(4)在所有可能的理想数据集中，只有一个 $d_0\in\mathscr{H}_d$ 是由实际地球模型 m_0 唯一地确定的：

$$d_0=Gm$$

因此，地球数据中包含有由模型的简化带来的误差，即

$$d-d_0=G(m-m_0)=G\delta m \tag{4.11}$$

利用希尔伯特空间来表示理想地球数据带来的一个好处是可以利用正交基来表示数据，这就给研究有限数据和无限数据带来方便。设 $\mathscr{B}\subset\mathscr{H}_d$ 为数据空间中一个可列集，$\mathscr{B}=\{b_i\}$，$i=1,\cdots,\infty$，如果而且只在对所有 $i,j\geqslant 1$

$$(b_i,b_j)=\delta_{ij}$$

时 $\mathscr{B}$ 才称为规格化正交集。相应地 $\{b_i\}$ 称为 $\mathscr{H}_d$ 空间的规格化正交基。根据希尔伯特空间

中元的表现定理，数据空间中的每一个元 d 都可以用基 $\{b_i\}$ 表示为

$$d=\sum_{i=1}^{\infty}(d,b_i)b_i=\sum_{i=1}^{\infty}d_ib_i \tag{4.12}$$

其中 d_i 可由(4.9)式表示。与有限维情况不同的是，(4.12)式的无穷级数必须收敛才有意义。泛函分析还证明，只要数据空间 $\mathscr{H}_d$ 存在一组规格化正交集 $\mathscr{B}=\{b_i\}$，则有正交补 $\mathscr{B}^{\perp}=\varnothing$，而且对于任何 $d\in\mathscr{H}_d$ 都满足以下 Parseval 公式

$$\|d\|^2=\sum_{i=1}^{\infty}|(d,b_i)|^2 \tag{4.13}$$

(5)由于无限连续的观测实际上不可能，因此实际地球物理数据为有限集，记为 $\{d_i\}$，$i=1,\cdots,n$，其中 d_i 由(4.9)定义并且彼此线性无关，它可表示为 n 维向量空间中的向量。设 $\{b_i\}$，$i=1,\cdots,n$ 为此向量空间的规格化正交基，则此向量空间中的任一个元 $\bar{d}$ 都可由基表示为

$$\bar{d}=\sum_{i=1}^{n}(d,b_i)b_i=\sum_{i=1}^{n}d_ib_i \tag{4.14}$$

我们称 $\bar{d}$ 为理想地球物理数据 d 的有限采样集，或简称为有限地球数据。

(6)由于实际测量不可避免地包含噪音，假定信号和噪音具有可加性，则实际测量数据可表为

$$\hat{d}=\bar{d}+\delta d=\sum_{i=1}^{n}(d_i+\delta d_i)b_i \tag{4.15}$$

式中 $\delta d=\{\delta d_i\}$，$i=1,\cdots,n$，而 δd_i 为第 i 个数据的噪音。注意实际数据 $\hat{d}$ 要用希尔伯特概率空间才能完整地描述，而不再适合于确定型的函数空间的情况。

(三)连续地球模型反问题的研究内容

根据以上对模型的数据的描述，我们可以把反问题研究内容归纳为以下各点。

(1)解的存在性问题。例如，规定具体的模型 m，算子 G 和理想数据 d，(4.10)式的解是否存在。从数学的角度，这时解的存在性主要取决于算子 G 的性质。对于实际反演来说最重要是要研究给出实际数据 $\hat{d}$ 时能否找到实际地球模型 m_0，这时解的存在性就不仅与算子 G 的数学性质有关，而且与物理上的假定有关。如果把(4.9)代入(4.11)和(4.15)，有

$$\hat{d}=\sum_{i=1}^{n}(G_i(m)+\delta d_i)b_i=\sum_{i=1}^{n}\{G_i(m_0+\delta m)+\delta d_i\}b_i \tag{4.16}$$

此时 $\hat{d}$ 不仅包含有测量噪音，也包括规定地球模型 m 和相应的算子 G 没有准确地描述实际地球所造成的误差。因此，地球用简化的模型 m 来表示是否恰当、噪音的性质及数据的个数都与解的存在性有关。这种情况下解的存在性问题的研究对检验反问题的提法是否正确是很重要的。可惜的是，这方面的工作还研究得很不充分。

(2)解的唯一性问题，包括算子方程(4.10)的解如果存在，是否唯一的问题，以及用(4.14)式中的有限地球数据 $\bar{d}$ 代替理想数据 d 时解是否唯一。这个问题在 BG 理论中做了较深入的研究，得出了地球物理反问题解具有高度的非唯一性的结论。

(3)稳定性问题，指(4.10)式中的数据有微小误差 δd 时，方程(4.10)的解 m 是否会产生很大的误差。换句话说，(4.10)的解是否连续地依赖于数据，在数学上解的不稳定性与反问题的不适定性有关，主要取决于算子和模型空间的性质。

(4)求解估计的算法。由于地球物理反问题的非唯一性或不适定性，在绝大情况下不能

唯一地确定实际地球模型 m_0，但我们可以通过数值计算获得的 m_0 的某种逼近，它即是解估计 $\hat{m}$。那么，可以用那些准则来逼近 m_0？在什么条件下计算 $\hat{m}$ 的算法是稳定的？这些都是连续地球模型反问题要研究的内容。在第4～6节中我们要讨论 BG 理论构造和评价解估计的三个准则，它们主要从解的非唯一性研究导出。在正则化方法中解估计的构造准则主要是从问题的不适定性导出的，一般指赋范空间中 m_0 和 $\hat{m}$ 之间的距离小于某一任意小量。

(5)解估计的评价问题。由于地球物理反问题的解不唯一，可以找出无限多个反问题可接受的解估计 $\hat{m}$，就必须分辨各种解估计的优劣，评价其质量。BG 理论的要点不仅在于构造出一种解估计，更重要的是评价各种可接受的解估计的优劣，并力求在其中找出一个最理想的，即所谓最佳折衷的解估计来。

二、线性反问题解的非唯一性

本节讨论 BG 理论中的一个核心问题，即地球物理反问题解的非唯一性，讨论分理想数据集及有限数据集两种情况，并假定方程(4.10)中的算子 G 为线性算子。非线性算子的情况在下节讨论。

首先让我们回忆一下算子理论中的一些术语。如(4.10)假定 $d=Gm$，$d\in\mathscr{H}_d$ 和 $m\in\mathscr{H}_m$，实际上数据 d 可能只是数据空间中的一个子集，记为 $\mathscr{D}(G)$，并称为算子 G 的域值。此外，通过算子 G 映射到子集 $\mathscr{D}(G)$ 上去的模型 m 也可能是 $\mathscr{H}_m$ 空间中的一个子集，它称为算子 G 的定义域，并记为 $\mathscr{M}(G)$。于是，算子 G 是把模型空间中子集 $\mathscr{M}(G)\subset\mathscr{H}_m$ 中的所有元映射为数据空间中的映象集 $\mathscr{D}(G)\subset\mathscr{H}_d$ 的映射，记为 $G:\mathscr{H}_m\to\mathscr{H}_d$。如果对 $\mathscr{M}(G)$ 中任两个元 m_1 和 m_2 及任两个数量 α_1 和 α_2，元 $\alpha_1m_1+\alpha_2m_2$ 仍在 G 的定义域中，而且

$$G(\alpha_1m_1+\alpha_2m_2)=\alpha_1Gm_1+\alpha_2Gm_2$$

则称 G 为线性算子。若线性算子 G 的定义域是整个 $\mathscr{H}_m$ 空间，且存在某个常数 C 使得

$$\|Gm\|<C\|m\| \tag{4.17}$$

对所有 m 都成立，则称 G 为有界算子。线性有界算子是连续的。

(一)理想数据集

要研究方程(4.10)的解的存在性和唯一性，必然要涉及到线性算子 G 的逆是否存在和有界的问题。定义线性算子 G 的逆 $G^{-1}:\mathscr{H}_d\to\mathscr{H}_m$ 为当 $d=Gm$ 时满足 $m=G^{-1}d$ 的算子，算子理论中证明[47]，当 G 为有界线性算子时一般存在有界的逆算子 G^{-1}。但是，此时算子方程(4.10)的解却不一定唯一。

以 G 为 Fredholm 积分算子为例，

$$Gm=\int_a^b G(x,y)m(y)\mathrm{d}y=\mathrm{d}(x), \tag{4.18}$$

如果 $G(x,y)=x\sin y$，$a=0$ 而 $b=\infty$，$\mathrm{d}(x)=\frac{\pi}{2}x$，(4.18)变为

$$\int_0^\infty G(x,y)m(y)\mathrm{d}y=x\int_0^\infty \sin y\,m(y)\mathrm{d}y=x \tag{4.19}$$

显然不仅 $m=1/2$ 是方程(4.19)的解，而且 $m=\frac{\sin ky}{y^2}$，$k=2,3,\cdots$，等也是(4.19)的解，说明此 Fredholm 方程的解不唯一。

在 d 为理想数据的情况下，方程(4.10)及(4.18)的唯一性可以用算子 G 的零空间来说明，它定义为由方程

$$Gm=0 \tag{4.20}$$

的解组成的集，记为 $\mathcal{N}(G)\subset\mathcal{M}(G)\subset\mathcal{H}_m$。显然 $\mathcal{N}(G)$ 是一个线性子空间。下面将说明当 $\mathcal{N}(G)\neq\mathcal{O}$ 时(4.10)的解是不唯一的。设 m 是方程(4.10)的一个解，即满足 $Gm=d$，而 m' 是(4.20)的一个解 $Gm'=0$，即 $\mathcal{N}(G)\neq\mathcal{O}$ 时其中的一个元，则有

$$G(m+m')=Gm+Gm'-Gm=d$$

说明 $m+m'$ 也是(4.10)的解。换句话说，只有 $\mathcal{N}(G)=\mathcal{O}$ 时线性算子方程(4.10)才有唯一解。零空间 $\mathcal{N}(G)\neq\mathcal{O}$ 说明模型空间中有一个无法通过数据来了解的子空间。

在解非唯一的情况下，在(4.8)式中提到的正交分解非常有用。这时由(4.8)我们可以把模型空间表示为 G 的零空间 $\mathcal{N}(G)$ 和 $\mathcal{N}(G)$ 的正交补的直和

$$\mathcal{H}_m=\mathcal{N}(G)\oplus\mathcal{N}(G)_\perp \tag{4.21}$$

由于我们对零空间内引起多解的解估计不感兴趣，因此我们应该把算子 G 的定义域限制在正交补 $\mathcal{N}(G)_\perp$，求 m 的一种解估计。限制在 $\mathcal{N}(G)_\perp$ 上的算子 G 的逆算子称为广义逆，可记为

$$G^{-g}=(G|\mathcal{N}(G)_\perp)^{-1} \tag{4.22}$$

依一定准则构造 G^{-g}，可以得到某种“最佳”解估计 $\hat{m}=G^{-g}d$。当 $\mathcal{H}_m$ 和 $\mathcal{H}_d$ 退化为有限维空间时线性算子 G 变为矩阵，(4.10)退化为线性欠定或混定方程组，上一章讨论的有关广义反演方法就属于用正交分解求广义逆的特例。

对离散的线性反问题 $\boldsymbol{Gm=d}$，如果它的解不唯一，则存在一些非零的解 $\boldsymbol{m}_{\text{null}}$，使得齐次方程

$$\boldsymbol{Gm}_{\text{null}}=\boldsymbol{0}$$

成立，这些非零的向量 $\boldsymbol{m}_{\text{null}}$ 被称为反问题的零向量，它构成矩阵 $\boldsymbol{G}$ 的零空间，如果 $\boldsymbol{m}_{\text{par}}$ 是方程 $\boldsymbol{Gm=d}$ 的一个特解，则 $\boldsymbol{m}_{\text{par}}+\alpha\boldsymbol{m}_{\text{null}}$ 也是方程的一个解，其中 α 为任意非零实数。注意到只有有限个零向量才是线性无关的，设线性无关的零向量的个数为 $q\leqslant J$(J 为 $\boldsymbol{m}$ 的维数)，因此通解可表示为

$$\boldsymbol{m}_{\text{gen}}=\boldsymbol{m}_{\text{par}}+\sum_{i=1}^{q}\alpha_i\boldsymbol{m}_{i\text{null}}$$

有趣的一点是即使线性反问题不唯一，但是它的平均却可能是唯一的，把模型参数的平均记为

$$\bar{\boldsymbol{m}}=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{m}$$

其中 $\boldsymbol{w}^{\mathrm{T}}$ 表示用于平均时的权，叫平均向量，且满足 $\sum_{i=1}^{J}w_i=1$。对于通解作平均，则有

$$\bar{\boldsymbol{m}}=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{m}_{\text{gen}}=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{m}_{\text{par}}+\sum_{i=1}^{q}\alpha_i\boldsymbol{w}^{\mathrm{T}}\boldsymbol{m}_{i\text{null}} \tag{4.23}$$

由于 $\boldsymbol{w}$ 有 J 个元，而 $q\leqslant J$，我们最少可以找到一个这样的平均向量 $\boldsymbol{w}$，使得 $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{m}_{i\text{null}}$ 对所有的 i 都为零向量，从而把(4.23)式中的最后项抵消掉，使 $\hat{m}$ 唯一。因此，不去求 $\boldsymbol{m}$ 而去求它的某种平均乃是消除反问题非唯一性的一条途径。

(二)有限数据集

对于数据有限的情况下，$\bar{d}=\{d_i\}$，其中由(4.9)有

$$d_i=G_i(m),\quad i=1,\cdots,n \tag{4.24}$$

它定义了一个泛函方程组。设 G 为 Fredholm 算子，则由(4.18)式可把上式具体化为

$$\int_a^b G_i(y)m(y)\mathrm{d}y=d_i,\quad i=1,\cdots,n \tag{4.25}$$

由于$\{d_i\}$为有限，而函数 $m(y)$为无穷维，因此满足泛函方程组(4.25)的解 $m(y)$有无穷多个。实际上如(4.25)的第一类 Fredholm 积分方程的解有以下形式

$$m(y)=m_{\mathrm{par}}(y)+m_{\mathrm{null}}(y) \tag{4.26}$$

其中 $m_{\mathrm{par}}(y)$是(4.25)的特解，而 $m_{\mathrm{null}}(y)$为

$$\int_a^b G_i(y)m(y)\mathrm{d}y=0,\quad i=1,\cdots,n$$

的任一非平凡解，它是必定存在的。

由(4.23)的启示，在(4.25)的解为非唯一的情况下我们可以去求模型 $m(y)$的某种平均，它也许是唯一的。设 $\dot{m}(y_0)$为变量在 $y=y_0$时地球模型 $m(y)$的某种平均，即

$$\dot{m}(y_0)=\int_a^b A(y,y_0)m(y)\mathrm{d}y=(A(y,y_0),m) \tag{4.27}$$

其中 $A(y,y_0)$称为平均核或平均函数，在 BG 理论中它定义为数据核 $G_i(y)$的线性组合

$$A(y,y_0)=\sum_{i=1}^{n}\alpha_i(y_0)G_i(y) \tag{4.28}$$

其中$\{\alpha_i\}$，$i=1,\cdots,n$，为待定系数组。(4.28)代入(4.27)有

$$\begin{aligned}\dot{m}(y_0)&=\Big(\sum_{i=1}^{n}\alpha_i(y_0)G_i(y),m\Big)\\&=\Big(\sum_{i=1}^{n}\alpha_i(y_0)(G_i,m)\Big)\end{aligned}$$

代入(4.25)有

$$\dot{m}(y_0)=\sum_{i=1}^{n}\alpha_i(y_0)d_i \tag{4.29}$$

说明在平均意义上的解估计只要在待定系数组$\{\alpha_i\}$确定之后便可由有限数据集求得。因此，在数据有限时反问题解估计的构造转化为根据一定准则确定待定系数组$\{\alpha_i\}$的问题。由于(4.27)式中 A 为平均核，还必须满足所谓的单位模条件，即

$$\int_a^b A(y,y_0)\mathrm{d}y=1 \tag{4.30}$$

Backus 和 Gilbert 证明[17,18,19]，对于$[a,b]$中的任一个 y_0，由如(4.14)表示的有限数据集 $\bar{d}$ 可以取得的关于模型 $m(y_0)$的信息仅仅是如(4.27)表示的 $m(y)$在 y_0附近的平均 $\dot{m}(y_0)$，而不可能是 $m(y_0)$本身。换句话说，这时平均核 $A(y,y_0)$和 $m(y)$都是不确定的，只有它们的平均 $\dot{m}(y_0)$才可以唯一地确定，即此时能满足(4.25)式的 $\dot{m}(y)$都具有与(4.27)相同的形式，但相应在(4.28)式中的待定系数组$\{\alpha_i\}$对不同的 y_0是不同的。最后，当数据量 $n\to\infty$时 $A(y,y_0)\to\delta(y-y_0)$，相应地 $\dot{m}(y_0)\to m(y_0)$，解估计随数据量的增大而趋向于地球模型。

(三)实际数据集

如(4.15)所示的实际地球数据 $\bar{d}$ 是有限和不精确的。由数据有限造成的连续地球模型反问题解的非唯一性上面已经讨论过。数据的误差给连续地球模型反问题的影响可分两种情况进行讨论。一种情况是数据只含微小的误差，这种情况放在第七节正则化方法中讨论。

另一种情况是数据的误差服从高斯分布，这个问题可以用算子的谱分解来研究。

由希尔伯特空间中元的表现定理(4.12)可知，任一连续地球模型 m 可以写成：

$$m=\sum_{i=1}^{\infty} b_i m_i \tag{4.31}$$

其中$\{b_i\}$为模型空间 $\mathscr{H}_m$ 的规格化正交基。令集$\{\lambda_i\}$为算子 G 的本征值，即满足

$$Gm_i=\lambda_i m_i \qquad i=1,2,\cdots \tag{4.32}$$

则有

$$Gm=\sum_{i=1}^{\infty} b_i\lambda_i m_i=\sum_{i=1}^{\infty}(b_i,m)\lambda_i b_i \tag{4.33}$$

(4.33)式叫做算子 G 的谱分解。进行算子的谱分解要求用正交化方程找出(4.31)式中的本征向量$\{m_i\}$及相应规格化正交基$\{b_i\}$，下面就来看看数据带有高斯分布的误差的情况。

对于带有高斯分布的误差的实际数据$\{\tilde{d}_i\}$、如果用每个观测值的标准差归一化，则线性反问题(4.25)可以写为

$$\frac{\tilde{d}_i}{\sigma_i}=\int \frac{G_i(y)}{\sigma_i}m(y)\mathrm{d}y \qquad i=1,2,\cdots,n$$

或记为

$$\tilde{d}'_i=\int G'_i(y)m(y)\mathrm{d}y \qquad i=1,2,\cdots,n \tag{4.34}$$

其中 $\tilde{d}'_i$ 是具有单位方差的无量纲的量，$G'_i(y)=G_i(y)/\sigma_i$。如果按下式定义一个矩阵 Γ 的元

$$\Gamma_{ij}=\int G'_i(y)G'_j(y)\mathrm{d}y \tag{4.35}$$

则矩阵 Γ 便是对称正定的，因此可以用正交矩阵 $\boldsymbol{V}$ 把它对角化(参考第二章第四节)

$$\boldsymbol{V}^{\mathrm{T}}\Gamma\boldsymbol{V}=\boldsymbol{\Lambda}$$

其中

$$\boldsymbol{\Lambda}=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n) \tag{4.36}$$

而

$$\lambda_1\geqslant\lambda_2\cdots\geqslant\lambda_n>0$$

为矩阵 Γ 的本征值，它不等于零表示我们不考虑 $G_i(y)$和 $G_j(y)$线性相关的情况。全部本征值被称为在算子 G 中的谱。注意(4.34)是用数据的标准差加权的，因此 λ_i 不仅与数据核 $G(y)$有关，而且还和数据的误差估计有关。

获得了线性反问题(4.34)的谱之后，对于每一个本征值 λ_i，都可以构成一个函数

$$\psi_i(y)=\lambda_i^{-1/2}\sum_j V_{ji}G'_j(y) \qquad i=1,\cdots,n \tag{4.37}$$

于是，我们便得到了一个函数集$\{\psi_i\}$。由于 $\boldsymbol{V}$ 是正交矩阵，由(4.37)不难证明

$$\int \psi_i\psi_j\mathrm{d}y=\delta_{ij}$$

即$\{\psi_i\}$是规格化正交函数系。利用(4.31)对于问题(4.25)中的模型 m 作谱分解便有

$$m=\sum_{i=1}^{n} C_i\psi_i \tag{4.38}$$

如果有的本征值为零，如 $\lambda_n=0$，则应把上式改为

$$m=\sum_{i=1}^{n} C_i\psi_i+\psi^*$$

其中 ψ^* 表示对应 G 的零空间中满足 $\int \psi^* \psi_i dy = 0$ 的正交函数。同时，由(4.33)式和(4.37)式可知，(4.38)式中的系数应为

$$\begin{aligned} C_i &= \int \psi_i m dy \\ &= \int \lambda_i^{-1/2} \sum_j V_{ji} G'_j(y) \cdot m(y) dy \\ &= \lambda_i^{-1/2} \sum_j V_{ji} \int G'_j(y) m(y) dy \\ &= \lambda_i^{-1/2} \sum_j V_{ji} \bar{d}'_j \end{aligned} \tag{4.39}$$

利用线性反问题模型的谱分解(4.37)和(4.38)可知，对应系数 C_i 比较大的正交函数 ψ_i 表示了反演时可由数据准确确定的可靠成分，而对应系数很小的函数项表示了由数据不能肯定的成分，即统计上不可靠的成分。这样谱分解以明显的方式把反演时模型的确定成分和不可靠成分区分开。由于模型 m 在统计上不可靠的成分也就是有限位数值计算中造成算法不稳定的成分，因此谱分解是研究算法稳定性的一种常用的技术，1971年由 Gilbert 首先应用到地球物理中来[48]，然后又为 Jackson(1972)，Wiggins(1972)等人应用到广义反演中去[20-21]。

将模型 m 分解为正交函数系 $\{\psi_i\}$ 之后可以发现，最大的本征值与(4.38)中最圆滑的一些函数 ψ_i 相对应，而且随 i 的增加 $\psi_i(y)$ 变得越来越振荡。模型 $m(y)$ 中的最圆滑成分也就是可以由数据准确地确定的成分，因为它们的系数 $\{C_i\}$ 是最准确的；而对应后面各项的振荡成分便代表了统计上不可靠的成分。最后的一项中，振荡函数 ψ^* 是完全不能由数据确定的。因为它属于算子的零空间。Parker(1976)曾建议以零空间中最圆滑的振荡函数作为 $\psi^*(y)$，依此为准则研究 ψ^* 的形状[22]，但是目前对 ψ^* 的研究仍很不充分。此外，由于系数 C_i 取决于振荡函数 ψ_i 的性质，当本征值甚小时对应的 C_i 便不确定，估计模型 $m(y)$ 时便存在上述解析上的不稳定性，或者说反问题是不适定的。

为了在估计模型 m 时得到统计上的稳定性，显然要去掉一些对应方差很大的系数 C_i 的正交函数 ψ_i。这样做有点类似于频谱分析中去掉高频分量的"低通滤波"，得到的是真解 $m(y)$ 的圆滑后的解估计，但同时损失了一些分辨率。第三章讨论的 Wiggins 法也是这种谱截断的方法。随机逆等其它方法都对高频振荡成分进行了压制，以提高反演时数值计算的稳定性。

三、非线性反问题的线性化

上节中我们讨论了关于连续地球模型的线性反问题解的存在性、非唯一性及稳定性，对于非线性的反问题，方程(4.10)中的算子 G 为非线性算子，我们不得不涉及目前还不成熟的非线性算子理论。下面我们就从非线性算子的基本概念开始，讨论与非线性地球物理反问题有关的数值求解方法。

(一)非线性算子和 Fréchet 导数

非线性反问题在地球物理反演中是常见的。在位场和电磁资料反演时，所谓地下物体几何形状的反问题一般归结为非线性问题。举一个重力界面反演的例子，设下半空间中存在一

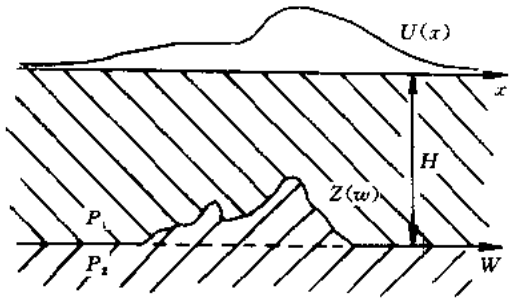


图 4.1 单界面起伏的重力反问题，地面为平面，界面 $Z(w)$ 两边密度均匀

个密度界面 $Z(w)$，w 为界面的水平坐标；其平均埋深为 H（见图4.1）。在二维情况下这个界面起伏产生的重力异常在归一化后为[57]

$$g(x)=\int_a^b \lg\frac{(x-w)^2+H^2}{(x-w)^2+[H-Z(w)]^2}\mathrm{d}w \tag{4.40}$$

反问题是由有限个 $g(x)$ 的观测值 $\{g_i\}$ 估计界面形状 $Z(w)$。这个反问题不仅在油气勘探和大地构造研究中有重要意义，而且也是不适定的非线性反问题的典型例子。把(4.40)改写为

$$A[Z(w)]=\int_a^b K[x,w,Z(w)]\mathrm{d}w-g(x) \tag{4.41}$$

则(4.40)可表示成为

$$AZ=0$$

其中 Z 由(4.40)式表示，A 表示对应的非线性算子，即作用于函数 $Z(w)$ 的非线性运算。在地球物理反演的许多情况下，设地球模型为 $m(x)$，$x\in\Omega\subset\mathscr{R}^3$，非线性算子 A 可以表为

$$A[m(x)]=\int_\Omega K[x,y,m(y)]\mathrm{d}y+g(x) \tag{4.42}$$

从应用的角度看非线性算子的研究主要与非线性微分或积分方法求解有关。与线性算子的研究不一样，非线性方程求解（包括数值解）问题远远没有解决。虽然对于某些非线性方程来说解的存在性、唯一性和稳定性问题已得到证明，但是工程上提出的许多非线性方程求解问题实际还没有完满的解答。非线性算子的应用研究可以从两个方面着手。一种方法与线性问题一样，用逐次逼近求解，其基础是非线性算子理论中的收缩映射原理（Contraction Mapping Principle）。另一种方法是把牛顿迭代算法推广到无穷维空间，这就要求对算子的“导数”给出合适定义。Fréchet 导数就是由此产生的[47]。

算子的导数可有多种定义，其中最简单而且最有用的一种叫 Fréchet 导数，它是微积分中导数定义推广的结果，在算子理论中起很重要的作用。

定义：设 $\mathscr{B}_1$ 和 $\mathscr{B}_2$ 是两个 Banach 空间，$\mathscr{D}$ 是 $\mathscr{B}_2$ 中的一个开子集而且函数 $f\in\mathscr{D}\subset\mathscr{B}_2$，再设 A 是一个把 $\mathscr{D}$ 映射为 $\mathscr{B}_1$ 的算子。如果存在一个线性算子 F 使得极限

$$\lim_{\|h\|\to 0}(\|A(f+h)-Af-Fh\|/\|h\|)=0 \tag{4.43}$$

则称算子 A 为 Fréchet 可微，而算子 F 为算子 A 对 f 的 Fréchet 导数，也可记为 $A'(f)$。

要补充说明的是上述定义中的 Banach 空间包含了希尔伯特空间（它是定义了内积的 Banach 空间）。设 $\mathscr{B}_2$ 为模型空间 $\mathscr{H}_m$，$\mathscr{B}_1$ 为数据空间 $\mathscr{H}_d$，上述 Fréchet 导数的定义就可用于 BG 理论中研究地球物理反演的场合。此外，(4.43)式要求当 h 以任何方式趋于零时有极限存在，即给定 $\varepsilon>0$，存在一个 $\delta>0$ 使得当 $\|h\|<\delta$ 时有

$$\|A(f+h)-Af-Fh\|\leqslant\varepsilon\|h\| \tag{4.44}$$

或者用无穷小符号记为

$$\|A(f+h)-Af-Fh\|=0(\|h\|) \tag{4.45}$$

回忆一下在微积分中一个函数的导数的定义可知 Fréchet 导数是它的自然推广。设 $\phi(y)$为某一单变量函数，它是 $\mathscr{R}^1\to\mathscr{R}^1$的映射，那么由导数定义知

$$\phi(y+\Delta y)-\phi(y)-\alpha\Delta y=0(|\Delta y|)$$

其中实数 α 是 $\phi(y)$的导数。比较(4.43)和(4.45)可见算子的 Fréchet 导数的来源。

下面来看看在几种情况下的 Fréchet 导数。

(1)设 $A:\mathscr{B}_1\to\mathscr{B}_2$为线性有界算子，则 A 对任意 $f\in\mathscr{B}_1$的 Fréchet 导数就是它自身，因为

$$\|A(f+h)-Af-Ah\|=0$$

(2)设 $\mathscr{D}$ 为 $\mathscr{R}^n$ 空间中的一个开子集而算子 $\phi:\mathscr{D}\to\mathscr{R}^n$，则 ϕ 的 Fréchet 导数就是 Jacobian 矩阵$(\partial\phi_i/\partial y_j)$。实际上，用向量形式记 $y\in\mathscr{D}$ 和 $\phi=(\phi_1(y),\cdots,\phi_n(y))^{\mathrm{T}}$，则由泰勒定理可知 $\phi'(y)$就是 Jacobian 矩阵。

对于无限维空间形如(4.41)的积分算子，其 Fréchet 导数要用(4.45)式来证明，文献〔47〕的第4～6章给出了一些例子。Fréchet 导数的引入给出相当于无限维微积分计算的一个基础，也有相应的中值定理，但要求有算子的范数等深一层的概念，这里不再详细讨论。

(二)非线性反问题的线性化

记 m_1和 m_2为模型空间 $\mathscr{H}_m$ 中对非线性算子 G 的定义域 $\mathscr{M}(G)$中的两个元，而且为变量 $y\in[a,b]$的函数，则 $\mathscr{H}_m$ 空间的内积可表为

$$(m_1,m_2)=\int_a^b m_1(y)m_2(y)\mathrm{d}y$$

考虑有限精确数据的情况，记 G_i 为对应数据 d_i 的非线性泛函，则有

$$d_i=G_i(m),\quad i=1,\cdots,n \tag{4.46}$$

令 δm 为模型 m 的微小扰动，即

$$m'=m+\delta m \tag{4.47}$$

也是 $\mathscr{H}_m$ 空间定义域 $\mathscr{M}(G)$中的一个单元，如果 G 为 Fréchet 可导的非线性算子，则由 Fréchet 导数的定义(4.44)有

$$G_i(m+\delta m)-G_i(m)=(F_i,\delta m)+\varepsilon(\|\delta m\|) \tag{4.48}$$

其中 $F_i=G'_i$ 为泛函 G_i 的 Fréchet 导数，注意它是线性泛函，由内积定义并可表为

$$(F_i,\delta m)=\int_a^b F_i(y)\delta m(y)\mathrm{d}y \tag{4.49}$$

在(4.48)式中当$\|\delta m\|\to0$时忽略无穷小量 $\varepsilon(\|\delta m\|)$，得

$$(F_i,\delta m)=G_i(m+\delta m)-G_im\qquad i=1,\cdots,n \tag{4.50}$$

记

$$\delta d_i=G_i(m+\delta m)-G_im \tag{4.51}$$

把(4.49)和(4.51)代入(4.50)得

$$\delta d_i=\int_a^b F_i(y)\delta m(y)\mathrm{d}y\qquad i=1,\cdots,n \tag{4.52}$$

此式是一个线性的泛函方程组，也就是非线性反问题线性化的结果。(4.52)式不难给出一个迭代格式，用实际数据与逼近的数据之差$\{\delta d_i\}$来求模型的修改 $\delta m(y)$，由于(4.52)式中方程组为有限个，而 $\delta m(y)$为一个有无限个取值的函数，因此方程组(4.52)的解不唯一。我们

可以把它的求解化为条件极值问题，用拉格朗日乘子法求解估计，即把(4.52)式作为约束，求使 $\delta m(y)$ 的模极小的解估计

$$\|\delta m\|^2=\int_a^b \delta m^2(y)\mathrm{d}y=\min \tag{4.53}$$

用拉格朗日乘子法求以上条件极值问题时，设 $\{\alpha_i\}$ 为拉格朗日乘子，然后求使

$$Q=\|\delta m\|^2+\sum_{i=1}^{n}\alpha_i[(F_i,\delta m)-\delta d_i] \tag{4.54}$$

取极小的解估计，这时条件极值问题已经化为无条件极值的问题。把(4.54)式展开为

$$Q=(\delta m,\delta m)+\Big(\sum_{i=1}^{n}\alpha_i F_i,\delta m\Big)-\sum_{i=1}^{n}\alpha_i\delta d_i \tag{4.55}$$

当取 Q 极小时有

$$\delta m=\sum_{i=1}^{n}\alpha_i F_i \tag{4.56}$$

为求待定系数组可将(4.56)回代到(4.50)式有

$$\begin{aligned}G_i(m+\delta m)-G_i(m)&=\Big(F_i,\sum_{k=1}^{n}\alpha_k F_k\Big)\\&=\sum_{k=1}^{n}\int_a^b F_i(y)\alpha_k F_k(y)\mathrm{d}y\end{aligned} \tag{4.57}$$

因此

$$\delta d_i=\sum_{k=1}^{n}(F_i,F_k)\alpha_k,\quad i=1,\cdots,n \tag{4.58}$$

此式是一个 n 阶线性方程组，由(4.58)解出 $\{\alpha_k\}$ 之后，由(4.56)式便可以求出 δm。与离散情况类似，我们可以构成以下迭代格式来逐步逼近模型 m：

$$m^{(q+1)}=m^{(q)}+\delta m^{(q)}=m^{(q)}+\sum_{i=1}^{n}\alpha_i^{(q)}F_i^{(q)} \tag{4.59}$$

(三)非线性问题反演的分辨率函数

由于非线性反问题在通过线性化后变为求解线性泛函方程组(4.52)

$$\delta d_i=\int_a^b F_i(y)\delta m(y)\mathrm{d}y\qquad i=1,\cdots,n$$

记对于 $y=y_0$ 处的模型修改量的估计为 $\delta m(y_0)$，相应地把数据误差用(4.49)和(4.52)式表为

$$\begin{aligned}\sum_{i=1}^{n}\alpha_i(y_0)\delta d_i&=\sum_{i=1}^{n}\alpha_i(y_0)(F_i,\delta m(y))\\&=\Big(\sum_{i=1}^{n}\alpha_i(y_0)F_i,\delta m(y)\Big)\\&=(A(y,y_0),\delta m(y))\\&=\int_a^b A(y,y_0)\delta m(y)\mathrm{d}y\\&=\delta\tilde{m}(y_0)\end{aligned} \tag{4.60}$$

因此对非线性反问题平均核为

$$A(y,y_0)=\sum_{i=1}^{n}\alpha_i(y_0)F_i(y) \tag{4.61}$$

如果 G 是线性算子，则泛函 G_i 的 Fréchet 导数就是它自身 G_i，因此(4.61)式化为上节讨论过的情况，即(4.28)式。

(四)应用例子

上述方法可直接用于非线性地球物理反问题。Parker 于1970年成功地用它研究了地幔电导率的反问题，成为最早应用 Backus-Gilbert 方法的一个范例[50]。法国学者讨论了广义最小二乘准则在求解非线性反问题时的应用，并给出了形如(4.40)的密度界面反演的数值结果[49]。他们的方法还说明了对于不适定的反问题，给出协方差函数的先验知识对数值求解的稳定性是很有益的。作为 Backus-Gilbert 方法应用的一个比较简单的例子，让我们来讨论一下由重力剖面观测值反演地下密度的例子[51](见图4.2)。在离散情况下这个反问题多用第三章第五节讨论过的线性规划方法求解[37]，因为这个问题导致的方程组总是欠定的。

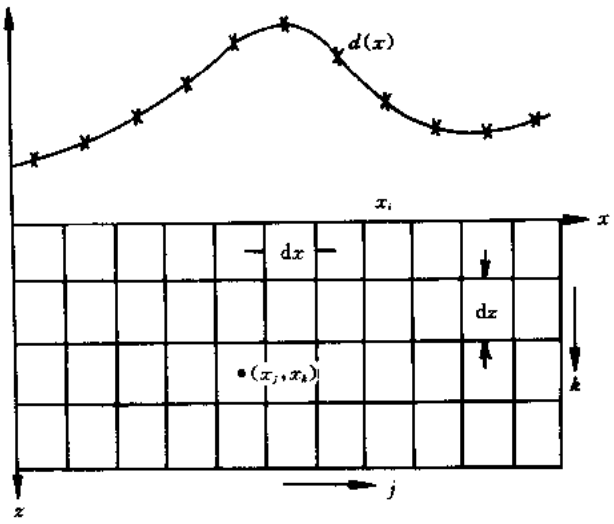


图 4.2 重力剖面反演

设观察在地面进行($z=0$)则合成重力数据可表为

$$\mathrm{d}(x)=\int_{-\infty}^{\infty}\int G(x,x',z')\rho(x',z')\mathrm{d}x'\mathrm{d}z' \tag{4.62}$$

其中数据核为

$$G(x,x',z')=\frac{2\gamma z'}{(x-x')^2+z'^2}$$

γ 为万有引力常数。把地下介质网格化为均匀的矩形网格，中心为(x_j,z_k)，$x_j=j\mathrm{d}x$，$z_k=k\mathrm{d}z$，$j=1,2,\cdots,J$，$k=1,2,\cdots,K$；对于有限数据 $\mathrm{d}(x_i)$，$i=1,\cdots,n$，(4.62)可写为以下方程组

$$d_i=\mathrm{d}(x_i)=\sum_j\sum_k G(x_i,x_j,z_k)\rho(x_j,z_k) \tag{4.63}$$

其中

$$G_{i,j,k}=\frac{2\gamma z_k\mathrm{d}x\mathrm{d}z}{(x_i-x_j)^2+z_k^2}$$

现在可以用第三章讨论的广义反演方法解线性方程组(4.63)，这个方程组总是欠定的。如果把密度 $\rho(x',z')$ 当作连续模型来讨论，可以用以下方法。

用 BG 方法求解估计时，由于 G 本身是线性的，因此它的 Fréchet 导数就是它本身。把(4.63)写为

$$d_i=(G_{ijk},\rho_{jk}) \qquad i=1,\cdots,n \tag{4.64}$$

这是求地下密度参数集 $m=\{\rho_{jk}\}$ 的线性反问题。用迭代格式给定模型的初始猜测 m_0，我们在约束条件(4.64)下求

$$\begin{aligned}\|\delta m\|^2&=\|m-m_0\|^2\\&=\int_v[\rho(x,z)-\rho_0(x,z)]^2\mathrm{d}v\end{aligned} \tag{4.65}$$

的极小来取得模型的修改 $\delta m=\rho(x,z)-\rho_0(x,z)$。记 $g_i(m_0)$ 为用 m_0 和(4.62)式计算的重力值，有

$$\begin{aligned}d_i-g_i(m_0)&=(G_i,m)-(G_i,m_0)=(G_i,\delta m)\\&=\int_v G_i\delta m\mathrm{d}v \qquad i=1,\cdots,n\end{aligned} \tag{4.66}$$

用拉格朗日乘子法求解以上条件极值问题，即是令下面的泛函取极小

$$\begin{aligned}Q&=\|m-m_0\|^2+\sum_{i=1}^{n}2\alpha_i[d_i-g_i(m_0)-(G_i,\delta m)]\\&=\int_v\delta m^2\mathrm{d}v-2\sum_{i=1}^{n}\alpha_i\int_v G_i\delta m\mathrm{d}v+2\sum_{i=1}^{n}\alpha_i[d_i-g_i(m_0)]\end{aligned}$$

其中 $\{2\alpha_i\}$ 为拉格朗日乘子，令 Q 对 δm 的导数为零，有

$$\int_v\delta m\mathrm{d}v=\int_v\sum_{i=1}^{n}\alpha_i G_i\mathrm{d}v \tag{4.67}$$

即

$$\delta m=\sum_{i=1}^{n}\alpha_i G_i \tag{4.68}$$

(4.68)式回代入(4.66)得

$$\begin{aligned}d_i-g_i(m_0)&=\int_v G_i\sum_{k=1}^{n}\alpha_k G_k\mathrm{d}v\\&=\sum_{k=1}^{n}\alpha_k(G_i,G_k),\quad i=1,\cdots,n\end{aligned} \tag{4.69}$$

此式中 (G_i,G_k) 为内积矩阵，因此它给出了求 n 个未知数 $\{\alpha_k\}$ 的 n 个线性方程组。求出待定系数之后便可代入(4.68)式求地下各处密度的修改值，从而构成迭代格式由对密度的初始猜测 m_0 获得地下构造。与广义反演不同之处是从(4.65)到(4.69)各式中 m 和 δm 是函数，也就是连续介质模型的情况。当然在实际计算时网格总是有限的，因此本例中 BG 方法的应用并没有充分体现 BG 理论的优点，这些优点主要体现在对解估计的评价上，这就是以下三节要讨论的内容。

四、脉冲度准则

在本章第二节中已讨论过，有限数据的地球物理反问题具有高度的非唯一性。在第三节中又讨论非线性反问题和利用 Fréchet 导数构造出解估计的一种方法。但是，由于反问题解不唯一，仅仅构造出一个解估计并不能说明它的质量，更重要的是对解估计进行评价。在

Backus-Gilbert 理论中，解估计的评价标准是分辨率和方差的大小。由于非线性反问题在用 Fréchet 导数线性化后得到的方程组与线性反问题相似，因此讨论线性评价就可以了。以下三节分别讨论 BG 线性评价的三条准则，即脉冲度准则，展布准则和折衷准则，先从脉冲度准则开始。脉冲度准则原来称为 δ 准则，又分为两种情况，下面分别讨论之。

(一)Direchlet 第一准则

从上几节已知，无论是线性反问题还是非线性反问题，按 Backus-Gilbert 的假设条件，都可以转化为求待定系数组$\{\alpha_i\}$，$i=1,\cdots,n$，然后由(4.29)得到

$$\hat{m}(y_0)=\sum_{i=1}^{n}\alpha_i(y_0)d_i \tag{4.70}$$

它是模型函数 $m(y)$在任意点 y_0处的某种平均估计值。对于非线性反问题只要用模型的修改估计 $\delta\hat{m}$ 代替 $\hat{m}$，用拟合差$\{\delta d_i\}$代替$\{d_i\}$，并用 Fréchet 核代替 G_i，则可用(4.70)式构成迭代格式。由于有限数据反问题的多解性，系数组$\{\alpha_i\}$也是不唯一的，它的确定取决于求解估计 $\hat{m}(y_0)$所用的准则。由于 y_0可变，因此任意稠密地选取 y_0可确定连续函数 $\hat{m}(y)$，这是与第三章讨论的离散模型所不同的地方。

Backus-Gilbert 首先提出的确定系数组$\{\alpha_i\}$的准则是从解估计的分辨率入手的。回忆第二节讨论的平均函数

$$A(y,y_0)=\sum_{i=1}^{n}\alpha_i(y_0)G_i(y) \tag{4.71}$$

对于非线性问题，$G_i(y)$表示 Fréchet 核。当数据为无限多时，分辨率函数 $A(y,y_0)$将变为 Dirac-δ 函数 $\delta(y-y_0)$，此时解估计 $\hat{m}(y)$变为 $m(y)$。因此，在要求高分辨率的解估计时，$A(y,y_0)$应尽量接近广义函数 $\delta(y-y_0)$，这就是脉冲度准则的出发点。

Direchlet 第一准则是根据最小平方原理使以下分辨率函数 A 和 $\delta(y-y_0)$的差的范数取极小

$$\begin{aligned}\psi_1&=\int_a^b[A(y,y_0)-\delta(y-y_0)]^2\mathrm{d}y\\&=\int_a^b\Big[\sum_{i=1}^{n}\alpha_i(y_0)G_i(y)-\delta(y-y_0)\Big]^2\mathrm{d}y\end{aligned} \tag{4.72}$$

令 $\partial\psi_1/\partial\alpha_k=0$，$k=1,\cdots,n$，则有

$$\frac{\partial\psi_1}{\partial\alpha_k}=\int_a^b 2\Big[\sum_{i=1}^{n}\alpha_i(y_0)G_i(y)-\delta(y-y_0)\Big]G_k(y)\mathrm{d}y=0$$

即

$$\sum_{i=1}^{n}\alpha_i(y_0)\int_a^b G_i(y)G_k(y)\mathrm{d}y=G_k(y_0) \tag{4.73}$$

把内积矩阵 Γ 记为

$$\Gamma_{ik}=\int_a^b G_i(y)G_k(y)\mathrm{d}y \tag{4.74}$$

则(4.73)式可表为线性方程组

$$\Gamma\alpha=g \tag{4.75}$$

其中

$$g^{\mathrm{T}}=(G_1(y_0),\cdots,G_n(y_0)) \tag{4.76}$$

为数据核或 Fréchet 核在 y_0处取值组成的向量。

上面讨论了以分辨率函数尽量趋近于 δ 函数为准则构造反问题解估计的方法，这里用的准则被称为脉冲度(Deltaness)准则。但是，由于函数 $\delta^2(y-y_0)$在 $y=y_0$时为无穷大，我们无法由(4.72)计算 A 对 $\delta(y-y_0)$的接近程度 ϕ_1，因此不能用 ϕ_1直接评价解估计的分辨率。为了近似地估计分辨率，可以假定分辨率函数是限带的 δ 函数 $\tilde{\delta}(t)$，它表为(参数见图4.3)

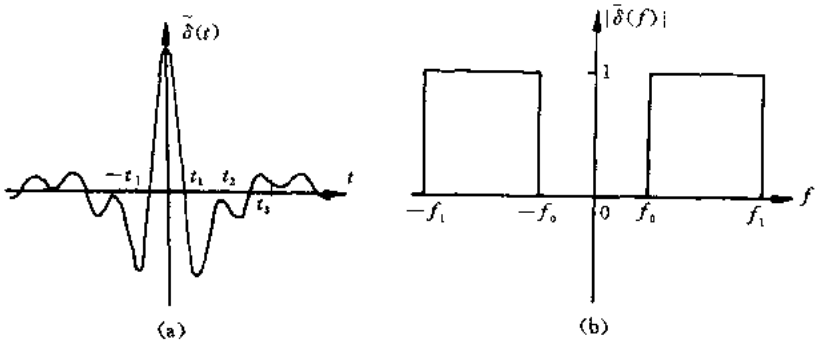


图 4.3 限带 δ 函数

(a)时域表示；(b)振幅谱

$$\begin{aligned}\tilde{\delta}(t)&=[\sin 2\pi f_1 t-\sin 2\pi f_0 t]/\pi t\\&=2\{\sin[\pi(f_1-f_0)t]\cos[\pi(f_1+f_0)t]\}/\pi t\end{aligned}\tag{4.77}$$

由(4.77)式不难看出 $\tilde{\delta}(t)$在 $t=0$时具有极大值

$$\tilde{\delta}(0)=2(f_1-f_0)$$

同时，$\tilde{\delta}(t)$函数最接近原点的零点为

$$t_1=\pm[2(f_1-f_0)]^{-1}\tag{4.78}$$

由于峰值 $\tilde{\delta}(0)$越大及 t_1越小表示 $\tilde{\delta}(t)$越接近于 δ 函数，因此，用以下商的百分比

$$W=\frac{f_1-f_0}{f_1+f_0}(\%)$$

可以衡量 $\tilde{\delta}(t)$对 $\delta(t)$的接近程度。由上式定义的百分比 W 称为百分比宽度，它与 $\tilde{\delta}(t)$的极大值成正比

$$W=\frac{\tilde{\delta}(0)}{2(f_1+f_0)}$$

为了衡量分辨率函数 $A(y,y_0)$与 $\delta(y-y_0)$的接近程度，可以由(4.78)式利用第一个零点的半宽度来表示 $A(y,y_0)$的局部化程度

$$T(y_0)=\frac{1}{2(f_1+f_0)}=\frac{W}{A(y,y_0)}$$

由于 W 是一个接近于1的常数❶，因此可取

$$T(y_0)\simeq A^{-1}(y_0,y_0)\tag{4.79}$$

(二)阶跃函数

基于分辨率函数向 δ 函数逼近的第一准则可以利用 δ 函数和阶跃函数的关系转化为新

❶ 例如在地震勘探中可取 $f_0=8$Hz，$f_1=128$Hz，因此 $W=0.88$。

的形式，这就是 Direchlet 第二准则。表示间断的阶跃函数有以下三种形式：

(1)Heaveside 阶跃函数

$$H(x)=\begin{cases}0, & x<0\\ 1, & x>0\end{cases} \tag{4.80}$$

当 $x=0$时 $H(x)$函数没有定义。Heaveside 阶跃函数是 Dirac-δ 函数的以下不定积分

$$H(x-x_0)=\int_{-\infty}^{x}\delta(u-x_0)\mathrm{d}u \tag{4.81}$$

因为

$$H(x-x_0)=\begin{cases}0, & x<x_0;\\ \int_{-\infty}^{\infty}\delta(u)\mathrm{d}u=1, & x>x_0\end{cases}$$

(2)单位阶跃函数

$$U(x)=\begin{cases}0 & x<0\\ 1/2 & x=0\\ 1 & x>0\end{cases} \tag{4.82}$$

单位阶跃函数与单位脉冲函数相对应，而且可用以下极限形式表示

$$U(x)=\lim_{\alpha\to\infty}\left\{\frac{1}{2}+\frac{1}{\pi}\arctan(\alpha x)\right\} \tag{4.83}$$

(3)符号函数

$$\operatorname{sgn}x=\begin{cases}-1 & x<0\\ 0 & x=0\\ 1 & x>0\end{cases} \tag{4.84}$$

容易证明

$$\operatorname{sgn}x=2U(x)-1$$

而且符号函数还可用以下极限形式表为

$$\operatorname{sgn}x=\lim_{\alpha\to\infty}\left\{\frac{2}{\pi}\arctan(\alpha x)\right\}$$

(三)Direchlet 第二准则

根据脉冲度准则我们希望分辨率函数 $A(y,y_0)$尽量接近 δ 函数，由于 $\delta(y-y_0)$是 Heaveside 阶跃函数 $H(y-y_0)$的不定积分，因此第一准则等效于令 $A(y,y_0)$的不定积分尽量接近于 $H(y-y_0)$。同样由最小二乘原理，可选目标函数

$$\psi_2=\int_a^b\left[\int_a^y A(u,y_0)\mathrm{d}u-H(y-y_0)\right]^2\mathrm{d}y \tag{4.85}$$

取极小，这就是 Direchlet 第二准则。

将(4.71)式代入(4.85)式，由于

$$\begin{aligned}\int_a^y A(u,y_0)\mathrm{d}u&=\int_a^y\sum_{i=1}^{n}\alpha_i(y_0)G_i(u)\mathrm{d}u\\ &=\sum_{i=1}^{n}\alpha_i(y_0)U_i(y)\end{aligned}$$

其中

$$U_i(y)=\int_a^y G_i(u)\mathrm{d}u \tag{4.86}$$

因此有

$$\psi_2=\int_a^b\left[\left(\sum_{i=1}^n\alpha_iU_i\right)^2-\left(2\sum_{i=1}^n\alpha_iU_i\right)H(y-y_0)+H^2(y-y_0)\right]\mathrm{d}y$$

$$=\sum_{i=1}^n\sum_{k=1}^n\alpha_i\alpha_k\int_a^bU_i(y)U_k(y)\mathrm{d}y-2\sum_{i=1}^n\alpha_i\int_{y_0}^bU_i(y)\mathrm{d}y+(b-y_0) \tag{4.87}$$

把上式写成矩阵形式有

$$\psi_2=\alpha^{\mathrm{T}}\Gamma\alpha-2\alpha^{\mathrm{T}}\beta+c$$

其中内积矩阵Γ的元为

$$\Gamma_{ik}=\int_a^bU_i(y)U_k(y)\mathrm{d}y$$

及

$$c=(b-y_0)$$

$$\beta_i=\int_a^bU_i(y)\mathrm{d}y$$

而向量α包含待定系数组$\{\alpha_i\}$，$i=1,\cdots,n$。令$\partial\psi_2/\partial\alpha^{\mathrm{T}}=0$便得

$$\Gamma\alpha=\beta \tag{4.88}$$

利用上述方法求出待定系数组$\{\alpha_i\}$之后代入(4.70)式可以构造出一种分辨率取极大的解估计。然而，经验表明，这时分辨率函数$A(y,y_0)$不是单位模的，即不满足总权为单位的条件：

$$\int_a^bA(x,x_0)\mathrm{d}x=1 \tag{4.89}$$

加上单位模条件后可以证明，如果用ψ_2极小求解估计则可以得到相应于(4.79)式的分辨宽度

$$T(y_0)=\frac{1}{12A(y_0,y_0)} \tag{4.90}$$

与(4.79)式比较可知，只要将(4.85)式中ψ_2乘以12再取极小，则(4.90)式化为(4.79)式，δ准则的两种变种都可以用相同的分辨宽度$T(y_0)$来评价局部化程度及解估计的分辨率。

本节讨论的δ准则和第三章第三节中提过的用分辨率矩阵的展布求离散模型的广义逆是对应的。但是 Backus-Gilbert 发现类似于第三章定义的展布并不能全面表现平均函数对δ函数的偏离，因此又把展布的定义作了推广，这就是在下节中要讨论的 BG 展布。

五、BG 展布准则

(一)BG 展布函数

在第三章中已讨论了展布的定义，它定义为分辨率矩阵与单位矩阵的差的范数2，但是必须注意到，这种展布的定义并不完全是分辨率好坏的最合适的度量，因为它对所有分辨率矩阵$\boldsymbol{R}$的非对角线元素给于相等的权：

$$\mathrm{spread}(\boldsymbol{R})=\sum_{i=1}^j\sum_{j=1}^j[R_{ij}-I_{ij}]^2$$

实际上，非对角线元素的展布不仅与 R_{ij} 的大小有关，还和它们与对角线元素的间距有关。在图4.4中(a)(b)二种情况下依上式计算的展布是一样的，而显然非对角线元素的散布情况却不一样。如果我们把不为零的非对角线元素称为"侧瓣"，如侧瓣越靠近对角线元素(情况a)，则分辨率越好，因为这时解估计对应于邻近元素的局部平均。侧瓣离的远时便不再是局部的平均。如果侧瓣与对角线元素反符号则说明 $\boldsymbol{R}$ 的行对模型起的是"反平均"的作用，这是在找广义逆时不希望出现的情况。

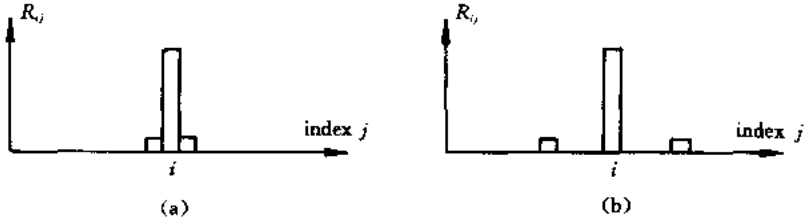


图 4.4　局部化的分辨率矩阵(a)和非局部化的分辨率矩阵(b)

Backus 和 Gilbert 注意到了展布的局部化问题，并建议用相对对角线元素的"距离"对非对角线元素加权，即用 $W(i,j)=(i-j)^2$ 对 $(R_{ij}-I_{ij})^2$ 加权，然后再定义展布；这样选择的权因子对于对角线元素的权为零，而对于其它元素的权总是正的，而且对称于 i 和 j。在离散情况下，BG 展布函数写为

$$\begin{aligned}\text{spread}(\boldsymbol{R}) &= \sum_i \sum_j W(i,j)[R_{ij}-I_{ij}]^2 \\ &= \sum_i \sum_j W(i,j)R_{ij}^2 \end{aligned} \tag{4.91}$$

相应在连续模型的情况下，平均核的展布定义为

$$\text{spread}(A)=12\int_a^b (y-y_0)^2[A(y,y_0)]^2\mathrm{d}y \tag{4.92}$$

(4.92)在实际上是分辨率函数的二阶矩的公式。所谓 BG 第二准则就是在单位模的约束条件(4.89)之下使上述展布取极小。(4.92)式中乘上因子12是为了便于计算分辨宽度并与 δ 准则比较。

(二)BG 展布准则的解估计

令平均核的 BG 展布取极小来求构成解估计的待定系数组 $\{\alpha_i\}$ 时，只要把(4.28)式代入(4.92)，并用拉格朗日乘子法及单位模约束条件(4.89)构成以下泛函

$$\psi_3(y_0)=12\int_a^b (y-y_0)^2\Big(\sum_{i=1}^n \alpha_i G_i(y)\Big)^2\mathrm{d}y+\lambda\Big(1-\int_a^b \sum_{k=1}^n \alpha_k G_k(y)\mathrm{d}y\Big) \tag{4.93}$$

式中 λ 为拉格朗日乘子。展开平方项后有

$$\psi_3(y_0)=\sum_{i=1}^n \sum_{k=1}^n H_{ik}(y_0)\alpha_i(y_0)\alpha_k(y_0)+\lambda\Big(1-\sum_{k=1}^n \alpha_k(y_0)u_k\Big) \tag{4.94}$$

其中

$$H_{ik}=H_{ki}=12\int_a^b (y-y_0)^2 G_i(y)G_k(y)\mathrm{d}y \tag{4.95}$$

和

$$u_k=\int_a^b G_k(y)\mathrm{d}y \tag{4.96}$$

对(4.94)式 ψ_3 取极小，即令 $\partial\psi_3/\partial\alpha_k=0$ 得方程组

$$\sum_{i=1}^{n}\alpha_i H_{ik}=\lambda u_k, k=1,2,\cdots,n \tag{4.97}$$

上式写成矩阵形式为

$$\boldsymbol{H\alpha}=\lambda\boldsymbol{u} \tag{4.98}$$

为了求出拉格朗日乘子 λ，可令 $\partial\psi_3/\partial\lambda=0$，得

$$\sum_{i=1}^{n}\alpha_i u_i=1$$

或写为矩阵形式

$$\boldsymbol{u}^{\mathrm{T}}\boldsymbol{\alpha}=1 \tag{4.99}$$

如果内积矩阵的逆 $\boldsymbol{H}^{-1}$ 存在，则由(4.98)式可知

$$\boldsymbol{\alpha}=\lambda\boldsymbol{H}^{-1}\boldsymbol{u}$$

代入(4.99)式

$$\lambda\boldsymbol{u}^{\mathrm{T}}\boldsymbol{H}^{-1}\boldsymbol{u}=1$$

得

$$\lambda=(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{H}^{-1}\boldsymbol{u})^{-1} \tag{4.100}$$

将(4.100)式回代入(4.98)，有

$$\boldsymbol{H\alpha}=\frac{\boldsymbol{u}}{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{H}^{-1}\boldsymbol{u}}$$

因此待定系数组可写为

$$\alpha=\frac{\boldsymbol{H}^{-1}\boldsymbol{u}}{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{H}^{-1}\boldsymbol{u}} \tag{4.101}$$

求出待定系数组 $\{\alpha_i\}$ 之后代入(4.70)式由数据求最小 BG 展布的解估计 $\hat{m}$，同时把(4.100)和(4.101)代入(4.93)式求 ψ_3，它是评价解估计局部化的分辨率的主要指标。

如上所述利用平均核局部化导致 spread(A)的二阶矩取极小，与脉冲度准则相比它的解估计比较光滑，但是分辨率相对低一些。与正则化方法比较，由 BG 展布准则导出的解估计与正则化后的数值解比较接近(在吉洪诺夫正则化方法中微分正则化算子也起光滑作用)。注意到(4.95)式中的权$(y-y_0)^2$ 随 y_0 而变化，即每一个 y_0 值都要计算一次矩阵 $\boldsymbol{H}$，因此 BG 展布准则的解估计计算成本比脉冲度准则的情况高。

(三)应用 BG 展布准则的广义逆矩阵

以上讨论了对于对任意一个自变量 y_0 求反问题连续的解估计 $\hat{m}(y)$的方法。反过来，在广义反演的情况下 y 在一开始就被离散化，这时我们就可以在约束

$$\sum_{j=1}^{J}R_{ij}=1 \tag{4.102}$$

的条件下求 BG 展布最小的广义逆 G^{-g}。由(4.91)式把加权的分辨率矩阵的第 k 行展布写为

$$\begin{aligned}\phi_k &= \sum_{l=1}^{J}W(l,k)R_{kl}R_{kl}\\ &= \sum_{l=1}^{J}W(l,k)\Big[\sum_{i=1}^{n}\tilde{G}_{ki}^{g}G_{il}\Big]\Big[\sum_{j=1}^{n}\tilde{G}_{kj}^{g}G_{jl}\Big]\\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{G}_{ki}^{g}\tilde{G}_{kj}^{g}\Big[\sum_{i=1}^{J}W(l,k)G_{il}G_{jl}\Big]\end{aligned}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} G_{ki}^{-g} G_{kj}^{g} [S_{ij}]_k \tag{4.103}$$

其中

$$[S_{ij}]_k = \sum_{l=1}^{J} (l-k)^2 G_{il} G_{jl} \tag{4.104}$$

而对单位模条件(4.102)用广义逆表示有

$$\begin{aligned} \sum_{k=1}^{J} R_{ik} &= \sum_{k=1}^{J} \Big[\sum_{j=1}^{n} G_{ij}^{-g} G_{jk} \Big] \\ &= \sum_{j=1}^{n} G_{ij}^{-g} \Big(\sum_{k=1}^{J} G_{jk} \Big) \\ &= \sum_{j=1}^{n} G_{ij}^{-g} u_j \end{aligned} \tag{4.105}$$

其中

$$u_j = \sum_{k=1}^{J} G_{jk} \tag{4.106}$$

用拉格朗日乘子法对第 k 行定义目标函数为

$$\phi_k = \sum_{i=1}^{n} \sum_{j=1}^{n} G_{ki}^{-g} G_{kj}^{-g} [S_{ij}]_k - 2\lambda \Big[\sum_{j=1}^{n} G_{kj}^{-g} u_j - 1 \Big]$$

其中 2λ 为拉格朗日乘子。在对 ϕ_k 求极小时令

$$\frac{\partial \phi_k}{\partial G_{kp}^{-g}} = 2 \sum_{i=1}^{n} [S_{ij}]_k G_{ki}^{-g} - 2\lambda u_p = 0 \tag{4.107}$$

于是我们得到以下$(J+1)$个线性方程组

$$\begin{pmatrix} [S_{ij}]_k & \boldsymbol{u} \\ \boldsymbol{u}^{\mathrm{T}} & 0 \end{pmatrix} \begin{pmatrix} G_{kp}^{-g} \\ \lambda \end{pmatrix} = \begin{pmatrix} \boldsymbol{0} \\ 1 \end{pmatrix} \tag{4.108}$$

其中第 k 行 G_{kp}^{-g}为向量。解此方程组可求得广义逆 G^{-g}的第 k 行元素组成的向量 G_{kp}^{-g}及拉格朗日乘子 2λ。假定(4.108)的系数矩阵的逆可以用分块矩阵表示为

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{b} \\ \boldsymbol{b}^{\mathrm{T}} & c \end{pmatrix}$$

则知道了逆便可以求出广义逆矩阵的元及 λ：

$$\begin{pmatrix} G_{kp}^{-g} \\ \lambda \end{pmatrix} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{b} \\ \boldsymbol{b}^{\mathrm{T}} & c \end{pmatrix} \begin{pmatrix} \boldsymbol{0} \\ 1 \end{pmatrix} \tag{4.109}$$

现在来求 $\boldsymbol{A}$、$\boldsymbol{b}$ 和 c。由矩阵逆的定义有

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{b} \\ \boldsymbol{b}^{\mathrm{T}} & c \end{pmatrix} \begin{pmatrix} [S_{ij}]_k & \boldsymbol{u} \\ \boldsymbol{u}^{\mathrm{T}} & 0 \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}[S_{ij}]_k + \boldsymbol{b}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{u} \\ \boldsymbol{b}^{\mathrm{T}}[S_{ij}]_k + c\boldsymbol{u}^{\mathrm{T}}\boldsymbol{b}^{\mathrm{T}}\boldsymbol{u} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0}^{\mathrm{T}} & 1 \end{pmatrix}$$

于是有

$$\boldsymbol{A} = [S_{ij}]_k^{-1} (\boldsymbol{I} - \boldsymbol{b}\boldsymbol{u}^{\mathrm{T}}) \tag{4.110}$$

$$\boldsymbol{A}\boldsymbol{u} = [S_{ij}]_k^{-1} (\boldsymbol{I} - \boldsymbol{b}\boldsymbol{u}^{\mathrm{T}})\boldsymbol{u} = 0$$

或

$$\boldsymbol{b} = \frac{[S_{ij}]_k^{-1} \boldsymbol{u}}{\boldsymbol{u}^{\mathrm{T}} [S_{ij}]_k^{-1} \boldsymbol{u}} \tag{4.111}$$

及

$$\boldsymbol{b}^{\mathrm{T}}[S_{ij}]_k+c\boldsymbol{u}^{\mathrm{T}}=\boldsymbol{0}^{\mathrm{T}}$$

或

$$c=\frac{-1}{\boldsymbol{u}^{\mathrm{T}}[S_{ij}]_k^{-1}\boldsymbol{u}} \tag{4.112}$$

把(4.110)—(4.112)代入(4.109)得 BG 展布的广义逆的元为

$$G_{kl}^{-k}=\frac{\sum_{i=1}^{n}[S_{ij}]_k^{-1}u_i}{\sum_{i=1}^{n}\sum_{j=1}^{n}u_iu_j[S_{ij}]_k^{-1}} \tag{4.113}$$

将(4.113)回代(4.103)可以求出对第 k 个模型参数 m_k 的 BG 展布，这样就可以在不依赖于数据的情况下评价解估计的分辨率。

(四)展布准则用于位场向下延拓的例子

类似于第三章第六节，展布准则可直接用于位场延拓问题[52]。如已知二维调和函数 $u(x,z)$ 在上半空间 n 个点 (x_i,z_i) 的观测值，$z_i>0$，求函数 $u(x,0)$。设 $z=0$ 为向下延拓平面，而场源位于 $z\leqslant 0$ 的半空间。这里地球数据为 $u(x_i,z_i)$，$i=1,\cdots,n$。

地球模型 $u(x,0)$ 为函数，数据可表为线性泛函

$$u(x_i,z_i)=\int_{-\infty}^{\infty}G_i(x)u(x,0)\mathrm{d}x \tag{4.114}$$

忽略了比例因子，数据核为

$$G_i(x)=\frac{z_i}{\pi[(x-x_i)^2+z_i^2]}$$

由 BG 理论可知，对任一个固定点 $(x_0,0)$，$u(x_0)$ 是不能唯一地确定的，我们只能得到它的一个平均 $\hat{u}(x_0)$，平均函数为

$$A(x,x_0)=\sum_{i=1}^{n}\alpha_i(x_0)G_i(x) \tag{4.115}$$

而

$$\hat{u}(x_0)=\int_{-\infty}^{\infty}A(x,x_0)u(x,0)\mathrm{d}x$$

用 BG 展布准则，我们要令 A 的二阶矩极小，然后求待定系数组 $\{\alpha_i\}$。由(4.93)可设目标函数为

$$\begin{aligned}\psi_3=&12\int_{-\infty}^{\infty}[(x-x_0)A(x,x_0)]^2\mathrm{d}x\\&+\lambda\left[\int_{-\infty}^{\infty}A(x-x_0)\mathrm{d}x-1\right]\end{aligned} \tag{4.116}$$

令

$$\frac{\partial\psi_3}{\partial\alpha_i}=\frac{\partial\psi_3}{\partial\lambda}=0$$

并将(4.115)式代入

$$\begin{aligned}\frac{\partial\psi_3}{\partial\alpha_i}=&24\sum_{j=1}^{n}\alpha_j\int_{-\infty}^{\infty}G_i(x)G_j(x)(x-x_0)^2\mathrm{d}x\\&+\lambda\int_{-\infty}^{\infty}G_j(x)\mathrm{d}x=0,\quad i=1,2,\cdots,n\end{aligned} \tag{4.117}$$

及

$$\frac{\partial \psi_3}{\partial \lambda} = \sum_{i=1}^{n} \alpha_i \int_{-\infty}^{\infty} G_i(x)\mathrm{d}x - 1 = 0$$

(4.117)式有 $n+1$ 个未知数，即 $\{\alpha_i\}$ 和 λ。求解后由(4.70)式可得解估计

$$\hat{u}(x_0,0) = \sum_{i=1}^{n} \alpha_i u(x_i,z_i) \tag{4.118}$$

实际上，将(4.114)式中的数据核表达式代入可以直接计算(4.117)式中的系数矩阵[52]，即

$$\int_{-\infty}^{\infty} G_i(x)G_j(x)(x-x_0)^2\mathrm{d}x$$
$$= \frac{(z_i+z_j)[(x_i-x_0)(x_j-x_0)+z_iz_j]+(x_i-x_j)[z_i(x_0-x_j)-z_j(x_0-x_i)]}{\pi[(z_i+z_j)^2+(x_i-x_j)^2]}$$

讨论到这里，读者也许会问：同是二维位场的延拓问题，第三章第六节讨论的广义反演方法与本节讨论的 BG 展布方法哪个优越。作者做过这方面的比较[6]。以一个垂直断层上的重力异常向下延拓为例(见图 4.5)，试验表明，上述广义线性反演法与用 BG 展布准则导出的延

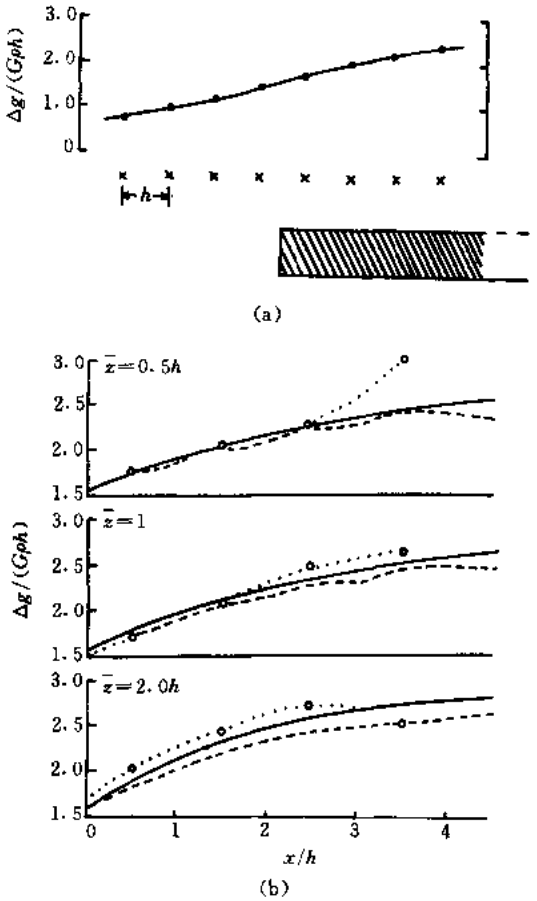


图 4.5　广义反演结果与 BG 展布准则的延拓结果的比较

(a)断层模型和重力异常，×号为取样点；(b)在深度 z 的延拓结果：实线为理论异常，虚线为展布准则的延拓结果[52]，圈点线为广义反演的结果(对 $z=0.5h$ 折衷参数 $\alpha=0$，其它深度 $\alpha=0.15$)

拓方法相比[52]，结果的精度差不多，而前者相对于大数据量的延拓处理而言，计算工作量小得多，因为这时系数矩阵只要做一次奇异值分解。图 4.5 的例子取自 Huestis 和 Parker 的文章[52]，实线为半无限水平板上方的理论重力异常，虚线为他们的延拓结果，圆圈和点线为广义反演的延拓结果。由图可见，在取最佳折衷的情况下这些方法的计算精度大致相同。在图 4.5b 的最上一组曲线上($\tilde{z}=0.5h$)，我们故意令 $\alpha=0$，以表示最小二乘解会导致较大的误差。对于广义反演的结果，由于采取了最佳折衷的措施，这种误差是可以控制的。注意这个试验的例子是在模型参数与数据相等的情况下取得的。如果数据少于模型参数，则基于连续地球模型假设的 BG 方法具有较大的优越性，而广义反演方法获得的结果不如图 4.5 所示的结果。

六、BG 折衷准则

上两节讨论如何根据算子 G 的性质构造和评价分辨率最高的一种解估计，考虑的是精确和有限的数据集。由于实际地球物理数据不可避免地带有测量噪音，还必须考虑由数据噪音给模型估计和评价带来的影响，这种影响包括导致解估计的方差及反问题求解的稳定性两个方面，这两个方面是相互联系的。也就是说，如果解估计方差的范数与数据方差的范数之比是指数关系，说明反问题求解不稳定。用方差来评价解估计隐含着数据和模型的误差属于高斯分布，这是 BG 理论中主要研究的课题。在正则化方法中，考虑在不假定数据和模型的误差属于高斯分布的一般情况下算子方程的稳定性，放到下节进行讨论。

(一)不精确的数据和折衷准则解估计

上面讨论了由于数据有限反问题的解不唯一，只能确定模型的平均 $\hat{m}(y)$。当数据有误差时，解估计的平均也不可能唯一地确定。设实际数据是合成数据 d_i 和误差 δd_i 的和，

$$\tilde{d}_i=\bar{d}_i+\delta d_i,\quad i=1,\cdots,n \tag{4.15}$$

并记数据的协方差矩阵为$\{e_{ij}\}$，则由(4.70)式知解估计可表为

$$\begin{aligned}\hat{m}(y_0)&=\sum_{i=1}^{n}\alpha_i(y_0)(\bar{d}_i+\delta d_i)\\&=\bar{m}(y_0)+\delta\bar{m}(y_0)\end{aligned} \tag{4.119}$$

其中 $\bar{m}(y_0)$为在 y_0 处模型的平均，不受数据误差的影响，而

$$\delta\bar{m}(y_0)=\sum_{i=1}^{n}\alpha_i(y_0)\delta d_i$$

表示由于数据的误差引起的解估计的不确定性。

假定数据的误差为高斯白噪的随机过程而且 δd_i 在统计上独立，则 $\delta\bar{m}(y_0)$的数学期望为零，$\hat{m}(y_0)$的数学期望为 $\bar{m}(y_0)$。但是方差

$$\begin{aligned}\mathrm{Var}[\hat{m}(y_0)]&=\mathrm{Var}[\bar{m}(y_0)+\delta\bar{m}(y_0)]\\&=\mathrm{Var}[\delta\bar{m}(y_0)]\\&=\mathrm{Var}\Big[\sum_{i=1}^{n}\alpha_i\delta d_i\Big]\\&=\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i(y_0)\alpha_j(y_0)e_{ij}\end{aligned}$$

$$= \sum_{i=1}^{n} \alpha_i^2(y_0)\sigma_{di}^2 \tag{4.120}$$

可见解估计的方差同时取决于数据的方差 σ_{di}^2 及系数组 $\{\alpha_i\}$ 的性质。因此在确定待定系数组 $\{\alpha_i\}$ 时不仅要考虑使解估计的分辨率最高，而且还要使解估计的方差尽可能地小。实际计算表明，在按上两节取最高分辨率的解估计时求出的 $\{\alpha_i\}$ 使 $\hat{m}(y_0)$ 的方差达到极大，说明提高分辨率和减小方差是相互矛盾的。高分辨率和小方差两者不可兼得，而只能在二者之间取折衷。根据取折衷的思想，Backus 和 Gilbert 建议采用以下目标函数

$$\psi = \cos\theta\psi_k + \sin\theta \sum_{i=1}^{n} \alpha_i^2\sigma_{di}^2 \tag{4.121}$$

其中 $k=1,2,3$，并由式(4.72)、(4.85)及(4.93)定义 ψ_k；同时 $0 \leqslant \theta \leqslant \frac{\pi}{2}$，当 θ 很小时强调分辨率，而 θ 较大时强调减小方差，θ 称为折衷参数。例如，取数据为单位方差 $\sigma_{di}^2=1(i=1,\cdots,n)$ 以及 BG 展布准则(见 97 页)

$$\psi_k = \psi_3 = 12\int_a^b (y-y_0)^2(A(y,y_0))^2\mathrm{d}y \tag{4.122}$$

时，(4.121)式变为

$$\psi = \cos\theta\alpha^{\mathrm{T}}\boldsymbol{H}\alpha + \sin\theta\alpha^{\mathrm{T}}\alpha \tag{4.123}$$

令 ψ 在规一化模的约束

$$\alpha^{\mathrm{T}}\boldsymbol{u} = \boldsymbol{u}^{\mathrm{T}}\alpha = 1 \tag{4.99}$$

下取极小，便可实现在分辨率和方差之间的折衷，θ 为折衷参数。由(4.95)及(4.96)可计算矩阵的元：

$$H_{ik} = 12\int_a^b (y-y_0)^2 G_i(y)G_k(y)\mathrm{d}y$$

$$u_k = \int_a^b G_k(y)\mathrm{d}y$$

把(4.99)及(4.123)的条件极值问题化为无条件极值问题，有

$$\psi' = \cos\theta\alpha^{\mathrm{T}}\boldsymbol{H}\alpha + \sin\theta\alpha^{\mathrm{T}}\alpha + 2\lambda(1-\alpha^{\mathrm{T}}\boldsymbol{u}) = \min$$

其中 2λ 为拉格朗日乘子。令 $\partial\psi'/\partial\alpha^{\mathrm{T}}=0$ 得

$$\cos\theta\boldsymbol{H}\alpha + \sin\theta\alpha - \lambda\boldsymbol{u} = 0$$

即

$$(\cos\theta\boldsymbol{H} + \sin\theta\boldsymbol{I})\alpha = \lambda\boldsymbol{u}$$

因此待定系数组为

$$\alpha = \lambda(\cos\theta\boldsymbol{H} + \sin\theta\boldsymbol{I})^{-1}\boldsymbol{u} \tag{4.124}$$

令 $\partial\psi'/\partial\lambda=0$ 有

$$\boldsymbol{u}^{\mathrm{T}}\alpha = 1$$

$$\lambda\boldsymbol{u}^{\mathrm{T}}(\cos\theta\boldsymbol{H} + \sin\theta\boldsymbol{I})^{-1}\boldsymbol{u} = 1$$

代入(4.124)得

$$\lambda = \frac{1}{\boldsymbol{u}^{\mathrm{T}}(\cos\theta\boldsymbol{H} + \sin\theta\boldsymbol{I})^{-1}\boldsymbol{u}}$$

最后，回代(4.124)得

$$\alpha=\frac{(\cos\theta\boldsymbol{H}+\sin\theta\boldsymbol{I})^{-1}\boldsymbol{u}}{\boldsymbol{u}^{\mathrm{T}}(\cos\theta\boldsymbol{H}+\sin\theta\boldsymbol{I})^{-1}\boldsymbol{u}} \tag{4.125}$$

当 $\theta=\frac{\pi}{2}$ 时(4.125)式化为(4.101)式。

(二)BG 线性评价

在(4.125)式中,对于不同的折衷参数 θ,可以计算出不同的待定系数组,从而可由(4.70)式得到不同的解估计,这些解估计都在某种程度上可以接受。众多解估计的存在反映了实际数据反演解的非唯一性,重要的是研究所有可接受的解估计的共同性质,它们也是可以由数据获得的唯一的信息,这就是解估计的展布和方差的关系,它是用所谓的折衷曲线表示出来的。

令 θ 从 0 到 $\frac{\pi}{2}$ 之间逐渐变化,对于任一个给定的 θ 值,我们都可以由(4.125)和(4.70)式获得相对于 y_0 处的解的估计 $\hat{m}(y_0)$。然后用(4.79)、(4.90)计算分辨宽度,更常用的是用(4.122)计算平均核的 BG 展布 spread(A),可以知道这个解估计的分辨率。同时,利用(4.120)式我们可以计算出解估计的方差。当数据不完全符合高斯分布时,把(4.120)式写为

$$\sigma^2(y_0)=\mathrm{Var}[\hat{m}(y_0)]=\sum_i\sum_j\alpha_i(y_0)\alpha_j(y_0)e_{ij} \tag{4.126}$$

也可以近似地计算数据协方差矩阵为非对角阵时解估计的方差。于是,对于不同的 θ 值可以作出如图 4.6 所示的折衷曲线。它包括了我们在 y_0 点处由数据可以获得的评价解估计的所有信息。

折衷曲线一般为单调地下降的曲线,它的开始部分一般很陡,说明方差随 θ 的加大而减小很快。在靠近原点的一块区域(如图 4.6 中斜线所示)是折衷曲线不可能达到的区域,它说明想求分辨率和精度同时达到最高的解估计是不可能实现的,最合适的折衷参数 θ 应该最接近坐标原点,因为坐标原点代表分辨宽度和方差都为零的情况,相应最接近原点的 θ 得到的解估计 $\hat{m}(y_0)$ 可以称为最佳折衷解估计。但是取决于核函数 $G_i(y)$,折衷曲线可能变化平缓,这时很难精确地确定最佳折衷参数 θ,它只能由研究者根据上述原则来选。

既然我们在获得模型函数的估计的同时还得到了它的分辨宽度和方差,而且它们又是反映解估计质量的重要参数,因此在表示反演结果时要把它们同时表示出来。于是,反演的

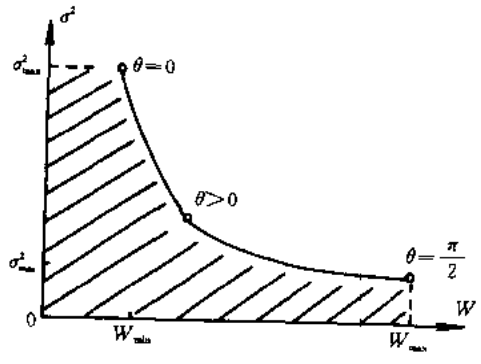


图 4.6 折衷曲线的示意图

斜线所占的区域为不能达到的区域,当取(4.121)式中 $k=1$,2 时横坐标 W 为分辨宽度 $T(y_0)$,当 $k=3$ 时,$W=\mathrm{spread}(A)$

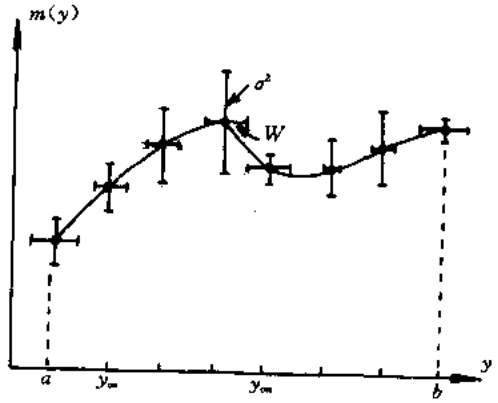


图 4.7 线性反演结果的综合表示

竖线段表示解估计的方差,横线段表示解估计的分辨宽度或展布

结果就可以表示成图 4.7 所示的形式，即用垂直线段表示方差，水平线段表示分辨宽度，它们的交点表示最佳折衷解估计 $\hat{m}(y_0)$。用图(4.7)的形式给出的线性反演结果科学地反映了可以由数据确定的地球模型的全部信息，并综合了反问题解的估计和评价，一目了然地表示出各个模型参数的估计值和它的质量。正如第三章第六节讨论过的，对于不同的反演算法(如在(4.121)式中取不同的 k)，折衷曲线也可用于鉴别各种这些算法的优劣，即看那种方法的折衷曲线最接近原点。同时，增加数据量 n 也可以使折衷曲线向原点靠近，因此折衷曲线还可以用来研究观测系统的合理设计。

最后要强调一点是，BG 理论的折衷思想和取与数据拟合最好的最小二乘思想是有区别的。当取(4.121)式中 $k=1$ 及 $\theta=0$ 时得到的分辨率最高的解估计相当于最小二乘解估计，而 BG 理论认为我们必须牺牲一些解的分辨率，而减少解估计的方差。因为从数据中可以唯一地确定的只是解的某种平均，在一般情况下，解的平均并不是与数据拟合最好的最小二乘解估计。因此，BG 折衷思想比最小二乘思想更全面地揭示了地球物理反演的本质，并包括了最小二乘法作为它的特殊情况。

(三)应用 BG 折衷准则的广义逆矩阵

在离散模型的情况下，评价矩阵方程 $\boldsymbol{d}=\boldsymbol{Gm}$ 的解估计时也可以按 BG 折衷准则选取。设折衷参数为 β，$0\leqslant\beta\leqslant1$，令目标函数为

$$\begin{aligned}\phi&=\beta\,\mathrm{spread}(\boldsymbol{R})+(1-\beta)\,\|\sigma^2\|\\&=\beta\sum_{i=1}^{J}\sum_{j=1}^{J}(i-j)^2R_{ij}^2+(1-\beta)\sum_{i=1}^{J}\sigma_i^2\end{aligned}\tag{4.127}$$

其中 $\boldsymbol{R}$ 为分辨率矩阵，σ_i^2 为模型方差，见式(4.126)。为求 BG 折衷准则的广义逆 $\boldsymbol{G}^{-g}$，把对第 k 行元素中的表达式写为

$$\phi_k=\beta\sum_{l=1}^{J}(k-l)^2R_{kl}^2+(1-\beta)\sum_{l=1}^{J}\sigma_l^2$$

代入(3.43)及(4.126)有

$$\begin{aligned}\phi_k&=\beta\sum_{i=1}^{n}\sum_{j=1}^{n}G_{ki}^{-g}G_{kj}^{-g}[S_{ij}]_k+(1-\beta)\sum_{i=1}^{n}\sum_{j=1}^{n}G_{ki}^{-g}G_{kj}^{-g}e_{ij}\\&=\sum_{i=1}^{n}\sum_{j=1}^{n}G_{ki}^{-g}G_{kj}^{-g}[S'_{ij}]_k\end{aligned}\tag{4.128}$$

其中

$$[S'_{ij}]_k=\beta\sum_{l=1}^{J}(k-l)^2G_{il}G_{jl}+(1-\beta)e_{ij}\tag{4.129}$$

类似于上节(4.103)～(4.113)展布准则的推导，令 $\partial\phi_k/\partial G_{kl}^{-g}=0$，得到 BG 折衷的广义逆的元为

$$G_{kl}^{-g}=\frac{\sum_{i=1}^{n}[S'_{il}]_k^{-1}u_i}{\sum_{i=1}^{n}\sum_{j=1}^{n}u_iu_j[S'_{ij}]_k^{-1}}\tag{4.130}$$

比较(4.130)式和上节(4.113)式可知，它们的不同仅在于内积矩阵的规定上。也就是说应用 BG 折衷准则得到的矩阵 $[S'_{ij}]_k$ 与折衷参数 β 有关，也与数据的协方差 e_{ij} 有关。改变 β 的值并作出折衷曲线对解估计进行评价，最后可以找到最佳折衷的解估计及其分辨率和方差。

(四)小结

通过以上六节的讨论可以看出,Backus-Gilbert 的反演理论首先用函数空间来描述地球模型和资料,并从算子零空间的存在揭示了反问题解的非唯一性。对于有限而精确数据的反演,导出了反问题解肯定不唯一,但是解的平均即可以是唯一的结论,因此解估计的分辨率可以被用作构造和评价解估计的一个准则。由于分辨率还与平均函数是否局部化有关,Backus 和 Gilbert 进一步推广了展布准则以改善解估计的质量。对于有限和带误差的数据,反问题的解可能不存在,即使存在也不唯一,这时要同时用分辨率和方差来评价解估计。但是,由于解估计的方差和分辨宽度不可能同时达到极小,只能通过修改折衷参数的办法求在分辨率和方差之间取最佳折衷的解估计。对于非线性反问题,可以用对非线性算子的 Fréchet 导数把反问题线性化,并构成迭代格式求反问题的解估计。鉴于实际地球物理数据反问题解的高度的非唯一性,对解估计进行评价是不可少的。Backus 和 Gilbert 不仅给出了对反问题解估计评价的理论准则,而且还指出了实现这种评价的具体方法。这些方法不仅可以应用于连续的地球模型的反演,而且还可应用于离散地球模型从而构造出 BG 广义逆。由于 Backus 和 Gilbert 比较系统和严格地研究了地球物理反演的理论和方法,为地球物理反演成为一门新的边缘学科奠定了坚实的基础。

然而,BG 理论的建立还没有解决地球物理反演的所有问题。例如

(1)BG 方法只解决了数据处在泛函方程组右端时解的评价问题,它只是地球物理反问题中的一类。直接从偏微分方程出发提出反问题更能概括地球物理反演的所有情况。

(2)对于非线性反问题,BG 迭代格式的收敛性取决于地球模型的性质和初始猜测,并不能保证收敛。

(3)在 BG 理论中只考虑了服从高斯分布的误差的情况,没有深入探讨解估计对于其它类型数据的依赖关系。

(4)应用 BG 方法评价解估计要求巨大的计算成本,造成它在工程问题上的应用的困难。

比较第三章讨论的广义反演方法和 BG 方法可知,广义线性反演和 BG 线性评价之间存在着一定的对应关系。从历史上看,BG 理论发展在先,地球物理广义反演发展在后,广义反演中的许多思想(如分辨率、折衷参数等等)都是从 BG 理论延伸到广义反演中的。因此,广义线性反演方法实际上是 BG 方法对于离散化模型的变种。下面可以从实用的观点比较一下广义反演和 BG 方法的特点。

对于非线性的反问题,无论是广义反演还是 BG 方法都必须分别按 3.1 节或 4.3 节所述的方法线性化,然后求解估计和进行评价。然而,广义反演方法除了要线性化外,还要把模型参数化。离散化的好处之一是得到一个容易计算的导数矩阵,而在连续情况有时很难确定(4.56)式中要求的 Fréchet 导数。然而,如果离散化做得不适当或者过于粗糙,就会造成灾难性后果,如解的存在性可能受到破坏,模型的自由度受到限制,以及解估计的分辨本领下降等等。这些影响是很难从理论上估计的,因此,BG 理论不仅比较完美,而且用 BG 准则求解和评价更反映了反问题的本质。同时,BG 理论反演结果可以是样点稠密的连续函数。

尽管如此,目前广义反演方法的应用要比 BG 方法普遍得多,其原因之一就在于利用折衷准则求解和评价要求巨大的计算的工作量。由(4.125)式可知,由于 $\alpha_i(y_0)$ 是 y_0 的函数,因此,对模型 $m(y)$ 在每一个 $y=y_0$ 处的估计,其待定系数组 $\{\alpha_i\}$ 都是不同的,这就意味着对

每个 y_0 我们都必须解一次类似于(4.124)的 n 阶方程组(n 为数据量)。如果要用 J 个值表示模型,就要解 J 次(n 阶的)方程组。再者,根据折衷准则,还要在$[0,\pi/2]$之间取不同的折衷参数值。假定 θ 的取样间距为 5°,则要取 19 个 θ 值,要解方程组的次数就变成了 $19\times J$,而在相应广义线性反演的情况下,我们只需要做一次奇异值分解或线性规划就行了。对于非线性问题,要经过迭代才能逼近解的平均 $\bar{m}(y)$。设迭代次数为 10,$J=50$,用广义反演只要分解 10 次奇异值,而且折衷准则要解 $10\times19\times50=9500$ 次 n 阶方程组,这么大的计算量即使对目前的计算机也是不现实的,对于要反复使用的应用地球物理软件是不实际的。相比之下广义反演方法虽然也要求较大的计算工作量,但比折衷准则来评价解估计的计算量还是小得多,用目前的计算机还是可行的。

当然,在数据比较少的时候,对于纯地球物理领域内一次性的科学计算用 BG 方法还是值得的。此外,虽然 BG 理论对线性反问题的解释是比较严密的,但是对于非线性反问题它还没有对迭代的收敛性做出估计,我们也不知道最终求出的解估计在多大程度上依赖于初值和叠加算法。因此对非线性问题还需要进一步深入研究。

从地球物理反问题研究得出的 BG 方法,不仅应用于地球物理,而且推动了计算数学的发展。近今的许多计算数学专著都有专门章节介绍 BG 反演方法,这是地球物理学家的骄傲。然而近 20 年来地球物理学家并没有满足于以 BG 理论为代表的反问题研究的成果,开拓了偏微分方程反问题、非线性反演等许多新领域的研究,在以下的五章中我们将对此逐一讨论。

第五章　地震层析成像

本章将讨论地震层析成像。在科学与工程中许多问题都归结为不可见物体的内部成像问题，计算机层析成像技术(或简称CT技术)，是这一领域的最新进展。层析的意思是指把不可分割的对象假想地切成一系列薄片，分别给出每一个薄片上的物体的图像。如果研究的目标只是在遮挡层里夹的物体形状而不是物体的内部，则称为轮廓成像，它可认为是层析成像的一种特殊情况。常用于物体轮廓成像的所谓全息成像技术(holography)，它也可认为是CT技术中的一种特殊类型。

地震层析成像(或简称地震CT)原先指利用来自不同方向的地震波走时对地球内部波速结构的成像技术，目前它包括所有利用地震波场信息对地球内部的成像技术。

地震层析成像的内容包括数据采集、数据处理和层析成像计算(重建)三部分，我们首先在第一节中简要介绍数据采集的几种方式，并将它们与医学上典型的人体CT作比较。关于数据处理的内容具有一般性，将放到第七章与其它地球物理数据处理一起介绍。

利用地震直达波走时数据进行层析成像计算已经很成熟，其基础之一为射线追踪。第二节讨论射线追踪方法，第三节和第四节分别讨论连续与离散的图像重建。图像重建的理论基础是Radon变换与广义Radon变换，它们将分别在第三节和第五节予以介绍。

在最后两节我们将详细讨论利用波场逆散射原理进行地震层析成像的方法与技术。

一、地震层析成像的数据采集

本节讨论应用于地球内部结构研究的地震层析成像中的数据采集和地球模型，以便进一步找到从许多领域中抽象出来的应用数学理论和地球物理应用之间的连结点。正如BG理论那样，必须对地球物理数据和地球模型作出确切的数学描述才能把反问题提得准确。遗憾的是，由于地震层析成像研究还处于初始阶段，对于数据有限性和不精确性的表示还不像BG理论那样深入，基本上还只是针对理想数据在理论上进行成像方法探讨。因此，我们的讨论也只能从理想数据入手，用自由度来估计数据和模型的特征。下面首先从不同的数据采集系统开始讨论地震层析成像的应用环境。

1. 跨孔和坑道地震CT(图5.1)

利用主频为10kHz～1MHz的声波在井孔间和坑道内测量波的走时和振幅，然后可用CT技术对井孔或坑道之间的介质成像，其应用环境与医学上的CT技术最为接近。因此它们发展得最早[63,64,65]，它们在圈定地下断层破碎带、岩溶、地下埋藏物，和在石油及煤田勘探中取得较好的效果。与地震CT平行发展起来的电磁CT技术利用1～40MHz的电磁波振幅衰减对吸收系数成像，在圈定核废料基地、可采油页岩、煤矿和油气藏内的流体分布等方面也取得了明显的效果[66]。

设图5.1中震源位置为r_s，而接收器位置为$r_g, r_s, r_g \in \mathcal{R}^2$，则接收的地震数据可记为

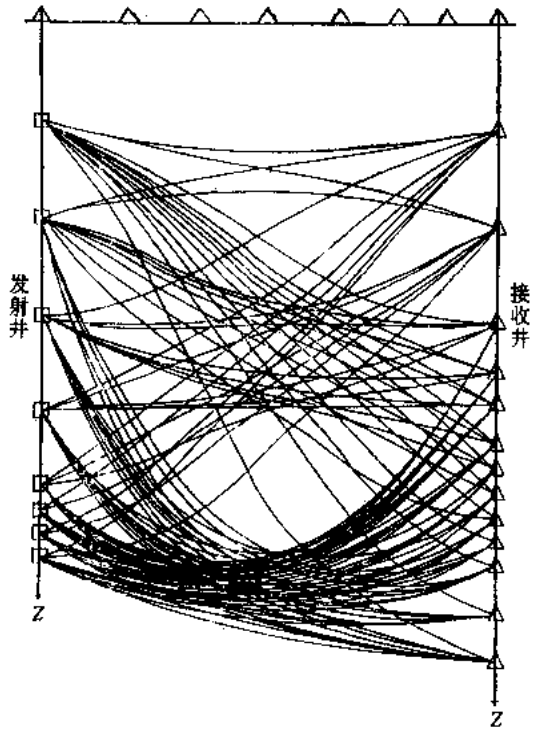


图 5.1　跨孔地震的测量系统
图中细直线表示射线

$u(\boldsymbol{r}_g,\boldsymbol{r}_s,t)$，它为透射声波的波场，可为位移速度或者孔中泥浆对于水听器的压力，对于三分量接收器来说 $\boldsymbol{u}$ 是向量。跨孔和坑道的数据采集装置的特点是 $\boldsymbol{r}_s$ 和 $\boldsymbol{r}_g$ 实际上沿一条曲线（或析线）分布，因此，它们在给定坐标系中可分别表示为角度 θ 和 ϕ 的函数。因此透射波场可写为 $u(\theta,\phi,t)$，自由度为 3；透射波的走时可写为 $\tau(\theta,\phi)$，自由度为 2。和医学上的 X 射线 CT 相比，走时 τ 相当于投影函数，研究区域内的波慢 $N(\boldsymbol{r})$，$\boldsymbol{r}\in\Omega\subset\mathscr{R}^2$，相当于图像函数，它是声速 $C(\boldsymbol{r})$ 的倒数。要注意的是 X 射线 CT 中 X 射线是平面入射的，而在跨孔（或坑道）CT 中震源为三维空间中的点源。此外，由于测量和发射波固定在井孔或坑道中，θ 和 ϕ 一般较小，属于 CT 技术中不完全投影的图像重建问题。假定 $N(\boldsymbol{r})$ 为连续函数，它的自由度为 2，因此不完全投影的 $\tau(\theta,\phi)$ 数据的图像重建问题的解不唯一，因为此时 $\tau(\theta,\phi)$ 的自由度小于 2。另外，也已证明投影重建图像的问题是不适定的数学问题。

X 射线 CT 与地震 CT 之间的最重要区别在于它们所用的波长与探测目标尺寸的比例相差悬殊。在医学应用中，X 射线的平均波长小于 0.1μm，约为探测目标的尺寸的千分之一左右，因此波的衍射效应可以忽略。在跨孔或坑道地震 CT 中声波的平均波长在几厘米以上，这时波的衍射和散射效应起重要作用（图 5.1 为计算射线的一个实例）。因此，在高分辨率地球物理成像的场合下常常不能使用 X 射线中常用的直射线模型。采用更高频率的超声波或应用衍射层析成像技术是提高成像分辨率的主要途径[67]。

广义的跨孔地震层析成像并不要求炮与接收器位于两条平行的直线上，而可呈折线或任意曲线形状。例如，在探测水电站大坝时，大坝的横截面就呈现折线边界（见第十章的例子），而接收器只能放在下游方向大坝的斜坡上，不规则与不均匀的炮一检波器组合并不会给射线追踪与成像计算带来很多的困难，但是这时成像的分辨率是不均匀的，必须在设计炮位和检波器位置时考虑到。如果地球波速随深度增加较快，则大炮检距对应的射线将穿透比较大的深度，而小间距的射线穿过浅部，利用层析成像的原理也可以在剖面上对地球局部地段进行波速成像。这时，如果成像部位是在山丘下方，炮线与检波器线分别位于山的两侧，则炮线与检波器线的交角等于山坡两侧的坡度之和。

跨孔层析成像在应用上的主要问题是数据采集效率不高，因此成本较高。以油气田开发中的应用为例，井距 300m，炮与接收器间距为 5m，设计 100 炮和 100 个接收器分置于两个井孔，共有一万个炮—接收器对，最高可取得 $5\times5m^2$ 分辨率的波速图像。但如果一炮一炮地放采集最少要 20 天左右。油田生产井停产 20 天意味着巨大的经济损失。最近，美国有的公司发明了一种“飞行放炮”的采集系统，多炮同时下井并自动控制放炮的时间间隔，可把采集的时间缩短为三天。这种系统已基本上满足油田开发和强化采油监测的需要。

2. 垂直地震剖面（VSP）成像

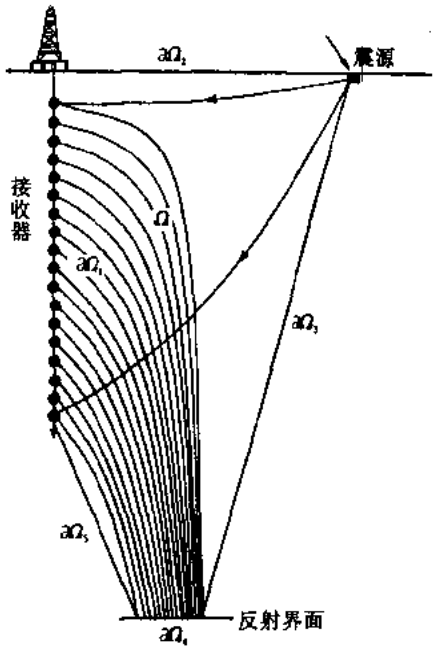


图 5.2 带偏移距的垂直地震剖面

带箭头的曲线表示透射波（下行波）的射线，不带箭头的实线表示来自井孔下方反射波（上行波）的射线，成像的区域包括透射波和反射波经过的区域

如图 5.2 所示，在某一口井中放置接收器，在离井几十到几千米处放置震源（通常为可控震源，产生压力波场），这种地震采集装置叫带偏移距的垂直地震剖面（offset VSP）。当震源固定时接收到的 $u(z,t)$ 为井中水听器检测的压力波场，它包括来自震源的透射波 $u_t(z,t)$ 及来自深层的反射波 $u_r(z,t)$ 两部分，它们可以通过地震处理分解开。对于同时有透射波（也称为下行波）及反射波（上行波）通过的区域 Ω 来说，u_t 和 u_r 相当于两组不同入射方向的透射波场，可用于区域内波慢 $N(r)$，$r\in\Omega$ 的成像。offset VSP 装置的最大优点就是由同时接收透射波和反射波带来的，它可以通过对比上行波的走时和波形变化来了解地层波速和衰减特性的变化[68]。

VSP 装置中采用的震源和接收器一般和反射地震一样，地震记录的通频带在 8～128Hz 之间，比跨孔 CT 装置的频带小得多，相应成像的分辨率也低得多，因为此时地层中的平均波长在 10 米以上。此外，由于同时考虑了反射波，相应的地球模型为带有间断面的连续介质模型，波慢 $N(r)$ 要用广义函数描述（见本节第二段）。这时，由于数据 $u(z,t)$ 的自由度远小于模型的自由度，VSP 成像问题具有高度的非唯一性。

由于 VSP 是介于反射地震和地震测井之间的一种地震方法，它比反射地震的分辨率高，又比地震测井的研究范围大，70 年代以后在油气田的评价和开发研究中应用广泛。

3. 反射地震测量[69]

反射地震法是用于勘查地球内部物质分布的地震学方法中应用得最广而且技术上最成熟的方法。反射地震勘探可定义为"利用从介质间断面上返回地面的初始反射波信号探测地球内部的一种地震学方法。"由于反射地震法传统上只研究纵波的初次反射，而把其他类型的波（如面波，转换波和多次波等）一律当作干扰而在数据处理阶段尽力加以消除，这就使得反射地震勘探在原理上和海洋声纳探测极为相似，并在建立数学物理模型时也常常以声波方程作为出发点。当然，反射地震记录中的折射波、转换波甚至面波等信号也包含有地球内部构造的信息而可以被利用，但目前我们还认为这方面的研究属于散射地震法、折射法或转换波法的研究内容。

根据目前地震数据采集技术的发展和应用环境的区别，反射地震法的典型装置可分为以下三种：

(1)海洋地震勘探：埋在海水中的位移速度型震源和压力型水听器。海上震源一般为气枪或电火花发生器（单极），在海水中产生质点位移速度的扰动，这种扰动在地球中传播并反射回海水层，被由压电材料制成的水听器接收。水听器的输出电压与海水中地震扰动的压力成正比。实际上，水听器并不直接与海水接触，而是一串串地装在充满煤油或其它液体的软管内，呈拖缆形式拖在水面下十多米处。因此它检测的是软管内流体的静水压力，只不过在制作时尽力使软管材料的波阻抗与海水接近，使管内流体的压力与管外海水的压力尽可能一致。

(2)可控震源的陆上地震勘探：放在地面的压力型震源和位移速度型检波器。地面上的震动车向地下发生垂向压力扰动，与介质中发生的无方向性震源（单极）不同，这种震源相当于数学上的偶极源。检波器多是由小质量、磁铁、弹簧线圈的电路组成的换能器，输出电压正比于地球表面的垂向位移速度 V_z。目前西方陆上地震勘探的震源多用此种类型。

(3)炸药震源陆上地震勘探：位移速度型震源和速度型检波器。此时由于炸药多放在浅井中或水中，属于点极型震源，产生位移速度的扰动。

上述用于油气勘探的反射地震震源和接收仪器的频带在 8～128Hz 范围内，用于煤田及工程勘查的高分辨率地震的通频带可扩大到 300Hz 左右。70 年代以来，由于数学地震仪已发展到上千道，数学处理能力也大大地加强，因此三维地震测量得到了迅速的发展。三维反射地震装置可以采集数据量最多的地震数据，其典型的数据采集装置如图 5.3 所示，由于勘探船或地面装置不断地在地面或水面上移动，因此在区域测量的情况下其震源点 r_s 及接收点 r_g 的自由度可认为是 2，接收的波场（海中为压力而陆上为垂向位移速度）可记为 $u(r_s, r_g, t)$，自由度为 5。经过处理后，二维地震数据可以分别表示为其炮点道集 $u(r_g, t)$，其中点道集 $u(r_s, t)$ 或零炮检距的道集，它们的自由度都为 3。如果只利用初次反射走时的数据，则它的自由度比波场减少 0.5，这是由于忽略了振幅信息所致。例如，三维地震反射走时可以表示为 $\{\tau_i(r_g, r_s)\}$，$i=1,\cdots,n$，其中 n 为反射界面总数。由于 r_g 和 r_s 在地面上的自由度为 2，则可以认为 $\{\tau_i\}$ 的自由度为 4.5。

4. 利用天然地震体波记录的层析成像[70]

全球性分布的天然地震台网已经接收了大量地震波记录，它们可用来研究地球内部波

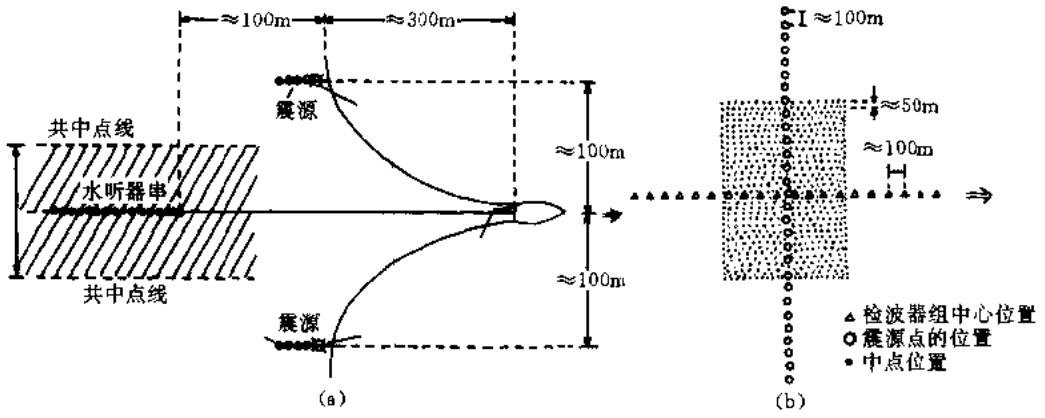


图 5.3　典型的三维地震测量示意图

(a)海上，调查区域为其中虚线之间的阴影区；(b)陆上，调查区域为中点组成的阴影区

速结构。如图 5.4 所示，设 r_g 为沿地球表面分布的地震台位置，自由度为 2。又设 r_s 为天然地震的震中坐标，由于震中环绕研究地区分布，在给定的坐标系内可以表示为角度 ϕ 的函数。如果假设所有选出的地震都发生在同一时刻，对时间归一化后的地震波场(位移速度的三个分量)可写为 $u(r_g,\phi,t)$，自由度为 4。实际上，目前多用直达的透射波走时作为天然地震层析成像的数据，它可记为 $\tau_p(r_g,\phi)$ 或 $\tau_s(r_g,\phi)$ 等，自由度各为 3。由于天然地震记录的通频带较低，主频在几赫兹以内，因此成像的分辨率不高，主要用于地球地幔或地壳内大尺度的横向不均匀性研究。除体波之外，还可以用地震面波作全球性的波速分层成像[70]。由于利用面波对地球内部结构成像的分辨率低，我们不来讨论。

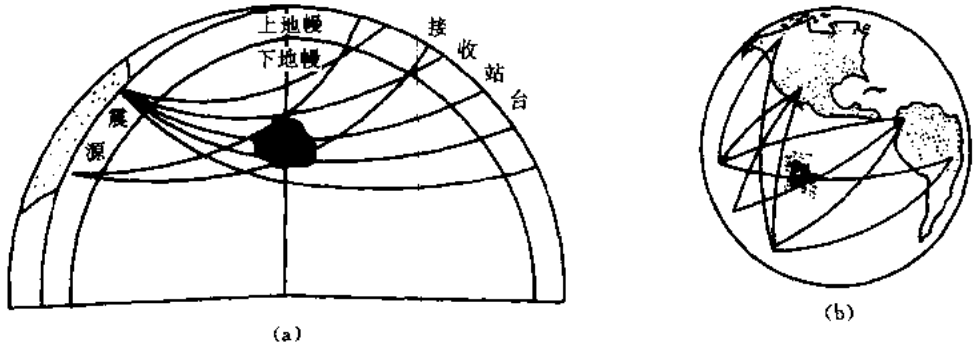


图 5.4　利用天然地震记录研究地幔波速结构的示意图

(a)地球截面图，研究区域位于震源区和接收台站之间；(b)震源和测站在地球表面分布的示意图

上面我们讨论了用于层析成像的四种主要的地震采集系统以及理想地震数据的自由度。与上一章讨论 BG 理论时一样，我们假定这些数据都是可以由求解微分方程正问题计算得到的。然而，实际的地震数据和上述理想数据有很大差别。例如：

(1)实际地震记录中信号的振幅不仅与地球内部结构有关，而且也与接收仪器类型、激发和接收条件等其它因素有关，这就使直接利用波场数据反演地球内部结构发生了困难。利用地震走时数据可以绕开这种困难，但是走时数据的自由度低，因此利用走时反演的层析成

像问题常常具有高度的非唯一性。

(2)实际地震记录中包含有大量干扰,它们不仅是随机的测量误差,而且还包含有大量相干性的干扰(如反射地震中的面波、多次波和绕射波,VSP 中的井筒波)。虽然现代地震数字处理技术提供了许多抑制干扰的方法以提高信噪比,但经过处理的地震数据中的有用信号也受到畸变。

(3)实际地震记录总是有限频带的,说明由于震源和接收仪器的限制,许多反演必须的有用信息根本就没有记录到。实际地震数据中信息量的不足乃是许多理论上可行的反演方法失败的最常见原因之一。有用信息太少常常是地球成像时遇到的最头痛的问题。

(4)实际地震数据总是在地球表面有限范围内经过离散取样获得的,而不像理想情况下的无限稠密取样集。换句话说,实际的数据采集系统都是所谓的"有限曝光"或"有限孔径"的观测系统。因此和 BG 理论中讨论过的情况一样,实际地震数据的反演成像问题的解具有高度的非唯一性。

表 5.1 中总结了医学 CT 和地震 CT 应用的数据采集环境的主要差别[9],这些差别的存在说明地震层析成像远比医学中对人体病灶成像困难,这就是为什么把地震 CT 称为"明天的 CT 技术"的原因。在讨论了地震数据之后,我们将接下去讨论地球模型的数学描述。

表 5.1 医学和地震学 CT 技术应用环境比较表

应用环境	医学 CT	地震体波 CT
1. 波的类型	X 射线或超声波 透过波,散射波	弹性波,包括 P 波、S 波、转换波面波 透过波、散射和多次散射波、绕射波
2. 频率/波长(数量级)	X 射线: 10^{17}Hz,0.01μm 超声波: 10^{6}Hz,mm	地下:10^{4}kHz,cm 地面反射法:$10^{1}\sim10^{2}$Hz,10m 大地构造研究: 1~20Hz,km
3. 探测目标大小(数量级)	cm	地下:m 油气层:10~100m 大地构造:10^{2}m~10^{2}km
4. 周围介质	空气;均匀介质	不均匀岩石介质
5. 源—接收器间距(数量级)	m	地下:$(n\times10\sim n\times10^{2})$m 油气:几百米~几公里 大地构造:$(n-n\times10^{2})$km
6. 源特性	可控单色源	地下:可控单色源 油气:限带的激发源 大地构造:限带的激发源,参数可能无法准确确定
7. 接收信息	X 射线强度、声波走时和振幅	地面(包括人工表面)的振动或振动速度

二、射线追踪方法

地震走时是地震图中最有用的信息之一。与地震图的振幅和波形信息相比,地震走时具有以下特点:(1)它主要取决于地球的波速结构,而与激发-接收条件和仪器性能关系不密

切，因此通常具有较高的信噪比；(2)走时是地震解释员根据地质和地球物理知识在地震图中划分震相之后取得的，因此走时数据本身带有一定的主观随意性；(3)对球面波和柱面波的传播，其振幅衰减的规律是不同的，但走时规律相同．因此利用走时反演地球波速结构的方法可同时用于二维、三维及点源加上二维介质的所谓二维半的情况。地震走时的这些特点使利用走时的层析成像方法具有简单、直观和通用的优点，但是也说明只有在地震解释员正确地分辨地震相和拾取走时的条件下，走时反演才可能成功。可惜的是，地震图中常有不同震相发生干涉的现象，这时准确地拾取走时就比较困难。

地震走时的含义是地震波沿着某种路径旅行所需要的时间．这里"某种路径"在物理上称为射线，即波阵面的法线方向的轨迹。我们首先来讨论射线理论，看看如何把波动方程化为走时满足的程函方程，又如何通过射线追踪确定射线的轨迹。

(一)射线方程

射线是几何光学中导出的概念，它是用射线长度的仿射参数(affine parametes)s 来描述的。射线仿射参数 s 和射线长度 l 之间的关系为

$$l=\int_0^s \| \boldsymbol{n}(\boldsymbol{r}') \| \mathrm{d}s' \tag{5.1}$$

其中

$$\boldsymbol{n}(\boldsymbol{r})=\mathrm{d}\boldsymbol{r}/\mathrm{d}s \tag{5.2}$$

为描述在空间点 $\boldsymbol{r}$ 处射线方向的向量，当(5.1)式沿射线积分时，$\| \boldsymbol{n}(\boldsymbol{r}') \|=1$；否则 $\| \boldsymbol{n}(\boldsymbol{r}') \|=\cos\theta$，$\theta$ 为射线方向与 $\mathrm{d}s'$ 的积分方向的夹角。假定空间波速函数为 $C(\boldsymbol{r})\subset\mathscr{C}(\mathscr{R}^3)$，则射线的走时为

$$\tau(s)=\int_0^s \frac{\| \boldsymbol{n}(\boldsymbol{r}') \|}{C(\boldsymbol{r}')}\mathrm{d}s' \tag{5.3}$$

如设射线的出发点为 A，终点为 B，则(5.3)也可表示为线积分

$$\tau_{AB}=\int_A^B \frac{\cos\theta(\boldsymbol{r}')}{C(\boldsymbol{r}')}\mathrm{d}s' \tag{5.4}$$

无论是用惠更斯原理还是从费马原理都可以导出空间任一点射线仿射参数 s 满足的微分方程。例如，从费马原理出发，在直角坐标系中射线元的长度表为：

$$\mathrm{d}s=(\mathrm{d}x^2+\mathrm{d}y^2+\mathrm{d}z^2)^{1/2}$$

将沿射线点的坐标表示为参数方程

$$x=x(u),\ y=y(u)\ 及\ z=z(u)$$

有

$$\mathrm{d}s=\sqrt{\left(\frac{\mathrm{d}x}{\mathrm{d}u}\right)^2+\left(\frac{\mathrm{d}y}{\mathrm{d}u}\right)^2+\left(\frac{\mathrm{d}z}{\mathrm{d}u}\right)^2}\ \mathrm{d}u \tag{5.5}$$

设波慢为 $N(\boldsymbol{r})=N(x,y,z)=1/C(\boldsymbol{r})$，则从点 A 到点 B 的走时为

$$\tau=\int_A^B N(\boldsymbol{r})\sqrt{\left(\frac{\mathrm{d}x}{\mathrm{d}u}\right)^2+\left(\frac{\mathrm{d}y}{\mathrm{d}u}\right)^2+\left(\frac{\mathrm{d}z}{\mathrm{d}u}\right)^2}\mathrm{d}u$$

根据费马原理，即沿射线走时取极小，应有走时的变分为零

$$\delta\tau=\delta\int_{u_A}^{u_B} F\left(x,y,z,\frac{\mathrm{d}x}{\mathrm{d}u},\frac{\mathrm{d}y}{\mathrm{d}u},\frac{\mathrm{d}z}{\mathrm{d}u}\right)\mathrm{d}u=0 \tag{5.6}$$

其中

$$F=N(x,y,z)\sqrt{\left(\frac{\mathrm{d}x}{\mathrm{d}u}\right)^2+\left(\frac{\mathrm{d}y}{\mathrm{d}u}\right)^2+\left(\frac{\mathrm{d}z}{\mathrm{d}u}\right)^2} \tag{5.7}$$

将(5.7)式代入(5.6)式并展开，得到

$$\int_{u_A}^{u_B}\left(\frac{\partial F}{\partial x}\delta x+\frac{\partial F}{\partial y}\delta y+\frac{\partial F}{\partial z}\delta z+\frac{\partial F}{\partial f}\delta f+\frac{\partial F}{\partial g}\delta g+\frac{\partial F}{\partial h}\delta h\right)\mathrm{d}u=0 \tag{5.8}$$

其中

$$\left.\begin{aligned} &f=\mathrm{d}x/\mathrm{d}u,\ g=\mathrm{d}y/\mathrm{d}u,\ h=\mathrm{d}z/\mathrm{d}u\\ &\delta f=\delta\left(\frac{\mathrm{d}x}{\mathrm{d}u}\right)=\frac{\mathrm{d}}{\mathrm{d}u}\delta x\\ &\delta g=\delta\left(\frac{\mathrm{d}y}{\mathrm{d}u}\right)=\frac{\mathrm{d}}{\mathrm{d}u}\delta y\\ &\delta h=\delta\left(\frac{\mathrm{d}z}{\mathrm{d}u}\right)=\frac{\mathrm{d}}{\mathrm{d}u}\delta z \end{aligned}\right\} \tag{5.9}$$

由(5.5)可知

$$\mathrm{d}s=\sqrt{f^2+g^2+h^2}\mathrm{d}u \tag{5.10}$$

为简化(5.8)的表达式，将(x,y,z)记为(x_i)，$i=1,2,3$；同时把(f,g,h)记为(f_i)，$i=1,2,3$。这时(5.8)式简写为

$$\int_{u_A}^{u_B}\left(\sum_{i=1}^{3}\frac{\partial F}{\partial x_i}\delta x_i+\sum_{i=1}^{3}\frac{\partial F}{\partial f_i}\frac{\mathrm{d}}{\mathrm{d}u}\delta x_i\right)\mathrm{d}u=0$$

利用分部积分法上式化为

$$\sum_{i=1}^{3}\frac{\partial F}{\partial f_i}\delta x_i\Big|_{u_A}^{u_B}+\int_{u_A}^{u_B}\sum_{i=1}^{3}\left\{\left[\frac{\partial F}{\partial x_i}-\frac{\mathrm{d}}{\mathrm{d}u}\left(\frac{\partial F}{\partial f_i}\right)\right]\delta x_i\right\}\mathrm{d}u=0 \tag{5.11}$$

由于在A、B点$\delta x_A=\delta x_B=0$，同时δx_i是任意的，因此上式左边为零，即第一项为零，而且第二项被积函数也为零，后者表为

$$\frac{\partial F}{\partial x_i}-\frac{\mathrm{d}}{\mathrm{d}u}\left(\frac{\partial F}{\partial f_i}\right)=0,\quad i=1,2,3 \tag{5.12}$$

代入(5.7)及(5.10)式把上式变为

$$\frac{\mathrm{d}s}{\mathrm{d}u}\frac{\partial N(x,y,z)}{\partial x_i}=\frac{\mathrm{d}}{\mathrm{d}u}\left(\frac{N(x,y,z)f_i}{\mathrm{d}s/\mathrm{d}u}\right)$$

由(5.9)式知$f_i=\mathrm{d}x_i/\mathrm{d}u$，因此有射线方程

$$\left.\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}s}\left(N(\boldsymbol{r})\frac{\mathrm{d}x}{\mathrm{d}s}\right)=\frac{\partial N(\boldsymbol{r})}{\partial x}\\ &\frac{\mathrm{d}}{\mathrm{d}s}\left(N(\boldsymbol{r})\frac{\mathrm{d}y}{\mathrm{d}s}\right)=\frac{\partial N(\boldsymbol{r})}{\partial y}\\ &\frac{\mathrm{d}}{\mathrm{d}s}\left(N(\boldsymbol{r})\frac{\mathrm{d}z}{\mathrm{d}s}\right)=\frac{\partial N(\boldsymbol{r})}{\partial z} \end{aligned}\right\} \tag{5.13}$$

或写成向量形式

$$\frac{\mathrm{d}}{\mathrm{d}s}\left(\frac{1}{C(\boldsymbol{r})}\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}s}\right)=\nabla_r\left(\frac{1}{C(\boldsymbol{r})}\right) \tag{5.14}$$

射线方程(5.13)及(5.14)表示了在空间每一点$\boldsymbol{r}$上波速与射线仿射参数s的关系，它的解就是用s表示的射线上的点的位置$\boldsymbol{r}(s)$。把(5.14)对s积分两次并求出积分常数可得

$$\boldsymbol{r}(s)=\boldsymbol{r}_0+\boldsymbol{n}_0\int_0^s\frac{C(\boldsymbol{r}')}{C(\boldsymbol{r}_0)}\mathrm{d}s'+\int_0^s C(\boldsymbol{r}')\int_0^{s'}\nabla_{r''}\left(\frac{1}{C(\boldsymbol{r}'')}\right)\mathrm{d}s''\mathrm{d}s' \tag{5.15}$$

其中 r_0 为震源位置，在坐标原点上有 $s=0$，n_0 为在原点射线方向的单位向量。

由于射线方程由几何光学原理所导出，因此在对地震波传播作几何光学近似时才成立，近似的条件为波速 $C(r)$ 的空间变化在一个波长范围内很小。在几何光学近似的条件下声波方程可以化为程函方程和迁移方程，它们将在下面讨论。

(二)程函方程和迁移方程

现在来考虑声波方程

$$\nabla^2 u(r,\omega)+\frac{\omega^2}{C^2(r)}u(r,\omega)=0 \tag{5.16}$$

的一种近似解。当 C 为常数时，我们知道

$$u=A\mathrm{e}^{\mathrm{i}\omega|r|/C} \tag{5.17}$$

乃是(5.16)式的一个解，这一点只需将(5.17)回代(5.16)便可证明。如果 ω 为上半空间的复数，而 $C(r)$ 是随空间缓慢变化的连续函数，可近似地认为不同频率 ω 的谐波虽然具有不同的振幅 $A_j(r)$，但具有与振幅无关的相位。这时方程(5.16)的近似解应有以下形式

$$u(r,\omega)\simeq\omega^{\beta}\mathrm{e}^{\mathrm{i}\omega\tau(r)}\sum_{j=0}^{\infty}\frac{A_j(r)}{(\mathrm{i}\omega)^j} \tag{5.18}$$

式中 $\tau(r)$ 为走时，β 为待定整数。形如(5.18)式的解通常称为 WKBJ 解，它是由 G. Wentzel, H. Kramers, L. Brillouin 和 H. Jeffreys 在处理不同问题时各自提出的，具有共同的 $\mathrm{e}^{\mathrm{i}\omega\tau(r)}$ 形式。

将(5.18)式回代(5.16)得

$$\sum_{j=0}^{\infty}\{(\mathrm{i}\omega)^{2-j}A_j[(\nabla\tau)^2-C^{-2}]+(\mathrm{i}\omega)^{1-j}[2\nabla\tau\cdot\nabla A_j+A_j\nabla^2\tau]+(\mathrm{i}\omega)^{-j}\nabla^2 A_j\}=0 \tag{5.19}$$

此式中 ω 的最高次项为 ω^2。如果选 τ 满足

$$(\Delta\tau)=\left(\frac{\partial\tau}{\partial x}\right)^2+\left(\frac{\partial\tau}{\partial y}\right)^2+\left(\frac{\partial\tau}{\partial z}\right)^2=N^2(x,y,z) \tag{5.20}$$

则对任何非零的 A_0 有 ω^2 的系数项为零，满足(5.19)式。方程(5.20)称为程函方程(eikonal equation)，它说明在几何光学中波阵面仅仅取决于折射率 $N=1/C$，而与振幅 A_j 无关。选 τ 满足(5.20)之后(5.19)式左边整个第一项级数都变为零，再令 ω 的各幂次系数项为零得

$$2\nabla\tau\cdot\nabla A_0+A_0\nabla^2\tau=0 \tag{5.21}$$

及

$$2\nabla\tau\cdot\nabla A_j\nabla^2\tau=-\nabla^2 A_{j-1},\ j=1,2,\cdots\cdots \tag{5.22}$$

方程(5.21)和(5.22)称为迁移方程(transport equation)，最常用的是(5.21)，它通常称为第一迁移方程。迁移方程表示沿射线谐波振幅的变化，反映了波传播的动力学特性。

在非均匀介质中几何光学近似的条件是波速在一个波长范围内的相对变化 $\Delta C/C$ 很小。假定地震勘探使用的主频率为 30Hz，地球平均波速为 3000m/s，则波长大约为 100m。如果取 $\Delta C/C=0.05$，则可知在介质波速的梯度小于每米 1.5m/s 时几何光学近似是允许的。由于实际地震波不是单频波，几何光学近似只对地震波的高频成分成立，因此由(5.18)式给出的波动方程近似解实际上是高频近似解。换言之，对于波速突变的间断面和波长很长的低频成分，几何光学假定及程函方程(5.20)都不再成立。

(三)离散射线追踪的角增量法

射线追踪指的是给定源和接收器位置及地球波速 $C(r)$ 求从源到接收器的射线轨迹。在几何光学近似的条件下，由费马原理可知源-检波器之间的射线只有一条，即射线追踪问题的解是唯一的。在地震层析成像中常常首先把地球介质用固定网格离散化，这时射线追踪问题转化为求射线通过的网格单元，坐标，射线段长度及走时。

为了解射线追踪的基本步骤，让我们来看看垂向不均匀二维介质中的射线追踪过程，此时设 $C=C(z)$。用步长分别为 Δx 和 Δz 的等间距网格划分地球，设震源位于网格(0.0)之中，并为(x,y)直角坐标系的原点，接收点位于$(x,0)$。由 Snell 定理可知射线参数 p 为

$$p=\frac{\sin i_0}{C_0}=\frac{\sin i_n}{C_n}$$

式中 i_0 为震源处射线出射角，C_0 为震源所在网格内的纵波速度，i_n 为第 n 层的入射角，C_n 为第 n 层的波速。如果网格足够密以使每一个网格之内的射线为直线，有

$$\tan i_n=\frac{\mathrm{d}x_n}{\mathrm{d}z_n}$$

式中 $\mathrm{d}x_n$ 和 $\mathrm{d}z_n$ 分别为第 n 层内射线的水平投影及垂直投影。射线长度和走时分别为

$$\mathrm{d}s_n=(\mathrm{d}x_n^2+\mathrm{d}z_n^2)^{\frac{1}{2}} \tag{5.23}$$

和

$$\mathrm{d}\tau_n=\mathrm{d}s_n/C_n \tag{5.24}$$

根据上述关系可将射线追踪的步骤归纳如下：

(1)对任一个射线的出射角 i_0 或射线参数 p_0 计算(0.0)网格的射线出射点(x_0,z_0)，它是射线

$$x=z\tan i_0$$

和 $|x|=\Delta x$ 或 $z=\Delta z$ 的交点。此时，可能发生以下三种情况：

A：如果 $x_0=\Delta z\tan i_0$ 而且 $|x_0|<\Delta x$，则为底(顶)边出射，有 $z_0=\Delta z$；

B：如果有 $z_0=\Delta x\cot i_0$ 满足 $z_0<\Delta z$ 则为侧边出射，有 $x_0=\Delta x$；

C：如果有 $|x_0|=\Delta x$ 及 $z_0=\Delta z$，则射线为对角出射。

(2)用(5.23)及(5.24)计算射线元 $\mathrm{d}s$，走时 $\mathrm{d}\tau$。注意由于源位于原点，即(0.0)网格的左上角，出射点坐标为 x_0 和 z_0。

(3)确定射线进入下一个网格元(i,j)及入射点。以第 n 次追踪为例$(n=1,2,\cdots)$，设第$(n-1)$次射线追踪的网格元为(k,l)，则对上述三种情况，第 n 次追踪时射线所在的网格元号也有三种情况

$$\mathrm{A}:j=l,i=\begin{cases}k+1 & \text{当}|i_{n-1}|\leqslant 90^\circ\\ k-1 & \text{当}|i_{n-1}|>90^\circ,\end{cases}$$

$$\mathrm{B}:i=k,j=\begin{cases}l+1 & \text{当}\ i_{n-1}\geqslant 0\\ l-1 & \text{当}\ i_{n-1}<0,\end{cases}$$

C：$i=k\pm1$ 及 $j=l\pm1$，符号同上。

设第$(n-1)$次追踪时射线的出射点为(x_{n-1},z_{n-1})，则在相对当前网格(i,j)的左上角点的坐标系中，入射点为

$$\mathrm{A}:\tilde{x}_{n-1}=x_{n-1}-(j-1)\Delta x,\tilde{z}_{n-1}=0;$$

$$\mathrm{B}:\tilde{x}_{n-1}=0,\tilde{z}_{n-1}=z_{n-1}-(i-1)\Delta z;$$

$$\mathrm{C}:\tilde{x}_{n-1}=0,\tilde{z}_{n-1}=0。$$

(4)重复(1)—(3)的步骤确定射线在(i,j)网格元的出射点$x_n=x_{n-1}+\mathrm{d}x_n$及$z_n=z_{n-1}+\mathrm{d}z_n$;其中入射角$i_n=\arcsin(p\cdot C_n)$,而$\mathrm{d}x_n$及$\mathrm{d}z_n$分三种情况由步骤(1)中的方法求出,然后由(5.23)及(5.24)求射线元$\mathrm{d}s_n$及走时$\mathrm{d}\tau_n$。

(5)当$z_n=0$时射线返回地表,停止对参数为p的射线的追踪。

(6)当$z_n=0$时,判断x_n是否等于接收点坐标x_R,如果$x_n\neq x_R$,则用一个角增量Δi_0修改入射角i_0,重复步骤(1)—(6),直至x_n与x_R的误差可以忽略为止。

上述离散射线追踪的方法称为角增量法。对于满足几何光学近似条件的二维非均匀介质$C(x,z)$,可以利用射线方程(5.13)导出用角增量法实现射线追踪的公式。和上面一样假定每一个网格内射线为直线,它与x轴夹角为θ,则由(5.23)式可知

$$\frac{\mathrm{d}x}{\mathrm{d}s}=\cos\theta,\frac{\mathrm{d}z}{\mathrm{d}s}=\sin\theta$$

上式代入(5.13)式有

$$\cos\theta\frac{\mathrm{d}N}{\mathrm{d}s}-N\sin\theta\frac{\mathrm{d}\theta}{\mathrm{d}s}=\frac{\partial N}{\partial x} \tag{5.25}$$

$$\sin\theta\frac{\mathrm{d}N}{\mathrm{d}s}+N\cos\theta\frac{\mathrm{d}\theta}{\mathrm{d}s}=\frac{\partial N}{\partial z} \tag{5.26}$$

将(5.25)乘以$-\sin\theta$及(5.26)乘以$\cos\theta$然后相加得

$$\frac{\mathrm{d}\theta}{\mathrm{d}s}=\frac{1}{N}\left(\cos\theta\frac{\partial N}{\partial z}-\sin\theta\frac{\partial N}{\partial x}\right) \tag{5.27}$$

这是关于网格内入射角θ的一阶常微分方程,可用数值方法求解。对于第n条射线经过的元,$\partial N/\partial x$及$\partial N/\partial z$为所在单元波慢的水平及垂直梯度,给定地球波速模型后可以计算。如用Runge-Kutta法解方程(5.27),便可由给定的θ_{n-1}角求出θ_n,并由下式得到射线矢径的坐标

$$x_n=x_{n-1}+\mathrm{d}s\cos\theta_n$$
$$z_n=z_{n-1}+\mathrm{d}s\sin\theta_n$$

式中$\mathrm{d}s$由(5.23)式计算。

离散射线追踪的角增量法虽然是地震学中应用广泛的一种射线追踪方法,但是有明显的缺陷。在理论上看,当相邻网格元的波速发生变化时网格内射线应该是曲线,用(5.23)式把它内射线作直线处理只是一种粗略的近似,另外,对不同网格用(5.27)作射线追踪时存在误差不断积累的问题,因此角增量法的精度较差。在实用上它要求不断用试错法修改入射角增量$\Delta\theta_0$,才能使射线最后趋近于接收点,因此要进行许多无意义的射线追踪计算,成本较高。这些缺陷在下面讨论的Langan等人1985年专门为地震层析成像提出的射线追踪方法中得到了改进[79]。

(四)精确和快速的射线追踪

Langan等人进行射线追踪的出发点从射线方程的积分解(5.15)式出发,并假定介质的波速梯度为常数。此时,波速的空间分布可以表示为

$$C(\boldsymbol{r})=C(0)+\boldsymbol{\lambda}\cdot\boldsymbol{r} \tag{5.28}$$

其中$C(0)$为坐标原点处的波速,$\boldsymbol{\lambda}$为波速梯度$\nabla\boldsymbol{r}[C(\boldsymbol{r})]$。于是波慢变为

$$N(\boldsymbol{r})=\frac{1}{C(0)+\boldsymbol{\lambda}\cdot\boldsymbol{r}}$$

它可以作泰勒级数展开并取到二阶项

$$N(\boldsymbol{r})=\frac{1}{C(0)}\left\{1-\left(\frac{\boldsymbol{\lambda}\cdot\boldsymbol{r}}{C(0)}\right)+\left[\frac{\boldsymbol{\lambda}\cdot\boldsymbol{r}}{C(0)}\right]^2\right\}+0(\boldsymbol{\lambda}^3) \tag{5.29}$$

其中 $0(\boldsymbol{\lambda}^3)$ 为三阶项，在满足

$$\left|\frac{\boldsymbol{\lambda}\cdot\boldsymbol{r}}{C(0)}\right|<1$$

即小梯度 $\boldsymbol{\lambda}$ 时这种射线追踪的方法才成立。对(5.29)两边作微分得

$$\nabla_r\left(\frac{1}{C(\boldsymbol{r})}\right)=\frac{1}{C(0)}\left\{-\frac{\boldsymbol{\lambda}}{C(0)}+\frac{2\boldsymbol{\lambda}}{C(0)}\left(\frac{\boldsymbol{\lambda}\cdot\boldsymbol{r}}{C(0)}\right)\right\}+0(\boldsymbol{\lambda}^3) \tag{5.30}$$

将(5.28)～(5.30)代入(5.15)式可得到射线追踪的方程。注意 s 为射线仿射参数，$\boldsymbol{n}_0$ 为 $s=0$ 的坐标原点射线方向的单位向量。当取波速分布的参考点为射线入射点 γ_0 时，有 $C(0)=C_0$ 为源点处的波速，此时由(5.28)有

$$\frac{\boldsymbol{\lambda}\cdot\boldsymbol{r}_0}{C_0}=\frac{C(\boldsymbol{r}_0)-C_0}{C_0}=0 \tag{5.31}$$

再考虑小梯度条件 Langan 等人最后得到[79]

$$\boldsymbol{r}(s)\simeq\boldsymbol{r}_0-\boldsymbol{n}_0 s\left(1+\frac{s}{2C_0}(\boldsymbol{\lambda}\cdot\boldsymbol{n}_0)\right)-\frac{\boldsymbol{\lambda}s^2}{2C_0}-\frac{\boldsymbol{n}_0}{6C_0^2}s^3[\lambda^2-(\boldsymbol{\lambda}\cdot\boldsymbol{n}_0)^2] \tag{5.32}$$

(5.32)式给出了射线参数 s 和射线点坐标 $\boldsymbol{r}(s)$ 之间的关系。由于 $\mathrm{d}\boldsymbol{r}/\mathrm{d}s$ 为射线的方向，由(5.32)对 s 微分便可得射线方向单位向量的方程

$$\boldsymbol{n}(s)\simeq\boldsymbol{n}_0\left(1+\frac{\boldsymbol{\lambda}\cdot\boldsymbol{n}_0}{C_0}s\right)-\frac{\boldsymbol{\lambda}s}{C_0}-\frac{\boldsymbol{n}_0 s^2}{2C_0^2}[\lambda^2-(\boldsymbol{\lambda}\cdot\boldsymbol{n}_0)^2] \tag{5.33}$$

它的元为射线的方向余弦。

沿射线的走时 τ 可利用(5.3)的积分式求出。将(5.29)和(5.33)代入(5.3)并注意 $\|\boldsymbol{n}(\boldsymbol{r})\|=1$，可得

$$\tau(s)\simeq\frac{s}{C_0}\left\{1+\frac{s^2}{6C_0^2}[\lambda^2+(\boldsymbol{\lambda}\cdot\boldsymbol{n}_0)^2]-(\boldsymbol{\lambda}\cdot\boldsymbol{n}_0)\frac{s}{2C_0}\right\} \tag{5.34}$$

公式(5.32)～(5.34)便给出了快速计算三维介质中射线位置、方向和走时的代数方程式，它们都使用射线参数 s 作变量。与以射线方向(如角增量法的入射角)为变量的射线追踪方法相比，由于 s 的误差不会引起计算射线坐标 $\boldsymbol{r}$ 的较大误差，所以避免了角增量法作射线追踪时有大误差积累的问题。此外，这种射线追踪方法不仅可用于二维，而且也可用于三维介质，因此用于走时数据的地震层析成像是比较理想的。

对于二维非均匀介质的射线追踪，我们可以把介质分割为许多单元，并把常数速度梯度定义在元的中心。这时每个单元内部各点上的波速不是常数，而要用线性内插的方法求；相应地元内的射线为圆弧而不是直线，这就可以取较大的网格单元，以减少计算成本。网格元可取为三角形或矩形，取为三角形时由线性内插求出的波速在越过边界时是连续的，当波速分布较复杂时具有优越性。取矩形网格时计算方便，但是波速在单元边界上不连续，要进行圆滑。

将射线追踪公式(5.32)～(5.34)应用于矩形网格时，设网格与坐标轴平行，对其中某一网格单元的入射点为 $\boldsymbol{r}_0$，入射角为 $\boldsymbol{n}_0$(图 5.5)。当网格元很小而且 $\boldsymbol{\lambda}$ 也很小时有

$$\frac{\|\boldsymbol{\lambda}\|(a+b)}{C_0}\ll 1$$

其中 a 和 b 为单元的边长，因此(5.32)～(5.34)中 λ 的二次项可以忽略。此时相应公式化为

$$\boldsymbol{r}(s)=\boldsymbol{r}_0+\boldsymbol{n}_0 s\left(1+\frac{s}{2C_0}(\boldsymbol{\lambda}\cdot\boldsymbol{n}_0)\right)-\frac{\boldsymbol{\lambda}s^2}{2C_0}+0(\lambda^2) \tag{5.35}$$

$$\boldsymbol{n}(s)=\boldsymbol{n}_0\left(1+\frac{(\boldsymbol{\lambda}\cdot\boldsymbol{n}_0)}{C_0}s\right)-\frac{\boldsymbol{\lambda}s}{C_0}+0(\lambda^2) \tag{5.36}$$

和

$$\tau(s)=\frac{s}{C_0}\left(1-\frac{(\boldsymbol{\lambda}\cdot\boldsymbol{n}_0)}{2C_0}s\right)+0(\lambda^2) \tag{5.37}$$

把(5.35)改写为对 s 的二次向量方程

$$\boldsymbol{r}(s)=\frac{1}{2C_0}[\boldsymbol{n}_0(\boldsymbol{\lambda}\cdot\boldsymbol{n}_0)-\boldsymbol{\lambda}]s^2+\boldsymbol{n}_0 s+\boldsymbol{r}_0+0(\lambda^2) \tag{5.38}$$

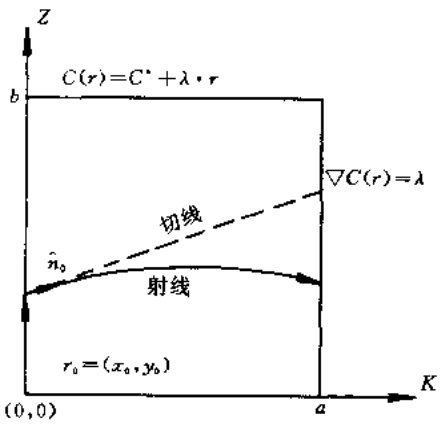


图 5.5 通常恒波速梯度的网格单元的射线

对单元的每一个边(5.38)转化为标量方程，有 2 个根，因此对四个边可以求出关于 s 的八个可能的解。由费马原理我们应挑出 s 的最小正实数解。

例如在图 5.5 的情况下射线由单元的左边射入，由 $\boldsymbol{n}_0$ 知它的切线在右边，这里 $x=a$。如果网格很小，我们知道射线穿过单元的右边，则(5.38)在元右边的标量方程变为

$$a=\frac{1}{2C_0}(n_{0x}(\boldsymbol{\lambda}\cdot\boldsymbol{n}_0)-\lambda_x)s^2+n_{0x}s+r_{0x} \tag{5.39}$$

其中下角标 x 表示 x 分量，解二次方程(5.39)求出 s，然后代入下式求射线在右边的交点

$$z_i=\frac{1}{2C_0}[n_{0x}(\boldsymbol{\lambda}\cdot\boldsymbol{n}_0)-\lambda_x]s^2+n_{0x}s+r_{0x} \tag{5.40}$$

其中下角标 z 表示向量 z 分量。如果此时求出的 $z_i<0$ 或 $z_i>b$，则说明射线不通过右边，而要找出相应边，用(5.38)重新求出射点坐标。

用反射走时作层析成像时要考虑地球中存在波速间断，这时必须假定网格单元内间断面连续而且光滑。假定单元内间断面可以用立方样条函数描述，

$$z=A+Bx+Gx^2+Dx^3 \tag{5.41}$$

其中 A、B、G 和 D 为已知常数，反射点位于此界面上，单元内的射线被反射点分成两段。将射线方程(5.38)写为分量形式并代入(5.41)式得

$$\begin{aligned}0=&(A+Bx_0+Gx_0^2+Dx_0^3-z_0)+s[n_{0z}(B+2Gx_0+3Dx_0^2)-n_{0x}]\\&+s^2[\alpha(B+2Gx_0+3Dx_0^2)+n_{0x}^2(G+3Dx_0)-\beta]\\&+s^3n_{0x}[2\alpha(G+3Dx_0)+Dn_{0x}^2]+s^4(3D\alpha n_{0x}^2)\end{aligned} \tag{5.42}$$

其中 (x_0,z_0) 为入射点，而

$$\alpha=\frac{1}{2C_0}[n_{0x}(\boldsymbol{\lambda}\cdot\boldsymbol{n}_0)-\lambda_x]$$

和

$$\beta=\frac{1}{2C_0}[n_{0z}(\boldsymbol{\lambda}\cdot\boldsymbol{n}_0)-\lambda_z]$$

为了求出有界面反射的射线参数 s，假定反射面的斜率变化不大，即 G 和 D 远小于 A 和 B，这时(5.42)式最后两项可忽略，而变成 s 的二次方程。求出 s 之后回代入(5.38)求出出射点

的 x 坐标，然后用(5.41)式求出射点的 z 坐标。

上述两种射线追踪的方法都是从射线方程(5.14)导出。利用程函方程(5.20)也可以导出一些射线追踪的方法，如特征值法等，可参考文献[80]。目前，计算最快的方法是直接应用费马原理加逻辑判断的方法。这时，网格内的介质分布被细分为若干个模板，判断当前网格属于那一种模板，就可立即得到出射点的位置和射线方向。由于计算速度快，这种方法在叠前深度偏移等大型工程计算中也得到了应用。下面我们开始讨论波速图像的计算机重建问题，包括连续图像重建技术和离散图像重建技术两部分。连续图像重建的数学基础为 Radon 变换，我们首先从最简单直射线情况出发讨论。

三、Radon 变换和连续图像重建

(一)Radon 变换[81]

一个二维的连续图像在数学上用函数 $f(x,y)$ 来描述，$f(x,y)$ 称为图像函数，在图像域 D 内各点上取确定的值，这个值叫(印刷象点的)密度，在黑白图上即为图像的灰度。当坐标原点取为图像中心时，对方形图像有

$$f(x,y)=0,\ |x/2|\text{或}|y/2|\geqslant E$$

而对圆形图像有

$$f(x,y)=0,\ \sqrt{x^2+y^2}\geqslant E$$

E 为图像的半边长或半径。在大多数情况下 $f(x,y)$ 是平方可积函数，因此 $f\in\mathscr{L}^2(\mathscr{R}^2)$，否则取连续函数，即 $f\in\mathscr{C}(\mathscr{R}^2)$，详见第五章第二节关于地球模型的讨论部分。

现在来看看如图 5.6 所示的典型的 CT 测量系统，它在医学上是 X 射线 CT 装置，在地震学中与用超声波作跨孔 CT 装置类似。当入射波的波长远小于介质的不均匀性时可认为平面透射波通过介质时沿直射线传播，因此在成像物体另一端的接收信号

$$p(\xi,\theta)=\int_L f(x,y)\mathrm{d}l=\int_{-\infty}^{\infty} f_\theta(\xi,\eta)\mathrm{d}\eta \tag{5.43}$$

式中 p 称为图像的投影函数，L 为沿直射线的积分路程，垂直于 ξ 轴方向，θ 为坐标系 (x,y) 和 (ξ,η) 之间的夹角，f_θ 为 $f(x,y)$ 经坐标旋转 θ 角的取值。对于跨孔 CT 的情况令(5.3)式中 $\|\boldsymbol{n}(\boldsymbol{r}')\|=1$ 即为直射线的情况，相应 $f(x,y)$ 表示波慢图像 $N(x,y)$，走时 $\tau(\xi,\theta)$ 为投影函数。

由(5.43)式表示的函数 $f(x,y)$ 沿直线 L 的积分称为(经典的)Radon 变换，它一般写为

$$[\mathscr{R}f](\xi,\theta)=p(\xi,\theta)=\int_L f[x(\xi,\theta),y(\xi,\theta)]\mathrm{d}l \tag{5.44}$$

其中

$$\left.\begin{aligned}x&=\xi\cos\theta-\eta\sin\theta\\ y&=\xi\sin\theta+\eta\cos\theta\\ \theta&=\arctan(y/x)\end{aligned}\right\} \tag{5.45}$$

起名为 Radon 变换是因为奥地利数学家 J. Radon 于 1917 年证明了图像函数 $f(x,y)$ 可由它的无穷多个 Radon 变换 $(\mathscr{R}f)$ 唯一地重建，因此首先给出了用这种 CT 装置作图像重建的数学基础。

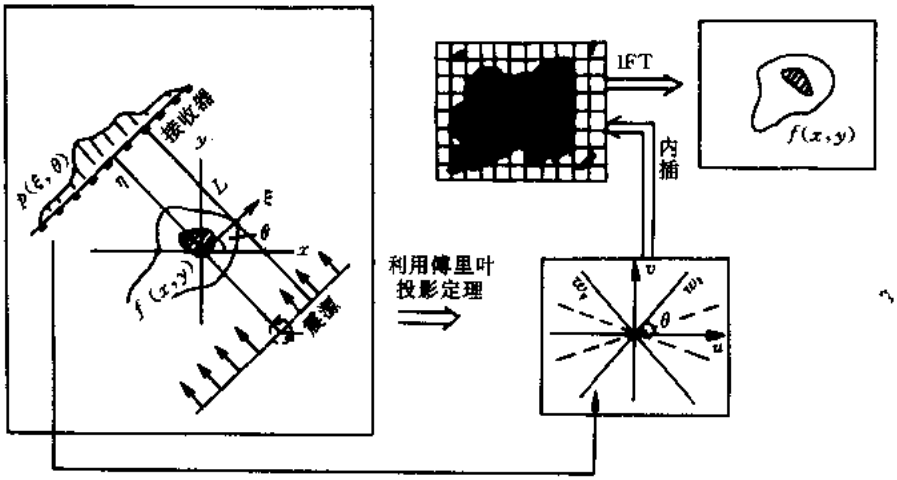


图 5.6　典型的 CT 测量和利用 Fourier 投影定理作图像重建过程

假设 $r=(x,y)$为图像中一个点，$s=(\cos\theta,\sin\theta)$为表示射线垂直方向的单位向量，则 $r\cdot s=\xi$ 表示坐标原点到射线 L 的垂直距离，即位于 L 上的点的 ξ 坐标。代入(5.44)式可以把 Radon 变换用 δ 函数表示为

$$[\mathscr{R}f](\xi,s)=\int_{-\infty}^{\infty}f(r)\delta(\xi-s\cdot r)\mathrm{d}r \tag{5.46}$$

此式便于把 Radon 变换推广到高维($r\in\mathscr{R}^n$)的情况。在二维情况下由于 $s\cdot r=x\cos\theta+y\sin\theta$，(5.46)的标量形式为

$$[\mathscr{R}f](\xi,\theta)=\iint_{-\infty}^{\infty}f(x,y)\delta(\xi-x\cos\theta-y\sin\theta)\mathrm{d}x\mathrm{d}y \tag{5.47}$$

如果再假定 $t=(-\sin\theta,\cos\theta)$为沿射线方向的单位向量，则有 $\eta=t\cdot r$ 为射线方向的坐标，且 $r=\xi s+\eta t$。此时(5.43)式可改写为

$$[\mathscr{R}f](\xi,\theta)=\int_{-\infty}^{\infty}f_\theta(\xi s+\eta t)\mathrm{d}\eta \tag{5.48}$$

如果 $f(x,y)$为圆对称函数，即 $f(x,y)=f(r)$，$r=\sqrt{x^2+y^2}$；不失一般性取 $\theta=0$，有 $s=(1,0)$和 $t=(0,1)$及 $\xi=x$ 和 $\eta=y$，由于 $\mathrm{d}y=r\mathrm{d}r/\sqrt{r^2-x^2}$，(5.48)式变为

$$\begin{aligned}[\mathscr{R}f](x)&=\int_{-\infty}^{\infty}f(xs+yt)\mathrm{d}y\\&=2\int_x^{\infty}\frac{rf(r)}{(r^2-x^2)^{1/2}}\mathrm{d}r,(x,y>0)\end{aligned} \tag{5.49}$$

(5.49)是数学上著名的 Abel 积分方程，由投影重建圆对称图像 $f(r)$的问题化为解 Abel 积分方程的问题。

最后，我们来看看在极坐标中的 Radon 变换式，它对于 Radon 逆变换的推导比较方便。在极坐标中图像函数表为 $f(r,\phi)$。如图 5.7 所示，对射线 L 上的点(r,ϕ)，有 $\xi=l$ 和 $\eta=-z$，因此

$$r=\sqrt{l^2+z^2}$$

$$\phi=\theta-\arctan(-z/l)=\theta+\arctan(z/l)$$

代入(5.44)式有

$$[\mathscr{R}f](l,\theta)=\int_{-\infty}^{\infty} f[\sqrt{l^2+z^2},\theta+\arctan(z/l)]\mathrm{d}z \tag{5.50}$$

当 $l=0$ 时射线过坐标原点，此时有

$$[\mathscr{R}f](l,\theta)=\int_{-\infty}^{\infty} f(z,\theta+\pi/2)\mathrm{d}z \tag{5.51}$$

从(5.43)到(5.50)给出了 Radon 变换的各种表示形式，都说明算子 $\mathscr{R}$ 将图像域(r,ϕ)上的函数 f 和投影域(l,θ)中的另一个函数$(\mathscr{R}f)$联系在一起[11]。这种联系表现为投影域的实数对(l,θ)对应于图像域(r,ϕ)中的一条直线 L，而图像域中通过一个点(r,ϕ)的所有直线对应于投影域中 $l=r\cos(\theta-\phi)$的余弦曲线，即此曲线上不同的点代表图像域以不同入射角 θ 穿过此点的射线。因此，千万不要把投影域中的一个实数对(l,θ)理解为平面上一个点的极坐标。此外，还要注意到 Radon 变换具有以下周期性及对称性，

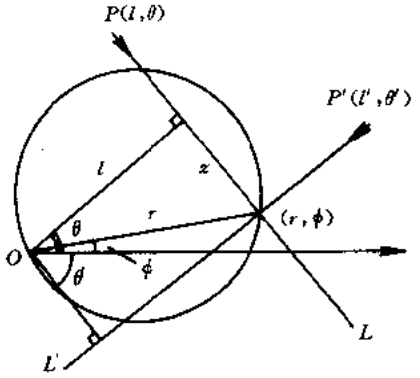


图 5.7　极坐标的 Radon 变换和反投影

$$[\mathscr{R}f](l,\theta)=[\mathscr{R}f](-l,\theta+\pi)=[\mathscr{R}f](l,\theta+2\pi) \tag{5.52}$$

因此 Radon 变换不是一一对应的。

(二)图像重建和 Radon 逆变换

一个连续图像重建算法是输入有限对(l,θ)的投影函数测量值 $p(l,\theta)=[\mathscr{R}f](l,\theta)$，求图像函数 $f(r,\phi)$的估计值。由于联系图像函数和投影函数的是 Radon 变换，重建图像时要找到它的逆算子 $\mathscr{R}^{-1}$，使它满足

$$\mathscr{R}^{-1}[\mathscr{R}f]=f \tag{5.53}$$

以实现图像 $f(r,\phi)$的重建。其中 f 是连续和有界的，当 $r\geqslant E$ 时 $f(r,\phi)=0$。

满足(5.53)式的 Radon 逆变换算子已被推导出来，推导过程可参考文献[82]及文献[11]的 5.2 及 11.3 两节。对于用极变量表示的投影函数 $p(l,\theta)$，它的 Radon 逆变换表示为

$$[\mathscr{R}^{-1}p](r,\phi)=\frac{1}{2\pi^2}\int_0^{\pi}\int_{-\infty}^{\infty}\frac{p'(l,\theta)}{r\cos(\theta-\phi)-l}\mathrm{d}l\mathrm{d}\theta \tag{5.54}$$

式中 $p(l,\theta)=\partial p/\partial l$，为 $p(l,\theta)$关于 l 的偏导数。为了理解逆变换算子的性质，我们可以把(5.54)式作以下分解

$$\mathscr{R}^{-1}=-\frac{1}{2\pi}\mathscr{B}\mathscr{H}_l\mathscr{D}_l \tag{5.55}$$

其中 $\mathscr{D}_l$ 为对第一极变量 l 的偏导数算子，即

$$[\mathscr{D}_l p](l,\theta)=\lim_{\Delta l\to 0}\frac{p(l+\Delta l,\theta)-p(l,\theta)}{\Delta l} \tag{5.56}$$

并假定其中极限存在。

式(5.55)中的第二个算子 $\mathscr{H}_l$ 是对两个极变量的函数 $q(l,\theta)$的第一变量 l 作希尔伯特变换，其中 $q(l,\theta)=[\mathscr{D}_l q](l,\theta)$。根据希尔伯特变换的定义

$$[\mathscr{H}_l q](l',\theta)=-\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{q(l,\theta)}{l'-l}\mathrm{d}l \tag{5.57}$$

它是一个反常积分，因为在 $l=l'$ 时被积函数发散，因此这个积分通常在柯西主值的意义下计算，即取

$$[\mathscr{H}_l q](l',\theta)=-\frac{1}{\pi}\lim_{\varepsilon\to 0}\left\{\int_{-\infty}^{l'-\varepsilon}\frac{q(l,\theta)}{l'-l}\mathrm{d}l+\int_{l'+\varepsilon}^{\infty}\frac{q(l,\theta)}{l'-l}\mathrm{d}l\right\} \tag{5.58}$$

并假定上式右端的极限存在。

式(5.55)中最后一个算子 $\mathscr{B}$ 被称为反投影算子。假定 $t(l',\theta)=[\mathscr{H}_l q](l',\theta)$ 为两个极变量的函数，而且 $l'=r\cos(\theta-\phi)$ 时，它定义为

$$[\mathscr{B}t](r,\phi)=\int_0^{\pi}t(r\cos(\theta-\phi),\theta)\mathrm{d}\theta \tag{5.59}$$

起到的作用是在图像域中把通过 (r,ϕ) 点沿不同入射角 θ 的射线取得的投影，沿着各自的射线“聚焦”回到点 (r,ϕ) 中去，从而取得图像在此点的密度值(如图 5.7 的箭头所示)。如上讨论可知，图像域中通过一个点 (r,ϕ) 的所有不同 θ 角的射线对应于投影域 (l',θ) 中的余弦曲线 $l'=r\cos(\theta-\phi)$，因此作反投影时要把 l' 代换成 $r\cos(\theta-\phi)$，然后对 0 到 π 的所有入射角 θ 进行积分。这个过程就是(5.59)式表示的运算，也是算子 $\mathscr{B}$ 取名为反投影算子的原因。

由(5.55)式给出的连续图像重建或 Radon 逆变换的过程包括：

(1)把投影函数 $p(l,\theta)$ 对 l 求偏导数求得函数 $q(l,\theta)$。

(2)把 $q(l,\theta)$ 对 l 进行希尔伯特变换(取主值)，求得函数 $t(l',\theta)$。

(3)令 $l'=r\cos(\theta-\phi)$ 然后由(5.59)式进行反投影，最后乘以归一化因子 $-1/2\pi$，便得到图像函数 $f(r,\phi)$。

Radon 逆变换式(5.54)在理论上给出了重建连续函数 $f(r,\phi)$ 的图像的方法，它要求射线入射角 θ 从 0 到 π 连续地变化，而且投影函数 $p(l,\theta)$ 的观测完整和准确。在实际测量中这些假设常不能满足，而且分三步计算也是不合算的。下面我们讨论利用 Fourier 变换的图像重建技术，它可以避免直接用 Radon 逆变换来重建图像的一些缺点。

(三)Fourier 变换图像重建

作为线性变换的一种形式，Radon 变换与 Fourier 变换之间有一定联系；如果找到这种关系，我们便可以利用快速 Fourier 变换的算法。让我们回到图 5.6 所示的直角坐标系中来，求图像函数 $f(x,y)$ 和投影函数 $p(\xi,\theta)$ 在波数域的关系。对 $p(\xi,\theta)$ 作关于 ξ 的 Fourier 变换并代入(5.43)有

$$p(\omega_\xi,\theta)=\int_{-\infty}^{\infty}p(\xi,\theta)\mathrm{e}^{-\mathrm{i}\omega_\xi\xi}\mathrm{d}\xi=\iint_{-\infty}^{\infty}f_\theta(\xi,\eta)\mathrm{e}^{-\mathrm{i}\omega_\xi\xi}\mathrm{d}\xi\mathrm{d}\eta \tag{5.60}$$

由于在 (ω_ξ,ω_η) 域中图像函数 $f(\xi,\eta)$ 的二重 Fourier 变换为

$$f(\omega_\xi,\omega_\eta)=\iint_{-\infty}^{\infty}f(\xi,\eta)\mathrm{e}^{-\mathrm{i}(\omega_\xi\xi+\omega_\eta\eta)}\mathrm{d}\xi\mathrm{d}\eta \tag{5.61}$$

比较(5.60)和(5.61)可见，当 $\omega_\eta=0$ 时(5.56)式表示的 $f(\omega_\xi,\omega_\eta)$ 与 $p(\omega_\xi,\theta)$ 相等。注意到 $\omega_\eta=0$ 表示的是在波数域以角度 θ 穿过原点的一条直线，说明投影函数 $p(\xi,\theta)$ 关于 ξ 的 Fourier 变换等于图像函数 $f(\xi,\eta)$ 的二重 Fourier 变换在以 θ 角穿过原点的直线上的取值，并可

写为

$$p(\omega_\xi,\theta)=f_\theta(\omega,0) \tag{5.62}$$

其中 f_θ 的下角 θ 表示坐标之间的旋转。如果设 $f(u,v)$ 为 $f(x,y)$ 的二重 Fourier 变换，由图 5.6 所示的坐标旋转关系可知

$$u=\omega_\xi\cos\theta,v=\omega_\xi\sin\theta \tag{5.63}$$

因此由(5.60)及(5.62)两式又有

$$\int_{-\infty}^{\infty}p(\xi,\theta)e^{-i\omega_\xi\xi}d\xi=f(u,v)\Big|_{\substack{u=\omega_\xi\cos\theta\\v=\omega_\xi\sin\theta}} \tag{5.64}$$

这就是所谓的 Fourier 投影定理(也有人称它为中央切片定理)，说明投影函数 $p(\xi,\theta)$ 关于 ξ 的 Fourier 变换等于图像函数的二重 Fourier 变换在过原点的直线上的切片。如果对所有入射角 θ 得到的投影都作关于 ξ 的 Fourier 变换，我们便得到沿极坐标分布的图像函数的"谱" $f(u,v)$。用缩写 RT 表示 Radon 变换，FT 表示 Fourier 变换，IFT 表示 Fourier 逆变换，则它们的关系可以表示为

$$\begin{array}{ccc} f(x,y) & \xrightarrow{\text{RT}} & p(\xi,\theta) \\ \uparrow & & \downarrow \\ f(u,v) & \xleftarrow{\text{IFT}} & \text{FT} \end{array} \tag{5.65}$$

因此，由(5.64)式表示的 Fourier 投影定理也同时说明了 Radon 变换与 Fourier 变换的关系。

利用 Fourier 投影定理可以得到连续图像重建的以下算法：

(1)对不同 θ 的投影函数 $p(\xi,\theta)$ 作关于 ξ 的一维 Fourier 变换，得到如图 5.6 所示的沿极坐标分布的 $f_\theta(\omega_\xi,0)$。

(2)将极坐标上的 $f_\theta(\omega_\xi,0)$ 内插到方格网上求出 $f(u,v)=f(\omega_\xi\cos\theta,\omega_\xi\sin\theta)$。

(3)对 $f(u,v)$ 作二重 Fourier 逆变换求得图像数 $f(x,y)$。

在上述图像重建算法中，第二步插值计算是很关键的，因为 $f_\theta(\omega_\xi,0)$ 的数据分布很不均匀，在原点处密集，而向外变得稀疏。插值的方法可分为利用局部 f_θ 值进行的局部插值法和利用全部 f_θ 数据的插值法两大类，前者计算速度快而精度低，后者精度高而速度慢。采用那一类插值方法要依数据的分布形式和成像的分辨率要求而定。

(四)褶积滤波图像重建

既然我们可以利用 Fourier 投影定理通过转换到波数域重建图像，那么就有可能找到与它对应的空间域算法，不通过波数域的转换直接重建图像。这种利用投影定理直接在空间域重建图像的技术叫褶积滤波算法，它比上面讨论的 Fourier 变换法计算成本更低。现在就利用 Fourier 变换的褶积定理来推导褶积滤波算法。

将 Fourier 投影定理(5.64)式右边写为二重 Fourier 变换的形式

$$p(\omega_\xi,\theta)=\iint_{-\infty}^{\infty}f(x,y)e^{-i(xu+yv)}dxdy\Big|_{\substack{u=\omega_\xi\cos\theta\\v=\omega_\xi\sin\theta}} \tag{5.66}$$

$$=\iint_{-\infty}^{\infty}f(x,y)e^{-i\omega_\xi(x\cos\theta+y\sin\theta)}dxdy \tag{5.67}$$

由于在极坐标中 $dudv=|\omega_\xi|d\omega_\xi d\theta$，令 $\psi(\omega)=|\omega_\xi|$，则图像函数 $f(x,y)$ 应是如下 Fourier 逆变换

$$f(x,y)=\frac{1}{4\pi^2}\int_0^{\pi}d\theta\int_{-\infty}^{\infty}p(\omega_\xi,\theta)\psi(\omega_\xi)e^{i\omega_\xi(x\cos\theta+y\sin\theta)}d\omega_\xi \tag{5.68}$$

上式右边对 ω_ξ 作变换的函数为 p 和 ψ 的乘积，由 Fourier 变换的褶积定理，它等于空间域 $p(\xi,\theta)$和 $\psi(\xi)$的褶积，即

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}p(\omega_\xi,\theta)\psi(\omega_\xi)e^{i\omega_\xi\xi}d\omega_\xi=\int_{-\infty}^{\infty}p(t,0)\psi(\xi-t)dt \tag{5.69}$$

其中 $\xi=x\cos\theta+y\sin\theta$，因此有

$$f(x,y)=\frac{1}{2\pi}\int_0^{\pi}d\theta\int_{-\infty}^{\infty}p(t,\theta)\psi(x\cos\theta+y\sin\theta-t)dt \tag{5.70}$$

公式(5.70)乃是褶积滤波法重建图像的基本公式，其中函数 $\psi(\xi)$实质上是个滤波函数。由于实际上在(5.70)式中对 t 的积分不可能无穷大，只能是有限宽的窗口，$\psi(\xi)$可被选来抑制有限窗宽效应，而不固定取为 $\psi(\omega_\xi)=|\omega_\xi|$。

将入射角 θ 离散为 $\theta_j=(j-1)\pi/N$，$j=\theta,\cdots,N$；同时把 ξ 离散化为 ξ_k，$k=0,\cdots,M$，则投影函数的取样集可表为 $p(\xi_k,\theta_j)$，(5.71)可写为以下离散褶积形式

$$F(x,y)=\frac{1}{MN}\sum_{j=0}^{N}\sum_{k=0}^{M}p(\xi_k,\theta_j)\psi(x\cos\theta_j+y\sin\theta_j-\xi_k) \tag{5.71}$$

上式被用来作高速褶积图像重建。显然，滤波器 $\psi(\xi)$的选择对褶积计算的精度和稳定性影响很大。$\psi(\xi)$的选择与投影数据的噪音水平和成像的窗宽有关。文献[82]对不同的滤波器作了比较，并认为以下两种选法最为适宜。第一种滤波器被称为 Shepp-Logan 滤波器，表为

$$\psi(0)=\frac{M^2}{\pi};\psi(\xi_k)=-\frac{M^2}{\pi}(4k^2-1)^{-1},k=1,2,3\cdots, \tag{5.72}$$

第二种滤波器称为修正 Shepp-Logan 滤波器，它在用(5.72)式计算 $\psi(\xi_k)$的基础上再对 $\psi(\xi_k)$作以下圆滑

$$\bar{\psi}(\xi_k)=0.4\psi(\xi_k)+0.3\psi(\xi_{k+1})+0.3\psi(\xi_{k-1}) \tag{5.73}$$

然后用 $\bar{\psi}$ 代入(5.71)式进行褶积计算。对于噪音较大的投影数据，用修正的 Shepp-Logan 滤波器效果更好一些。

本书讨论的图像重构方法都是在假定射线为直线的情况下进行的，因此在地震层析成像中的应用有一定的局限性。另外，尽管算法在计算机上执行时要把图像离散化，但是我们在推导计算公式中都把图像作为连续函数 $f(x,y)$处理，因此属于连续图像重建的技术。注意到实际投影数据是有限的，而 $f(x,y)$是连续函数，因此作为连续图像反问题的解，$f(x,y)$不是唯一的，在褶积法中滤波函数可有多种选择就说明解的非唯一性。此外，和其它反问题一样，Radon 逆变换也是不适定的，在投影数据有误差时还要根据正则化准则求图像函数的某种估计[11,84]。除了连续图像重建外，类似于广义反演方法，还可以从一开始就把图像离散化为一系列象元，然后求图像函数 $f(x,y)$每一个象元内的平均值，这就是下面要讨论的离散图像重建技术。

四、离散图像重建技术

(一)离散图像重建问题

一幅黑白的图像指的是在其中每一点上灰度 $f(r,\phi)$ 都有定义的区域。把一张图像划分成 J 个不重叠的小区域(称为象元),并以象元内总灰度不变为准则定义象元内均匀的灰度,就得到了图像的数字化版。离散图像重建问题是由投影函数的有限取样数据求图像的数字化版。

假设在极坐标 (r,ϕ) 中成像区域为 $|r|<a,\phi<2\pi$,它被划分为 J 个不重叠的象元。对应每一个象元可以定义它的基为

$$b_j=\begin{cases}1,\text{当}(r,\phi)\text{在第 }j\text{ 个象元内}\\0,\text{其它}\end{cases} \tag{5.74}$$

则图像的基表示为 J 维向量 $\boldsymbol{b}^{\mathrm{T}}=(b_1,\cdots,b_J)$。根据有限维向量空间元的表现定理可知(见第四章第一节),图像 $f(r,\phi)$ 的数字化版为基的以下线性组合

$$f(r,\phi)=\sum_{j=1}^{J}f_jb_j(r,\phi) \tag{5.75}$$

其中 f_j 为 $f(r,\phi)$ 在第 j 个象元内的平均值,而向量

$$\boldsymbol{f}^{\mathrm{T}}=(f_1,\cdots,f_J)$$

被称为图像向量。如果象元足够小,则沿第 i 条射线图像的投影函数可表为

$$\mathscr{R}_if\simeq\mathscr{R}_if=\sum_{j=1}^{J}f_j(\mathscr{R}_ib_j) \tag{5.76}$$

根据 Radon 变换的定义(5.44)和(5.74)可知,

$$\mathscr{R}_ib_j=\begin{cases}\int_{L_i}\mathrm{d}l=a_{ij},\text{当射线 }i\text{ 经过象元 }j\\0\qquad\qquad,\text{其它}\end{cases} \tag{5.77}$$

其中 a_{ij} 为第 j 个象元内射线 i 的长度。将(5.77)代回(5.76)得

$$\mathscr{R}_if=\sum_{j=1}^{J}a_{ij}f_j+e_j,i=1,2,\cdots,I \tag{5.78}$$

其中 e_i 为误差项,I 为射线总数。于是我们便得到了投影函数的级数展开式,它也给出了重建图像数字化版的方程组。

在地震层析成像的情况下,投影数据为地震波走时 τ_i,图像向量为象元内波慢的平均值,因此(5.78)可写为以下矩阵方程

$$\boldsymbol{\tau}=\boldsymbol{A}\boldsymbol{f}+\boldsymbol{e} \tag{5.79}$$

离散图像重建问题转化为给出 $\boldsymbol{\tau}$ 估计图像向量 $\boldsymbol{f}$。

(5.78)和(5.79)都是线性方程组,在第二和第三两章我们已讨论了用于线性问题的各种有效算法,表面上看求解没有什么问题。其实不然,以前讨论过的各种直接算法多数不适用于层析成像中方程(5.79)的求解。首先,一张图包含的象元数目一般很大,常见 $J\geqslant10^4$ 的情况,这时用直接算法解方程成本太高,要求机器内存太大。此外,矩阵 $\boldsymbol{A}$ 是极其稀疏的,因为它的每一行有 J 个元素,而射线只通过所有 J 个象元中的一小部分,因此 $\boldsymbol{A}$ 的其它大部分元素为零。最后,由于 Radon 逆变换是不适定问题,方程组(5.79)的求解也是不适定的,

要进行正则化处理和解估计的评价。

根据系数矩阵稀疏的特点，对方程(5.79)多采用迭代方法求解，如代数重建(ART)，联合迭代重建，以及其轭斜量法等，将在下面分别进行讨论。最简单和粗糙的离散图像重构方法之一叫反投影技术(Back projection technigue 简写为 BPT)，它是一种非迭代方法。将走时沿射线分配给每一个元；分配时以象元内的第 i 条射线长度 a_{ij} 与射线总长度 $\sum_j a_{ij}$ 之比为权，然后把通过 j 象元在加权后的走时对所有射线总加，并除以总射线长度求得波慢的

$$\hat{f}_j = \sum_{i=1}^{n} \left[a_{ij}\tau_i / \sum_{j=1}^{m} a_{ij} \right] / \sum_{j=1}^{n} a_{ij} \tag{5.80}$$

估计式中 m 为第 i 条射线通过的象元总数，n 为通过 j 象元的射线总数。BPT 用于波慢成像时计算简单快速，但是分辨率极低，仅在个别场合下使用。

(二)代数重建技术(ART)

代数重建是按射线依次修改有关象元的图像向量的一类迭代算法。在方程(5.79)中略去误差项并令图像向量产生一增量 Δf，有

$$\Delta\tau_i = \sum_j \Delta f_j a_{ij}, i = 1, 2, \cdots, I \tag{5.81}$$

式中对射线经过的象元 j 求和。作为迭代算法要根据第 i 条射线的走时差 $\Delta\tau_i$ 求波慢的修改增量 Δf_j。由于方程(5.81)是高度欠定的，可以用它作为约束求 Δf_j 的 L^2 模的极小解。由拉格朗日乘子法令目标函数为

$$Q = \sum_j (\Delta f_j^2 - \lambda \Delta f_j a_{ij}) + \lambda \Delta\tau_i = \min \tag{5.82}$$

其中 λ 为拉格朗日乘子。由 $\partial Q / \partial(\Delta f_j) = 0$ 得

$$\Delta f_j = \frac{\lambda}{2} a_{ij} \tag{5.83}$$

回代入(5.81)有

$$\Delta\tau_i = \sum_j \frac{\lambda}{2} a_{ij}^2$$

即得

$$\frac{\lambda}{2} = \Delta\tau_i / \sum_j a_{ij}^2$$

因此，再由(5.83)式便可写出对第 i 条射线及第 j 个像元求波慢修改增量的公式

$$\Delta f_j = a_{ij}\Delta\tau_i / \sum_j a_{ij}^2 \tag{5.84}$$

这就是所谓加法修正的公式，或称为 ART1。

当然，我们不一定非取 Δf 的 L^2 模极小不可，也可以取任意阶的模，如 L^{2p} 模极小来求波慢的修改增量，其中 $p = 1, 2, \cdots, \infty$。由向量范数定义(2.5)及(2.6)式可知，此时目标函数可取为

$$Q = \left(\sum_j |\Delta f_j|^{2p} \right)^{1/2p} + \lambda \left(\Delta\tau_i - \sum_j a_{ij}\Delta f_j \right)$$

令 $\partial Q / \partial(\Delta f_j) = 0$ 及 $\partial Q / \partial\lambda = 0$ 最后可得

$$\Delta f_j = \frac{a_{ij}^{1/(2p-1)}}{\sum_j a_{ij}^{2p/(2p-1)}} \Delta\tau_i$$

此式虽然可以用来作迭代修改，但是当 $p>1$时涉及开方运算，速度太慢，一般很少采用。只是在令 $p\to\infty$时，可导出 ART 迭代的最简单修正公式

$$\Delta f_j=\Delta\tau_i/\sum_j a_{ij} \tag{5.85}$$

它说明走时差平均地分配给每一个射线 i 通过的象元，而不考虑象元内射线的长短。这种 ART 修正方法虽然粗糙，但由于修正一次只用到一拍的计算量，仍不失为 ART 的一种常用方法。

ART 迭代过程可以归纳如下。把射线依次编号，从第一条射线($i=1$)到最后一条射线($i=I$)的像元波慢修改称为一轮迭代，下一轮迭代从 $i=1$重新开始。记 $\hat{f}_j^{q,i}$ 为第 q 轮迭代时用第 i 条射线对第 j 个象元的波慢估计算，而第 i 条射线的理论走时为

$$\hat{\tau}_i^q=\sum_j a_{ij}\hat{f}_j^{q,i}$$

对下一条射线($i+1$)，如果它通过像元 j，则由(5.84)式对波慢作如下修改

$$\hat{f}_j^{q,i+1}=\hat{f}_j^{q,i}+\mu\frac{a_{ij}}{\sum_j a_{ij}^2}(\tau_i-\hat{\tau}_i^q) \tag{5.86}$$

式中 μ 为弛豫参数，$0<\mu\leqslant 1$，当 $\mu=1$时即为(5.84)的典型修正公式。把修改增量 Δf_j^q 乘以弛豫参数 μ 的目的是增加计算的稳定性，因为图像重建是不适定问题。如果用 L^∞ 模取极小作修正，则由(5.85)式有

$$\hat{f}_j^{q,i+1}=\hat{f}_j^{q,i}+\mu(\tau_i-\hat{\tau}_i^q)/\sum_j a_{ij} \tag{5.87}$$

迭代过程由给定初值$\{\hat{f}_j^0\}$($j=1,\cdots,J$)开始，然后按 $i=1,\cdots,I$ 由公式(5.86)或(5.87)作第一轮($q=1$)修改，如果所有射线修改完后计算出走时差的模$\|\Delta\tau^q\|<\varepsilon$，$\varepsilon$ 为给定允许误差，则停止迭代，否则再从 $i=1$开始进行 $q+1$轮迭代。弛豫参数 μ 由试验选取，以保证迭代收敛为原则。

除了(5.86)及(5.87)的修正方法外，还有多种 ART 修正方法的变种，可以记为 ARTk，k 表示 ART 算法的变种编号，$k=1,2,\cdots$。这些方法都是按逐条射线依次对象元波慢进行修改的，因此可以把这一类算法数字表示如下。令 $\boldsymbol{r}_i$ 为矩阵 $\boldsymbol{A}$ 的第 i 行向量的转置，它的元为第 i 条射线在各个象元的长度。凡是迭代步骤满足以下表示的离散图像重建算法都称为 ART 算法

$$\boldsymbol{f}^{i+1}=\mathrm{ART}k(\boldsymbol{f}^i,\boldsymbol{r}_i,\tau_i) \tag{5.88}$$

例如，(5.86)的迭代算法表示为

$$\mathrm{ART1}(\boldsymbol{f},\boldsymbol{r},\tau)=\begin{cases}\boldsymbol{f}+\dfrac{\tau-(\boldsymbol{r},\boldsymbol{f})}{(\boldsymbol{r},\boldsymbol{r})}\boldsymbol{r}, & (\boldsymbol{r},\boldsymbol{r})\neq 0\\ \boldsymbol{f} & (\boldsymbol{r},\boldsymbol{r})=0\end{cases} \tag{5.89}$$

式中用内积表示的

$$(\boldsymbol{r},\boldsymbol{r})=\sum_j a_{ij}^2$$

$(\boldsymbol{r},\boldsymbol{r})=0$表示对射线不通过的象元，而

$$(\boldsymbol{r},\boldsymbol{f})=\sum_j a_{ij}f_j$$

为理论计算的走时。

(5.88)式说明，对不论哪一种 ART 算法，它对 $\hat{f}_j^{i-1}$ 进行修改后立刻以 $\hat{f}_j^i$ 存回原存储单元，以供第 $i+1$ 次迭代使用。计算修改值 f_j^i 时只用到 r_i。因为对每个迭代步不必要求对整个矩阵 A 作运算，ART 类方法的这些特点说明它们要求的内存少，计算速度快，但是迭代的收敛性能较差，并依赖于初值选择。当然，收敛性能的改善可通过合适选取弛豫因子 μ 来改善。

(三)最优化准则和联合迭代重建技术(SIRT)

与 ART 类重建技术不同，典型的 SIRT 算法不利用第 i 条射线的修改值 $\hat{f}^i$ 来求慢度 $\hat{f}^{i+1}$ 的修改值，而是把第 q 轮中由所有射线得到修改值保存下来，在本轮对射线迭代结束时求某种平均 $\Delta\bar{f}_j^q$，然后由

$$\hat{f}_j^{q+1}=\hat{f}_j^q+\Delta\bar{f}_j^q$$

对每个象元的波慢作修改，并留作下一轮迭代使用。在上面关于 ART 的讨论中已知，ART 是方程(5.79)的一种最小范数解，那么 SIRT 必然与其中最优化准则联系，让我们回到(5.79)方程求解的优化问题作进一步的讨论。

在方程 $\tau=Af+e$ 中，简单地令误差向量 e 为零，有可能使方程 $\tau=Af$ 无解或有无穷多个解。为此我们只能用一个或多个优化准则使得 τ 的解估计唯一。如果假设 τ 和 e 都是高斯分布的随机向量，由第三章第二节和第三节可知，(5.79)式的最小二乘解估计是找使得

$$\|e\|^2=\|\tau-Af\|^2=(\tau-Af)^{\mathrm{T}}(\tau-Af)$$

极小的解估计，而最大似然解估计是找使得

$$(\tau-Af)^{\mathrm{T}}C_\tau^{-1}(\tau-Af) \tag{5.90}$$

极小的解估计，即以数据的协方差矩阵的逆 C_τ^{-1} 加权的最小二乘解估计。相应地，最小方差解估计是找使得

$$(f-\bar{f})^{\mathrm{T}}C_f^{-1}(f-\bar{f}) \tag{5.91}$$

极小解估计，其中 $\bar{f}$ 是数字化图像的平均密度，C_f 为图象向量 f 的协方差矩阵。当 $\bar{f}$ 固定时，(5.91)取极小等效于找加权的最小范数解估计，或同时使 $\|f\|_2$ 和 $f^{\mathrm{T}}C_f^{-1}f$ 取极小的解估计。于是，可以去找使

$$f^{\mathrm{T}}C_f^{-1}f+f^{\mathrm{T}}f=f^{\mathrm{T}}(C_f^{-1}+I)f \tag{5.92}$$

极小的解估计。根据折衷准则，即分辨宽度和方差不能同时达到最小，我们可以组合(5.90)和(5.92)，将优化准则写为使得

$$\alpha(\tau-Af)^{\mathrm{T}}C_\tau^{-1}(\tau-Af)+f^{\mathrm{T}}(I+\gamma C_f^{-1})f \tag{5.93}$$

取极小，其中折衷参数 α 和 γ 为非负实数。以(5.93)为准则求出的方程(5.79)的解估计叫最小二乘最小范数解估计，当 $\gamma=0$ 时它为最大似然解估计，$\alpha=0$ 时为加权最小范数解估计。

优化准则(5.93)还可以进一步推广到更一般的情况，即求 f 使

$$Q(f)=\alpha(\tau-Af)^{\mathrm{T}}W(\tau-Af)+(f-f_0)^{\mathrm{T}}(\beta B+\gamma C^{-1})(f-f_0) \tag{5.94}$$

取极小，式中，α、β、γ 为非负实数，W 为对称矩阵，B 为非负定矩阵，C 为对称正定距阵，f_0 为与 f 同维的向量，它与我们对图像密度的某种先验知识有关。(5.94)式是二次最优化准则的普遍公式，包括了以前我们讨论的(5.90)到(5.93)作为特殊情况。矩阵 W，B，C 和 f_0 要根据实际问题来选择，如对最小二乘最小范数解估计(5.93)，选 $W=C_\tau^{-1}$，$B=I$ 及 $C=C_f$，f_0 可作为初始猜测(如果对图像向量毫无所知则只能取为零向量)。

由于图像重建问题不适定，(5.94)取极小的解估计仍然可能不唯一。因此，我们还希望

由(5.94)式作二次优化时解估计 $\hat{f}$ 处于解空间的紧集之内。类似于某种正则化的约束条件，可以要求解估计还要满足

$$\|\boldsymbol{D}^{-1}\boldsymbol{f}\|=\min \tag{5.95}$$

的条件，其中 $\boldsymbol{D}$ 为某一对称正定矩阵。于是，图像重建问题变为在满足(5.94)的所有解估计中找同时使(5.95)取极小的解估计。

为导出同时满足(5.94)和(5.95)的图像向量的估计，首先让我们来消掉(5.94)中的一个逆矩阵 $\boldsymbol{C}^{-1}$。由于 $\boldsymbol{C}$ 是正定矩阵，必存在一个矩阵 $\boldsymbol{C}^{1/2}$ 使得

$$\boldsymbol{C}^{1/2}\boldsymbol{C}^{1/2}=\boldsymbol{C}$$

其中 $\boldsymbol{C}^{1/2}$ 也是正定矩阵，且也存在它的逆 $\boldsymbol{C}^{-1/2}$ 使 $\boldsymbol{C}^{-1/2}\boldsymbol{C}^{-1/2}=\boldsymbol{C}^{-1}$。这样我们可以引入一个新向量

$$\boldsymbol{u}=\boldsymbol{C}^{-1/2}(\boldsymbol{f}-\boldsymbol{f}_0) \tag{5.96}$$

使得

$$\boldsymbol{f}=\boldsymbol{C}^{1/2}\boldsymbol{u}+\boldsymbol{f}_0 \tag{5.97}$$

同时把(5.94)式变为

$$\begin{aligned}Q(\boldsymbol{C}^{1/2}\boldsymbol{u}+\boldsymbol{f}_c)=&\alpha[\boldsymbol{\tau}-\boldsymbol{A}(\boldsymbol{C}^{1/2}\boldsymbol{u}+\boldsymbol{f}_0)]^{\mathrm{T}}\boldsymbol{W}[\boldsymbol{\tau}-\boldsymbol{A}(\boldsymbol{C}^{1/2}\boldsymbol{u}+\boldsymbol{f}_0)]\\&+(\boldsymbol{C}^{1/2}\boldsymbol{u})^{\mathrm{T}}(\beta\boldsymbol{B}+\gamma\boldsymbol{C}^{-1})(\boldsymbol{C}^{1/2}\boldsymbol{u})\\=&\boldsymbol{u}^{\mathrm{T}}[\alpha\boldsymbol{C}^{1/2}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A}\boldsymbol{C}^{1/2}+\beta\boldsymbol{C}^{1/2}\boldsymbol{B}\boldsymbol{C}^{1/2}+\gamma\boldsymbol{I}]\boldsymbol{u}\\&-2\alpha\boldsymbol{u}^{\mathrm{T}}\boldsymbol{C}^{1/2}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}(\boldsymbol{\tau}-\boldsymbol{A}\boldsymbol{f}_0)+\alpha(\boldsymbol{\tau}-\boldsymbol{A}\boldsymbol{f}_0)^{\mathrm{T}}\boldsymbol{W}(\boldsymbol{\tau}-\boldsymbol{A}\boldsymbol{f}_0)\\=&\boldsymbol{u}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{u}-2\boldsymbol{u}^{\mathrm{T}}\boldsymbol{z}-\omega\end{aligned} \tag{5.98}$$

式中

$$\left.\begin{aligned}\boldsymbol{P}&=\alpha\boldsymbol{C}^{1/2}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A}\boldsymbol{C}^{1/2}+\beta\boldsymbol{C}^{1/2}\boldsymbol{B}\boldsymbol{C}^{1/2}+\gamma\boldsymbol{I}\\\boldsymbol{z}&=\alpha\boldsymbol{C}^{1/2}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}(\boldsymbol{\tau}-\boldsymbol{A}\boldsymbol{f}_0)\\\omega&=\alpha(\boldsymbol{\tau}-\boldsymbol{A}\boldsymbol{f}_0)^{\mathrm{T}}\boldsymbol{W}(\boldsymbol{\tau}-\boldsymbol{A}\boldsymbol{f}_0)\end{aligned}\right\} \tag{5.99}$$

于是，令 $\partial Q/\partial\boldsymbol{u}^{\mathrm{T}}=0$，即当 Q 取极小时有

$$\hat{\boldsymbol{u}}=\boldsymbol{P}^{-1}\boldsymbol{z} \tag{5.100}$$

回代到(5.97)得到关于图像向量的二次最优化解估计

$$\hat{\boldsymbol{f}}=\boldsymbol{C}^{1/2}\hat{\boldsymbol{u}}+\boldsymbol{f}_0 \tag{5.101}$$

上面我们已经把二次优化的问题转化为(5.100)的方程求解问题。当矩阵 $(\beta\boldsymbol{B}+\gamma\boldsymbol{C}^{-1})$ 为正定时(即 $\boldsymbol{B}$ 为非负定)，$\boldsymbol{P}$ 为非负定矩阵，方程组(5.100)有唯一解。如果 $\boldsymbol{B}$ 不是非负定矩阵，则可以用附加条件(5.95)求最小范数解，其结果也可以用(5.100)表示，只不过此时 $\boldsymbol{P}$ 与矩阵 $\boldsymbol{D}$、$\boldsymbol{W}$ 和 $\boldsymbol{A}$ 有关。但是由于矩阵 $\boldsymbol{P}$ 是 $J\times J$ 的大矩阵，其中 J 为象元总数，我们要采用迭代方法求解。

在联合迭代重建方法中，迭代过程每一步都使方程(5.100)中的误差同时得到改正，即令

$$\boldsymbol{u}_{q+1}=\boldsymbol{u}_q+\mu_q(\boldsymbol{z}-\boldsymbol{P}\boldsymbol{u}_q) \tag{5.102}$$

其中 μ_q 为第 q 轮迭代时的弛豫参数。如(5.102)的迭代格式在数值分析中被称为 Richadrson 方法。利用(5.97)式，第 q 轮迭代时

$$\boldsymbol{f}_q=\boldsymbol{C}^{1/2}\boldsymbol{u}_q+\boldsymbol{f}_0 \tag{5.103}$$

因此，将(5.102)及(5.99)代入(5.103)得

$$\begin{aligned}\boldsymbol{f}_{q+1}=&\boldsymbol{C}^{1/2}[\boldsymbol{u}_q+\mu_q(\boldsymbol{z}-\boldsymbol{P}\boldsymbol{u}_q)]+\boldsymbol{f}_0\\=&\boldsymbol{f}_q+\mu_q[\alpha\boldsymbol{C}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}(\boldsymbol{\tau}-\boldsymbol{A}\boldsymbol{f}_q)+(\beta\boldsymbol{C}\boldsymbol{B}+\gamma\boldsymbol{I})(\boldsymbol{f}_0-\boldsymbol{f}_q)]\end{aligned} \tag{5.104}$$

这就是SIRT典型的迭代修改公式，任意式中并不需要计算 C^{-1} 或 $C^{1/2}$，而只要知道矩阵 C 就可以了。在这些公式中，矩阵 B 和 C 称为平滑矩阵[11]。

从(5.104)式可以看出，SIRT重建算法的特点是在某一轮迭代中，所有象元上的图像函数的平均值(即图像向量 f^{q+1})都用前一轮的近似值 f^q 来修改，而不象ART算法那样逐条射线进行修改。为了和ART算法的修改公式(5.86)或(5.87)进行比较，可考虑最简单情况下对 j 象元的SIRT迭代修改公式。令(5.104)中 $\beta=\gamma=0$，适当选取矩阵 W 和 C 和重新定义不随迭代变化的弛豫参数 μ，对 j 象元的迭代公式可写为

$$\hat{f}_{j,q+1}=\hat{f}_{j,q}+\frac{\sum_i a_{ij}(\tau_i-\hat{\tau}_{i,q})/\sum_i a_{ij}}{\mu+\sum_i a_{ij}} \tag{5.105}$$

其中 $\hat{\tau}_{i,q}$ 为根据慢度近似值 $\hat{f}_{j,q}$ 计算的走时，见(5.86)式；下脚标 i 表示射线号。由(5.105)可见慢度修改增量为通过 j 象元的所有射线的修改增量的加权平均。

SIRT算法虽然有要求机器内存大的缺点，但是收敛性好，是跨孔CT中常用的方法。详见第五节的例子。除了ART和SIRT外，还有利用最大熵准则的图像重建方法及利用先验概率信息的Bayesian重建算法[11-84]，它们在地震层析成像中不常用，而较常用是下面要讨论的共轭梯度法。

(四)阻尼LSQR法

线性层析成像的重建算法主要是求解以下大型稀疏矩阵方程

$$\boldsymbol{Ax}=\boldsymbol{f}, \tag{5.106}$$

其中 A 是由象元内射线段长度组成的稀疏矩阵，x 为象元平均波慢向量，而 f 含走时数据。用最小二乘原理方程(5.106)的法方程可写为

$$\boldsymbol{Bx}\equiv\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{f}\equiv\boldsymbol{b} \tag{5.107}$$

在SIRT的Richardson方法中，对应(5.107)的迭代格式为

$$\boldsymbol{x}_{i+1}=(\boldsymbol{I}-\omega\boldsymbol{B})\boldsymbol{x}_i+\omega\boldsymbol{B} \tag{5.108}$$

其中下标 i 为迭代次数，ω 为松弛因子，I 为单位矩阵。如果取

$$\boldsymbol{x}_0=\boldsymbol{0}$$

则由(5.108)式有

$$\boldsymbol{x}_{i+1}=(\alpha_0\boldsymbol{I}+\alpha_1\boldsymbol{B}+\cdots,\alpha_i\boldsymbol{B}^i)\boldsymbol{b} \tag{5.109}$$

其中 α_i 为常数。(5.109)式表明，迭代输出序列 x_i 处于Kryloy子空间 $K_i(B,b)$ 中。我们希望 x_i 在 i 很大时接近(5.107)式中 x 在 K_i 子空间中的正交投影，于是就有投影类方法发展起来[76, 77]。

对于对称矩阵 B 和向量 b，Lanczos在1950年就提出一种方法产生一组正交投影，而且用它可把 B 简化为三对角阵的形式[78]。将这组正交基记为矩阵形式

$$\boldsymbol{V}_k\equiv[\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_k]$$

由正交性有 $\boldsymbol{V}_i^{\mathrm{T}}\boldsymbol{V}_i=\boldsymbol{I}$，则Lanczos的迭代过程为

$$\begin{aligned}
&\beta_1\boldsymbol{v}=\boldsymbol{b},\boldsymbol{v}_0=0\\
&\text{对 } i=1,2,\cdots\\
&\boldsymbol{w}_i=\boldsymbol{B}\boldsymbol{v}_i-\beta_i\boldsymbol{v}_{i-1}\\
&\alpha_i=\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{w}_i\\
&\beta_{i+1}\boldsymbol{v}_{i+1}=\boldsymbol{w}_i-\alpha_i\boldsymbol{v}_i
\end{aligned} \tag{5.110}$$

式中 β_i 与残差向量 r_{i+1} 及 r_i 的模的比值有关，只取正值。向量 $v_1, v_2, \cdots, v_i$ 称为 Lanczos 向量，它构成了 $K_i(B,b)$ 子空间的规格化正交基，为求解方程还需求出 $Bx=b$ 在此子空间的投影 $x_i \in K_i$。记三对角线矩阵[76,77]

$$T_k \equiv \text{tridiag}(\beta_i, \alpha_i, \beta_{i+1})$$

可得

$$BV_k = V_k T_k + \beta_{k+1}(0,0,\cdots,0,v_{k+1}) \tag{5.111}$$

此式两边乘以任意向量 y_k 得

$$BV_k y_k = V_k T_k y_k + \beta_{k+1}(0,0,\cdots,0,v_{k+1}) y_k \tag{5.112}$$

令

$$x_k = V_k y_k, T_k y_k = \beta_1 (1,0,\cdots,0)^T = \beta_1 e_1^T \tag{5.113}$$

并由(5.110)式中 $b=\beta_1 v_1$ 有

$$Bx_k = b + \eta_k \beta_{k+1} u_{k+1} \tag{5.114}$$

当 $\eta_k \beta_{k+1}$ 小到可忽略时，便可求出方程组(5.107)的近似解，根据 $V_i \perp K_i$ 的正交投影条件可知，$x_k \in K$ 即为方程组(5.107)在此子空间的投影。

对于方程(5.106)的阻尼最小二乘问题，我们令

$$\left\| \begin{bmatrix} A \\ \lambda I \end{bmatrix} x = \begin{bmatrix} b \\ 0 \end{bmatrix} \right\|_2 \tag{5.115}$$

取极小，其中实常数 λ^2 称为阻尼因子。换句话说，要求解方程

$$\begin{bmatrix} I & A \\ A^T & -\lambda^2 I \end{bmatrix} \begin{bmatrix} r \\ x \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix} \tag{5.116}$$

其中 $r=b-Ax$ 为残差向量。利用 Lanczos 的迭代格式(5.110)来求解方程(5.116)便可以构成带阻尼的最小二乘新算法。仿照(5.110)到(5.113)的推导，经过 $2k+1$ 次迭代后有

$$\begin{bmatrix} I & B_k \\ B_k^T & -\lambda^2 I \end{bmatrix} \begin{bmatrix} t_{k+1} \\ y_k \end{bmatrix} = \begin{bmatrix} \beta_1 & e_1 \\ 0 & 0 \end{bmatrix}$$

和

$$\begin{bmatrix} r_k \\ x_k \end{bmatrix} = \begin{bmatrix} U_{k+1} & 0 \\ 0 & v_k \end{bmatrix} \begin{bmatrix} t_{k+1} \\ y_k \end{bmatrix} \tag{5.117}$$

其中 B_k 是 $(k+1)\times k$ 的下双角阵。这里 y_k 是另一个阻尼最小二乘问题

$$\min \left\| \begin{bmatrix} B_k \\ \lambda I \end{bmatrix} y_k - \begin{bmatrix} \beta_1 & e_1 \\ 0 & 0 \end{bmatrix} \right\|$$

的解，即

$$B_k y_k = \beta (1,0,\cdots,0)^T$$

的最小二乘解，可以很方便地用 QR 分解法求解，这便是带阻尼的最小二乘 QR 算法(LSQR)的原理。

注意到阻尼因子与矩阵 B_k, U_{k-1} 及 V_k 无关，即当 $\lambda=0$ 时这些矩阵的计算应完全一样。对于大扰动的非均匀介质的成像，λ 可随修改模型的非线性迭代(不是上述解方程的迭代)而减小，以提高成像的分辨率，参见文献[5～6]。

带阻尼的 LSQR 算法可归纳如下：

(1)初始化

$$\beta_1\boldsymbol{u}_1=\boldsymbol{b},\alpha_1\boldsymbol{v}_1=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}_1,\boldsymbol{w}_1=\boldsymbol{v}_1,\boldsymbol{x}_0=\boldsymbol{0},\boldsymbol{d}_1=\boldsymbol{0}$$

$$\gamma_1=0,\bar{\gamma}_1=0,\bar{\varphi}_1=\beta_1,\bar{\rho}=\alpha_1$$

其中 $\boldsymbol{b}_1,\boldsymbol{u}_1$ 为 m 维向量，$\boldsymbol{d}_1,\boldsymbol{v}_1,\boldsymbol{w}_1,\boldsymbol{x}_0$ 为 n 维向量，$\beta_1,\alpha_1,\gamma_1,\varphi_1,\rho_1$ 为实数。

(2)对 $i=1,2,3\cdots$，做以下各步

(3)QR 分解

(a)$\gamma_{i+1}\boldsymbol{u}_{i+1}=\boldsymbol{A}\boldsymbol{v}_i-\alpha_i\boldsymbol{u}_i$

(b)$\bar{\gamma}_{i+1}\boldsymbol{d}_{i+1}=\lambda\boldsymbol{v}_i-\alpha_i\boldsymbol{d}_i$

(c)$\beta_{i+1}=(\gamma_{i+1}^2+\bar{\gamma}_{i+1}^2)^{1/2}$

(d)$\boldsymbol{u}_{i+1}=\gamma_{i+1}\boldsymbol{u}_{i+1}/\beta_{i+1}$

(e)$\boldsymbol{d}_{i+1}=\bar{\gamma}_{i+1}\boldsymbol{d}_{i+1}/\beta_{i+1}$

(f)$\boldsymbol{v}_{i+1}=\lambda\boldsymbol{d}_{i+1}-\beta_{i+1}\boldsymbol{v}_i$

(g)$\alpha_{i+1}\boldsymbol{v}_{i+1}==\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}_{i+1}-\boldsymbol{v}_{i+1}$

(4)修改参数

(a)$\rho_i=(\bar{\rho}_i^2+\beta_{i+1}^2)^{1/2}$

(b)$c_i=\bar{\rho}_i/\rho_i$

(c)$s_i=\beta_{i+1}/\rho_i$

(d)$\theta_{i+1}=s_i\alpha_{i+1}$

(e)$\bar{\rho}_{i+1}=-c_i\boldsymbol{d}_{i+1}$

(f)$\varphi_i=c_i\bar{\varphi}_i$

(g)$\bar{\varphi}_{i+1}=s_i\varphi_i$

(5)迭代求解

(a)$\boldsymbol{x}_i=\boldsymbol{x}_{i-1}+(\varphi_i/\rho_i)\boldsymbol{w}_i$

(b)$\boldsymbol{w}_{i+1}=\boldsymbol{v}_{i+1}-(\theta_{i+1}/\rho_i)\boldsymbol{w}_i$

(6)收敛判别。

以上迭代算法和 LSQR 不同之处在步骤(3)。事实上，阻尼因子 λ^2 对线性方程组 $\boldsymbol{Ax}=\boldsymbol{b}$ 的解的影响也在步骤(3)得到了体现。在步骤(3)，当 $\lambda=0$ 时，则与 LSQR 相同；若 $\lambda\neq0$ 时，$\boldsymbol{Ax}=\boldsymbol{b}$ 的解得到相应的圆滑。关于带阻尼 LSQR 算法的误差分析可参考文献[2,3]。

带阻尼的最小二乘 QR 分解算法在观测数据有较大误差时仍能控制解估计的分辨率及方差。对于大扰动的非均匀介质成像(非线性反问题)，需要经过逐步修改地球模型和射线路径，更要求不断调整每次逼近时控制线性方程求解方差的参数，而带阻尼的 LSQR 方法实现了用阻尼因子 λ^2 来控制逼近过程的思想。

五、广义 Radon 变换问题

(一)走时反演和广义 Radon 变换

在本节中我们从沿弯曲射线的地震波走时的角度，考虑如何把 Radon 变换推广到弯曲射线的情况，以实现波慢的成像。这时，类似于直射线情况下的 Radon 变换式(5.43)，在二维情况下，透射地震波走时可以写为

$$\tau(\xi,\theta)=\int_{L_{\xi,\theta}} N(x,y)\mathrm{d}\sigma$$

式中 $L_{\xi,\theta}$为弯曲射线的路径，$\mathrm{d}\sigma$ 为此曲线的弧长微分，θ 为入射角，ξ 为垂直入射方向走时观测器的坐标，研究区域$(x,y)\in\Omega\subset\mathscr{R}^2$。用图像函数和投影来表示上式，则有

$$[\mathscr{R}_a f](\xi,\theta)=\int_{L_{\xi,\theta}} f(x,y)\mathrm{d}\sigma \tag{5.118}$$

在这里 $f(x,y)$为图像函数，$\mathscr{R}_a$ 为沿曲线 $L_{\xi,\theta}$的广义 Radon 变换。由于投影函数$[\mathscr{R}_a f](\xi,\theta)$取决于积分路径 $L_{\xi,\theta}$，因此必须对积分路径作出具体的描述和规定。

回忆在直射线情况下的 Radon 变换式(5.46)的情况，在那时直射线表示为以下直线族

$$\phi(\boldsymbol{r},\boldsymbol{s})=\boldsymbol{s}\cdot\boldsymbol{r}=x\cos\theta+y\sin\theta$$

其中 $\boldsymbol{r}=(x,y)$为表示图像中点的向量，$\boldsymbol{s}=(\cos\theta,\sin\theta)$为表示射线方向的单位向量，因此 $\phi(\boldsymbol{r},\boldsymbol{s})$确定了区域 Ω 上的以下曲线族

$$\{\boldsymbol{r}\in\Omega|_{\xi=\phi(\boldsymbol{r},\boldsymbol{s})},\xi\in\mathscr{R}^1,\cos^2\theta+\sin^2\theta=1\}$$

然而在弯曲射线情况下，射线已经不能用内积 $\boldsymbol{s}\cdot\boldsymbol{r}$ 表示，而是满足一定条件的实值函数，而且射线方向也不再能用 θ 的三角函数表示，而要表示成向量 $\boldsymbol{\omega}=(\omega_1,\omega_2)$，其满足单位模 $\omega_1^2+\omega_2^2=1$，或$\|\boldsymbol{\omega}\|=1$。于是，用 $\phi(\boldsymbol{r},\boldsymbol{\omega})$表示并以 γ 为参变量的射线族可写为

$$H_{\xi,\gamma}=\{\boldsymbol{r}\in\Omega|_{\xi=\phi(\boldsymbol{r},\gamma)},\xi\in\mathscr{R}^1,\|\gamma\|=1\} \tag{5.119}$$

现在，我们可以参照(5.46)式把广义 Radon 变换用 δ 函数表为

$$[\mathscr{R}_a f](\xi,\boldsymbol{\omega})=\int_{\Omega} f(\boldsymbol{r})\delta(\xi-\phi(\boldsymbol{r},\boldsymbol{\omega}))\mathrm{d}\boldsymbol{r} \tag{5.120}$$

或者写为标量形式

$$[\mathscr{R}_a f](\xi,\boldsymbol{\omega})=\iint_{-\infty}^{\infty} f(x,y)\delta(\xi-\phi(x,y,\omega_1,\omega_2))\mathrm{d}x\mathrm{d}y \tag{5.121}$$

其中 $\omega_2^2=1-\omega_1^2$。

值得注意的是，上面我们只说明了射线弯曲时，透射地震波的走时可以用介质波慢的广义 Radon 变换来表示，并未对广义 Radon 变换作出严格的定义(它在本节后面将讨论)。和直射线情况一样，图像重建问题是根据投影$(\mathscr{R}_a f)$求广义 Radon 变换的逆；由于这个问题是十分复杂的问题，我们先来看看(5.119)属于某些比较简单的曲线族的情况。

(二)简单射线情况下广义 Radon 变换的逆

由上所述在二维的情况下，图像函数 $f(x,y)$沿曲线族 L 的广义 Radon 变换可由(5.118)式(5.120)式表示。如果射线是一些比较简单的曲线族，则可以利用复变函数中共形映射的方法把它们映射为直线族，把广义 Radon 变换求逆的问题化为标准的 Radon 变换来求逆[12]。用 $z=x+\mathrm{i}y$ 表示 z 平面上的点，图像函数 $f(x,y)$定义在其中的区域 $\boldsymbol{D}_z$。又设 $w=u+\mathrm{i}v$ 为另一平面，而函数 $z=F(w)$构成 w 平面上的区域 $\boldsymbol{D}_w$ 到 z 平面区域 $\boldsymbol{D}_z$ 的共形映射，则 w 平面上的直线族

$$l_{\xi,\theta}:\quad u\cos\theta+v\sin\theta=\xi$$

在函数 $z=F(w)$的作用下映射为 z 平面的曲线族

$$L_{\xi,\theta}:\quad u(x,y)\cos\theta+v(x,y)\sin\theta=\xi$$

由共形映射的条件可知，$u(x,y)$和 $v(x,y)$必须是一对共轭调和函数。

现设在 w 平面上直线 $l_{t,\theta}$ 的弧长微分为 ds，则由共形映射的性质有

$$d\sigma=\left|\frac{dz}{dw}\right|ds$$

代入(5.118)式得

$$\begin{aligned}[\mathscr{R}_a f](\xi,\theta)&=\int_{l_{t,\theta}} f(F(w))\left|\frac{dz}{dw}\right|ds\\&=\int_{l_{t,\theta}} f(w)ds\end{aligned}\tag{5.122}$$

其中

$$f(w)=f(F(w))\left|\frac{dz}{dw}\right|\tag{5.123}$$

(5.122)式说明，在 z 平面图像 $f(z)$ 沿曲线族 L 的广义 Radon 变换就是在 w 平面上函数 $\hat{f}(w)$ 沿直线 l 的标准 Radon 变换。因此，可以用标准 Radon 变换求逆的方法求得广义 Radon 变换的逆，具体步骤如下。

1. 由保形映射 $z=F(w)$ 找出 u,v 和 x,y 的关系并计算导数 $|dz/dw|$。

2. 将投影函数 $[\mathscr{R}_a f]_L$ 代入(5.122)式左边，用本章第二节讨论的标准 Radon 变换的方法求逆 $\hat{f}(w)$

3. 由(5.123)式可知，图像函数

$$f(x,y)=f(z)=f(w(z))/\left|\frac{dz}{dw}\right|\tag{5.124}$$

上述利用保形映射的方法求广义 Radon 变换的逆时要求找出解析函数 $F(w)$，同时因子 $\left|\frac{dz}{dw}\right|$ 不宜过大，其理由见下面的例子。此外注意到在平面 z 上 $f(x,y)$ 所在的图形区域 D_w 不再是正方形，在进行上述的步骤2时要考虑非方形区域的标准 Radon 变换求逆问题。

现在看看保形映射应用于广义 Radon 变换求逆的一个例子。例如，考虑射线族为过一定点 $z=i$ 的圆周族，此定点相当于地震中的自发自收点，则将 w 平面上的直线映射为 z 平面上过此点圆周的映射为

$$z=\frac{\pi+iw}{\pi i+w}\tag{5.125}$$

或

$$w=\frac{\pi(1-zi)}{z-i}\tag{5.126}$$

设 $w=u+iv$ 及 $z=x+iy$，则由上两式有

$$\left.\begin{aligned}u&=\frac{2\pi x}{x^2+(y-1)^2} \quad v=\frac{\pi(1-x^2-y^2)}{x^2+(y-1)^2}\\x&=\frac{2\pi u}{u^2+(\pi+v)^2} \quad y=\frac{u^2+v^2-\pi^2}{u^2+(\pi+v)^2}\end{aligned}\right\}\tag{5.127}$$

及

$$\left|\frac{dz}{dw}\right|=\frac{x^2+(y-1)^2}{2\pi}=\frac{2\pi}{u^2+(\pi+v)^2}\tag{5.128}$$

由于(5.125)式把 w 平面上的直线 $l_{t,\theta}=u\cos\theta+v\sin\theta$ 映照为 z 平面上过点 $z=(0,i)$ 的圆周，它的圆心位于

$$\left(\frac{\pi\cos\theta}{\xi+\pi\sin\theta},\frac{\xi}{\xi+\pi\sin\theta}\right)$$

而半径为

$$\frac{\pi}{|\pi\sin\theta+\xi}$$

再由(5.122)式

$$[\mathscr{R}_a f](\xi,\theta)=\int_{l_{\xi,\theta}} f(u,v)\mathrm{d}s$$

利用上二节讲过的连续图像重建方法便可以求出 $f(u,v)$。注意此时图像函数 $f(x,y)$的取值应限制在 $x\in[-1,1]$及 $y\in[-2,0]$的正方形区域范围内，这时由(5.128)有

$$0.159=\frac{1}{2\pi}\leqslant\left|\frac{\mathrm{d}z}{\mathrm{d}w}\right|\leqslant\frac{10}{2\pi}=1.59$$

于是由(5.124)到(5.127)式便有图像函数

$$f(x,y)=f\left(\frac{2\pi x}{x^2+(y-1)^2},\frac{\pi(1-x^2-y^2)}{x^2+(y-1)^2}\right)\Big/\left(\frac{x^2+(y-1)^2}{2\pi}\right)$$

这时由于$|\mathrm{d}z/\mathrm{d}w|$的值与1相差不大，算法的成像精度便可以保证。否则由于图像重建属不适定问题，$|\mathrm{d}z/\mathrm{d}w|$值过小时便会把 f 中由数据带来的误差放得很大，而且 $f(x,y)$的取样点分布变得极不均匀。

除了本例中 $L_{\xi,\theta}$属于圆的广义 Radon 变换求逆之外，凡是满足 $u(x,y)$和 $v(x,y)$为共轭调和函数的曲线$\{L_{\xi,\theta}\}$都可用共形映射的方法求逆。然而，满足这一条件的曲线族是(5.119)式的 $H_{\xi,\gamma}$中有限的类型，对于复杂波慢结构的介质其射线族很少能满足这种条件。在地震层析成像的一般情况下，我们必须讨论更广泛的射线类型的情况。此时广义 Radon 变换的求逆问题乃是 Beylkin 于1984年定义的[86]。

(三)广义 Radon 变换和它的对偶变换[86,87]

由上面的讨论可知，广义 Radon 变换取决于(5.119)式中射线族的性质，Beylkin(1984)考虑了更为一般的情况，即对 n 维欧几里德空间 $\mathscr{R}^n$ 中的图像函数 $f(\boldsymbol{r})$的广义 Radon 变换。相应地，他用 $\phi(\boldsymbol{r},\boldsymbol{\omega})$来定义 $\mathscr{R}^n$ 空间某个开集 $\mathscr{X}$ 内的曲面。规定 $\phi(\boldsymbol{r},\boldsymbol{\omega})$满足以下四个条件。

(1)$\phi(\boldsymbol{r},\boldsymbol{\omega})$是在 $\mathscr{X}$ 上的实值 $\mathscr{C}^\infty$ 函数(即无穷阶可微函数)。

(2)$\phi(\boldsymbol{r},\boldsymbol{\omega})$是关于 $\boldsymbol{\omega}$ 的齐次和一次函数(简称为齐一次函数)，即对实数 λ 有 $\phi(\boldsymbol{r},\lambda\boldsymbol{\omega})=\lambda\phi(\boldsymbol{r},\boldsymbol{\omega})$。

(3)ϕ 对 $\boldsymbol{r}$ 的微分 $\mathrm{d}_r\phi$ 在开集 $\mathscr{X}$ 上不同时为零。

(4)在 $\mathscr{X}$ 内存在着用行列式定义的以下函数。

$$h(\boldsymbol{r},\boldsymbol{\omega})=\det\left(\frac{\partial^2\phi(\boldsymbol{r},\boldsymbol{\omega})}{\partial r_j\partial\omega_k}\right)>0 \tag{5.129}$$

其中 $j=1,\cdots,n$。在这样的规定下，对于常规 Radon 变换的情况有 $\phi(\boldsymbol{r},\boldsymbol{s})=\boldsymbol{r}\cdot\boldsymbol{s}$，其中"·"是 $\mathscr{R}^n$ 空间的内积。

在这样规定 ϕ 之后，相应于(5.119)的在域 $\mathscr{X}$ 内的$(n-1)$维曲面族就是

$$H_{\xi,\gamma}=\{\boldsymbol{r}\in\mathscr{X}|_{\xi=\phi(\boldsymbol{r},\gamma)},\xi\in\mathscr{R}^1,\gamma\in S^{n-1}\} \tag{5.130}$$

式中 S^{n-1}表示$(n-1)$维 Sobolev 空间，$|\gamma|=1$。

在二维情况下，$\mathscr{X}$ 为平面上的一个区域，相应于(5.119)的曲线族中函数 ϕ 满足的条件简化为

(1)$\phi(r,\omega)$为平面区域$\mathscr{X}$上的$\mathscr{C}^{\infty}$实值函数，$r=(x,y)$及$\omega=(\omega_1,\omega_2)$。

(2)$\phi(r,\omega)$为关于ω的齐一次函数。

(3)$\partial\phi/\partial x$和$\partial\phi/\partial y$在$\mathscr{X}$上不同时为零，即

$$\left(\frac{\partial\phi}{\partial x}\right)^2+\left(\frac{\partial\phi}{\partial y}\right)^2\neq 0$$

(4)在$\mathscr{X}$上

$$h(r,\omega)=\begin{vmatrix}\dfrac{\partial^2\phi}{\partial x\partial\omega_1} & \dfrac{\partial^2\phi}{\partial x\partial\omega_2}\\ \dfrac{\partial^2\phi}{\partial y\partial\omega_1} & \dfrac{\partial^2\phi}{\partial y\partial\omega_2}\end{vmatrix}>0 \tag{5.131}$$

由上述规定可知，由(5.131)规定的ϕ属于一大类比较一般的曲线，可以用于比较多的射线类型；而(5.129)和(5.130)规定的曲线族便于地球三维波速结构的研究。

在规定了ϕ及相应曲线族之后，Beylkin把属于$\mathscr{C}_0^{\infty}(X)$空间的图像函数$f(r)$的广义Radon变换定义为

$$[\mathscr{R}_a f](\xi,\omega)=\int_{H_{\xi,\omega}} f(r)a(r,\omega)\varphi \tag{5.132}$$

式中$a(r,\omega)$是一个可供选取的正值函数，叫密度函数，并假定$a\in\mathscr{C}^{\infty}$，$r\in\mathscr{R}^n$，$\omega\in\mathscr{S}^{n-1}$，$a(r,\omega)=a(r,-\omega)$。$\varphi$是曲面微分元，它包含多项微分元。

例如在二维情况下，

$$\varphi=\frac{\dfrac{\partial\phi}{\partial x}\mathrm{d}y-\dfrac{\partial\phi}{\partial y}\mathrm{d}x}{\left(\dfrac{\partial\phi}{\partial x}\right)^2+\left(\dfrac{\partial\phi}{\partial y}\right)^2} \tag{5.133}$$

广义Radon变换的一般定义(5.132)可以化为我们已经讨论过的(5.118)沿曲线族(5.119)的第一型曲线积分的形式。这时，选取

$$a(r,\omega)=\left(-\operatorname{sign}\frac{\partial\phi}{\partial y}\right)\sqrt{\left(\frac{\partial\phi}{\partial x}\right)^2+\left(\frac{\partial\phi}{\partial y}\right)^2} \tag{5.134}$$

由于这时在(5.119)定义的曲线族$H_{\xi,\omega}$上有

$$\frac{\partial\phi}{\partial x}\mathrm{d}x+\frac{\partial\phi}{\partial y}\mathrm{d}y=0$$

即

$$\mathrm{d}y=-\frac{\partial\phi}{\partial x}\mathrm{d}x\Big/\left(\frac{\partial\phi}{\partial y}\right)$$

代入(5.133)和(5.134)得

$$a(r,\omega)\varphi=\sqrt{\mathrm{d}x^2+\mathrm{d}y^2}=\mathrm{d}\sigma \tag{5.135}$$

其中$\mathrm{d}\sigma$为沿弧长的微分。把(5.135)式代入(5.132)便有

$$[\mathscr{R}_a f](\xi,\omega)=\int_{H_{\xi,\omega}} f(r)a(r,\omega)\varphi=\int_{H_{\xi\omega}} f(r)\mathrm{d}\sigma$$

这就是(5.118)式。

虽然我们已说明由(5.132)式定义的广义Radon变换可以化为以前我们熟悉的(5.118)式，但是表面看起来它似乎和以前二维情况下由射线导出的其它表示形式不一样。

其实，(5.132)式亦可用 δ 函数表示为

$$[\mathscr{R}_a f](\xi,\omega)=\int_{\mathscr{D}^n} f(r)a(r,\omega)\delta(\xi-\phi(r,\omega))\mathrm{d}r \tag{5.136}$$

上式左边为投影函数，$f(r)$为图像函数，a 为密度函数，由 δ 函数知右边的积分沿着 $\mathscr{R}^n$ 空间中 $\xi=\phi(r,\omega)$的超曲面进行。因此，变换(5.136)为函数 $f(r)$在超曲面 $\xi=\phi(r,\omega)$上乘权函数 $a(r,\omega)$的积分。比较(5.136)式和(5.120)式，令 $a=1$和 $n=2$两者便完全相同。这时如果再令 $\phi=r\cdot\omega$ 则变为标准 Radon 变换。当(5.136)式中取密度 $a(r,\omega)=e^{m(r,\omega)}$，其中 $m(r,\omega)$为无穷可微函数，则(5.136)式变为所谓指数衰减的 Radon 变换，它是考虑介质吸收作用的所谓发射型 CT 图像重建的基本公式，也是广义 Radon 变换的一种特殊情况。实际上，由于广义 Radon 变换是新近才被重视的研究课题，它的表现形式是不断发展的，Beylkin 本人就在文献[86]中对(5.136)式作了进一步的推广。

对于层析成像来说，我们最关心的是广义 Radon 变换求逆的方法，即希望象标准 Radon 变换那样，找到一个类似于(5.53)那样满足(5.51)式的逆 $\mathscr{R}_a^{-1}$。然而，广义 Radon 变换的逆要比标准 Radon 变换复杂得多，而且用如同(5.53)中那种一般的微积分算子是表示不出来的，而必须用到象 Fourier 积分算子那样的新的数学工具。因此，我们将在下一章讲完 Fourier 积分算子的概念之后，再讨论广义 Radon 变换的求逆问题。这里，仅讨论一下广义 Radon 变换的对偶变换。

由(5.132)及(5.136)定义的广义 Radon 变换的对偶变换 $\mathscr{R}_a^*$ 不等于逆 $\mathscr{R}_a^{-1}$，只是逆的一部分，相当于标准 Radon 变换式(5.63)中的反投影算子。因此，$\mathscr{R}_a^*$ 又称为广义反投影算子或拟逆。设 $g(\xi,\omega)$是一个属于 $\mathscr{E}^\infty$ 空间的函数，其中 $\xi\in\mathscr{R}^1$，$\omega\in\mathscr{S}^{n-1}$，$g(\xi,\omega)=g(-\xi,-\omega)$，则对偶变换定义为

$$[\mathscr{R}_a^* g](r')=\int_{\|\omega\|=1} b(r',\omega)g(\xi,\omega)\big|_{\xi=\phi(r',\omega)}\mathrm{d}\omega \tag{5.137}$$

式中左边为 g 的对偶变换$[\mathscr{R}_a^* g]$，从层析成像的观点看，它是图像空间的函数，但不等于图像函数 $f(r)$，而右边的

$$b(r',\omega)=h(r',\omega)/a(r',\omega) \tag{5.138}$$

a 为(5.132)及(5.136)式中的密度函数，h 是由(5.129)或(5.131)定义为 ϕ 的偏导数的行列式。

如果在(5.137)式中取 g 为投影函数$[\mathscr{R}_a f]$的如下积分

$$g(\xi',\omega)\big|_{\xi'=\phi(r',\omega)}=\int_{-\infty}^{\infty}[\mathscr{R}_a f](\xi,\omega)K(\xi-\xi')\big|_{\xi'=\phi(r',\omega)}\mathrm{d}\xi \tag{5.139}$$

其中积分核

$$K(\xi)=\frac{1}{2(2\pi)^n}\int_{-\infty}^{\infty}U(\gamma)|\gamma|^{n-1}e^{i\gamma\xi}\mathrm{d}\gamma \tag{5.140}$$

由(5.139)取得的 $f^*(r')=[\mathscr{R}_a^* g](r')$就是广义 Radon 变换$[\mathscr{R}_a f]$的所谓拟逆，其中 $U(\gamma)$为可供选取的无限可微实函数。拟逆 f^* 和图像函数 $f(r)$相比差一连续函数，因此两者的间断面是相同的。因此利用对偶变换公式(5.137)～(5.140)可以由投影重建图像的间断面(点)，这种类型的图像重建反问题被称为奇性反演。

(四)拟逆的计算

以上介绍了 Beylkin 利用 Fourier 积分算子在理论上得出的求广义 Radon 变换拟逆的

公式(5.137)～(5.140)，它们在实际应用时存在一些困难。首先，当(5.140)式中函数$U(\gamma)$衰减得不够快时，$K(\xi)$是一个广义函数，(5.139)式是关于广义函数$K(\xi)$的褶积，作数值计算有困难。再者(5.140)式出现因子$e^{i\gamma\xi}$，因此属振荡积分，不能用一般的数值求积方法计算，否则误差难以控制。为克服这些困难，李世雄等提出了在二维情况下计算拟逆的以下算法[12]。

在二维广义 Radon 变换的情况下，参照共形映射的情况把射线$\phi(r,\omega)=\xi$记为以下形式

$$\alpha(x,y)\cos\theta+\beta(x,y)\sin\theta=\xi \tag{5.141}$$

其中$r=(x,y)$，$\omega=(\omega_1,\omega_2)$，$\omega_1=\cos\theta$，$\omega_2=\sin\theta$；和共形映射时不同，$\alpha(r)$和$\beta(r)$不限于一对共轭调和函数，而只要求$\phi(r,\omega)$满足(5.131)规定的四个条件即可。令$n=2$，并设(5.140)式中$U(\gamma)=1$，(5.139)代入(5.137)并变换积分次序得拟逆的一种形式

$$f^*(r)=\frac{1}{4\pi^2}\int_0^{2\pi}b(r,\omega)\mathrm{d}\theta\left\{\int_0^{\infty}w(\omega,\gamma)e^{-i\phi(r,\omega)\gamma}\gamma\mathrm{d}\gamma\right\} \tag{5.142}$$

其中

$$w(\omega,\gamma)=\int_{\xi_1}^{\xi_2}[\mathscr{R}_af](\xi,\omega)e^{i\gamma\xi}\mathrm{d}\xi$$

式(5.142)中的$\phi(r,\omega)$是由(5.141)定义的，对应$\xi=\phi(r,\omega)$的限制，$b(r,\omega)$由(5.138)式定义，其中密度函数a由(5.134)式计算，h由(5.131)式计算。由于(5.142)方括号内的积分都是振荡积分，计算时要作以下处理。

将(5.142)式中的指数项展开后，它们中的积分可归为以下形式的振荡积分

$$I_1(\gamma)=\int_{\xi_1}^{\xi_2}[\mathscr{R}_af](\xi)\cos(\gamma\xi)\mathrm{d}\xi \tag{5.143}$$

和

$$I_2(\gamma)=\int_{\xi_1}^{\xi_2}[\mathscr{R}_af](\xi)\sin(\gamma\xi)\mathrm{d}\xi \tag{5.144}$$

当ξ很大时，I_1和I_2的被积函数剧烈振荡，致使计算误差很大。为克服这一困难，可将离散的投影函数$[\mathscr{R}_af]$的观测值用一多项式来拟合，要求此多项式不仅函数值逼近而且导数值也逼近，才能使I_1和I_2的计算比较精确。拟合数据的多项式可选择以下三次样条多项式。

在区间$[\xi_1,\xi_2]$上取分点$\xi_i=s_0<s_1<\cdots<s_n=\xi_2$，则可把区间$[s_k,s_{k+1}]$上的三次样条多项式表为

$$g(s)=P_k(s)=\alpha_ks^3+\beta_ks^2+\psi_ks+\varphi_k,\quad k=0,1,\cdots,n-1$$

于是(5.143)中的积分可近似表为

$$I_1(\gamma)\simeq\int_{\xi_1}^{\xi_2}g(s)\cos(\gamma\xi)\mathrm{d}\xi=\sum_{k=0}^{n-1}\int_{S_{k-1}}^{S_k}P_k(s)\cos(\gamma s)\mathrm{d}s$$

通过分部积分法可得I_1的解析表达式

$$\begin{aligned}I_1(\gamma)=\sum_{k=0}^{n-1}\Bigg\{&\left[\frac{P'_k(s_k)}{\gamma^2}-\frac{P'''_k(s_k)}{\gamma^4}\right]\cos(\gamma s_k)\\&-\left[\frac{P'_k(s_{k-1})}{\gamma^2}-\frac{P'''_k(s_{k-1})}{\gamma^4}\right]\cos(\gamma s_{k-1})\\&+\left[\frac{P_k(s_k)}{\gamma}-\frac{P''_k(s_k)}{\gamma^3}\right]\sin(\gamma s_k)\end{aligned}$$

$$-\left(\frac{P_k(s_{k-1})}{\gamma}-\frac{P''_k(s_{k-1})}{\gamma^3}\right)\sin(\gamma s_{k-1})\Bigg\} \tag{5.145}$$

其中 $P'''_k(s_k)=P'''_k(s_{k-1})=6\alpha_k$。

对(5.144)式中的 $I_2(\gamma)$也可以作相似的推导。将(5.145)代入(5.142)式对 γ 积分，可分解为以下三种类型的数值积分计算

$$\int_0^\infty \frac{\cos k\gamma}{\gamma}\mathrm{d}\gamma,\ \int_0^\infty \frac{\sin k\gamma}{\gamma^2}\mathrm{d}\gamma,\ \int_0^\infty \frac{\cos k\gamma}{\gamma^3}\mathrm{d}\gamma$$

它们在 $\gamma\to 0$时发散。但是可以证明，当投影$[\mathscr{R}_\omega f](\xi)$有对 ξ 的三阶导数时这些无穷积分是收敛的，可采取以下数值计算公式

$$\int_\varepsilon^\infty \frac{\cos k\gamma}{\gamma}\mathrm{d}\gamma=-(0.577216+\ln|k|+\ln\varepsilon)$$

$$\int_\varepsilon^\infty \frac{\sin k\gamma}{\gamma^2}\mathrm{d}\gamma=k(1-\ln|k|-\ln\varepsilon-0.577216)$$

$$\int_\varepsilon^\infty \frac{\cos k\gamma}{\gamma^3}\mathrm{d}\gamma=\frac{k^2}{2}\left(\ln|k|+\ln\varepsilon+0.577216-\frac{3}{2}\right)+\frac{1}{2\varepsilon^2}$$

在 $\varepsilon\to 0$时，与 $\ln\varepsilon$ 及$1/\varepsilon^2$有关的奇异项可以相互抵消，因此计算时可以把这些项略去。以上求广义 Radon 变换的算法的优点在于不仅避免了计算振荡积分引起的误差的发散，而且由于解析表达式的导出而使(5.142)中方括号的两重积分计算化为代数运算，大大减少了计算成本。

六、衍射层析成像[102~106]

在上一节已经讨论过当射线弯曲时利用地震走时对波幔成像的方法，它实际上属于衍射 CT 的范畴。当探测波长比典型的探测目标大得不多时，波在介质性质发生变化处产生衍射及散射等现象，而不再沿直线传播，这个现象叫衍射。在地震 CT 中所谓的衍射层析成像(Diffraction Tomography)实际上指的是在考虑波非直线传播的情况下，用各种透射地震方法收集的数据对地球介质成像，之所以不用透射 CT 这个词主要是要强调它与(5.64)式讨论的射线 CT 技术的区别。除了考虑非直线传播的情况之外，衍射 CT 考虑的地球模型一般是无间断面的连续变化的介质，即波速 $C(\boldsymbol{r})\in\mathscr{C}^\infty(\mathscr{R}^3)$，这是和反射地震数据成像的不同之处。正如图5.1所示，在跨孔地震采集系统及 VSP 的情况下我们一般要考虑射线弯曲的情况，这时用衍射 CT 技术对波速成像要准确一些。本节讨论用波场数据成像的衍射 CT 技术，它是在医学上的超声 CT 的基础上发展起来的。让我们首先从与 Fourier 投影定理(见(5.64)式)对应的所谓 Fourier 绕射投影定理开始讨论。

考虑如下声波方程的反问题

$$\left.\begin{aligned}&\left\{\nabla^2-\frac{1}{C^2(\boldsymbol{r})}\frac{\partial^2}{\partial t^2}\right\}u(\boldsymbol{r},\boldsymbol{r}_s,t)=-\delta(\boldsymbol{r}-\boldsymbol{r}_s),\ \boldsymbol{r}\in\Omega\subset\mathscr{R}^3\\&u=\partial u/\partial t=0,\qquad t<0\\&u(\boldsymbol{r},\boldsymbol{r}_s,t)=f(\boldsymbol{r}_g,\boldsymbol{r}_s,t),\qquad \boldsymbol{r}_g\in\partial\Omega\end{aligned}\right\} \tag{5.146}$$

式中 $\boldsymbol{r}_s$ 为声源位置，$\boldsymbol{r}_g$ 为接收器坐标，反问题为给定 $f(\boldsymbol{r}_g,\boldsymbol{r}_s,t)$求声速 $C(\boldsymbol{r})$。如果我们假定对给定的源-接收器系统声波通过某种比较简单的标准介质 $C_0(\boldsymbol{r})$时声场为已知，然后把同样的系统放到某种非标准介质 $C(\boldsymbol{r})$中去，则声场将发生变化。在两种不同介质$C_0(\boldsymbol{r})$和 $C(\boldsymbol{r})$

中声场的变化叫声散射；反之，通过观测声场的变化求介质 $C_0(\boldsymbol{r})$和 $C(\boldsymbol{r})$之间的差异称为声场的反散射。对于比声波复杂的地震波，可以进一步用纵波方程(变密度标量波动方程)代替(5.146)中的声波方程，此时反散射问题还包括由波场的变化推测标准介质密度 $\rho_0(\boldsymbol{r})$和非标准介质 $\rho(\boldsymbol{r})$之间的差异。

令(5.146)式中系数项

$$\frac{1}{C^2(\boldsymbol{r})}=\frac{1}{C_0^2}(1-\alpha(\boldsymbol{r})) \tag{5.147}$$

其中标准介质的波速 C_0可以是常数，也可以是随空间变化缓慢的连续函数，如在 $\mathscr{C}_0^\infty$空间的已知函数；$\alpha(\boldsymbol{r})$表示了标准介质和非标准介质之间波速的相对变化，

$$\alpha(\boldsymbol{r})=1-\frac{C_0^2}{C^2(\boldsymbol{r})} \tag{5.148}$$

将(5.147)代入(5.146)并对 t 作 Fourier 变换便可把它改写为

$$\left\{\nabla^2+\frac{\omega^2}{C_0^2}(1-\alpha(\boldsymbol{r}))\right\}u(\boldsymbol{r},\boldsymbol{r}_s,\omega)=-\delta(\boldsymbol{r}-\boldsymbol{r}_s) \tag{5.149}$$

求 $C(\boldsymbol{r})$的反问题变成确定 $\alpha(\boldsymbol{r})$的问题。有趣的是，方程(5.149)与描述量子散射问题 Schrödinger 方程非常相似。由于在70年代后期量子反散射问题在理论物理学中已经有了相当系统的研究，而且有许多现成的算法，因此一些地球物理学家便致力于把量子反散射方法应用于地震波速成像的研究，并发展成本章后几节将要讨论的各种地震 CT 技术。

把(5.149)式两边对 t 进行 Fourier 变换，便有频率域的对应方程

$$(\nabla^2+k^2)u(\boldsymbol{r},\boldsymbol{r}_s,\omega)=-\delta(\boldsymbol{r}-\boldsymbol{r}_s)-k^2\alpha(\boldsymbol{r})u(\boldsymbol{r},\boldsymbol{r}_s,\omega) \tag{5.150}$$

其中 $k=\omega/C_0$。式(5.146～5.150)都是表示点源激发的情况，点源位置在 $\boldsymbol{r}_s$。如果是平面波入射的情况，则应把这些公式中 δ 函数项略去，并把总波场 u 分解为入射平面波场与衍射波场两部分。

(一)Fourier 衍射投影定理[103-104]

由(5.58)式表示的 Fourier 投影定理表明，在直射线的情况下，投影函数 $P(\xi,\theta)$关于 ξ 的 Fourier 变换等于图像函数 $f(\xi,\eta)$的二重 Fourier 变换在以 θ 角穿过原点的直线上取值。相应地在介质不均匀引起衍射的情况下(如图5.1的情况)，已证明投影函数 $P(\xi,\theta)=\mathscr{R}f(\xi,\theta)$与介质 $f(\xi,\eta)$的关系可用以下 Fourier 衍射投影定理表示。

定理：当介质 $f(x,y)$被平面波照射而产生弱散射时，在垂直入射波方向测量的波场的一维 Fourier 变换正比于介质 $f(x,y)$的二维 Fourier 变换在一条半圆弧上的取值(图5.8)。

为导出 Fourier 衍射投影定理，使用如图5.8(a)所示的坐标系。从声散射的观点看，取标准介质的波速为 $C_0=\text{const}$，则入射的平面波场可写为

$$u_i(\boldsymbol{r})=u_0e^{-ik_0s_0\cdot r} \tag{5.151}$$

其中 $\boldsymbol{r}=(x,y)$，u_0为入射波的振幅(初相位不为零时为复数)，$k_0=\omega/C_0$，$\boldsymbol{s}_0$为入射波方向的单位向量，即 $\boldsymbol{s}_0=(\cos\theta,\sin\theta)$。波通过非标准介质 $f(\boldsymbol{r})$时发生散射，其中 $f(\boldsymbol{r})=N^2(\boldsymbol{r})-1=\frac{C_0^2}{C^2(\boldsymbol{r})}-1$，$N$ 为折射指数。此时介质内任一点的总波场等于入射波与散射波之和

$$u(\boldsymbol{r})=u_i(\boldsymbol{r})+u_s(\boldsymbol{r}) \tag{5.152}$$

把(5.150)式用于平面波入射的情况($|\boldsymbol{r}_s|\to\infty$)，知波场满足以下非齐次 Helmholtz 方程

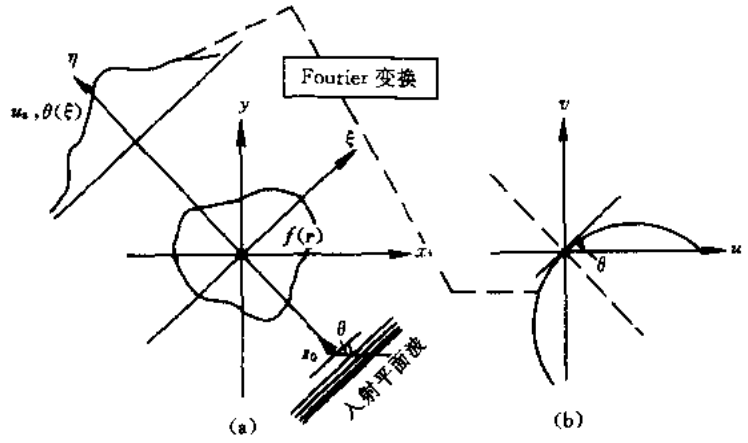


图 5.8 Fourier 衍射投影定理的图示

(a)空间域;(b)波数域

$$(\nabla^2+k_0^2)u(\boldsymbol{r})=k_0^2\alpha(\boldsymbol{r})u(\boldsymbol{r}) \tag{5.153}$$

在弱散射的情况下,有$|u_s|\ll|u_i|$,将(5.153)式右端的$u(\boldsymbol{r})$用$u_i(\boldsymbol{r})$表示并利用(5.151),则得到所谓 Born 近似的方程

$$(\nabla^2+k_0^2)u(\boldsymbol{r})\approx k_0^2\alpha(\boldsymbol{r})u_0\mathrm{e}^{-\mathrm{i}k_0\boldsymbol{s}_0\cdot\boldsymbol{r}} \tag{5.154}$$

如果考虑到散射场对总场还有附加的相位畸变$\phi_s(\boldsymbol{r})$,用

$$u(\boldsymbol{r})=u_0\mathrm{e}^{-\mathrm{i}k_0\boldsymbol{s}_0\cdot\boldsymbol{r}+\phi_s(\boldsymbol{r})}$$

代替(5.153)式右边的$u(\boldsymbol{r})$,则得到所谓的 Rytov 近似方程

$$(\nabla^2+k_0^2)u(\boldsymbol{r})\approx k_0^2\alpha(\boldsymbol{r})u_0\exp\{-\mathrm{i}k_0\boldsymbol{s}_0\cdot\boldsymbol{r}+\phi_s(\boldsymbol{r})\} \tag{5.155}$$

由于我们后面讨论的算法是基于 Rytov 近似的,这里顺便对 Rytov 近似作一点解释。把(5.153)式中波场u作对数变换[105]

$$\psi=\ln u$$

再代回(5.153),便得到

$$\nabla^2\psi+(\nabla\psi)^2+k_0^2=k_0^2\alpha(\boldsymbol{r}) \tag{5.156}$$

这便是著名的 Riccat 微分方程。用摄动法把$\alpha(\boldsymbol{r})$当作微扰并设ε为分数形式的参量,用$\varepsilon\alpha(\boldsymbol{r})$代替(5.156)中的$\alpha(\boldsymbol{r})$,则(5.156)的解可表示成$\varepsilon$的级数

$$\psi=\psi_0+\psi_1\varepsilon+\psi_2\varepsilon^2+\cdots\cdots$$

回代入(5.156)后可以导出一级近似的方程式

$$\nabla^2\psi_1+\mathrm{i}2k_0\boldsymbol{s}_0\cdot\nabla\psi_1=k_0^2\alpha(\boldsymbol{r}) \tag{5.157}$$

注意到$u=\exp(\psi)$便得出如(5.155)的 Rytov 近似并可确定$\phi_s(\boldsymbol{r})$。

在超声 CT 中用 Born 或 Rytov 近似对 Fourier 衍射投影定理的证明见文献[103]。下面我们结合地震层析成像来证明这个定理。由以上讨论已知,基本方程式(5.153)的解可用 Green 函数表示,这时 Green 函数为二维均匀介质中的平面波场,可用零阶第一类 Hankel 函数表示,即

$$G(\boldsymbol{r},\boldsymbol{r}')=-\frac{\mathrm{i}}{4}\mathrm{H}_0^{(1)}(k_0|\boldsymbol{r}-\boldsymbol{r}'|)$$

于是(5.153)的解表为

$$u(\boldsymbol{r},\omega,\boldsymbol{s}_0)=U_0(\omega)\mathrm{e}^{\mathrm{i}k_0\boldsymbol{s}_0\cdot\boldsymbol{r}}-\mathrm{i}\frac{k_0^2}{4}\int\alpha(\boldsymbol{r}',\omega)u(\boldsymbol{r}',\omega,\boldsymbol{s}_0)\mathrm{H}_0^{(1)}(k_0|\boldsymbol{r}-\boldsymbol{r}'|)\mathrm{d}\boldsymbol{r}' \tag{5.158}$$

其中，入射波振幅 U_0 是用频率域表示的。相应在时域的入射波形为

$$U_0(\boldsymbol{r},t,\boldsymbol{s}_0)=\frac{1}{2\pi}\int_{-\infty}^{\infty}U_0(\omega)\mathrm{e}^{\mathrm{i}k\boldsymbol{s}_0\cdot\boldsymbol{r}}\mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}\omega \tag{5.159}$$

散射问题为给定 $\alpha(\boldsymbol{r}',\omega)$ 求解第二类 Fredholm 积分方程(5.158)，原则上不存在数学上的困难，因为第二类 Fredholm 积分方程的求解是适定的。

对于层析成像来说我们关心的是逆散射问题，即给出 $u(\boldsymbol{r},\omega,\boldsymbol{s}_0)$ 的观测值求 $\alpha(\boldsymbol{r}',\omega)$。由于在(5.158)式右边积分号内 $u(\boldsymbol{r}',\omega,\boldsymbol{s}_0)$ 是未知的，介质内的波速变化 $\alpha(\boldsymbol{r}'\omega)$ 的精确解不能直接求得。如上所述，可以用 Born 近似或 Rytov 近似求反散射问题的近似解。考虑研究区域内介质波速的变化比较缓慢的弱散射情况，用 Born 近似法时可以将(5.151)式的入射波场近似代替(5.158)式右边积分号内的 $u(\boldsymbol{r}',\omega,\boldsymbol{s}_0)$，从而求得

$$u(\boldsymbol{r},\omega,\boldsymbol{s}_0)=U_0(\omega)\mathrm{e}^{\mathrm{i}k_0\boldsymbol{s}_0\cdot\boldsymbol{r}}-\mathrm{i}\frac{k_0^2}{4}U_0(\omega)\int\alpha(\boldsymbol{r}',\omega)\mathrm{e}^{\mathrm{i}k_0\boldsymbol{s}_0\cdot\boldsymbol{r}'}\mathrm{H}_0^{(1)}(k_0|\boldsymbol{r}-\boldsymbol{r}'|)\mathrm{d}\boldsymbol{r}' \tag{5.160}$$

考虑到跨孔地震和全偏移距 VSP 的数据采集系统的场合，在此式中我们假定观测是在 $x=l_0=\mathrm{const}$。的井孔中进行的，此时(5.160)中对数据点有 $\boldsymbol{r}=(x=l_0,y)$。如果用方程(5.155)表示的 Rytov 近似，则常用以下复相位换算观测数据

$$\Psi(\boldsymbol{r},\omega,\boldsymbol{s}_0)=\ln\frac{u(\boldsymbol{r},\omega,\boldsymbol{s}_0)}{U_0(\omega)} \tag{5.161}$$

即

$$u(\boldsymbol{r},\omega,\boldsymbol{s}_0)=U_0(\omega)\exp[\Psi(\boldsymbol{r},\omega,\boldsymbol{s}_0)]$$

这样，由 Rytov 近似式(5.155)或(5.157)可得

$$\Psi(\boldsymbol{r},\omega,\boldsymbol{s}_0)=\mathrm{i}k_0\boldsymbol{s}_0\cdot\boldsymbol{r}-\mathrm{i}\frac{k_0^2}{4}\mathrm{e}^{-\mathrm{i}k_0\boldsymbol{s}_0\cdot\boldsymbol{r}}\cdot\int\alpha(\boldsymbol{r}',\omega)\mathrm{e}^{\mathrm{i}k_0\boldsymbol{s}_0\cdot\boldsymbol{r}'}\mathrm{H}_0^{(1)}(k_0|\boldsymbol{r}-\boldsymbol{r}'|)\mathrm{d}\boldsymbol{r}' \tag{5.162}$$

其中 $\boldsymbol{r}'=(x',y')$ 为图像函数 $f(\boldsymbol{r}')$ 中的象元点。在进行反散射图像重建时，把数据统一表为

$$D(y,\omega,\boldsymbol{s}_0)=\mathrm{i}\frac{\alpha}{k_0^2}\begin{cases}\dfrac{u(\boldsymbol{r}_0,\omega,\boldsymbol{s}_0)}{U_0(\omega)}-\mathrm{e}^{\mathrm{i}k_0\boldsymbol{s}_0\cdot\boldsymbol{r}_0} & (\text{Born 近似})\\ [\Psi(\boldsymbol{r}_0,\omega,\boldsymbol{s}_0)-\mathrm{i}k_0\boldsymbol{s}_0\cdot\boldsymbol{r}_0]\mathrm{e}^{\mathrm{i}k_0\boldsymbol{s}_0\cdot\boldsymbol{r}_0} & (\text{Rytov 近似})\end{cases} \tag{5.163}$$

式中 $\boldsymbol{r}_0=(l_0,y)$ 为 $x=l_0$ 时直井中的接收点。这样，反散射公式(5.160)及(5.162)可统一表示为

$$D(y,\omega,\boldsymbol{s}_0)=\frac{1}{2}\int\alpha(\boldsymbol{r}',\omega)\mathrm{e}^{\mathrm{i}k_0\boldsymbol{s}_0\cdot\boldsymbol{r}'}\mathrm{H}_0^{(1)}(k_0|\boldsymbol{r}_0-\boldsymbol{r}'|)\mathrm{d}\boldsymbol{r}' \tag{5.164}$$

现在来考虑由(5.164)表示的数据 $D(y,\omega,\boldsymbol{s}_0)$ 对 y 的一维 Fourier 变换。首先用 Hankel 函数的展开式[104]

$$\mathrm{H}_0^{(1)}(k_0|\boldsymbol{r}-\boldsymbol{r}'|)=\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\mathrm{d}k}{\gamma}\mathrm{e}^{\mathrm{i}[k(y-y')+\gamma(l_0-x')]} \tag{5.165}$$

其中

$$\gamma=\begin{cases}\sqrt{k_0^2-k^2}, & \text{当 } |k|\leqslant k_0\\ \mathrm{i}\sqrt{k^2-k_0^2}, & \text{当 } |k|>k_0\end{cases}$$

把(5.165)代入(5.164)并对 y 作 Fourier 变换

$$D(k,\omega,s_0)=\int_{-\infty}^{\infty}D(y,\omega,s_0)\mathrm{e}^{-iky}\mathrm{d}y$$

得

$$D(k,\omega,s_0)=\gamma^{-1}\exp(\mathrm{i}\gamma l_0)\alpha(\gamma-k_0s_{ox},k-k_0s_{oy},\omega) \tag{5.166}$$

式中 $-k_0\leqslant k\leqslant k_0$，$s_0=(s_{ox},s_{oy})$，而 $\alpha(\boldsymbol{r},\omega)$ 的二维 Fourier 变换为

$$\alpha(\boldsymbol{k},\omega)=\iint\alpha(\boldsymbol{r},\omega)\mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{r}}\mathrm{d}\boldsymbol{r} \tag{5.167}$$

比较上两式可知，数据的一维 Fourier 变换 $D(k,\omega,s_0)$ 正比于图像函数的二维 Fourier 变换 $\alpha(\boldsymbol{k},\omega)$ 在某半圆弧上的取值，比例因子为 $\gamma^{-1}\exp(\mathrm{i}\gamma l_0)$。为看清楚这一点，设 x 和 y 为坐标方向的单位向量而 $\boldsymbol{k}=(u,v)$，则(5.166)中 k 的轨迹方程可以写为

$$\boldsymbol{k}=(\gamma-k_0s_{ox})\boldsymbol{x}+(k-k_0s_{oy})\boldsymbol{y}$$

设单位向量

$$\boldsymbol{s}=k_0^{-1}[\gamma\boldsymbol{x}+k\boldsymbol{y}]$$

则有

$$\boldsymbol{k}=k_0(\boldsymbol{s}-\boldsymbol{s}_0) \tag{5.168}$$

同时(5.166)变为

$$D(k,\omega,s_0)=\gamma^{-1}\exp(\mathrm{i}\gamma l_0)\alpha[k_0(\boldsymbol{s}-\boldsymbol{s}_0),\omega] \tag{5.169}$$

方程式(5.168)规定了一个圆心在 $\boldsymbol{k}=-k_0\boldsymbol{s}_0$ 而半径为 k_0 的圆周。注意到在图5.8中 $\boldsymbol{k}=(u,v)$ 由于 $\gamma=\sqrt{k_0^2-k^2}$ 要求 $u\geqslant-k_0s_{ox}$，因此(5.168)式实际上定义了一个半圆弧。这样，(5.166)式便在弱散射的情况下证明了如图5.8所示的 Fourier 衍射投影定理。

(二)Fourier 衍射投影定理的应用[184,185,186]

和第三节讨论的 Fourier 投影定理一样，衍射投影定理也从原则上给出了射线弯曲时连续图像函数 $\alpha(x,y)$ 的如下重建方法：

(1)对不同入射角 s_0 的数据 $D(y,t,s_0)$ 作关于 t 及 y 的 Fourier 变换，获得在波数域沿半圆弧族上分布的波速扰动 $\alpha(k,k_0,s_0)$。

(2)将 $\alpha(k,k_0,s_0)$ 内插到方格网上求出 $\alpha(\boldsymbol{k})=\alpha(u,v)$。

(3)对 $\alpha(u,v)$ 作二重 Fourier 逆变换获得图像函数 $\alpha(x,y)$。

虽然这种在弯曲射线情况下的 Fourier 变换算法原理上是可行的，但是应用在地震勘探时却发生困难。与医学上的超声 CT 不同，地震观测只能在地面和井孔中进行，这时由 s_0 表示的入射角 θ 只在较小的范围内变化。换句话说，我们只有不完全投影的数据，它的 Fourier 变换 $\alpha(k,k_0,s_0)$ 只分布在 (u,v) 平面的极其有限的区域，这时把它内插为全平面的 $\alpha(u,v)$ 是极其困难的。为了详细了解这个困难，让我们来看看各种地震 CT 观测系统的资料在波数域分布的情况。

下面来考虑最理想的情况下跨孔的和 VSP 的数据采集情况。首先来看看对不同入射角 θ 的轨迹半圆弧的位置。当 $\theta=0°$ 时 $s_0=(1,0)$，圆心位于 $u=-k_0$，$v=0$，轨迹半圆弧为过原

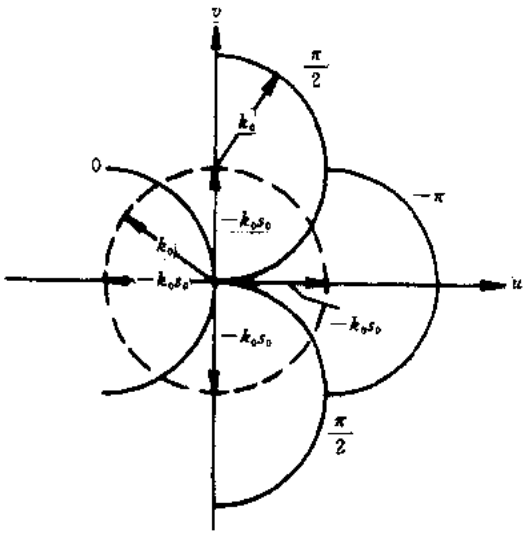


图 5.9 决定投影剖面变换的半圆弧

所有半圆的圆心都位于半径为 k_0 的虚线上，图中只展示了 $\theta=0,\pi/2,-\pi/2$ 及 $-\pi$ 四种情况

点与 v 轴相切的半圆弧，对称于 u 轴(见图5.9)。类似地当 $\theta=-\pi,-\pi/2$ 及 $\pi/2$ 弧度时轨迹半圆弧都示于图5.9，它们位于波数域 $u>0$ 的半平面内。其它所有的半圆弧的圆心都在以原点为圆心并以 k_0 为半径的圆周上。频率越高，$k_0=\omega/C_0$ 便越大，在波数域的轨迹半圆弧的半径就越大，说明频率扫描相当于在波数域增加了数据点的密度(图5.10)。

在图5.1所示的跨孔 CT 的情况下，入射角 θ 的最大的变化范围为 $-\pi/2$ 到 $\pi/2$(根据互换原理，即发射井和接收井互换观测结果相同)，因此在用 Fourier 衍射投射定理变换到波数域后资料点应分布在如图5.10所示的两个圆之内，圆内部的资料点是频率扫描的结果。实际上由于井深有限及频带宽度有限，这时资料仅为图5.10所示的圆内的一部分。

对于全偏移距 VSP 数据收集系统，上行波的 θ 角虽然在 $-\pi/2\sim0$ 之间变化，但在井孔中接收到的只是其中很少的一部分，即图5.11(a)中左下角的花瓣形区域，而主要的资料点集中在 $\pi/2\leqslant\theta\leqslant\pi$ 的各半圆内(右上角区域)，相应于井中间的接收点上下行波的入射方向

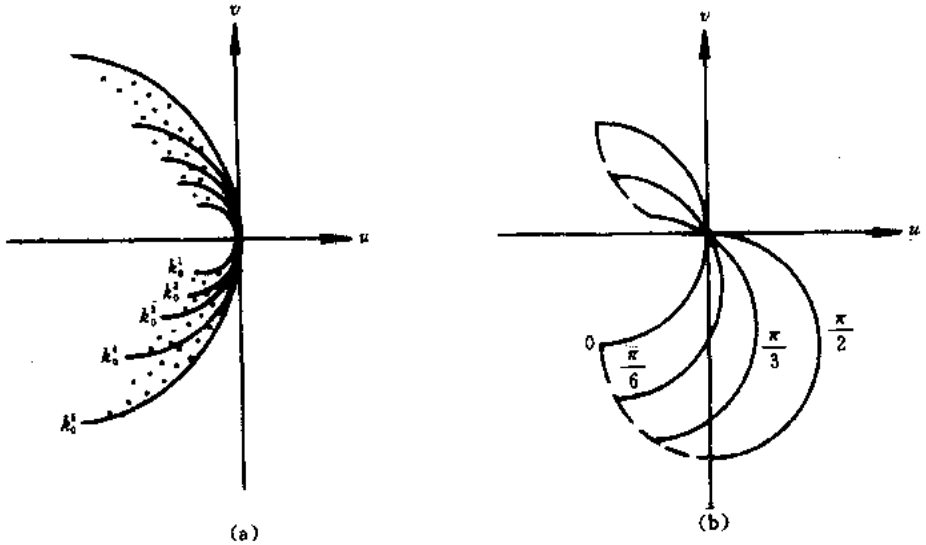


图 5.10 用频率扫描加大波数域数据点密度的示意图

(a)阴影区间表示随 ω 增大加密的范围；(b)随 θ 的变化

从$-\pi/2$变化到$\pi/2$。如果同时接收波场的垂直分量和水平分量，即在复数波场的情况下，我们可得到如图5.12(a)中R_0^*的共轭复数区域R_0，它与R_0的并$R_0 \cup R_0^*$为VSP资料点的最大分布区域。当然，在实际上炮点离井孔不可能无穷远，井深也有限，资料点只分布在图5.12所示的$R_0 \cup R_0^*$内的局部区域。

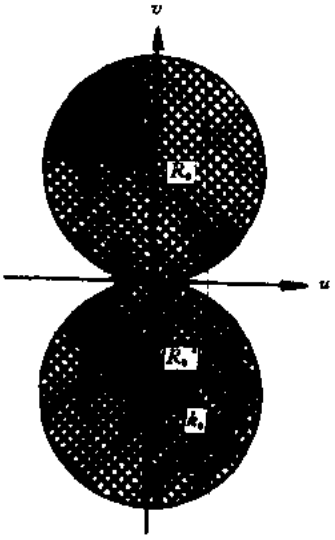


图 5.11 跨孔CT装置在波数域的资料点分布

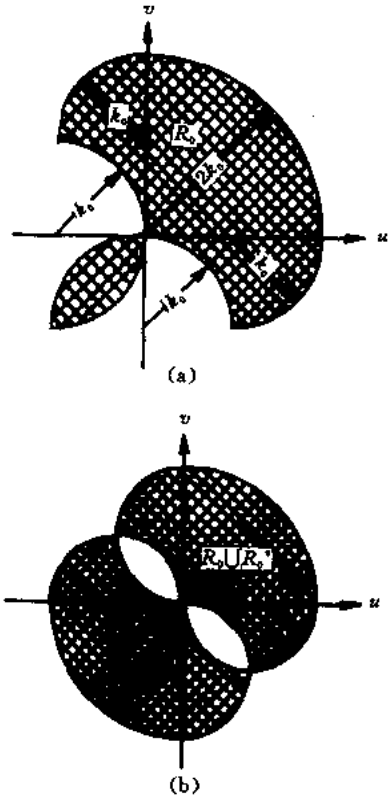


图 5.12 全炮检距VSP在波数域资料点的分布
(a)单分量实数波场接收；(b)双分量复数波场接收

上面关于地震衍射CT资料点在波数的分布说明，无论是跨孔地震CT还是VSP在波数域都存在资料空白的区域(盲区)，这就给图像重建带来了困难。在资料空白区较大的情况下，常用的波数域内插的方法并非总是可以成功的，处理盲区的问题常常成为不完全投影层析成像成败的关键，对成像的分辨率影响极大。和第三节中提到过的一样，利用全部的$\alpha(k, k_0, s_0)$数据进行插值求出的$\alpha(u,v)$精度较高，但是计算工作量很大。此外，也可以采用加数字滤波器的办法摈除没有数据点的局部区域的影响，但这样做导致解的非唯一性。在下一段讨论的衍射CT图像重建方法中就采用了加滤波器的方法，其优点是计算速度快并便于在机器上实现。

上面我们只讨论了平面波入射时的Fourier衍射投影定理和成像问题。而地震CT主要是点震源入射的情况，似乎不能肯定上述讨论在地震应用上的合法性。实际上，吴如山等人在考虑点震源的情况下得到了和上面讨论相同的结果[106]。

(三)衍射CT重建算法[104,105]

为了用不完全投影数据重建图像，必须考虑如下滤波器

$$F_0(\boldsymbol{k})=\begin{cases}1 & \text{如果 } \boldsymbol{k}\in R_0\\ 0 & \text{其它}\end{cases} \tag{5.170}$$

其中 R_0 为波数域数据点覆盖的空间。由于 $F_0(\boldsymbol{k})\alpha(\boldsymbol{k},\omega)$ 表示了波数域可以由数据取得的图像信息，而在 R_0 之外为零，这一乘积的 Fourier 逆变换

$$\hat{\alpha}(\boldsymbol{r},\omega)=\frac{1}{(2\pi)^2}\int F_0(\boldsymbol{k})\alpha(\boldsymbol{k},\omega)\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}}\mathrm{d}\boldsymbol{k} \tag{5.171}$$

便是图像的一种估计，因为 $\hat{\alpha}(\boldsymbol{r},\omega)$ 的二维 Fourier 变换在 R_0 上等于 $\alpha(\boldsymbol{k},\omega)$，而在 R_0 外为零。然而，由于在区域 R_0 外没有用数据来定义，从原则上对这些区域变换值是自由的，可以有无穷多个图像的估计满足(5.171)式。我们可以把 $\hat{\alpha}(\boldsymbol{r},\omega)$ 加上如下形式的任何函数

$$\Delta(\boldsymbol{r},\omega)=\frac{1}{(2\pi)^2}\int[1-F_0(\boldsymbol{k})]\Delta(\boldsymbol{k},\omega)\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}}\mathrm{d}\boldsymbol{k}$$

则

$$\hat{\alpha}(\boldsymbol{r},\omega)=\alpha(\boldsymbol{r},\omega)+\Delta(\boldsymbol{r},\omega)$$

也是可以产生与数据 $D(y,\omega,s_0)$ 拟合的图像估计。由于不完全投影成像问题解的非唯一性，我们可借助于最优化准则求广义解。实际上，(5.171)表示的解估计就是使

$$E=\int|\alpha^2(\boldsymbol{r},\omega)|^2\mathrm{d}\boldsymbol{r}$$

取极小的最小能量解估计。

现在考虑如图5.10和5.11的跨孔和 VSP 地震观测的情况，并假定数据 D 为复函数，它在波数域的覆盖区域设为 R_0(即图中的 $R_0\cup R_0^*$)，这时(5.171)式可写为

$$\hat{\alpha}(\boldsymbol{r},\omega)=(2\pi)^{-2}\int_{R_0}\alpha(\boldsymbol{k},\omega)\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}}\mathrm{d}\boldsymbol{k} \tag{5.172}$$

把积分变量 $\boldsymbol{k}=(u,v)$ 作如下变换

$$\left.\begin{aligned}u&=\gamma-k_0s_{0x}=\gamma-k_0\cos\theta,\gamma=\sqrt{k_0^2-k^2}\\ v&=k-k_0s_{0y}=k-k_0\sin\theta\end{aligned}\right\} \tag{5.173}$$

由于此变换的 Jacobian 行列式为

$$k_0|k\cos\theta-\gamma\sin\theta|/\gamma$$

因此有

$$\mathrm{d}\boldsymbol{k}=\frac{k_0}{\gamma}|k\cos\theta-\gamma\sin\theta|\mathrm{d}\theta\mathrm{d}k \tag{5.174}$$

如在 θ 固定时对 k 积分，上限应由半圆弧

$$\boldsymbol{k}=k(\boldsymbol{s}-\boldsymbol{s}_0)$$

的切点和积分区域 R_0 的边界限定，即

$$\begin{aligned}&-k_0\leqslant k\leqslant k_0 && \text{对}\quad -\pi/2\leqslant\theta\leqslant\pi/2\\ &-k_0\leqslant k\leqslant -k_0\sin\theta && \text{对}\quad -\pi\leqslant\theta\leqslant-\pi/2\end{aligned} \tag{5.175}$$

因此可定义滤波函数为

$$F_1(k,\theta)\begin{cases}0 & \text{对 } k\geqslant -k_0\sin\theta \text{ 及 } -\pi\leqslant\theta\leqslant-\dfrac{\pi}{2}\\ 1 & \text{其它}\end{cases} \tag{5.176}$$

这样(5.172)式可写为

$$\hat{\alpha}(\boldsymbol{r},\omega)=\frac{1}{(2\pi)^2}\int_{\varepsilon_0}\mathrm{d}\theta\int_{-k_0}^{k_0}\frac{k_0}{\gamma}|k\cos\theta-\gamma\sin\theta|F_1(k,\theta)\alpha[k_0(s-s_0)]$$
$$\exp[\mathrm{i}k_0(s-s_0)\cdot\boldsymbol{r}]\mathrm{d}k \tag{5.177}$$

其中 ε_0 对全偏移距 VSP 为 $-\pi\leqslant\theta\leqslant 0$，跨孔地震时为 $-\pi/2\leqslant\theta\leqslant\pi/2$。此外，对跨孔 CT 取 $F_1=1$。回忆在(5.168)中有式 $\boldsymbol{s}=(\gamma\boldsymbol{x}+k\boldsymbol{y})/k_0$，并将(5.169)式代入上式得到图像重建公式

$$\hat{\alpha}(\boldsymbol{r},\omega)=\frac{k_0}{(2\pi)^2}\int_{\varepsilon_0}\mathrm{d}\theta\mathrm{e}^{-\mathrm{i}k_0\boldsymbol{s}_0\cdot\boldsymbol{r}}\int_{-k_0}^{k_0}|k\cos\theta-\gamma\sin\theta|$$
$$\times F_1(k,\theta)D(k,\omega,\boldsymbol{s}_0)\mathrm{e}^{\mathrm{i}(\gamma(x-l_0)+ky)}\mathrm{d}k \tag{5.178}$$

式中 l_0 为钻孔位置。当单分量测量数据 $D(y,\omega,s_0)$ 为实数时，用(5.178)式积分后取实部再乘以2也可计算图像的估计。

上述衍射 CT 的图像重建算法在超声 CT 上称为滤波反传播算法(filtered backpropagation algorithm)，反传播的意思指的是在 Rytov 近似的意义下用波位相的公式把滤波后的数据回推到图像空间去。由于算法的基础是在 Ryfov 近似下求 Riccati 方程(5.155)源项的解估计，因此适用条件是 $\alpha(\boldsymbol{r})$ 很小而产生弱散射的情况。在本节中我们只介绍了文献[104]提出的算法，它虽然涉及了不完全投影数据成像时解的非唯一性，但没有考虑带误差的数据和成像问题的不适定性及算法的稳定性。总之，衍射 CT 中图像重建问题也是目前正处在探讨和不断完善过程中的 CT 技术。不完全投影的成像，考虑散射时的成像及算法稳定性的方法还有待进一步深入研究。

七、逆散射地震层析成像

以上几节我们从量子反散射讨论到它在地震勘探上应用的各种途径，基本上是数学家们在理论上的探讨，我想读者一定会感到不够满意，因为这些理论还没有水乳交融地渗入地震学之中，成为地震学的一部分，因此也没有取得比地震偏移更好的效果。在1986—1988年期间，我的研究工作致力于找到反散射理论、广义 Radon 变换和地震学之间的接合点，找出高分辨率地球成像的新路子，终于取得了初步的成功。与其它作者不同的是，我从地震学出发，把反射地震波和散射地震波区分开。用反射波走时确定参考波速场，用散射波确定波速的扰动，即散射体。由于散射发生在散射体的尺寸在主波长的1/4以下，这样地震层析成像的分辨率便可以高于主波长的1/4，即在主频为20～30Hz 时小于20～30m。

(一)地震反散射理论和地球模型的结合

在原始的地震记录中包含有各种地震波，如初次和多次反射波，面波，折射波，转换波和散射波等等。在地震数据处理中，除了初次反射波外，其它所有的波都被当作干扰而尽量加以消除。我们认为，为了对尺寸小于1/4主波长的物体(如储油层，粘泥透镜体等)成像，散射波应作为信号，在地震记录中保留和提取出来。

从地震学的观点看，所谓散射体是沿波传播方向上尺寸小于1/4波长的异常波阻抗物体。由于散射体的两个侧面产生的二次波总是同相的，因此它们在波场中可视为一组次生点震源，产生的二次及多次波称为散射波。在另一方面，反射体在波传播方向上的尺寸大于1/4波长，因此其两侧面产生的二次波不同相，而成为可分离的反射波。对反射地震和 VSP 装置而言，由于人工震源总是在地面上，地震波的传播方向趋向于垂直，因此岩层的厚度可用于

作为分辨散射体和反射体的一种粗糙的度量。但是还必须注意到，地球中还存在着另一类散射体，它们是由断层、岩层尖灭等构造中的奇异点组成的，也产生散射波。

我们用地球的波速场 $C(\boldsymbol{x})$ 来描述地球模型。虽然称为 $C(\boldsymbol{x})$ 为场，但它只表示地球物质波速的空间分布，而不是物理存在的具有能量的场。由波的散射理论，$C(\boldsymbol{x})$ 可表示为参考波速场 $C_0(\boldsymbol{x})$ 和波速扰动 $\delta C(\boldsymbol{x})$ 之和，这里 $\boldsymbol{x}$ 表示三维空间的一个点。我们定义波速扰动即为散射体，而参考波速场为地震波传播的介质。通常，波速扰动用因子 $\alpha(\boldsymbol{x})$ 表示，它满足

$$\frac{1}{C^2(\boldsymbol{x})}=\frac{1\pm\alpha(\boldsymbol{x})}{C_0^2(\boldsymbol{x})} \tag{5.179}$$

式中的符号可以任选，我们选加号。从此式可知，波速扰动要比参考波速小，而且分布在局部范围内，即满足条件

$$-1\leqslant|\alpha(\boldsymbol{x})|\leqslant 1\text{和}\ C_0(\boldsymbol{x})=C(\boldsymbol{x})\text{当}\ |\boldsymbol{x}|\to\infty \tag{5.180}$$

参考波速场等于常数的特殊情况和连续函数表示的情况，前人已经进行了不少层析反演的研究，取得了一些理论成果[63,107-111]。现在要讨论它用广义函数表示的情况，以便使假设的地球模型更接近真实地球。在大多数情况下，所研究的地球具有层状结构。考虑二维层状地球模型，则参考波速在某一层内等于层速度，即波速垂向平均值。令层与层之间的界面为光滑函数，记为 $z=H_i(x)$，则 H_i 表示第 i 层和 $(i+1)$ 层之间的界面，而参考波速场可用以下广义函数表示：

$$C_0(x,z)=\begin{cases}C_i(x) & \text{当}\quad (x,z)\notin H_i\\ \infty & \text{当}\quad (x,z)\in H_i\end{cases} \tag{5.181}$$

其中 $C_i(x)$ 为第 i 层的层速度，$i=1,2,\cdots,N$，N 为界面总数。这样，由参考波速场产生的所谓参考波场就包括除散射波以外的其它各种波，记为 $u_0(\boldsymbol{x},t)$。由波速扰动 $\alpha(\boldsymbol{x})$ 产生的直接散射波记为 $u_{s1}(\boldsymbol{x},t)$。直接散射波在界面 H_i 上还会产生反射波，而反射波在到达散射体后也会产生新的散射，这些高次散射波记为 $u_{s2}(\boldsymbol{x},t)$。于是，某一点震源产生的总波场可以表为

$$u(\boldsymbol{x},t)=u_0(\boldsymbol{x},t)+u_{s1}(\boldsymbol{x},t)+u_{s2}(\boldsymbol{x},t) \tag{5.182}$$

假定脉冲点震源位于点 $\boldsymbol{s}$，则对应此地球模型声波方程写为

$$\left(\nabla^2-\frac{1}{C(\boldsymbol{x})^2}\frac{\partial^2}{\partial t^2}\right)u(\boldsymbol{x},\boldsymbol{s},t)=-\delta(\boldsymbol{x}-\boldsymbol{s})\delta(t) \tag{5.183}$$

将(5.179)和(5.182)代入(5.183)并在频率域表示出来，便有

$$\left(\nabla^2+\frac{\omega^2}{C_0^2(\boldsymbol{x})}\right)u_0(\boldsymbol{x},\boldsymbol{s},\omega)=-\delta(\boldsymbol{x}-\boldsymbol{s}) \tag{5.184}$$

和

$$\begin{aligned}&\left(\nabla^2+\frac{\omega^2}{C_0^2(\boldsymbol{x})}\right)(u_{s1}(\boldsymbol{x},\boldsymbol{s},\omega)+u_{s2}(\boldsymbol{x},\boldsymbol{s},\omega))\\&=-\alpha(\boldsymbol{x})\frac{\omega^2}{C_0^2(\boldsymbol{x})}(u_0(\boldsymbol{x},\boldsymbol{s},\omega)+u_{s1}(\boldsymbol{x},\boldsymbol{s},\omega)+u_{s2}(\boldsymbol{x},\boldsymbol{s},\omega))\end{aligned} \tag{5.185}$$

由(5.180)的条件进一步假定 $|\alpha(\boldsymbol{x})|\ll 1$，则可得

$$u_{s2}(\boldsymbol{x},\boldsymbol{s},\omega)\ll u_{s1}(\boldsymbol{x},\boldsymbol{s},\omega)\ll u_0(\boldsymbol{x},\boldsymbol{s},\omega) \tag{5.186}$$

此式代入(5.184)和(5.185)后我们便得到散射波的 Born 近似式

$$\left(\nabla^2+\frac{\omega^2}{C_0^2(\boldsymbol{x})}\right)u_{s1}(\boldsymbol{x},\boldsymbol{s},\omega)\approx-\alpha(\boldsymbol{x})\frac{\omega^2}{C_0^2(\boldsymbol{x})}u_0(\boldsymbol{x},\boldsymbol{s},\omega) \tag{5.187}$$

和

$$\left(\nabla^2+\frac{\omega^2}{C_0^2(\boldsymbol{x})}\right)u_{s2}(\boldsymbol{x},\boldsymbol{s},\omega)\approx-\alpha(\boldsymbol{x})\frac{\omega^2}{C_0^2(\boldsymbol{x})}u_{s1}(\boldsymbol{x},\boldsymbol{s},\omega) \tag{5.188}$$

由于微分方程(5.187)的解可用 Green 函数表为

$$u_{s1}(\boldsymbol{s},\boldsymbol{r},\omega)\approx\omega^2\int\frac{G(\boldsymbol{x},\boldsymbol{s},\omega)G(\boldsymbol{x},\boldsymbol{r},\omega)}{C_0^2(\boldsymbol{x})}\alpha(\boldsymbol{x})\mathrm{d}^3\boldsymbol{x} \tag{5.189}$$

其中向量 $\boldsymbol{r}$ 为接收器位置，而 Green 函数满足

$$\nabla^2G(\boldsymbol{x},\boldsymbol{y},\omega)+\frac{\omega^2}{C_0^2(\boldsymbol{x})}G(\boldsymbol{x},\boldsymbol{y},\omega)=-\delta(\boldsymbol{x}-\boldsymbol{y}) \tag{5.190}$$

其中向量 $\boldsymbol{y}$ 是 $\boldsymbol{s}$ 和 $\boldsymbol{r}$ 其中任一个的代号。为了取得用广义 Radon 变换表示的直接散射波的积分式，方程(5.189)还可以进一步用几何光学近似来简化，即假定

$$G(\boldsymbol{x},\boldsymbol{y},\omega)=A(\boldsymbol{x},\boldsymbol{y})\mathrm{e}^{\mathrm{i}\omega\tau(\boldsymbol{x},\boldsymbol{y})} \tag{5.191}$$

其中 $\tau(\boldsymbol{x},\boldsymbol{y})$ 为走时，满足程函方程

$$\nabla\tau\cdot\nabla\tau=C_0^{-2}(\boldsymbol{x}) \tag{5.192}$$

而振幅 $A(\boldsymbol{x},\boldsymbol{y})$ 满足迁移方程

$$2\nabla\tau\cdot\nabla A+A\nabla^2\tau=0 \tag{5.193}$$

实际上方程(5.192)是用射线追踪求解的。至于振幅函数，物理上与球面扩散，通过层面时的穿透系数和反射系数有关。在声波方程中不考虑转换波和介质粘弹性引起的能量损失。

采用以下简化记号

$$a(\boldsymbol{x},\boldsymbol{s},\boldsymbol{r})=A(\boldsymbol{x},\boldsymbol{s})A(\boldsymbol{x},\boldsymbol{r})/C_0^2(\boldsymbol{x}) \tag{5.194}$$

和

$$\phi(\boldsymbol{x},\boldsymbol{s},\boldsymbol{r})=\tau(\boldsymbol{x},\boldsymbol{s})+\tau(\boldsymbol{x},\boldsymbol{r})$$

并将(5.191)代入(5.189)便得到直接散射波的解

$$u_{s1}(\boldsymbol{s},\boldsymbol{r},\omega)\approx\omega^2\int a(\boldsymbol{x},\boldsymbol{s},\boldsymbol{r})\mathrm{e}^{\mathrm{i}\omega(\phi(\boldsymbol{x},\boldsymbol{s},\boldsymbol{r}))}\alpha(\boldsymbol{x})\mathrm{d}^3\boldsymbol{x} \tag{5.195}$$

返回到时间域就得到用广义 Radon 变换表示的公式

$$\begin{aligned}u_{s1}(\boldsymbol{s},\boldsymbol{r},t)&\approx\frac{\partial^2}{\partial t}\int a(\boldsymbol{x},\boldsymbol{s},\boldsymbol{r})\alpha(\boldsymbol{x})\delta[t-\phi(\boldsymbol{x},\boldsymbol{s},\boldsymbol{r})]\mathrm{d}^3\boldsymbol{x}\\&=-\int a(\boldsymbol{x},\boldsymbol{s},\boldsymbol{r})\alpha(\boldsymbol{x})\delta''[t-\phi(\boldsymbol{x},\boldsymbol{s},\boldsymbol{r})]\mathrm{d}^3\boldsymbol{x}\end{aligned} \tag{5.196}$$

由于这是由声波方程导出的公式，因此也被称为声的广义 Radon 变换式。因此，对于波速扰动的成像，也就是求声广义 Radon 变换的逆，即对散射波场沿等时面 S_{x0} 的加权积分：

$$\begin{aligned}\alpha(\boldsymbol{x}_0)&=\int\mathrm{d}W(\boldsymbol{x}_0,\boldsymbol{s},\boldsymbol{r})u_{s1}(\boldsymbol{s},\boldsymbol{r},t=\phi(\boldsymbol{x}_0,\boldsymbol{s},\boldsymbol{r}))\\&=\int_{S_{x0}}u_{s1}(\boldsymbol{s},\boldsymbol{r},t)\mathrm{d}W(\boldsymbol{x}_0,\boldsymbol{s},\boldsymbol{r})\end{aligned} \tag{5.197}$$

其中 $\boldsymbol{x}_0$ 表示成像的点，即象元的中心点。假设 $\boldsymbol{d}$ 为散射波场的数据空间中的任一点，则等时面可表示为以下的点集

$$S_{x0}=\{\boldsymbol{d}:t=\phi(\boldsymbol{x}_0,\boldsymbol{s},\boldsymbol{r})\} \tag{5.198}$$

在(5.197)式中的权函数要由广义 Radon 变换的逆导出，留在本节的下一段讨论。

我们先来看看上两节介绍的 Beylkin 的理论可以怎样应用到这里的散射地震层析成像

问题。应用(5.136)的广义 Radon 变换的定义，可以把(5.195)的散射波表示为

$$u_{s1}(\boldsymbol{s},\boldsymbol{r},\omega)\approx\omega^2\int_{-\infty}^{\infty}\int a(\boldsymbol{x})a(\boldsymbol{x},\boldsymbol{s},\boldsymbol{r})\delta[t-\phi(\boldsymbol{x},\boldsymbol{s},\boldsymbol{r})]\mathrm{e}^{\mathrm{i}\omega t}\mathrm{d}^3\boldsymbol{x}\mathrm{d}t$$

$$=\omega^2\int_{-\infty}^{\infty}[\mathscr{R}_a a](\boldsymbol{s},\boldsymbol{r},t)\mathrm{e}^{\mathrm{i}\omega t}\mathrm{d}t \tag{5.199}$$

其中广义 Radon 变换为

$$[\mathscr{R}_a a](\boldsymbol{s},\boldsymbol{r},t)=\int a(\boldsymbol{x})a(\boldsymbol{x},\boldsymbol{s},\boldsymbol{r})\delta[t-\phi(\boldsymbol{x},\boldsymbol{s},\boldsymbol{r})]\mathrm{d}^3\boldsymbol{x} \tag{5.200}$$

由(5.199)可知，u_{s1}/ω^2即为波速扰动的广义 Radon 变换的 Fourier 变换，因此有

$$v(\boldsymbol{s},\boldsymbol{r},t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\omega^{-2}u_{s1}(\boldsymbol{s},\boldsymbol{r},\omega)\mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}\omega \tag{5.201}$$

由(5.137)可知，广义 Radon 变换的对偶变换即为反投影算子，由它可以求出广义 Radon 变换的伪逆。因此，与(5.197)等价的一种反散射成像可用伪逆 $\mathscr{R}^*$ 表示为

$$a(\boldsymbol{x}_0)=\mathscr{R}_a^*[\mathscr{R}_a a](\boldsymbol{s},\boldsymbol{r},t)\equiv\mathscr{R}_a^* v(\boldsymbol{s},\boldsymbol{r},t) \tag{5.202}$$

其中 v 为由(5.201)计算的数据。对应反投影成像的公式[8][15]

$$a(\boldsymbol{x}_0)=\int_{S_0}\int_{S_r}b(\boldsymbol{x}_0,\boldsymbol{s},\boldsymbol{r})v(\boldsymbol{s},\boldsymbol{r},t)\big|_{t=\phi(\boldsymbol{x}_0,\boldsymbol{s},\boldsymbol{r})}\mathrm{d}^2\boldsymbol{r}\mathrm{d}^2\boldsymbol{s} \tag{5.203}$$

其中 S_0是地震震源分布的曲面，只考虑点震源时对它的积分消失；S_r 为接收器分布的面。权函数表为

$$b(\boldsymbol{x}_0,\boldsymbol{s},\boldsymbol{r})=\frac{h(\boldsymbol{x}_0,\boldsymbol{s},\boldsymbol{r})}{a(\boldsymbol{x}_0,\boldsymbol{s},\boldsymbol{r})}f(\boldsymbol{x}_0,\boldsymbol{s},\boldsymbol{r}) \tag{5.204}$$

其中 f 为滤波器，用于使

$$h(\boldsymbol{x}_0,\boldsymbol{s},\boldsymbol{r})f(\boldsymbol{x}_0,\boldsymbol{s},\boldsymbol{r})\geqslant 0$$

的条件成立。假定固定点震源的情况，象点表为 $\boldsymbol{x}_0=(x_1,x_2,x_3)$，而接收点对反射地震为 $\boldsymbol{r}=(r_1,r_2,0)$，对 VSP 为 $\boldsymbol{r}=(r_1,0,r_2)$，而

$$h(\boldsymbol{x}_0,\boldsymbol{r})=\begin{vmatrix}\partial\phi/\partial x_1 & \partial\phi/\partial x_2 & \partial\phi/\partial x_3\\ \partial^2\phi/\partial x_1\partial r_1 & \partial^2\phi/\partial x_2\partial r_1 & \partial^2\phi/\partial x_3\partial r_1\\ \partial^2\phi/\partial x_1\partial r_2 & \partial^2\phi/\partial x_2\partial r_2 & \partial^2\phi/\partial x_3\partial r_2\end{vmatrix}$$

再一种成像的方法是应用上节讨论的奇性反演。把(5.201)和(5.204)两式代入(5.203)有

$$a(\boldsymbol{x}_0)=\int_{-\infty}^{\infty}\int_{S_0}\int_{S_r}\frac{h(\boldsymbol{x}_0,\boldsymbol{s},\boldsymbol{r})f(\boldsymbol{x}_0,\boldsymbol{s},\boldsymbol{r})}{\omega^2 a(\boldsymbol{x}_0,\boldsymbol{s},\boldsymbol{r})}$$

$$\times\mathrm{e}^{-\mathrm{i}\omega\phi(\boldsymbol{x}_0,\boldsymbol{s},\boldsymbol{r})}u_{s1}(\boldsymbol{s},\boldsymbol{r},\omega)\mathrm{d}^2\boldsymbol{r}\mathrm{d}^2\boldsymbol{s}\mathrm{d}\omega \tag{5.205}$$

再将(5.195)式代入上式得如下 Fourier 积分算子的表示公式

$$a(\boldsymbol{x}_0)=\int_{-\infty}^{\infty}\int_{S_0}\int_{S_r}\int\frac{h(\boldsymbol{x}_0,\boldsymbol{s},\boldsymbol{r})f(\boldsymbol{x}_0,\boldsymbol{s},\boldsymbol{r})a(\boldsymbol{x},\boldsymbol{s},\boldsymbol{r})}{\omega^2 a(\boldsymbol{x}_0,\boldsymbol{s},\boldsymbol{r})}\times$$

$$\mathrm{e}^{\mathrm{i}\omega(\phi(\boldsymbol{x},\boldsymbol{s},\boldsymbol{r})-\phi(\boldsymbol{x}_0,\boldsymbol{s},\boldsymbol{r}))}a(\boldsymbol{x})\mathrm{d}^2\boldsymbol{x}\mathrm{d}^2\boldsymbol{r}\mathrm{d}^2\boldsymbol{s}\mathrm{d}\omega \tag{5.206}$$

注意上式中 $\boldsymbol{x}$ 为地球模型中的任一点，而 $\boldsymbol{x}_0$为象点。应用泰勒展开

$$\phi(\boldsymbol{x},\boldsymbol{s},\boldsymbol{r})-\phi(\boldsymbol{x}_0,\boldsymbol{s},\boldsymbol{r})=(\boldsymbol{x}-\boldsymbol{x}_0)\nabla_{x_0}[\phi(\boldsymbol{x},\boldsymbol{s},\boldsymbol{r})-\phi(\boldsymbol{x}_0,\boldsymbol{s},\boldsymbol{r})]+0(|\boldsymbol{x}-\boldsymbol{x}_0|^2) \tag{5.207}$$

则得到以下 Fredholm 第一类积分方程

$$a(\boldsymbol{x}_0)=\int_{-\infty}^{\infty}\int_{S_0}\int_{S_r}\int\mathrm{e}^{\mathrm{i}\omega(\boldsymbol{x}-\boldsymbol{x}_0)\nabla_{x_0}(\phi(\boldsymbol{x},\boldsymbol{s},\boldsymbol{r})-\phi(\boldsymbol{x}_0,\boldsymbol{s},\boldsymbol{r}))}\omega^{-2}$$

$$\times h(x_0,s,r)f(x_0,s,r)\alpha(x)\mathrm{d}^3x\mathrm{d}^2r\mathrm{d}^2s\mathrm{d}\omega \tag{5.208}$$

用伪逆 $\alpha(x_0)$为数据求解此积分方程得到 $\alpha(x)$，它应是突出波速扰动边界的渐近解。

(二)声广义 Radon 变换的近似逆

在第五节中我们讨论过广义 Radon 变换的问题，本章又讨论了它在奇性反演中的应用，使我们了解到求广义 Radon 变换的逆是一个很困难的问题。现在我们通过与经典 Radon 变换对比的方法，来求广义 Radon 变换的一级近似逆。

由(5.46)可知，某个三维函数 $\alpha(x)$的经典 Radon 变换可表为沿平面 $p=\xi\cdot x$ 的积分

$$[\mathscr{R}\alpha](\xi,p)=\int\alpha(x)\delta(p-\xi\cdot x)\mathrm{d}^3x \tag{5.209}$$

类似于(5.52)式，在三维情况下经典 Radon 变换的逆可表示为

$$\alpha(x_0)=-\frac{1}{8\pi^2}\int\mathrm{d}^2\xi\left(\frac{\partial^2[\mathscr{R}\alpha]}{\partial p^2}\right)_{p=\xi\cdot x_0} \tag{5.210}$$

代入(5.209)有

$$\alpha(x_0)=-\frac{1}{8\pi^2}\int\mathrm{d}^2\xi\int\mathrm{d}^3x\delta''[\xi\cdot(x_0-x)]\alpha(x) \tag{5.211}$$

其中用了公式

$$\frac{\partial^2}{\partial p^2}\int\mathrm{d}^3x\delta(p-\xi\cdot x)\alpha(x)=\int\mathrm{d}^3x\delta''(p-\xi\cdot x)\alpha(x)$$

现在来重写由(5.197)式表示的声广义 Radon 变换。把积分中的坐标原点移到 x_0，令 $x=x_0+y$，得

$$u_{sl}(s,r,t)\approx-\int a(x_0+y,s,r)\delta''[t-\phi(x_0+y,s,r)]\alpha(x_0+y)\mathrm{d}^3y \tag{5.212}$$

利用泰勒展开并代入(5.194)式有

$$\begin{aligned}\phi(x_0+y,s,r)&=\phi(x_0,s,r)+[\nabla_x\phi(x,s,r)]_{x=x_0}\cdot y\\&=\tau_0+[\nabla_x\tau(x_0,s)+\nabla_x\tau(x_0,r)]\cdot y\end{aligned} \tag{5.213}$$

再代入程函方程(5.192)

$$[\nabla_x\tau(x_0,s)]^2=C_{\mathrm{in}}^{-2}(x_0)$$

$$[\nabla_x\tau(x_0,r)]^2=C_{\mathrm{out}}^{-2}(x_0)$$

其中 $C_{\mathrm{in}}(x_0)$是在 x_0点靠入射波方向邻域的参考波速，而 $C_{\mathrm{out}}(x_0)$为向接收器方向的邻域的参考波速。于是，位相函数的空间梯度可表为

$$\nabla_x\phi(x_0,s,r)=\left(\frac{1}{C_{\mathrm{in}}(x_0)}+\frac{1}{C_{\mathrm{out}}(x_0)}\right)\cos\gamma\xi(x_0,s,r) \tag{5.214}$$

其中 $\xi(x_0,s,r)$为指向梯度方向的单位向量，γ 为入射和出射射线之间的半角。

假设振幅函数在 x_0附近不变化，将(5.213)和(5.214)代入(5.212)便将广义 Radon 变换化为沿平面的经典 Radon 变换式

$$u_{sl}(s,r;t)\approx-a(x_0,s,r)\int\alpha(x_0+y)\delta''[t-\tau_0-k\xi(x_0,s,r)\cdot y]\mathrm{d}^3y \tag{5.215}$$

其中

$$k=\frac{\cos\gamma}{C_{\mathrm{out}}(x_0)}\left[1+\frac{C_{\mathrm{out}}(x_0)}{C_{\mathrm{in}}(x_0)}\right] \tag{5.216}$$

现在来对散射走时面上的点 x_0 计算散射波，由于在 x_0 上 $t=\tau_0=\tau(x_0,s,r)$，而且 $\delta''(-ax)=\delta''(x)/a^3$，有

$$u_{s1}(s,r,t=\tau_0)\approx-\frac{a(x_0,s,r)}{|k|^3}\int a(x_0+y)\delta''[\xi(x_0,s,r)\cdot y]\mathrm{d}^3y \tag{5.217}$$

最后把坐标原点搬回 $x=x_0+y$，得

$$u_{s1}(s,r,t=\tau_0)\approx-\frac{a(x_0,s,r)}{|k|^3}\int a(x)\delta''[\xi(x_0,s,r)\cdot(x-x_0)]\mathrm{d}^3x \tag{5.218}$$

再将(5.218)代入(5.197)式便有

$$a(x_0)\approx-\frac{a(x_0,s,r)}{|k|^3}\int \mathrm{d}W(x_0,s,r)\int a(x)\delta''[\xi(x_0,s,r)\cdot(x-x_0)]\mathrm{d}^3x \tag{5.219}$$

由于广义 Radon 变换在这里已经简化为经典 Radon 变换式，只要拿(5.219)式和(5.211)式比较易得

$$\mathrm{d}W(x_0,s,r)=\frac{|k|^3}{8\pi^2a(x_0,s,r)}\mathrm{d}^2\xi \tag{5.220}$$

最后，上式回代(5.197)便得到声广义 Radon 变换的一种近似逆，也就是近似层析成像公式

$$a(x_0)\approx\frac{1}{8\pi^2}\left(1-\frac{C_{\mathrm{out}}(x_0)}{C_{\mathrm{in}}(x_0)}\right)^3\int\frac{|\cos\gamma(x_0,s,r)|^3}{a(x_0,s,r)C_{\mathrm{out}}^3(x_0)}u_{s1}(s,r,t=\tau_0)\mathrm{d}^2\xi \tag{5.221}$$

注意到对固定的象元 x_0，对 $\xi(x_0,s,r)$ 的积分就是对分布在曲面上的源-接收器对的积分如果源也是固定的，如在共炮点道集及 VSP 的情况，那么(5.221)便只对接收器积分。

(三)参考波速场的确定

由(5.181)式的定义，参考波速场包括层速度 $C_n(x)$ 和界面深度函数 $H_n(x)$，$n=1,\cdots,N$，N 为反射界面总数。根据广义反演的原理，可构成以下求层速度和界面深度的联合反演方法，用来求参考波速场[10]。假设条件为层速度和界面深度都是随 x 缓慢变化的函数。

反演用的是走时数据，取自共中点(CMP)反射地震记录。我们知道，在 CMP 道集中第 n 个反射面的走时在小炮检距 x 时可近似表为

$$t_n^2(x)\approx t_n^2(0)-x^2/V_{\mathrm{NMO},n}^2 \tag{5.222}$$

其中 $t_n(0)$ 为零炮检距的走时，V_{NMO} 为 NMO 速度，满足

$$V_{\mathrm{NMO},n}^2=\frac{1}{t_n(0)\cos^2\beta_{0,n}}\sum_{i=1}^{n}C_i^2\Delta t_i\prod_{k=1}^{i-1}\left(\frac{\cos^2\alpha_{k,n}}{\cos^2\beta_{k,n}}\right) \tag{5.223}$$

其中 $\Delta t_i=t_{i-1}(0)-t_i(0)$，$\alpha_{k,n}$ 和 $\beta_{k,n}$ 为对到达 n 界面的法射线在 k 界面的入射角和折射角，$k<n$。用走时 $t_n(x)$ 为数据，用最小二乘法解方程(5.222)求得 $t_n(0)$ 和 V_{NMO}，再用它们为数据求解(5.223)。当界面水平时 NMO 速度就是均方根速度 V_{RMS}，因此(5.223)褪化为线性方程组，求解方法已在第三章中已讨论过了。

采用矩阵方程的记法，把(5.223)写成线性方程组

$$\boldsymbol{b}=\boldsymbol{A}\boldsymbol{w} \tag{5.224}$$

其中

$$\boldsymbol{A}=\begin{bmatrix}\Delta t_1 & & & \\ \Delta t_1 & \Delta t_2 & & \\ \vdots & \vdots & \ddots & \\ \Delta t_1 & \Delta t_2 & \cdots & \Delta t_n\end{bmatrix}$$

$$w_i=C_i^2\prod_{k=1}^{i-1}\left\{\frac{\cos^2\alpha_{k,n}}{\cos^2\beta_{k,n}}\right\},\quad i=1,2,\cdots,N \tag{5.225}$$

$$b_j=V_{\mathrm{NMO},j}^2 t_j(0)\cos^2\beta_{0,j}$$

和

$$\beta_{0,n}=\arcsin\left\{\frac{V_{\mathrm{NMO},1}}{2}\frac{\mathrm{d}t_n(0)}{\mathrm{d}x}\right\} \tag{5.226}$$

这个方程组的解 w 是与层速度 C_i 有关的中间参数，要用逆射线追踪才能确定层速度。所谓逆射线追踪乃是用射线定理由上至下逐层求层速度和射线参数，如入射角，折射角及层内射线段长度等。当 $n=1$时，由(5.225)有

$$C_1=\sqrt{w_1},\quad a_1=\beta_{0,1}R_{1,1}=0.5C_1\Delta t_1$$

其中 a_i 为界面 i 在射线入射处的倾角。当 $n=2$时由(5.225)式有

$$a_{1,2}=\beta_{0,2}-a_1$$

和

$$\frac{\cos^2\alpha_{1,2}}{\cos^2\beta_{1,2}}=\frac{w_2}{C_2^2}$$

而由 Snell 定理有

$$\frac{\sin\alpha_{1,2}}{\sin\beta_{1,2}}=\frac{C_1}{C_2}$$

解联立方程便可求得 C_2和 $\beta_{1,2}$。如此递推到 $n=N$ 时便可以获得所有层速度及射线参数。对每一个 CMP 道集都这样做，画出射线路径并内插射线转折点便可求出反射界面的深度函数。逆射线追踪的优点是有递推公式可用，计算快速。但是，注意到在离散走时数据情况下(5.226)式中的偏导数 $\mathrm{d}t_i(0)/\mathrm{d}x$ 很难求得准，即初始入射角的误差可能逐渐积累，这种方法在层数多时还要进一步改善。

(四)算例

通过下面的算例，我们可以了解上述反散射层析成像的方法步骤和分辨率。理论的地球模型示如图5.13，剖面长1260m，原点在左上角，z 轴向下。震源共8个，位于地面($z=0$)和 $x=1260,1000,760,480,220,-300,-560,-820$m 处(后三个震源在图5.13中未画出)。对反射装置来说接收器位于 $x=0$到 $x=1100$m 之间，间距20m。对 VSP 而言接收器位于 $x=110$m 的井中，深度为 $z=200$m 到 $z=1120$m，$\Delta x=20$m。注意所有炮点都不在井口。

根据(5.181)式我们选用以下反正切函数定义地球内各层的层速度

$$C_i(x)=C_{0i}+\Delta C_i\left\{\arctan\frac{x-x_{ci}+d_i}{d_i}-\arctan\frac{x-x_{ci}-d_i}{d_i}\right\} \tag{5.227}$$

其中 $i=1,\cdots,6$式中的参数示如下表：

i	C_{0i}(m/s)	ΔC_i(m/s)	x_{ci}(m)	d_i(m)
1	1500	0.0	0.0	0.0
2	2000	0.0	0.0	0.0
3	2600	−200	450	700
4	2300	−150	400	600
5	2800	−400	350	400
6	3000	0.0	0.0	0.0

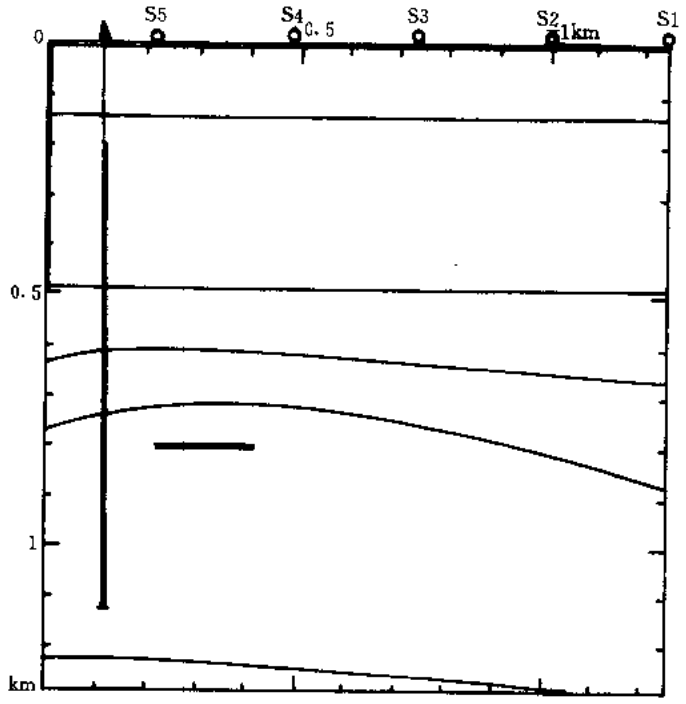


图 5.13 算例用的地球模型和装置

圆圈为点震源，铁轨线为接收器所在位置

在地球模型中有一个背斜，描写它的深度函数也是用反正切函数规定的

$$H_i(x)=H_{0i}+\Delta H_i\left(\arctan\frac{x-x_{hi}+b_i}{b_i}-\arctan\frac{x-x_{hi}-b_i}{b_i}\right) \tag{5.228}$$

其中界面总数为5个，各参数示如下表

i	H_{0i}(m)	ΔH_i(m)	x_{hi}(m)	b_i(m)
1	150	0.0	0.0	0.0
2	490	0.0	0.0	0.0
3	700	−50	450	500
4	900	−120	350	350
5	1350	−80	0.0	700

波速扰动为一个水平透镜体，波速差为400m/s，相应 α 大约为0.286左右，中心深度为812m；长度为220m，即 x 在230m 和450m 之间。整个地球波速场用彩色示如图版 I (a)。由于接收器间距为20m，因此成像时象元大小用$20\times20m^2$。

为了地球波速场的成像，要用到共炮点和共中点的反射地震记录和 VSP 记录，它们都用 WKBJ 近似得来。首先我们用射线追踪求出反射波，散射波及下行波的走时，以及射线的入射角 α_i，然后计算各反射界面的反射系数

$$r_i(x,\alpha_i)=\frac{C_{i+1}(x)\cos\alpha_i-\sqrt{C_i^2(x)-C_{i+1}^2(x)\sin^2\alpha_i}}{C_{i-1}(x)\cos\alpha_i+\sqrt{C_i^2(x)-C_{i+1}^2(x)\sin^2\alpha_i}} \tag{5.229}$$

根据声波方程(5.183)的要求，地层的密度被假设为常数。按能量守恒定理，穿透系数为

$$T_i(x,\alpha_i)=\sqrt{1-r_i^2} \tag{5.230}$$

球面扩散因子本来易于由射线参数求出，但是为了显示清晰，计算时不予考虑。因此(5.191)式中的振幅函数为各层穿透系数的连乘，只是在反射面上用反射系数代替穿透系数而已。

散射波的振幅与较多的因素有关。由(5.218)可知，它取决于波速扰动、参考波速场和射线的角度，即散射波在入射的反方向强度最大。公式(5.218)离散化后即可用来计算直接散射波，多次散射波还要考虑乘上相应界面上的反射系数等等。震源子波用零相位子波，它的振幅谱选为

$$\begin{aligned}
A_s(f)&=0.0 && \text{当}\quad f\leqslant 8\text{Hz}\\
&=(f-8)/(f_p-8) && 8\text{Hz}<f\leqslant f_p\\
&=1.-0.2(f-f_p)/f_p && f_p<f\leqslant 2f_p\\
&=0.8(1.-(f-2f_p)/16) && 2f_p<f\leqslant 2f_p+16\text{Hz}\\
&=0.0 && f>2f_p+16\text{Hz}
\end{aligned} \tag{5.231}$$

其中 f_p 为高峰频率，选为25Hz，使地震波的主波长在第五层中大于82m。由 WKBJ 近似公式(5.191)计算出的是个脉冲系列，它与震源子波褶积之后成为地震道。对反射波而言，合成地震记录是初次反射波和散射波之和；而对 VSP 还要再加上下行波。

如上所述，参考波速场是由 CMP 道集的走时确定的，我们选用16个 CMP 道集取走时数据。每个 CMP 道集有24个炮检距，从最小炮检距100m 开始，每次递增20m。在图5.14中显示了5个这样的道集。用(5.222)到(5.226)所述的方法确定的参考波速场见图版 I (b)(见《地球物理反演与地震层析成象》)的背景。把它与图版 I (a)对比不难发现，地球中的各层内波速的变化和背斜构造都在反演波速图中很清楚的确定了，层速度的最大误差为160m/s；界面深度的误差一般小于5m，只有最下一层例外，达到十多米。

首先来看看利用反射记录对波速扰动成像的结果。在图5.15中展示当炮点 $x_s=1260$m 时的共炮点记录。五条同相轴反映5个反射面，最后两个反射面因构造复杂的波速横向变化产生的反射波也很复杂。直接散射波出现在1.0s 到1.2s 之间。由散射波再次在第五界面反射而成的多次散射波出现在1.4s 到1.6s 之间。对实际资料直接散射波可以用切除或滤波等方法预先提取。层析成像用的公式为(5.221)，它与用(5.203)式等效，但不必把数据按(5.201)变换到频率域，然后再变换回来。对于固定的象元 x_0 和单个震源，(5.221)和(5.203)式都褪变为对接收器的积分。

图5.16的(b)和(c)是对应震源位于1260m 和－820m 时对波速扰动的成像结果，而(a)是地球波速扰动的理论模型。对比(b)、(c)和(d)可见，单震源散射成像可以大致地反映出厚度小于1/4主波长的散射体的存在，但分辨率在垂直波传播方向上比较低。把八个震源的单独成像结果作平均结果示于图5.16(d)。现在把参考波速场和波速扰动的成像结果叠加在一起，用放大的彩色图表示在图版 II (b)(见《地球物理反演与地震层析成象》)中，并把它与图版(a)的模型比较，波速的彩色标度为50m/s，背斜构造和高波速干扰体的像是相当清晰准确的。

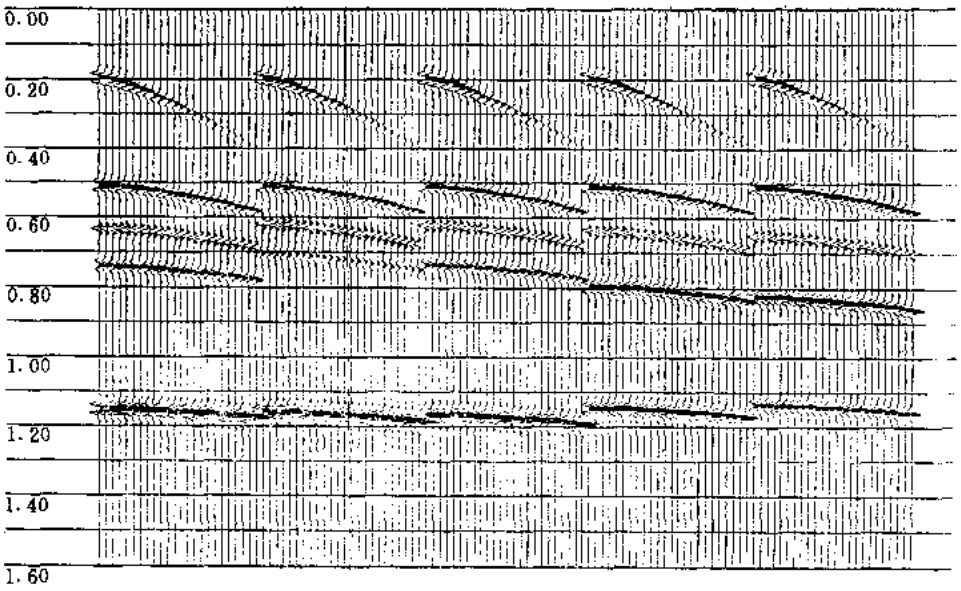


图 5.14　取走时数据用的5个 CMP 道集，相应于测站 $x=0,320,640,960,1280$m

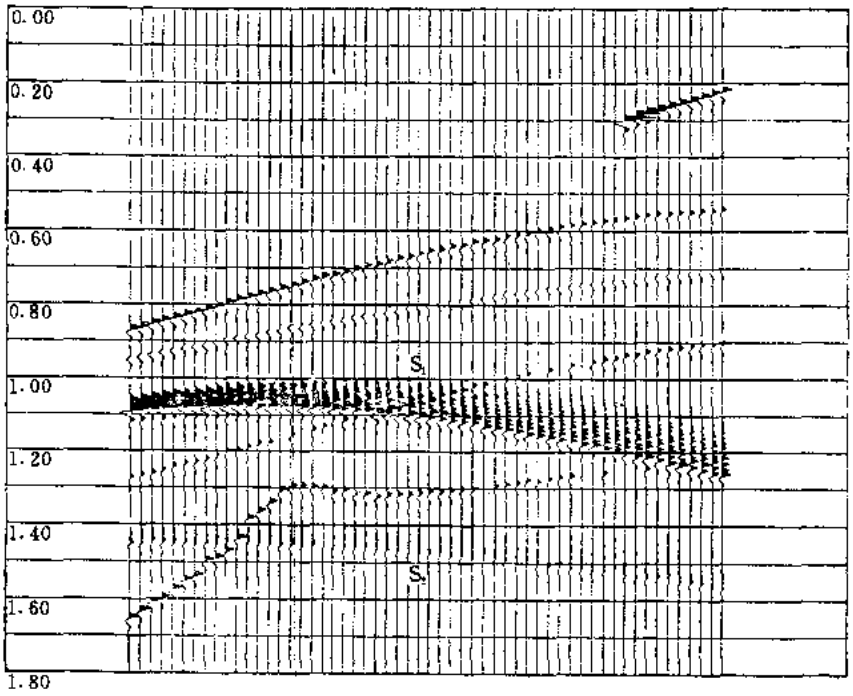
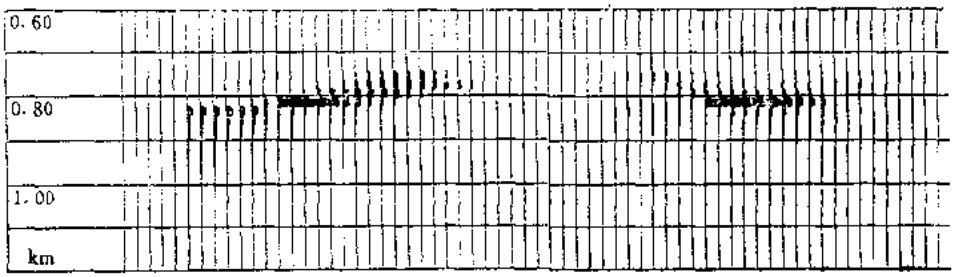


图 5.15　震源位于1260m 时的合成共炮点记录，S_1和 S_2分别标记直接散射波和多次散射波



(a) (b)



(c) (d)

图 5.16 波速扰动的成像结果(a)模型;(b)单震源 $x_s=1260$m;
(c)单震源,$x_s=-820$m;(d)八个震源成像结果的平均

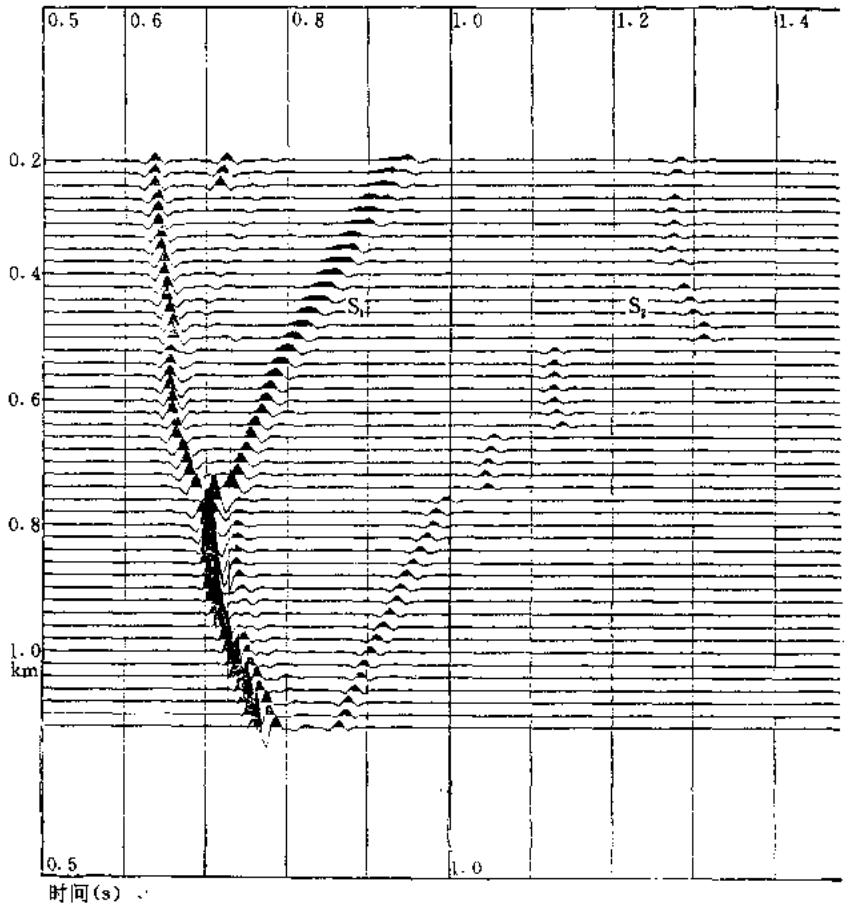


图 5.17 震源在1260m 时的 VSP 记录,S_1为直接散射波,S_2为二次散射波

现在来看看利用 VSP 数据对波速扰动成像的结果。在图5.17中显示出震源在1260m 时的 VSP 剖面。出现在0.6～0.8s 的下行波在740m 以下被直接散射波重叠，而在此深度以上直接散射波看上去类似于上行波，把第四界面的上行波全部盖着，多次散射波出现在0.9～1.3s(图5.17中的 s_2)，它部分地与第五界面的上行波重叠。这一上行波在各界面深度上有位移，是由于背斜构造和层内波速变化所致。

对波速扰动成像的算法和反射情况一样，不同的只是接收器位置在井中而不在地面。单个震源的成像结果示如图5.18的(a)～(c)，相应源位置为1260m，480m 和 −820m。这里我们再次看到成像的分辨率与射线方向有关。8个震源的平均成像结果示如图5.18(d)。综合参考波速反演和反散射成像的结果示如图版Ⅱ(c)(见《地球物理反演与地震层析成象》。比较图中的(a)(b)和(c)可见，在本例中用 VSP 数据的成像结果不如用反射数据，其原因可能是散射体为水平，而反射装置对水平物体敏感所致。最后，我们把如图版Ⅱ(b)和(c)的成像结果再次平均，然后得到的剖面波速场的像示如图版Ⅱ(b)。比较此图(a)中的地球模型和图(b)中的像，读者自己可以了解目前对地球成像可以精细到什么程度。

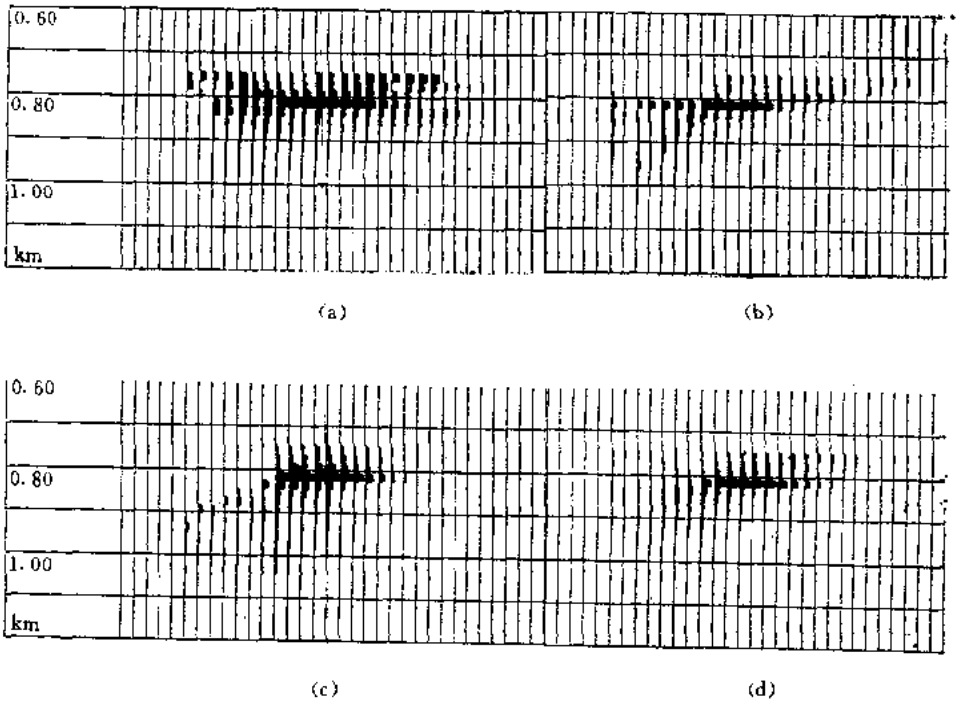


图 5.18　用 VSP 数据取得的波速扰动成像结果

(a)单震源，$x_s=1260$m；(b)单震源，$x_s=480$m；(c)单震源，$x_s=-820$m；(d)8个震源的平均

最后我们总结一下本节讨论的反散射地震层析成像。我们的目的是要把成像的分辨率提高到1/4主波长以上。首先，要把初次反射波和散射波在地震记录上区分出来。然后用初次反射波的走时为数据，反演层速度和反射面深度函数，并把它作为参考波速。最后我们用直

接散射波场为数据，根据广义 Radon 变换的近似逆重建波速扰动的像。数值计算的例子证明了这套方法的可行性。在此方法应用于实际地震数据之前，还要研究如何提取散射波，如何改善层速度反演结果等问题。可以相信，地震层析成像将有助于解决工业上的难题，如石油勘探上直接圈定储油层，在工程调查中发现粘泥层等等。

参考文献(上篇)

[1]马仲蕃等,数学规划讲义,中国人民大学出版社,1981。

[2]郑维敏,系统工程 Fortran 程序集,清华大学出版社,1986。

[3]王家映等,大地电磁测深的拟地震解释法,石油地球物理勘探,20卷1期,1985年。

[4]杨文采,反射地震道的反演——问题和限制,石油物探,25卷3期,1986年。

[5]杨文采,用于位场数据处理的广义反演技术,地球物理学报,29卷3期,1986年。

[6]邹振恒、杨文采,地震道广义线性反演技术,石油地球物理勘探,22卷4期,1987年。

[7]杨文采,粘弹性介质中反射地震道的合成,石油地球物理勘探,21卷6期,1986年。

[8]杨文采,地球物理反演与地震层析成像,地质出版社,1989年。

[9]杨文采,应用地震体波的 CT 技术和地球探查,中国地震,2卷3期,1986年。

[10]杨文采、焦富光,利用联合反演技术进行反射地震的波速成像,地球物理学报,30卷5期,1987年。

[11]赫尔曼,由投影重建图像:CT 的理论基础,科学出版社,1985年。

[12]李世雄等,沿曲线族的 Radon 变换,地球物理学报反演增刊,1988年。

[13]吴大猷,量子力学甲部(理论物理第六册),科学出版社,1984年。

[14]仇庆久等,傅里叶积分算子理论及其应用,科学出版社,1985年。

[15]栾文贵,地球构造反演问题的新途径,地球物理学报,29卷4期,1986年。

[16]Ben-Menahem, A., and Singh, S. J., 1981, Seismic waves and sources, Springer Verlag.

[17]Backus, G. E., and Gilbert, J. F., 1967, Numerical application of a formulism for geophysical inverse problems. Geophys. J. R. astr. Soc., 13, 247~276.

[18]Backus, G. E., and Gilbert, J. F., 1968, The resolving power of gross earth data. Geophys. J. R. astr. Soc., 16, 169~205.

[19]Backus, G. E., and Gilbert, J. F., 1970, Uniqueness in the inversion of inaccurate gross Earth data. Phil. Trans. Roy. Soc., London, Ser. A266, 123~192.

[20]Wiggins, R. A., 1972, The generalized linear inverse problem: Implication of surface waves and free oscillations for Earth structure. Rev. Geophys. Space Phys., 10, 251~285.

[21]Jackson, D. D., 1972, Interpretation of inaccurate, insufficient, and inconsistent data. Geophys. J. R. astr. Soc., 28, 97~109.

[22]Parker, R. L., 1976, Understanding inverse theory. Ann. Rev. Earth Planet. Sci., 5, 35~64.

[23]Anderson, D. L., and Dziewonski, A. M., 1984, Seismic tomography, Scientific American, 251(4), 58~66.

[24]Stewart, G. W., Introduction to matrix computation. Academic Press, 1973.

[25]Jennings, A., 1977, Matrix computation for engineers and scientists. John Wiley & Sons.

[26]Householder, A. S., 1964, The theory of matrices in numerical analysis. Blaisdell.

[27]Golub, G. H., 1965, Numerical methods for solving linear least squares problems. Numer. Math., 7, 206~216.

[28]Golub, G. H., and Kahan, W., 1965, Calculating the singular Values and pseudo-inverse of a matrix. SIAM Numer. Anal., 2, 202~224.

[29]Golub, G. H., and Reinsch, C., 1970, Singular value decomposition and least squares solution. Numer. Math., 14, 103~420.

[30]Forsythe, G. E., *et al.*, 1977, Computer methods for mathematical computations: Prentice-Hall Inc.

[31]Menke, W., 1984, Geophysical data analysis: discrete inverse theory. Academic Press Inc.

[32]Jackson, D., 1979, The use of a priori data to resolve non-uniqueness in linear inversion. Geophys. J. R. astr. Soc., 57, 137~157.

[33]Franklin, J. N., 1970, Well posed stochastic extension of ill-posed problem. J. Math. Appl., 31, 682~716.

[34]Horel, A. E., and Kennard, R. W., 1970, Ridge regression: biased estimation for nonorthogonal problems and applications to nonorthogonal problems. Technometrices, 12, 55~82.

[35]MPS/360 IBM, Mathematical programming system 360, linear and separable programming, application program:

edited by IBM.

[36]Sabatier, P. C., 1977, Possibility constraints in linear inverse problems. Geophys. J. R. astr. Soc., 48, 415～422 and 443～469.

[37]Safon, C. *et al.*, 1977, Some applications of linear programming to the inverse gravity problem. Geophys. 42, 1215～1229.

[38]Jupp, D. L. B., and Vozoff, K., 1975, Joint inversion of geophysical data. Geophys. J. R. astr. Soc., 42, 977～991.

[39]Pedersen, L. B., 1977, Interpretation of potential field data-A generalized inverse approach. Geophys. Pros., 25, 199～230.

[40]Grant, F. S., and West, G. F., 1965, Interpretation theory in applied geophysics, chapter 8～12. ter McGraw-Hill Book Co. Inc.

[41]Aki, K., and Richards, P. G., 1980, Quantitative seismology theory and methods, vol. I, chapter 12. Freeman and Company.

[42]Oldenburg, D., 1981, A comprehensive solution to the linear deconvolution problem. Geophys. J. R. astr. Soc., 65, 331～358.

[43]Lines, L. R., and Treitel, S., 1983, Digital filtering with the second moment norm. Geophys., 48, 505～514.

[44]Cooke, D. A., and Schneiner, W. A., 1983, Generalized inversion of reflection seismic data. Geophys., 48, 665～676.

[45]McAulay, A. D., 1985, Prestack inversion with plane-layer point source modelling. Geophys., 50, 77～89.

[46]Treitel, S., and Robinson, E. A., 1966, Seismic wave propagation in layered media in terms of communication theory. Geophys., 31, 17～32.

[47]Hutson, V., and Pym, J. S., 1980, Applications of functional analysis and operator theory. Academic Press.

[48]Gilbert, F., 1971, Ranking and windowing gross Earth data for inversion and resolution. Geophys. J. R. astr. Soc., 23, 125～128.

[49]Tarantola, A., and Valatte, B., 1982, Generalized nonlinear inverse problems solved using the least squares criterion. Review of geophys. space physics, 20, 219～232.

[50]Parker, R. L., 1970, The inverse problem of electrical conductivity in the mantle, Geophys. J. R. astr. Soc., 22, 121～138.

[51]Green, W. R., 1975, Inversion of gravity profiles by using a Backus-Gilbert approach. Geophys., 40, 363～372.

[52]Huestis, S. P., and Parker, R. L., 1979, Upward and downward continuation as inverse problem. Geophys. J. R. astr. Soc., 57, 137～157.

[53]Hadamard, J., 1923, Lectures on the Cauchy problem in linear partial differential equations, Yale University Press.

[54]Miller, K., 1970, SIAM J. Math. Anal., 1, 52～74.

[55]Tikhonov, A. N., and Arsennin, V. Y., 1977, Solution of illposed problems. John Wiley, New York.

[56]Hanson, R. J., and Phillips, J. L., 1975, An adaptive numerical method for solving linear Fredholm integral equation of the first kind. Numer. Math., 24, 291～307.

[57]Tikhonov, A. N., 1968, Continuation of potential fields in a layered medium by the regularlization method. Earth Physics, No. 12.

[58]Ricker, N., 1977, Transient waves in visco-elastic media. Elsevier, Amsterdam.

[59]Smith, N. C., and Vozoff, K., 1984, Two-dimensional DC resistivity inversion for dipole-dipole data. IEEE Trans. Geoscience and remote sensing, GE-22, 21～28.

[60]Bleistein, N., 1984, mathematical methods for wave Phenomena, Academic Press, Inc.

[61]Lavrentiev, M. M., *et al.*, 1970, Multidimensional inverse problems for differential equation, Lecture notes in math. Vol. 167, Springer-Verlag.

[62]Marchuk, G. I., 1975, Methods of numerical mathematics. Springer-Verlag, New York.

[63]Wong J., *et al.*, 1983, Crosshole seismology and seismic imaging in crystalline rocks. Res. Lett., 10,686～689.

[64]Ivansson, S., 1985. A study of methods for tomographic velocity estimation in the presence of low-velocity zones. Geophys., 50, 969～988.

[65]Peterson, J. E., *et al.*, 1985. Application of algebraic reconstruction techniques to crosshole seismic data. Geophys., 50, 1566～1580.

[66]Daily, W. D., *et al.*, 1982. Geotomography in oil shale. J. Geophys. Res., 87,B7, 5507～5515.

[67]Devaney, A. J., 1984. Geophysical diffraction tomography. IEEE trans. on geoscience and remote sensing, GE-22,3～13.

[68]Balch, A. H., and Myung, W. L., 1984. Vertical seismic profiling: techniques, applications and case histories. International Numan Res. Co., Boston.

[69]Sheriff, R. E., and Geldart, L. P., 1982. Exploration seismology, V. I. Cambridge Univ. Press.

[70]Fawcett, J. A., and Clayton, R. W., 1984, Tomographic reconstruction of velocity anomalies. Bull. Seism. Am. Soc., 74, 2201～2219.

[71]Kanwal, R. P., 1983, Generalized functions: Theory and technique. Academic Press.

[72]Born, M., and Wolf, E., 1975. Principles of optics, 5th ed. New York, Pergamon Press.

[73]Keller, J. B., and Paradakis, J. S., 1977, Wave Propagation and underwater acoustics. Springer Verlag.

[74]Claerbout, J. F., 1985. Imaging the Earth's interior. Blackwell Scientific Pub.

[75]Berkhout, A. J., 1985. Seismic migration. Elsevier.

[76]Paige, C. C. and Saunders, M. A., LSQR: An algorithm for sparse linear equations and sparse least squares, *AMC Transactions Math. Software*, 8, 43～71,1982.

[77]Paige, C. C. and Saunders, M. A., LSQR: Sparse linear equations and least squares problems, *AMC Transactions. Math.*, 8,195～209,1982.

[78]Lanczos, C., An iteration method for the solution of the eigenvalue problem of linear differential and integral operators, *J. Res. N. B. S.*, 45, 255～282,1950.

[79]Langan, R. T., *et al.*, 1985, Tracing of rays through heterogeneous media: An accurate and efficient procedure. Geophys., 50, 1456～1465.

[80]Andersen A. H., and Kak, A. C., 1982, Digital ray tracing in two-dimensional refractive fields. J. Acoust. Soc. Am., 172(5), 1593～1604.

[81]Deans, S. R., 1983, The Radon transform and some of its applications. John Wiely and Sons Cc.

[82]Rowland, S. W., 1979, computer implementation of image reconstruction formula: in "Image Reconstruction from Projections: Implementation and Application" ed. by G. T. Herman, 9～80, Springer-Verlag.

[83]Shepp, L. A., and Logan, B. F., 1974, The Fourier reconstruction of a head section, IEEE trans on Ns., NS-21, 21～38.

[84]Baltes, H. P., (ed.), 1980, Inverse scattering problems in optics. Springer-Verlag.

[85]Sales, J. A., 1987, Tomographic inversion via conjugate gradient method. Geophys., 52, 179～185.

[86]Beylkin, G., 1984, The inversion problem and applications of the generalized Radon transform. Comm. Pure & Applied Math., 37, 579～599.

[87]Beylkin, G., 1985, Imaging of discontinuities in the inverse scattering problems by inversion of a caused generalized Radon transform. J. Math. Physics, 26,99～108.

[88]Ivansson, S., 1983, Remark on an earlier proposed iterative tomographic algorithm. Geophys. J. Roy. astr. Soc., 75,855～860.

[89]Neumann-Denzau, G., and Behren, J., 1984, Inversion of seismic data using tomographic reconstruction techniques for investigations of laterally inhomogeneous media. Geophys. J. R. astr. Soc., 79, 305～315.

[90]Burmakov, J. A., *et al.*, 1984, Determination of three-dimensional velocity structure from observation of refracted body waves. Geophys. J. R. astr. Soc., 79, 285～292.

[91]Bishop, T. N., *et al.*, 1985, Tomographic determination of velocity and depth in laterally varying media. Geo-

phys., 50, 903～923.

[92]Chiu, S. K. L., *et al.*, 1986, Three-dimensional determination of structure and velocity by seismic tomography. Geophys., 51, 1559～1571.

[93]Zhu, T., and Brown, L. D., 1987, Two-dimensional velocity inversion and synthetic seismogram computation. Geophys., 552, 37～50.

[94]Tarantola, A., 1984, The seismic reflection data in the acoustic approximation. Geophys., 49,1259～1266.

[95]Gauthier, O., *et al.*, 1986, Two-dimensional nonlinear inversion of seismic waveforms: numerical results. Geophys., 51, 1387～1403.

[96]Tarantola, A., 1986, A strategy for nonlinear inversion of seismic reflection data. Geophys., 51,1893～1903.

[97]Gelfand, L. M., and Levitan, B. M., 1951, On the determination of a differential equation by its spectral function. Dokl. Akad. Nauk SSSR, 77,557～560(MR. 13,240).

[98]Prosser, R. T., 1980, The Jost-kohn algorithm for inverse scattering, In "Mathematical methods and applications of scattering theory" ed. by J. A. Desanto, 268～276, Springer-Verlag.

[99]Newton, G. R., 1981, Inversion of reflection data for layered media: a review of exact methods. Geophys. J. R. astr. Soc., 65, 191～215.

[100]Weglein, A. B., *et al.*, 1981, Obtaining three-dimensional velocity information directly from reflection seismic data: an inverse scattering formalism. Geophys., 46, 1116～1120.

[101]Weglein, A. B., 1982, Nearfield inverse scattering formalism for the three-dimensional wave equation: the inclusion of a prioi velocity information, J. Acoust. Soc. Am., 71,1179～1182.

[102]Raz, S., 1981, Three-dimensional velocity profile inversion from finite offset scattering data. Geophys., 46,837～842.

[103]Pan, S. X., and Kak, A. C., 1983, A computational study of reconstruction algorithms for diffraction tomography: interpolation versus filtered backpropogation. IEEE trans. Acoust. speech and signal processing, ASSP-31, 1262～1275.

[104]Devaney, A. J., and Oristaglio, M. L., 1984, Geophysical diffraction tomography. The 54th SEG Meeting, Atlanta.

[105]Mueller, R. K., *et al.*, 1979, Reconstructive tomography and applications to ultrasonics. Proceedings of the IEEE, 67, 567～587.

[106]Rushan Wu and Toksoz, M. N., 1987, Diffraction tomography and multisource holography applied to seismic imaging. Geophys., 52, 11～25.

[107]Cohen, J. K., and Bleistein, N., 1977, An inverse method for determining small variation in propogation speed. SIAM on Applied Math., 32, 784～799.

[108]Cohen, J. K., and Bleistein, N., 1979, Velocity inversion procedure for acoustic waves. Geophys., 44, 1497～1511.

[109]Bleistein, N., *et al.*, 1985, Computational and asymptotic aspects of Velocity inversion. Geophys., 50, 1253～1265.

[110]Cohen, J. K., *et al.*, 1986, Three-dimensional Born inversion with an arbitrary reference. Geophys., 51,1552～1558.

[111]Bleistein, N., *et al.*, 1987, Two and one-half dimensional Born inversion with arbitrary reference. Geophys., 52, 26～36.

[112]Weglein, A. B., *et al.*, 1986, Using multiparameter Born theory to obtain certain exact multi parameter inversion goals. Geophys., 51, 1069～1074.

[113]Clayton, R. W., and Stolt, R. H., 1981, A Born-WKBJ inversion method for acoustic reflection data. Geophys., 46, 1559～1567.

[114]Yang Wen-cai, 1987, Towards high-resolution velocity imaging: a perspective on seismic tomographic techniques. Earthquake research in China, 4(4).

[115]Dziewonski, A. M., and Woodhouse, J. H., 1987, Global images of the Earth's interior. Science, 236, 37～48.

[116]Yang Wen-cai, 1990. Seismic velocity imaging in an inhomogeneous layered medium. Chinese J. Geophys., 33, 30～46.

[117]Huang Lian-jie and Yang Wen-cai, 1991, Inversion of the acoustic wave equation by inverse scattering in a medium of linear reference velocity. Chinese J. Geophys., 34, 47～57.

[118]Huang Lian-jie and Yang Wen-cai, 1991, Approximation method of the inversion of the acoustic wave equation by inverse Scattering, Chinese J. Geophys., 34, 58～67.

[119]Yang Wencai and Du Jianyuan, 1992, Seismic Scattering tomography and application to crustal investigation. Earthquake Research in China, 6, 68～86.

[120]Du Jianyuan and Yang Wencai, 1991. Cascade approach for cross-hole seismic tomography. Chinese J. Geophys., 34, 87～96.

[121]Shen Yuan and Yang Wencai, 1992, Application of the generalized Radon transform to prestack seismic data processing. Chinese J. Geophys., 35, 97～110.

[122]Yang Wencai and Du Jianyuan, 1994, A new algorithm of seismic tomography with application to engineering detections. Chinese J. Geophys., 37, 331～337.

[123]Yang Wencai and Jiao Fuguang. 1992, A velocity inversion technique for gas exploration. Bulletin of Chinese Geophysical society. vol. I, 337～344.

下篇　地球物理反演精义

Part Ⅱ　Advanced Geophysical Inversion

"Hence we must believe that all the sciences are so interconnected, that it is much easier to study them all together than to isolate one from all the others. If, therefore, anyone wishes to search out the truth of things in serious earnest, he ought not to select one special science, for all the sciences are cojoined with each other and interdependent."

Rene Descartes

第六章　应用地球物理中的正问题

在应用地球物理中，给定地球模型和初始-边界条件，求地球物理场被称为正问题。提出准确和完备的正问题，并找出快速的算法，是地球物理反演的基础之一。地球物理学的各个分支，如地震学、重力学、大地电磁学等，都研究正问题，我们不作全面讨论，而只能对它们的共性作出简要的归纳。因此，我们将首先介绍应用地球物理中的通用方程式，它适用于地震法、重力法、电磁法、地热法和磁法勘探。

一、应用地球物理学的基本方程式

(一)基本方程式

应用地球物理学是探测地球内介质物理上的不均匀性的一门应用科学，目前探测的深度从几米到几百公里。根据探测对象的深度和尺度，可大致分为以下四类。

1. 浅层(1～100m)，目标尺度＜20m，应用于工程地基、地下建筑物探测及考古。

2. 中浅层(50m～1km)，目标尺度＜100m，应用于固体矿产、地下水等资源探查和地质构造调查。

3. 中层(0.5～10km)，目标尺度＜1km，石油天然气勘探及上地壳探查。

4. 深层(10～500km)，目标尺度＞1km，应用于地壳与岩石圈组构调查。

以下方程式包含了地震、电磁、重力、地磁和地热等各种地球物理方法的正问题，因此称为基本方程式。

$$Pu \equiv \nabla^2 u(\boldsymbol{x},t) - h(\boldsymbol{x})\frac{\partial^2 u(\boldsymbol{x},t)}{\partial t^2} - g(\boldsymbol{x})\frac{\partial u(\boldsymbol{x},t)}{\partial t} = f(\boldsymbol{x},t) \tag{6.1}$$

式中 u 表示地球物理场的一种，如声场，电磁场的某一分量，等等；f 为源函数，$\boldsymbol{x}$ 为空间的一个点，t 为时间。方程(6.1)为阻尼标量波动方程。系数 h 和 g 的物理含义对不同方法是不同的，详见表6.1。

表6.1　方程(6.1)中系数的物理意义

方　法	物理场 u	h	g
地震	声压或位移分量	$1/c^2$	$2\alpha/c$
瞬变电磁	磁场分量或电场分量	$\mu\varepsilon$	$\mu\sigma$
地质雷达	电场分量	$\mu\varepsilon$	0
低频电磁法	电磁场分量	0	$\mu\sigma$
地热法	地温场	0	$1/a^2$
电法、重力和地磁	位场	0	0

注：c 为波速，α 为波场衰减系数，μ 为磁导率，ε 为介电系数，σ 为电导率，a^2 为热扩散系数

方程(6.1)中的算子 P 是拟微分算子，而不是一般的偏微分算子，关于拟微分算子及其有关性质将在本节最后讨论。我们先来说明对于地震勘探和瞬变电磁法，(6.1)是如何导出

的。

(二)非广角反射地震波场的阻尼声波方程[1~5]

用于石油和浅层地质构造勘查的反射地震勘探，其炮检距一般小于探测目标的深度，因此转换波能量很弱。加上表层低速带的影响，地表接收信号以位移垂直分量为主，通常不观测水平分量，因此标量波动方程可以近似地描述其波场。然而，地震波的衰减和频散是十分明显的，这与地球介质为非弹性有关。当介质不是完全弹性时，弹性模量及波速不再是实数，而变为复数，它们的虚部与波能量的衰减有关。这时，Hooke 定律表示为

$$\sigma(\omega)=M(\mathrm{i}\omega)\varepsilon(\omega) \tag{6.2}$$

式中 σ 为应力，ε 为应变，ω 为角频率，$\mathrm{i}\omega$ 注记频率的复函数。对纵波而言，复弹性模量

$$M(\mathrm{i}\omega)=\lambda(\mathrm{i}\omega)+2\mu(\mathrm{i}\omega)=k(\mathrm{i}\omega)+\frac{4}{3}\mu(\mathrm{i}\omega) \tag{6.3}$$

其中 λ 和 μ 为拉梅系数，k 为体积模量。将复函数展开虚实部，有

$$M(\mathrm{i}\omega)=M'(\omega)+\mathrm{i}M''(\omega)=|M(\mathrm{i}\omega)|\mathrm{e}^{\mathrm{i}\delta(\omega)} \tag{6.4}$$

式中 δ 为耗损角，$\tan\delta(\omega)$ 称为耗损正切，而品质因数

$$Q(\omega)=[\tan\delta(\omega)]^{-1}=M'(\omega)/M''(\omega) \tag{6.5}$$

反射地震的频带为8～250Hz。在这个频带内，品质因数和其它宏观介质参数随频率的变化不明显，以下假设为常数，只随空间坐标而变化。

在没有体力存在时，牛顿第二定律表示为

$$\rho\frac{\partial^2 u}{\partial t^2}=\frac{\partial\sigma}{\partial x} \tag{6.6}$$

其中 ρ 为密度，u 为压力(对海洋勘探)或位移的垂直分量(地面反射)。上式代入(6.2)式有

$$\frac{\partial^2 u}{\partial t^2}-\frac{1}{v^2}\frac{\partial^2 u}{\partial t^2}=0 \tag{6.7}$$

其中复波速

$$v=[M/\rho]^{1/2}=[|M|/\rho]^{1/2}\mathrm{e}^{\mathrm{i}\delta/2} \tag{6.8}$$

在(6.7)式中波的衰减是由 v 的虚部表示的。(6.7)式的稳态解为

$$u(x,t)=u_0(\omega)\exp[\mathrm{i}\omega(t-x/v)] \tag{6.9}$$

设 c 为相速度，α 为衰减系数，都为实数。由(6.8)式代入(6.9)式，分开虚实部，便有

$$c=[|M|/\rho]^{1/2}\sec\frac{\delta}{2} \tag{6.10}$$

$$\alpha=[\rho/|M|]^{1/2}\omega\ \sin\frac{\delta}{2}=\omega\tan\frac{\delta}{2}/c \tag{6.11}$$

当 $Q\gg 0$ 时 $\delta\to 0$，$\tan\frac{\delta}{2}\to\frac{1}{2}\delta$，代入(6.4)有

$$\alpha(\omega)\simeq\frac{\omega}{2Qc} \tag{6.12}$$

将(6.8)及(6.12)代入(6.7)有

$$u(x,t)=u_0(\omega)\mathrm{e}^{-\omega x/2Qc}\mathrm{e}^{\mathrm{i}\omega(t-x/c)} \tag{6.13}$$

这一结果说明波随频率和距离指数衰减，是与现场实验结果符合的。

将复波速推广到三维标量波动方程的稳态情况，在频域有

$$\nabla^2 u(x,\omega)+\frac{\omega^2}{v^2(x)}u(x,\omega)=0 \tag{6.14}$$

代入(6.8)式有

$$\nabla^2 u(x,\omega)+\left[\frac{\omega^2}{c^2(x)}-\alpha^2(x)-\frac{2i\omega\alpha(x)}{c(x)}\right]u(x,\omega)=0 \tag{6.15}$$

或

$$\nabla^2 u(x,\omega)+\frac{\omega^2}{c^2(x)}\left[1-\frac{1}{4Q^2(x)}\right]u(x,\omega)-\frac{2i\omega\alpha(x)}{c(x)}u(x,\omega)=0 \tag{6.16}$$

对岩石 $Q\gg 1$有[2]

$$\nabla^2 u(x,\omega)+\frac{\omega^2}{c^2(x)}u(x,\omega)-i\omega\frac{2\alpha(x)}{c(x)}u(x,\omega)=0 \tag{6.17}$$

此式左边为拟微分算子的象征，相应在时间域的稳态方程可根据此象征写为

$$\nabla^2 u(x,t)-\frac{1}{c^2(x)}\frac{\partial^2 u(x,t)}{\partial t^2}-\frac{2\alpha(x)}{c(x)}\frac{\partial u(x,t)}{\partial t}=0 \tag{6.18}$$

此式对应数学物理中典型的方程之一，即阻尼波动方程，其中系数项的参数都是可测量的。这一方程简洁明了，用于反射地震合成和反演应是很方便的。要说明的是，(6.18)式中的衰减系数 α 是与频率成正比的，因为在(6.12)式中相速度 c 和品质因数 Q 都(在地震频段内)不频散。如果把表6.1的系数代入(6.18)便得到基本方程式(6.1)。

现在将上述结果与前人的研究作一对比。较早提出的 Maxwell 粘弹性模型偏向粘性介质，在地震学中应用较少。讨论较多的 Voigt 粘弹性体，用 $\eta\frac{d}{dt}+\mu$ 代替完全弹性体中的 μ，其中 η 为介质的粘滞系数，因此复波速为

$$v(i\omega)=[(\mu+i\omega\eta)/\rho]^{1/2} \tag{6.19}$$

$$\frac{1}{v}=\left[\frac{\rho(\mu+\sqrt{\mu^2+\omega^2\eta^2})}{2(\mu^2+\omega^2\eta^2)}\right]^{1/2}+i\left[\frac{\rho(\sqrt{\mu^2+\omega^2\eta^2}-\mu)}{2(\mu^2+\omega^2\eta^2)}\right]^{1/2} \tag{6.20}$$

与(6.8)对比可知有以下对应关系

$$c^2\leftrightarrow\frac{\mu\rho}{2(\mu^2+\omega^2\eta^2)}\left[\sqrt{1+\frac{\omega^2\eta^2}{\mu^2}}+1\right] \tag{6.21}$$

$$\alpha^2\leftrightarrow\frac{\mu\rho\omega^2}{2(\mu^2+\omega^2\eta^2)}\left[\sqrt{1+\frac{\omega^2\eta^2}{\mu^2}}-1\right] \tag{6.22}$$

这些结果说明本文的公式包含了 Voigt 粘弹性体模型的情况。

但是，(6.18)式与目前地震反射中常用的 Kelvin Voigt 粘弹性体模型所得的结果有明显区别。后者加上涨缩粘滞性的机制，用

$$\lambda+\eta_1\frac{\partial}{\partial t}\text{和 }\mu+\eta_2\frac{\partial}{\partial t}$$

代替完全弹性体中的 λ 和 μ，结果导致某些文章中常出现的以下方程

$$(\lambda+2\mu)\nabla^2 u+(\eta_1+2\eta_2)\frac{\partial}{\partial t}(\nabla^2 u)=\rho\frac{\partial^2 u}{\partial t^2} \tag{6.23}$$

此方程的阻尼项对 $\nabla^2 u$ 求偏导，而不象(6.18)式直接对 u 求 t 的偏导数。在一维情况下，(6.23)式的稳态解为

$$u(x,\omega)=u_0(\omega)e^{-\omega^2\beta x}e^{i\omega(t-x/c)} \tag{6.24}$$

其中 β 为与 η_1 及 η_2 有关的参数。(6.24)式中波场与频率的平方呈指数衰减，与现场实验结果不符，因此作者不主张用(6.23)及(6.24)式描述地震勘探的波场。

由于我们假定品质因数 Q 为常数，上面推导出的方程(6.17)和(6.18)仅适用于频率为数百赫兹以下的纵波勘探的情况。实际上，Q 值随频率的变化是存在的，规律比较复杂。以沉积盆地中常见的砂岩为例，Q^{-1} 值在几千赫的频段上可出现谐振峰，即选择性吸收，这个问题已在有关文章中作过描述，并给出相应波速和吸收系数随频率变化的定量关系[2]。数值计算表明，考虑非恒 Q 介质对地震合成结果影响不大。因此，反射地震合成用(6.17)及(6.18)式求有限元等数值解是可行的，但在全波声测资料合成时不宜使用。

(三)电磁法中的阻尼波动方程

这里参照文献[6]第四章中的推导，把电磁法中的阻尼波动方程与本文(6.18)式得出的结果作对比，用国际单位。

电磁场满足的 Maxwell 方程为(忽略位移电流)

$$\nabla \times \boldsymbol{E} + \frac{\partial \boldsymbol{B}}{\partial t} = 0 \tag{6.25}$$

$$\nabla \times \boldsymbol{H} - \frac{\partial \boldsymbol{D}}{\partial t} = \boldsymbol{j} \tag{6.26}$$

$$\nabla \cdot \boldsymbol{B} = 0 \tag{6.27}$$

$$\nabla \cdot \boldsymbol{D} = \rho \tag{6.28}$$

式中 $\boldsymbol{E}$ 为电场强度，$\boldsymbol{B}$ 为磁感应强度，$\boldsymbol{H}$ 为磁场强度，$\boldsymbol{D}$ 为电位移矢量，$\boldsymbol{j}$ 为电流密度，ρ 为自由电荷密度，在磁导率 μ、电导率 σ 和介电系数 ε 与时间无关并无频散时，有状态方程

$$\left.\begin{aligned} \boldsymbol{D} &= \varepsilon \boldsymbol{E} \\ \boldsymbol{B} &= \mu \boldsymbol{H} \\ \boldsymbol{j} &= \sigma \boldsymbol{E} \end{aligned}\right\} \tag{6.29}$$

将(6.25)和(6.26)式取旋度并代入(6.29)式得

$$\nabla \times (\nabla \times \boldsymbol{E}) + \mu \nabla \times \frac{\partial \boldsymbol{H}}{\partial t} = 0 \tag{6.30}$$

$$\nabla \times (\nabla \times \boldsymbol{H}) - \varepsilon \nabla \times \frac{\partial \boldsymbol{E}}{\partial t} = \sigma \nabla \times \boldsymbol{E} \tag{6.31}$$

在这里我们假定在取微分的小区间内 μ、σ 和 ε 为常数。同时，把(6.29)式代入(6.27)式及(6.28)式后有

$$\left.\begin{aligned} \nabla \cdot \boldsymbol{H} &= 0 \\ \nabla \cdot \boldsymbol{E} &= 0 \end{aligned}\right\} \tag{6.32}$$

其中隐含着均匀介质中不存在自由电荷的假定。

只要矢量函数 $\boldsymbol{H}$ 和 $\boldsymbol{E}$ 分段连续并二阶可微，则(6.30)式和(6.31)式中($\nabla \times$)和 $\partial/\partial t$ 的次序便可互换。再代入(6.25)(6.26)及(6.29)式有

$$\nabla \times \nabla \times \boldsymbol{E} + \mu\varepsilon \frac{\partial^2 \boldsymbol{E}}{\partial t^2} + \mu\sigma \frac{\partial \boldsymbol{E}}{\partial t} = 0 \tag{6.33}$$

$$\nabla \times \nabla \times \boldsymbol{H} + \mu\varepsilon \frac{\partial^2 \boldsymbol{H}}{\partial t^2} + \mu\sigma \frac{\partial \boldsymbol{H}}{\partial t} = 0 \tag{6.34}$$

利用恒等式

$$\nabla\times\nabla\times\boldsymbol{A}=\nabla\nabla\cdot\boldsymbol{A}-\nabla^2\boldsymbol{A} \tag{6.35}$$

并代入(6.32)式，可把(6.33)式和(6.34)式写为

$$\nabla^2\boldsymbol{E}-\mu\varepsilon\frac{\partial^2\boldsymbol{E}}{\partial t^2}-\mu\sigma\frac{\partial\boldsymbol{E}}{\partial t}=0 \tag{6.36}$$

$$\nabla^2\boldsymbol{H}-\mu\varepsilon\frac{\partial^2\boldsymbol{H}}{\partial t^2}-\mu\sigma\frac{\partial\boldsymbol{H}}{\partial t}=0 \tag{6.37}$$

这是关于电场和磁场的矢量阻尼波动方程，写成电场和磁场分量形式后与(6.18)式结构完全相同，其中系数的对应关系为

$$\mu\varepsilon\sim1/c^2;\ \mu\sigma\sim2a/c$$

因此，(6.36)和(6.37)也是基本方程式(6.1)的特殊情况。

(四)拟微分算子

基本方程(6.1)中的算子 P 为拟微分算子，因为其中的系数项 h 和 g 为 x 的函数，当系数 h 和 g 为常数时，P 才退化为偏微分算子(PDO)，以下注记为 D。拟微分算子同时与 D 与 x 有关，即

$$P\equiv P(D,x)。$$

从偏微分方程理论知，系数为常数的二阶偏微分方程可以通过 Fourier 变换化为代数方程求解。例如，当(6.1)中 $h=$const 和 $g=$const 时，在二维情况下它可化为

$$-k_x^2\tilde{u}-k_z^2\tilde{u}+\omega^2h\tilde{u}-\mathrm{i}\omega g\tilde{u}=\tilde{f}(k_x,k_z,\omega) \tag{6.38}$$

其中 k_x 和 k_z 为 x 或 z 方向的角波数，$\tilde{u}\equiv u(k_x,k_z,\omega)$。

当系数 h 和 g 不为常数时，(6.1)相应的 Fourier 变换为

$$-k_x^2\tilde{u}-k_z^2\tilde{u}-\frac{\partial^2}{\partial t^2}\iint h(x)u(x,z,t)\mathrm{e}^{-\mathrm{i}(k_xx+k_zx)}\mathrm{d}x\mathrm{d}z$$

$$-\frac{\partial}{\partial t}\iint g(x)u(x,z,t)\mathrm{e}^{-\mathrm{i}(k_xx+k_zx)}\mathrm{d}x\mathrm{d}z=\tilde{f}(k_x,k_z,\omega) \tag{6.39}$$

此式是积分方程式，不能简化为代数方程。由于如何求解过去的偏微分方程理论没有解决，因此发展了拟微分算子理论，以便对变系数偏微分方程进行计算。

所谓拟微分算子就是在系数可变时与古典偏微分算子 D 最接近的一种算子，它是这样定义的。

对(6.1)左边，如果有

$$P(D,\boldsymbol{x})u(\boldsymbol{x})=\frac{1}{(2\pi)^{n/2}}\iint\mathrm{e}^{\mathrm{i}\langle x-y,\theta\rangle}p(\boldsymbol{x},\theta)u(\boldsymbol{y})\mathrm{d}\boldsymbol{y}\mathrm{d}\theta \tag{6.40}$$

存在，而且其中新函数 $p(\boldsymbol{x},\theta)$ 可表示为一个多项式

$$\sum_{|\mu|\leqslant m}a_\mu(\boldsymbol{x})\theta^\mu \tag{6.41}$$

则称 $P(D,x)$ 为拟微分算子。式(6.40～6.41)中，μ、m 和 n 为整数，$\langle\boldsymbol{x}-\boldsymbol{y},\theta\rangle$ 表示 $x-y$ 和 θ 的齐一次函数。这样定义的拟微分算子 $P(D,\boldsymbol{x})$ 与多项式 $p(\boldsymbol{x},\theta)$ 之间存在一一对应的关系。由于多项式 p 是“好函数”，有关计算不会发生困难。

对以拟微分算子方程(6.1)，古典意义下的解析解不存在。举一个最简单的例子，两段密度不相同的弦的振动问题，即求解

$$\frac{\partial^2u(x,t)}{\partial t^2}=a^2(x)\frac{\partial^2u(x,t)}{\partial x^2} \tag{6.42}$$

其中

$$a(x)=\begin{cases}\alpha, & 0<x\leqslant b\\ \beta\neq\alpha, & b<x\leqslant l\end{cases} \tag{6.43}$$

l 为弦长。这个问题不存在古典解，因为波场有间断。但是，我们可以根据某种物理上的准则定义它的广义解。例如，记 $u_t=\partial u/\partial t$，$u_x=\partial u/\partial x$，弦振动系统动能与势能之差为

$$L(t)=\int_0^l\left(\frac{1}{2}au_t^2-\frac{T}{2}u_x^2\right)\mathrm{d}x \tag{6.44}$$

此式中 T 为常数。我们可以求 $L(t)$ 对时间积分取极小时 u 的广义解。这种求解方法是使用变分法的出发点。在本章第三节将进一步讨论这一问题。

二、用于波传播问题的有限差分法

由于非均匀复杂地球介质中基本方程(6.1)的解析解不存在，必须求助于数值计算方法。本节讨论反射地震数值模拟常用的有限差分法。虽然许多读者熟知有限差分法，但是它用于解波动问题时对波正向及反向传播有直观的解释，因此仍需在本篇中作概要讨论。

(一)反射地震中波传播方程

反射地震法是最重要的应用地球物理方法。它由点震源激发，在海洋中测量压力波，在陆上测量位移的垂直分量，记录可用标量波场 $u(x,t)$ 来表示。在一般情况下，反射地震排列(指震源和接收器串)长度比探测目标小许多。例如，深反射地震排列长度6km 左右，而探测对象 Moho 面深度在陆上一般在30km 以上。因此，波传播到界面上产生的转换波能量在地震记录上可以忽略。此时，描述地震波传播的弹性波方程可以简化为以下标量波动方程

$$\Delta u-\frac{1}{c^2}\frac{\partial^2u}{\partial t^2}-\frac{1}{\rho}\nabla\rho\cdot\nabla u=\delta(s)f(t) \tag{6.45}$$

式中 ρ 为密度，s 为点震源位置，$f(t)$ 为震源波形函数。如果假定密度 ρ 为常数和平面波入射，(6.45)进一步简化为一般标量波动方程

$$\nabla^2u_{xx}-\frac{1}{c^2}\nabla^2u_{tt}=0 \tag{6.46}$$

它是基本方程(6.1)中 $g=0$ 的特殊情况。此方程只能准确地反映波的相位，而不能准确地反映波在非均匀介质中传播时振幅的变化，在合成反射地震图时不宜使用。

在二维情况下，对(6.45)左边的算子作展开得

$$\frac{\partial^2u}{\partial t^2}=c^2\left[\left(\frac{\partial^2u}{\partial x^2}+\frac{\partial^2u}{\partial z^2}\right)-\frac{1}{\rho}\left(\frac{\partial\rho}{\partial x}\cdot\frac{\partial u}{\partial x}+\frac{\partial\rho}{\partial z}\cdot\frac{\partial u}{\partial z}\right)\right]+\delta(x_s,z_s)f'(t) \tag{6.47}$$

相应的初始条件为

$$u|_{t\leqslant0}=0\text{和}\frac{\partial u}{\partial t}|_{t\leqslant0}=0 \tag{6.48}$$

边界条件为地面($z=0$)上垂直方向应力等于0。

更复杂的地震波传播正问题将在本章第四节中讨论。

(二)有限差分法(FDM)

用如图6.1所示的网格划分地下介质，其中 h 为正方形

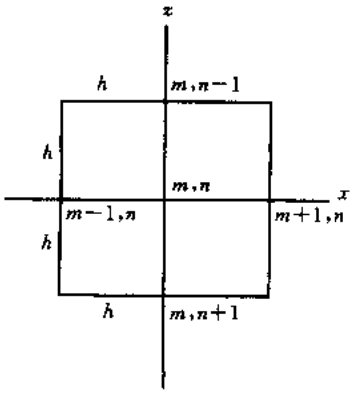


图6.1 有限差分网格结点标记

网格元的边长，m 和 n 为当前网格节点的横向及垂向编号。用 l 注记时间取样号，则可把(6.47)逐步离散化：

$$\frac{\partial^2 u_{m,n}}{\partial x^2}=\frac{1}{h}\left(\frac{\partial u_{m+h/2,n}}{\partial x}-\frac{\partial u_{m-h/2,n}}{\partial x}\right)=h^{-2}(u_{m+1,n}-2u_{m,n}+u_{m-1,n}) \tag{6.49}$$

$$\frac{\partial^2 u_{m,n}}{\partial z^2}=\frac{1}{h}\left(\frac{\partial u_{m,n+h/2}}{\partial z}-\frac{\partial u_{m,n-h/2}}{\partial z}\right)=h^{-2}(u_{m,n+1}-2u_{m,n}+u_{m,n-1}) \tag{6.50}$$

$$\frac{\partial^2 u_l}{\partial t^2}=\Delta t^{-2}(u_{l+1}-2u_l+u_{l-1}) \tag{6.51}$$

$$\frac{\partial \rho_{m,n}}{\partial x}=\frac{\rho_{m+1,n}-\rho_{m,n}}{h},\frac{\partial \rho_{m,n}}{\partial z}=\frac{\rho_{m,n+1}-\rho_{m,n}}{h} \tag{6.52}$$

将(6.49～6.59)代入(6.47)得

$$u_{m,n,l+1}=-u_{m,n,l-1}+C_1u_{m,n,l}+C_2u_{m+1,n,l}+C_3u_{m,n+1,l}+C_4u_{m-1,n,l}+C_5u_{m,n-1,l} \tag{6.53}$$

其中系数为

$$C_1=2-4K,C_2=K\left(1-\frac{\rho_{m+1,n}-\rho_{m,n}}{2\rho_{m,n}}\right), \tag{6.54}$$

$$C_3=K\left(1-\frac{\rho_{m,n+1}-\rho_{m,n}}{2\rho_{m,n}}\right),C_4=K\left(1+\frac{\rho_{m+1,n}-\rho_{m,n}}{2\rho_{m,n}}\right), \tag{6.55}$$

$$C_5=K\left(1+\frac{\rho_{m,n+1}-\rho_{m,n}}{2\rho_{m,n}}\right);K=\left(\frac{c\Delta t}{h}\right)^2 \tag{6.56}$$

对于差分方程(6.53)，由上至下、由左向右并随时标 l 加大计算空间任一点(m,n)计算波场 $u_{m,n,l+1}$，便得到波传播的图像，$u(m,0,l)$便是地面直达波和反射波场的合成记录图。有限差分计算必须满足的条件如下。

1. 时间取样率 $\Delta t(t=l\Delta t)$满足 $\Delta t\leqslant h/C$。

2. 震源信号的主周期 $T<10h/C$，否则有严重频散。

3. 由于地下介质无限，而计算网格有限，计算网格的边界必须是吸收边界。

4. 震源必须作专门处理，即在源点加入 $f(t)$信号。

(三)基本方程式的有限差分格式

类似于上边的推导，即把(6.49～6.52)代入(6.1)式可得(二维情况)

$$\begin{aligned}u_{m,n,j+1}=&-B_1u_{m,n,j-1}+B_2u_{m,n,j}+B_3[u_{m+1,n,j}+u_{m-1,n,j}\\&+u_{m,n+1,j}+u_{m,n-1,j}]-B_4f_{m,n,j}\end{aligned} \tag{6.57}$$

其中系数项为

$$\left.\begin{aligned}B_1&=\frac{h_{m,n}}{h_{m,n}+\Delta t\cdot g_{m,n}}\\B_2&=\frac{2h_{m,n}+\Delta t\cdot g_{m,n}-4(\Delta t/l)^2}{h_{m,n}+\Delta t\cdot g_{m,n}}\\B_3&=\frac{(\Delta t/l)^2}{h_{m,n}+\Delta t\cdot g_{m,n}}\\B_4&=\frac{\Delta t^2}{h_{m,n}+\Delta t\cdot g_{m,n}}\delta_{p,q}\end{aligned}\right\} \tag{6.58}$$

式中 l 为剖分网格边长，p 指网格任一结点，q 为人工场源点。

(四)波场延拓、反传与地震偏移

假定震源在地面上或其上方，地震波场将向下方及四周传播，即正向传播。以上讨论的方程式都是正向传播的方程式。

将(6.53)式移项，有

$$u_{m,n+1,l}=C_3^{-1}\{u_{m,n,l+1}+u_{m,n,l-1}-C_1u_{m,n,l}-C_2u_{m+1,n,l}-C_4u_{m-1,n,l}-C_5u_{m,n-1,l}\} \quad (6.59)$$

此式说明，如果给出任意时刻 T 在某平面 $z_1=0$ 上方的波场 $u(x,z>z_1,t\geqslant T)$，则可以计算此平面(及其下方)上一时刻的波场 $u(x,z\geqslant z_1,t<T)$。实际上，只要知道地面上($z=0$)的波场 $u(x,0,t)$及其垂向导数，则可以外推地下的波场 $u(x,z,t<T)$。这种计算算为波场向下延拓。但是，这种计算是极不稳定的，其误差随深度增加呈指数放大(见第七章第四节的详细讨论)。

将(6.53)移项还可以由其它项计算 $u_{m,n-1,l}$，即由下方的波场计算上方的波场。这种计算称为波的反传，即反向传播。波的反传是由 $u(x,z,t)$计算 $u'(x,z,t'<t)$，这时波场将会聚到震源或衍射体，从而给出衍射体的图像[7~9]。

假定震源是来自无穷远处垂直入射的平面波，那么零时刻 $u(x,z,t=0)$代表了所有震源(包括二次震源)同时激发的瞬间的波场，它直接反映反射界面、衍射体的图像。将地面接收的反射记录 $u(x,0,t)$换算为 $u(x,z,0)$的地震处理称为地震偏移[7]。由于 $u(x,0,t)$与 $u(x,z,0)$的自由度相同，因此地震偏移计算比波动向下延拓稳定一些。从原理上看，(6.59)式就可以用来作"逆时"地震偏移，成为有限差分法偏移的一种方法。但是，作为大型工程计算的一种方法，偏移需要更深入的理论和技巧。马在田先生的专著[10]对有限差分地震偏移作了精辟的分析和介绍。

三、有限元法简析

上节讨论的有限差分法具有简明快速的特点，但由于只能使用矩形网格，对复杂的地质构造不能准确地模拟。例如，对反射地震中常见的倾斜界面，在有限差分网格中变成了阶梯形状，其衍射模式与斜界面反射不尽相同。因此，需要有更为灵活的方法求解基本方程(6.1)。这时，有限元法(FEM)就是一种很有用而且很灵活的计算技术，最适用于与复杂介质结构有关的拟微分方程边值问题的数值解。

在第一节最后我们已提到拟微分方程的广义解，有限元法的基础就是来自这种用变分原理导出的广义解。注意，有限元法一般只对(6.1)式中的空间微分算子作逼近，而与时间微分有关的计算仍然多采用有限差分。因此，本节只对空间微分算子方程(Laplace 方程)的有限元作剖析。

(一)场问题的变分原理[11~13]

变分法的主要出发点之一，对与场能量有关的泛函极小化等效于直接求解相应场方程。在(6.44)式中，我们曾讨论过有关弦振动问题的例子。对于 Laplace 方程

$$\nabla^2u=0 \quad (6.60)$$

最小势能原理要求，在封闭区域 Ω 内势的分布使场能量达到最小(稳定保守场)。势场能量表为

$$F(u)=\frac{1}{2}\int_{\Omega}|\nabla u|^2 d\Omega \tag{6.61}$$

于是，满足 Laplace 方程的势场同时也是使(6.61)取极小的场。因此，我们有两大类不同的方法求解势场问题，或者用有限差分等方法直接求解(6.60)，或者将势场 u 展成某种简单函数和系数的线性组合，通过对能量 $F(u)$ 的极小化确定这些系数之后逼近势场。有限元法就属于后一种。

一般来说，在应用地球物理中涉及的多数拟微分方程都可以找到对应的泛函，并通过其极小化来逼近[12,13]。如 Poission 方程

$$\nabla\cdot(\nu\nabla u)=g \tag{6.62}$$

其对应泛函为

$$F(u)=\int_{\Omega}[\nu(\nabla u)^2/2+gu]d\Omega \tag{6.63}$$

对电磁法中常见的非齐次 Helmholtz 方程

$$\nabla\cdot(\nu\nabla u)+k^2u=g \tag{6.64}$$

其对应泛函为

$$F(u)=\frac{1}{2}\int_{\Omega}[\nu(\nabla u)^2-k^2u^2+2gu]d\Omega \tag{6.65}$$

将(6.46)中标量波动方程改写为

$$\nabla^2u+k^2u=0 \qquad k=\omega/c \tag{6.66}$$

其对应的泛函为

$$F(u)=\frac{1}{2}\int_{\Omega}[(\nabla u)^2-k^2u^2]d\Omega \tag{6.67}$$

对于使用微波段的电磁法(地质雷达)，其似稳电磁场方程可由(6.33)及(6.34)对 t 作 Fourier 变换得出

$$\nabla\times(\nabla\times\boldsymbol{H}/\varepsilon)=\omega^2\mu\boldsymbol{H}, \qquad \nabla\times(\nabla\times\boldsymbol{E}/\mu)=\omega^2\varepsilon\boldsymbol{E} \tag{6.68}$$

相应的泛函为

$$F(\boldsymbol{H})=\frac{1}{2}\int_{\Omega}[\mu\boldsymbol{H}^2-(\nabla\times\boldsymbol{H})^2/\omega^2\varepsilon]d\Omega \tag{6.69}$$

上述各泛函极小化与介质的灵活剖分相结合，就是有限元法的科学思想。介质剖分选择的单元以单纯形最为灵活。所谓单纯形，在平面上为三角形，三维空间为四面体。下面以平面问题为例介绍剖分的基本计算，重点在于用单纯形网格逼近任意复杂的边界与介质。

(二)一阶三角形单元

在用有限元法解二维 Laplace 方程时，研究区域用三角形单元剖分。由于三角元以公共边界及顶点连接成网，势的分布在穿过单元时保持连续。图6.2表示一个典型的三角元，并假定元内势可用线性(一阶)方程表为

$$V=a+bx+cy \tag{6.70}$$

因此，沿三角元边缘势 V 可以由相应两角点势值线性内插而来，如果两个三角元共用一条边，则位势在跨单元时保持连续。

为求(6.70)中的系数，设三个顶点上势为 V_1，V_2 和 V_3，代入后有

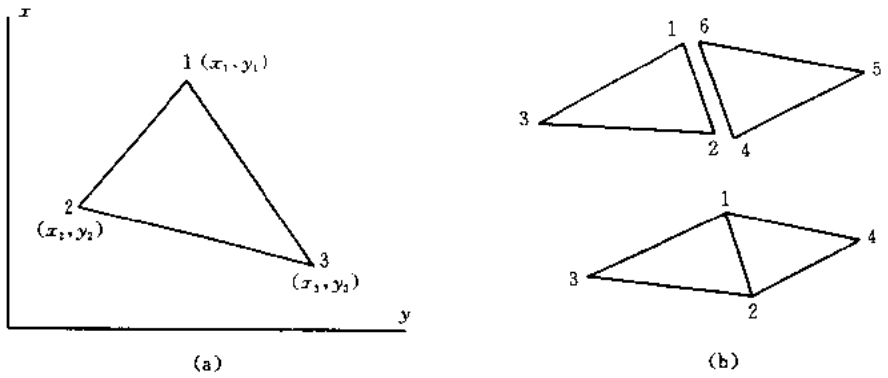


图 6.2　典型的三角形单元(a)和元的合并(b)

顶点1—2—3和4—5—6为合并前元的顶点号，合并后重编为1—2—3—4

$$\begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix} = \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} \tag{6.71}$$

用 Cramer 准则解线性方程组(6.71)，可求得系数 a、b 和 c 的表达式。将它们代回(6.70)便得三角元内任一点位势的一阶近似式

$$V = [1 \quad x \quad y] \begin{vmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{vmatrix}^{-1} \begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix} = \sum_{i=1}^{3} V_i \alpha_i(x, y) \tag{6.72}$$

式中的系数为

$$[\alpha_1 \quad \alpha_2 \quad \alpha_3] = [1 \quad x \quad y] \begin{vmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{vmatrix}^{-1} = [1 \quad x \quad y] \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \cdot \frac{1}{A} \tag{6.73}$$

式中 A 为三角元的面积。用伴随矩阵法求出 $\{b_{ij}\}$，便有

$$\left.\begin{aligned} \alpha_1 &= \left(\begin{vmatrix} x_2 & y_2 \\ x_3 & y_3 \end{vmatrix} - \begin{vmatrix} 1 & y_2 \\ 1 & y_3 \end{vmatrix} x + \begin{vmatrix} 1 & x_2 \\ 1 & x_3 \end{vmatrix} y \right) \Big/ 2A, \\ \alpha_2 &= \left(- \begin{vmatrix} x_1 & y_1 \\ x_2 & y_3 \end{vmatrix} + \begin{vmatrix} 1 & y_1 \\ 1 & y_3 \end{vmatrix} x - \begin{vmatrix} 1 & x_1 \\ 1 & x_3 \end{vmatrix} y \right) \Big/ 2A, \\ \alpha_3 &= \left(\begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} - \begin{vmatrix} 1 & y_1 \\ 1 & y_2 \end{vmatrix} x + \begin{vmatrix} 1 & x_1 \\ 1 & x_2 \end{vmatrix} y \right) \Big/ 2A; \end{aligned}\right\} \tag{6.74}$$

注意其中

$$\alpha_i(x_j, y_j) = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases} \tag{6.75}$$

在求出三角元内位势近似式(6.72)之后，考虑单个元内位势能的极小问题，此势能表为

$$F^{(e)} = 0.5 \int |\nabla V|^2 \mathrm{d}S = 0.5 \sum_{i=1}^{3} \sum_{j=1}^{3} V_i \int \nabla \alpha_i \cdot \nabla \alpha_j V_j \mathrm{d}S \tag{6.76}$$

其中

$$\nabla V(x,y)=\nabla\sum_{i=1}^{3}V_i\alpha_i=\sum_{i=1}^{3}V_i\nabla\alpha_i(x,y) \tag{6.77}$$

上角标(e)表示对一个单元。将(6.76)写成矩阵形式有

$$F^{(e)}=0.5\boldsymbol{V}^{\mathrm{T}}\boldsymbol{S}^{(e)}\boldsymbol{V},\text{其中 }S_{ij}^{(e)}=\int\nabla\alpha_i\cdot\nabla\alpha_j\mathrm{d}S \tag{6.78}$$

将(6.74)代入(6.78)便有

$$\left.\begin{aligned}S_{11}^{(e)}&=\frac{1}{4A}[(y_2-y_3)^2-(x_3-x_2)^2]\\S_{12}^{(e)}&=\frac{1}{4A}[(y_2-y_3)(y_3-y_1)+(x_3-x_2)(x_1-x_3)]\end{aligned}\right\} \tag{6.79}$$

其它的矩阵元也可用类似方法推得，可以由顶点的位势和坐标计算出来。至此，对单个元的近似已经完成，接下去要把所有单元一个个装配起来求解。

(三)三角元的组合与 Laplace 方程的数值解

根据最小势能准则，要把整个研究区域的势能极小化，而整个区域的势能为单个三角元势能之总和。由于上述三角元剖分具有连续的势能，知道两个元如何连接便可以将所有元逐次组合起来。此时，将一对元在连接前的顶点位势写成以下列向量

$$\boldsymbol{V}_{\mathrm{dis}}^{\mathrm{T}}=[V_1,V_2,V_3,V_4,V_5,V_6] \tag{6.80}$$

下脚标 dis 表示组合前不连接的二角元。对应这两个元的未组合能量写为矩阵

$$\boldsymbol{W}-0.5\boldsymbol{V}_{\mathrm{dis}}^{\mathrm{T}}\boldsymbol{S}_{\mathrm{dis}}\boldsymbol{V}_{\mathrm{dis}},\text{其中 }\boldsymbol{S}_{\mathrm{dis}}=\begin{pmatrix}S_{11}^{(1)}&S_{12}^{(1)}&S_{13}^{(1)}&&&\\S_{21}^{(1)}&S_{22}^{(1)}&S_{23}^{(1)}&&&\\S_{31}^{(1)}&S_{32}^{(1)}&S_{33}^{(1)}&&&\\&&&S_{44}^{(2)}&S_{45}^{(2)}&S_{46}^{(2)}\\&&&S_{54}^{(2)}&S_{55}^{(2)}&S_{56}^{(2)}\\&&&S_{64}^{(2)}&S_{65}^{(2)}&S_{66}^{(2)}\end{pmatrix} \tag{6.81}$$

按照图6.2所示的顶点连接后的顶点标号，我们知道这一对三角元在连接前后满足以下关系

$$\boldsymbol{V}_{\mathrm{dis}}\equiv\begin{pmatrix}V_1\\V_2\\V_3\\V_4\\V_5\\V_6\end{pmatrix}=\begin{pmatrix}1&&&\\&1&&\\&&1&\\&1&&\\&&&1\\1&&&\end{pmatrix}\begin{pmatrix}V_1\\V_2\\V_3\\V_4\end{pmatrix}\equiv\boldsymbol{C}\boldsymbol{V}_{\mathrm{con}} \tag{6.82}$$

下标 con 注记已联接，对应连接后的能量变为

$$\boldsymbol{W}=0.5\boldsymbol{V}_{\mathrm{con}}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{V}_{\mathrm{con}}=0.5\boldsymbol{V}_{\mathrm{con}}^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{S}_{\mathrm{dis}}\boldsymbol{C}\boldsymbol{V}_{\mathrm{con}} \tag{6.83}$$

这样，连续近似的势能分布被表示与元顶点位势向量有关的二次型。

Laplace 方程的有限元近似解要将连接网中的势能极小化。记 k 为网格节点(连接后多个三角形的顶点)的编号，则势能极小化为

$$\text{令}\qquad\frac{\partial\boldsymbol{W}}{\partial V_k}=0. \tag{6.84}$$

在求解的边值问题中，在给定研究区域的边界上位势(或其导数)是给定的。假如我们在对网格节点编号时首先编区域内部的号，最后对给定位势的边界点编号，则(6.84)可改写为

$$\frac{\partial W}{\partial V_k}=\frac{\partial}{\partial [V_f]_k}[V_f \quad V_p]\begin{bmatrix}S_{ff} & S_{fp}\\ S_{pf} & S_{pp}\end{bmatrix}\begin{bmatrix}V_f\\ V_p\end{bmatrix}=0, \tag{6.85}$$

其中下角标 f 表示区域内的节点，下脚标 p 表示边界上预给定位势的节点（不必对它们求导），矩阵 S 由(6.83)式给出。对(6.85)作微分便有

$$S_{ff}V_f=-S_{fp}V_p \tag{6.86}$$

解线性矩阵方程(6.86)便可求出研究区域内各节点的位势 V_f，此方程有唯一解，其精度取决于三角元的尺度。

在求解比 Laplace 方程更复杂的问题时，上述有限元方法同样可用于处理有关空间微分算子的逼近。在有限元法中并没有用到高于微积分的数学工具，但其中正确的思路与巧妙的计算技术相配合的思维，在不少工程计算中都可借鉴。

四、几种深入的波动问题

基本方程式(6.1)和变密度波动方程(6.45)反映了波动问题的普遍性规律，但对复杂地球的研究还是不够的。本节将考虑以下三种复杂情况，即波的衰减与色散、各向异性介质与多相介质中波的传播问题。

（一）波的衰减与色散

实际地震记录与合成计算的地震记录在振幅与波形上常常有很大的差别，对陆上采集的反射地震记录更是如此。Sheriff 和 Geldart 曾列出十余种影响反射地震波振幅的因素[14,15]，其中包括震源强度及其与地球的耦合关系、球面扩散、界面反射系数与入射角的关系，层间多次波与转换波，反射体的曲率和粗糙程度、介质的吸收与散射，检波器排列方向及其与大地的耦合，检波器灵敏度与一致性，接收仪器性能、环境噪音等。这些影响因素无法在地震合成中一一考虑，因此，用数据拟合差取极小为准则的地震反演也经常遇到困难。在以下讨论中，我们仅考虑岩石介质的吸收和频散。

大量实验表明，岩石对地震波能量的吸收是随频率增加而增加的。由于在传播路程中高频成分吸收较快，吸收与频散是同时出现和互为因果的。这种互为因果的关系可用 Kramers-krönig 线性关系式近似描述[2,5]。设一个线性因果系统的复传递函数的实部与虚部分别为 $R(\omega)$ 及 $X(\omega)$，则此关系式表为

$$\left.\begin{aligned}X(\omega)&=\frac{1}{\pi}P\int_{-\infty}^{\infty}\frac{R(\tau)}{\tau-\omega}\mathrm{d}\tau+X(\infty)\\ R(\omega)&=R(\infty)-\frac{1}{\pi}P\int_{-\infty}^{\infty}\frac{x(\tau)}{\tau-\omega}\mathrm{d}\tau\end{aligned}\right\} \tag{6.87}$$

其中 P 为积分主值。

基本方程(6.1)中已经包含了波的吸收（左边第三项），其中衰减系数 α 与岩石介质的品质因数近似成反比，见(6.12)式。于是，搞清地震波衰减的关键在于了解品质因数 Q 的性状，它与吸收的物理机制有关。

从室内物理实验出发，早期的研究者认为地震波的吸收主要与岩石内部晶格的位错运动、岩石颗粒表面或裂隙面上的滑动摩擦及热弹性张弛有关[16]。与此对应的介质吸收模型有两种。一种是由 Strick 和 Kjartanson 给的所谓恒 Q 模型。

$$M(\omega)=M_c\left|\frac{\omega}{\omega_c}\right|^{2\nu}e^{(i\pi\nu\mathrm{sgn}(\omega))} \tag{6.88}$$

$$c(\omega)=c_c\left|\frac{\omega}{\omega_c}\right|^{\nu} \tag{6.89}$$

$$\alpha(\omega)=k\left(\frac{\omega}{2\pi}\right)^{1-\nu} \tag{6.90}$$

式中 ω_c 为低频段的参考频率；M_c 和 c_c 分别为对应 ω_c 的弹性模量和相速度；ν 为在0和1之间变化的常数，与岩石构造有关；$\mathrm{sgn}(\omega)$ 表示 ω 的符号函数；k 为比例因子。

另一种近似恒 Q 模型是 Futterman 利用 Kramers-Krönig 关系式推导出的，它的一组常用公式是

$$c(\omega)=c_c\left[1-\frac{1}{\pi Q_c}\ln\frac{\gamma\omega}{\omega_c}\right]^{-1},\omega\geqslant\omega_c \tag{6.91}$$

$$\alpha(\omega)=\frac{\omega}{2Q_cc_c}\left[1-\exp\left(-\frac{\omega}{\omega_c}\right)\right],\omega\geqslant\omega_c \tag{6.92}$$

$$Q^{-1}(\omega)=Q_c^{-1}\left[1-\frac{1}{\pi Q_c}\ln\frac{\gamma\omega}{\omega_c}\right],\omega\geqslant\omega_c \tag{6.93}$$

式中 ω_c 为接近零的参考频率；c_c 和 Q_c 为对应 ω_c 的参考相速度和品质因数；$\ln\gamma$ 为欧拉常数，即 $\gamma=1.781\ 072\ 418$。

由于沉积岩中 Q 值谐振现象的发现，用上述模型来描述岩石的粘弹性便显得不够了。于是1983年 Meissner 和 Theilen 又提出一种模型，他们结合 Boit 先前对沉积岩吸收机制的理论分析[18,19]，用以下标准线性固体的应力-应变关系

$$\sigma+\tau_\sigma\dot{\sigma}=M(i\omega)(\varepsilon+\tau_\varepsilon\dot{\varepsilon}), \tag{6.94}$$

导出了谐振 Q 的基本关系式

$$Q^{-1}(\omega)=Q_0^{-1}\left(\frac{2\omega\tau_0}{1+\omega^2\tau_0^2}\right) \tag{6.95}$$

式中 τ_σ 为应力弛豫时间；τ_ε 为应变弛豫时间；M 为复弹性模量；Q_0 为谐振频率 ω_0 处的品质因数；τ_0 为谐振弛豫时间。Q_0 及 τ_0 等参数可以根据岩石孔隙度、渗透率、孔隙流体的密度和粘滞系数计算。

由(6.95)式给出的 Q 模型是不完善的，当 $\omega=0$ 时，它给出 $Q(0)$ 为无穷大的结果，与实验不符。根据上述关于岩石中引起地震波吸收的各种机制的分析，我认为波能量的损失可分为两类：可以用恒 Q 模型描述的机制(滑动摩擦等)和与谐振有关的机制(孔隙流体运动)。由于这两类机制表示岩石中不同的相(固相和流体相)的作用，因此两者之间的相互作用在近似表示时可以忽略，即可以用以下线性叠加表示岩石的品质因数

$$Q^{-1}(\omega)=Q_c^{-1}+Q_r^{-1}\frac{2\omega\tau_0}{1+\omega^2\tau_0^2}, \tag{6.96}$$

式中 Q_c 为在谐振频段之外某一参考频率 ω_c 处的常数品质因数值；Q_r 为单纯由谐振引起的品质因数的峰值。在谐振频率 ω_0 处，令 $\tau_0=1/\omega_0$，代入(6.96)式有

$$Q_r^{-1}=Q_0^{-1}-Q_c^{-1}=\frac{Q_c-Q_0}{Q_0Q_c},$$

式中 $Q_0\equiv Q(\omega_0)$ 为谐振频率处的 Q 值，上式表示 $\omega=\omega_0$ 时 $Q^{-1}(\omega_0)$ 为恒 Q_c^{-1} 和谐振 Q_r^{-1} 的叠加。令

$$\Delta Q = Q_c - Q_0 \geqslant 0, \tag{6.97}$$

则(6.96)改写为

$$Q(\omega) = Q_c\left[1+\frac{2\Delta Q}{Q_0}\frac{\omega\tau_0}{1+\omega^2\tau_0^2}\right]^{-1} \tag{6.98}$$

当 $Q_c^{-1}=0$时有 $\Delta Q=0$,(6.98)式退化为恒 Q 模型。对于考虑有谐振吸收机制的沉积岩,应选 $\omega_c \ll \omega_0$。

在 $Q \gg 1$的情况下,由[2]的推导可得复弹性模量的实部

$$M'(\omega) \simeq M'_c\left[1+\frac{2}{\pi Q_c}\ln\frac{\omega}{\omega_c}+\frac{4\Delta Q}{\pi Q_0 Q_c}\arctan\frac{(\omega-\omega_c)\tau_0}{1+\omega\omega_c\tau_0^2}\right] \tag{6.99}$$

复弹性模量的虚部

$$M''(\omega) \simeq \frac{\pi\omega \mathrm{d}M'(\omega)}{2\mathrm{d}\omega} = M'_c\left\{\frac{1}{Q_c}+\frac{2\omega\tau_0\Delta Q}{Q_0 Q_c}\cdot\frac{1+\omega_c^2\tau_0^2}{1+(\omega^2+\omega_c^2)\tau_0^2+\omega^2\omega_0^2\tau_0^4}\right\} \tag{6.100}$$

式中 M'_c 为对应参考频率 ω_c 处复弹性模量的实部(见附录)。相速度为

$$c(\omega)-c(\omega_c)\left[1+\frac{2\ln(\omega/\omega_c)}{\pi Q_c}+\frac{4\Delta Q}{\pi Q_0 Q_c}\arctan\frac{(\omega-\omega_c)\tau_0}{1+\omega\omega_c\tau_0^2}\right]^{1/2} \tag{6.101}$$

当 $\Delta Q=0$时,容易证明(6.101)式退变为(6.91)式,即当谐振不存在时变为 Futterman 模型;当 $Q_c=\infty$时,(6.96)式变为(6.95)式。因此,由(6.98)～(6.101)表示岩石粘弹性的谐振 Q 模型包含了前人提出的各种模型。

在文献[2]中还可看到,用粘弹性谐振 Q 模型计算出的 Q 随频率的变化与实验数据很符合。利用本节给出的计算相速度和衰减系数的公式代入基本方程(6.1)便可以合成各种频带下的地震图,尤其是将地震图与全波声测井的记录(频段在10kHz 以上)进行对比,进行声测井中的声速与地震波速(频带在10～200Hz 内)的匹配。

(二)地震波在各向异性介质中的传播

地下的许多介质具有各向异性,它们往往是人类感兴趣的探测目标。例如,裂隙型的油气储层、工程基础中的剪切带、地壳中的韧性剪切带和有流体活动的裂隙带等等。探讨地震波在各向异性介质中的传播对探测这些对象是很有意义的。

各向异性指的是在空间任一点上,介质的物理性质取决于其测量的方向。在这种介质中,质点振动的方向(及射线方向)可能与应力方向偏离(偏振),观测这种偏振现象,是了解测量地段介质各向异性的一种方法。此外,在下面要提到的横观各向异性介质中,当一组任意偏振的横波入射时,由于此时介质沿不同方向有不同的波速,因此穿过此介质时旅行时间也不同.它将沿对称轴分裂为两组方向正交的横波,这种现象称为横波分裂。在三分量的垂直地震剖面(VSP)记录中,常可以记录到横波分裂的现象,并用来揭示裂隙带的存在及区域构造应力方向[20]。

不同于各向同性介质,各向异性介质中应力与应变的关系用以下广义 Hooke 定律表示

$$\begin{pmatrix}\sigma_{xx}\\ \sigma_{yy}\\ \sigma_{zz}\\ \sigma_{xy}\\ \sigma_{yz}\\ \sigma_{zx}\end{pmatrix} = \begin{pmatrix}C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16}\\ C_{12} & C_{22} & C_{23} & C_{24} & C_{25} & C_{26}\\ C_{13} & C_{23} & C_{33} & C_{34} & C_{35} & C_{36}\\ C_{14} & C_{24} & C_{34} & C_{44} & C_{45} & C_{46}\\ C_{15} & C_{25} & C_{35} & C_{45} & C_{55} & C_{56}\\ C_{16} & C_{26} & C_{36} & C_{46} & C_{56} & C_{66}\end{pmatrix}\begin{pmatrix}\varepsilon_{xx}\\ \varepsilon_{yy}\\ \varepsilon_{zz}\\ \varepsilon_{xy}\\ \varepsilon_{yz}\\ \varepsilon_{zx}\end{pmatrix} \tag{6.102}$$

其中 σ 和 ε 分别为应力和应变，下标 ij 表示 i 方向的力作用于垂直 j 方向的平面上。矩阵 $\boldsymbol{C}$ 称为刚度张量，由于对称性，它有21个元，其中18个为独立的元。

如果各向异性在两个正交方向(如 x 与 y 轴)上相同，而在第三个正交方向(如 z 轴，它为对称轴)上不同，此类介质称为横观各向异性介质，它是最简单的各向异性介质。此时，刚度张量只有5个独立的元素，本构方程(6.102)简化为

$$\begin{pmatrix}\sigma_{xx}\\ \sigma_{yy}\\ \sigma_{zz}\\ \sigma_{xy}\\ \sigma_{yz}\\ \sigma_{zx}\end{pmatrix}=\begin{pmatrix}\lambda_{=}+2\mu_{=} & \lambda_{=} & \lambda_{\perp} & 0 & 0 & 0\\ \lambda_{=} & \lambda_{=}+2\mu_{=} & \lambda_{\perp} & 0 & 0 & 0\\ \lambda_{\perp} & \lambda_{\perp} & \lambda_{\perp}+2\mu_{\perp} & 0 & 0 & 0\\ 0 & 0 & 0 & \mu_{=} & 0 & 0\\ 0 & 0 & 0 & 0 & \mu^{*} & 0\\ 0 & 0 & 0 & 0 & 0 & \mu^{*}\end{pmatrix}\begin{pmatrix}\varepsilon_{xx}\\ \varepsilon_{yy}\\ \varepsilon_{zz}\\ \varepsilon_{xy}\\ \varepsilon_{yz}\\ \varepsilon_{zx}\end{pmatrix} \tag{6.103}$$

下标=表示水平方向，⊥表示 z 方向，对于横观各向异性弹性介质，在二维情况下波动方程写为

$$\frac{\partial^2 u}{\partial t^2}=\alpha_{=}^{2}\frac{\partial^2 u}{\partial x^2}+\beta^{*2}\frac{\partial^2 u}{\partial z^2}+(\alpha^{*2}-\beta^{*2})\frac{\partial^2 w}{\partial z\partial x} \tag{6.104}$$

$$\frac{\partial^2 w}{\partial t^2}=(\alpha^{*2}-\beta^{*2})\frac{\partial^2 u}{\partial x\partial z}+\beta^{*2}\frac{\partial^2 w}{\partial x^2}+\alpha_{\perp}^{2}\frac{\partial^2 w}{\partial z^2} \tag{6.105}$$

其中 u 为波场垂直分量，w 为水平分量，它们是相互耦合的。式中的系数与刚性张量的元的关系为

$$\begin{aligned}&\alpha_{=}^{2}=(\lambda_{=}+2\mu_{=})/\rho,\alpha^{*2}=(\lambda_{\perp}+2\mu^{*})/\rho,\\&\alpha_{\perp}^{2}=(\lambda_{\perp}+2\mu_{\perp})/\rho,\beta^{*2}=\mu^{*}/\rho\end{aligned} \tag{6.106}$$

它们与对应 P 波、SH 波和 SV 波速度的关系为

$$v_{P,Z}=\alpha_{\perp},v_{SH,Z}=\beta^{*};v_{P,xy}=\alpha_{=},v_{SV,xy}=\beta^{*},v_{SH,xy}=\sqrt{\mu_{=}/\rho} \tag{6.107}$$

式中下标后部分表示传播方向，即 Z 为沿 Z 轴传播，xy 为沿 xy 平面传播。在其它方向上波速为所有常数的复杂函数。利用有限差分或有限元等等方法解方程(6.104)及(6.105)，可以做二维横观各向异性介质中的地震合成。

在电磁场中，某些岩石的电导率 σ 与介电系数 ε 也与测量方向有关，因此可用张量表示为

$$(\sigma)=\begin{pmatrix}\sigma_{xy} & 0 & 0\\ 0 & \sigma_{xy} & 0\\ 0 & 0 & \sigma_{z}\end{pmatrix};(\varepsilon)=\begin{pmatrix}\varepsilon_{xy} & 0 & 0\\ 0 & \varepsilon_{xy} & 0\\ 0 & 0 & \varepsilon_{z}\end{pmatrix} \tag{6.108}$$

为计算方便，它们可以合并为复电导率张量

$$(s)=\begin{pmatrix}\sigma_{=} & 0 & 0\\ 0 & \sigma_{=} & 0\\ 0 & 0 & \sigma_{\perp}\end{pmatrix};\text{其中}\begin{matrix}\sigma_{=}=\sigma_{xy}+\mathrm{i}\varepsilon_{xy}\omega\\ \sigma_{\perp}=\sigma_{z}+\mathrm{i}\varepsilon_{z}\omega\end{matrix} \tag{6.109}$$

将 Maxwell 方程(6.25)对谐波推广为各向异性介质情况，有

$$\mathrm{curl}\boldsymbol{H}-(s)\boldsymbol{E},\mathrm{curl}\boldsymbol{E}=-\mathrm{i}\mu_0\omega\boldsymbol{H} \tag{6.110}$$

引入矢位 $\boldsymbol{A}$ 和标位 Ψ，得

$$\boldsymbol{H}=\mathrm{curl}\boldsymbol{A},\boldsymbol{E}=-\mathrm{i}\mu_0\omega\mathbf{A}-\mathrm{grad}\Psi \tag{6.111}$$

令

$$\mathrm{div}\boldsymbol{A}+\sigma_z\Psi=0 \tag{6.112}$$

便得到

$$\nabla^2A_x-(\sigma_x-\sigma_{xy})\frac{1}{\sigma_z}\frac{\partial}{\partial x}\mathrm{div}A-\frac{\sigma_z}{\sigma_{xy}}\gamma^2A_x=0 \tag{6.113}$$

其中

$$\nabla^2A_x-\gamma^2A_x=0,\nabla^2A_y-\gamma^2A_y=0;\gamma^2=\mathrm{i}\mu_0\omega(\sigma_x+\mathrm{i}\varepsilon_x\omega) \tag{6.114}$$

此两式表明了各向异性参数对电磁场的影响。

(三)多相介质与地震岩性模型

许多岩石是由固体矿物骨架和孔隙流体组成的混合物,其物理性质常因不同的孔隙流体可产生重要变化。就地震波传播而言,孔隙流体对波的衰减、色散和各向异性都有一定影响,因此考虑多相介质模型中波的传播可能导致更精确的地震或声波记录的合成。反过来,根据地震波传播的特征变化推测地层中的流体成分(如油、气、水)及裂隙性状(即地震岩性)对指导油气勘探或地壳探测亦有重要意义。

要注意的是地震波长与孔隙或裂隙的尺度问题。岩石中的孔隙与裂隙的尺度一般不大(在10cm 以下),对地面地震勘探而言,主波长一般为几十米,因此地震波反映的信息为几米以上地下物质的平均信息。它虽然含有孔隙流体的作用,但这种作用是以宏观影响出现的,如表现为频谱的变化,振幅的衰减加快等等。在全波声测井中,波长可达几个毫米,这时孔隙流体的作用就大大加强,不考虑多相介质就不可能全面了解声波的传播特征。

多相介质中地震波的研究早先是针对油气储层(如砂岩)提出来的,近年来大陆科学钻探表明,不仅在浅层的沉积盆地中,在深层的结晶岩石中,裂隙和其中的流体都普遍存在,而且地壳中的流体对许多地质作用都起到重要作用[21]。因此,多相介质中的地震波传播和流变学研究将会受到更广泛的关注。

多相介质中地震波的传播是一个很复杂的物理问题,不可能在反演专著中作详细讨论。以下我们仅就有关的物理参数和多相介质的特点作一介绍,详细讨论参见文献[18,19,21~23]。我们的讨论限定于油气勘探有关的沉积岩介质。

在沉积岩中,要区分出以下三个有关连的术语:

(1)固体矿物颗粒(在岩石中被胶结),用下标"s"表示;

(2)岩石格架:它包括矿物颗粒和空隙,但不含流体,以下用下标"m"注记;

(3)流体,用下标"f"表示。

相应这三种对象具有不同的密度 ρ、体积模量 k、切变模量 μ。由于流体中没有剪切运动,因此对岩石格架和岩石本身剪切模量是相同的。岩石的密度与孔隙度 ϕ 的关系为

$$\rho=\phi\rho_f+(1-\phi)\rho_s \tag{6.115}$$

假定在格架与流体之间不存在相对运动(这是很粗糙的 Gassmann 模型的情况),即流体粘度很大或者波长比孔隙尺度大得多时,由体积模量定义有

$$k=-\Delta p(\Delta V/V);\text{其中 }\Delta p=\Delta p_m+\Delta p_f,\Delta V=\Delta V_s+\Delta V_f \tag{6.116}$$

其中 p 为压力,V 为体积。对固体颗粒,压挤同时来自其它固体颗粒和流体,因此骨架体积变化

$$\Delta V_f/V=-\Delta p_f/k_f;\text{即},\Delta V_f=-\phi V\Delta p_f/k_f \tag{6.117}$$

$$\Delta V_m=-V_s\Delta p_f/k_s-V\Delta p_m/k_s;\text{其中 } k_s=k_m,V_s=(1-\phi)V \tag{6.118}$$

将(6.117、6.118)代入(6.116)有

$$\frac{\Delta V}{V}=-\left(\frac{\phi}{k_f}+\frac{1-\phi}{k_s}\right)\Delta p_f-\Delta p_m/k_s \tag{6.119}$$

从另一角度看,对稳态情况由定义我们有

$$\Delta V_m/V=-\Delta p_m/k_m \text{ 和 } \Delta V_f/V=-\Delta p_f/k_s \tag{6.120}$$

后一式说明由流体压力引起孔隙体积的压缩,也引起骨架体积的压缩。因此,体积压缩的总和为

$$\Delta V/V=-\Delta p_f/k_s-\Delta p_m/k_m \tag{6.121}$$

从(6.116)、(6.119)和(6.121)可以解出岩石的体积模量

$$k=k_m-\frac{(1-k_m/k_s)^2}{[\phi/k_f+(1-\phi)/k_s+k_m/k_s]^2} \tag{6.122}$$

据此,可计算多相介质中的张性模量 M 及纵横波波速

$$M=k+4\mu/3;c_P=\sqrt{M/\rho}\text{及 } c_S=\sqrt{\mu/\rho} \tag{6.123}$$

由于假定孔隙流体与格架之间不存在相对运动(无摩擦),上述多相介质模型不能说明地震波的衰减特性。为改善这种粗糙的模型,必须引入两个新参数,流体粘滞系数 η 和岩石渗透率 κ。今 v 为孔隙流体的运动速度,由定义有

$$v=\frac{\kappa\partial p}{\eta\partial x} \tag{6.124}$$

其中 $p(x,t)$ 为流体和格架之间相互作用产生的压力波场(在 Biot 模型中称为第二类压力波)。此压力波的变化速率为

$$\frac{\partial p}{\partial t}=-k_e\frac{\partial v}{\partial x},\text{或}\frac{\partial v}{\partial x}=-k_e^{-1}\frac{\partial p}{\partial t} \tag{6.125}$$

其中的系数称为有效体积模量,即

$$k_e=\left(\frac{\phi}{k_f}+\frac{1-\phi}{k_e}-\frac{k_m}{k_s^2}\right)^{-1}\cdot\frac{M_m}{M} \tag{6.126}$$

将(6.124)对 x 求导并代入(6.125)得

$$\frac{\partial^2 p}{\partial x^2}=\frac{\eta}{\kappa k_e}\frac{\partial p}{\partial t} \tag{6.127}$$

此式的解为第二类压力波,形如

$$p(x,t)=Ae^{-\gamma x}e^{i\omega t},\text{其中 } \gamma=(i\omega\eta/\kappa k_e)^{1/2} \tag{6.128}$$

其中 A 为压力波的幅度。将(6.128)代入(6.124)得

$$v(x,t)=\frac{\kappa\gamma}{\eta}Ae^{-\gamma x}e^{i\omega t} \tag{6.129}$$

便可求出孔隙流体的运动速率。

在本章第一节推导中,曾说过复波速的实部为相速度,而虚部代表波的衰减项。在(6.128)中出现了复衰减系数,它正好相反,实部为波的衰减,而虚部与频散的波速有关

$$\gamma=\beta_{p2}-i\omega/c_{p2} \tag{6.130}$$

用以上公式可得第二类压力波的衰减系数和波速为

$$\beta_{p2}=(\omega\eta/2\kappa k_e)^{1/2},c_{p2}=(2\omega\kappa k_e/\eta)^{1/2} \tag{6.131}$$

而第二类压力波的波阻抗为 p 与 γ 的比值。第二类波由于衰减很快，在地面地震勘探时一般接收不到，仅需在声测井中考虑。

以上述多相介质中的物理模型为基础，Biot 发展了多相介质中弹性波传播的理论[18,19]。用此理论可预言本节开头提到的地震波的谐振吸收。然而，在 Biot 模型中假定，固体颗粒为球形或椭球形，孔隙粘性流体可在任意长的连通孔隙中流动。对于裂隙来说这种模型并不能满足。Murphy 等人发现，在固体颗粒之间的细小通道或缝隙对地震波传播有相当重要的影响[22,23]。当弹性波传播到这些缝隙时它们会张开和压缩，而压缩时会将流体挤出缝隙到周围的孔隙中，造成地震波能量的可观的衰减。他们开始用多面体固体颗粒模型来描述岩石，多面体的界面表示长而狭窄的缝隙。这种模型可用于低孔隙率与低渗透率的岩石介质中。

多相介质中的电磁波也是频散的，此时复电导率也常用来描述它随频率的变化[24~27]。此外，在激发极化法中，欧姆定律被推广为

$$j=\gamma E,\gamma(\omega)=\sigma(\omega)-i\omega\varepsilon \tag{6.132}$$

其中 γ 也是复电导率。比较(6.132)和(6.130)我们可以再次看到地震与电磁方法在数学处理上的相似性。

(四)小结

本章的重点在于讨论各种地球物理探测方法正问题的共性，以及类同的数学处理方法，为反演打下基础。除了在第一节中讨论的通用方程之外，在第四节中也强调了波传播问题数学上的相似性。在第二节和第三节中讨论了解正问题两种常用计算方法：有限差分法与有限元法。比这两种方法更有前途的是新近出现的格子气自动机模型[28,29]。格子气自动机是一个由微观粒子构成的离散系统，其中粒子运动满足格子的 Boltzmann 方程，而粒子流的宏观统计平均量服从流体力学中的 Navier-Stokes 方程。格子气模型中粒子的分布存在局部平衡状态，而模型的宏观行为是由这种局部平衡态连接而成的。当此系统中粒子处于宏观速度为零的平衡态时，模型中粒子流密度扰动的传播满足声波方程。根据这样的理论分析进行波动模拟，可以揭示不同尺度和不同层次的波动现象，以及它们的相互联系，从而对很复杂的介质中波的衍射等现象作出精确的数学仿真。

正演问题的求解是反问题求解的重要基础之一。正问题解得好已走完了反演过程的一小半，因此反演研究必须首先从正演开始。当前，地震反演的一个重要难点就在地震合成记录的振幅与实际记录振幅不拟合上，这里不仅仅有理论与计算方法的问题，也有采集中的问题，需要考虑更多的因素。

第七章　地球物理数据处理

地球物理学家毕生与数据打交道。数据是反演的输入，而且数据处理本身也往往是广义的反问题。因此，本章有必要讨论地球物理数据处理。当然，不同的地球物理方法有自己特殊的数据处理问题，无法在本书中一一讨论。我们还是讨论地球物理数据处理中共同的问题。

地球物理数据处理共同的目的有以下三方面。

(1)使数据集尽量符合简化了的数学模型。例如，假定观测在平面上进行，就要通过静校正把观测点换算到基准面上。为消除炮与接收器与地下介质耦合得不均匀，要用道均衡等技术调整每个地震道的幅度，等等。

(2)提高数据的信噪比，如各种数字滤波器。

(3)提高信号的分辨率(包括时间分辨率和空间分辨率)，如反褶积、地震偏移等等。

一、数据与数据模型

(一)地球物理数据集

地球物理数据是地球物理信号与噪音的混合。地球物理信号 f 指的是某种地球物理场，它是满足指定偏微分方程的连续函数，如(6.1)式中的 u。地球物理噪音 n 包括环境噪音 n_a，观测误差 n_o；数字记录的量化误差 n_d；和所谓的数学误差 n_m，它是由不够精确的数学模型引起的(如用声波方程代替弹性波方程，又如当介质具有各向异性时用弹性波方程描述信号)。迄今为止，人们总是把信号和噪音作为各自独立的量来处理，因此认为数据是它们的和

$$d=f+n_a+n_o+n_d-n_m \tag{7.1}$$

在第四章中已经说过，给定方程之后，所有可能的地球物理场 f 组成一个 Hibert 空间，称为理想数据空间。与此对应的数据 d 组成一个 Hilbert 概率空间，如果我们能给出噪音的概率分布的话。

如果测量是在规则测线(或道)上进行的话，理想的地球物理数据可用梳状函数表为

$$f(nT)=f(x)\cdot\mathrm{comb}(x)=\sum_{n=-\infty}^{\infty}f(x)\delta(x-nT) \tag{7.2}$$

其中 T 为取样间隔，$k_0=2\pi/T$ 称为取样率。离散取样号信号的频谱变为以 k_0 为周期的周期函数。当 T 过大时发生假频，即频率或波数 k 的混淆。当没有假频时，连续信号 $f(x)$ 可用取样定理恢复。取样定理表述为：当信号的频谱 $F(k)$ 满足

$$|k|\geqslant k_c, k_c=k^*/2 \tag{7.3}$$

时

$$f(x)=\sum_{n=-\infty}^{\infty}f(nT)\frac{\sin k_c(x-nT)}{k_c(x-nT)} \tag{7.4}$$

当(7.3)不满足时取样过程产生假频。

由于测线有限长度,有限离散的数据集表为

$$(d_n, n=1, N)=\sum_{n=1}^{N} f(x)\delta(x-nT) \tag{7.5}$$

显然,(7.5)式给出的是一个数据向量 d,它是 f 在 N 维向量空间的投影。

类似地,在二维规则网上观测的理想数据表为

$$f(nx, my)=\sum_{n=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} f(x,y)\delta(x-nT_x)\delta(y-mT_y) \tag{7.6}$$

其中 T_x 和 T_y 分别为 x 及 y 方向的取样间隔,二维规则网上的数据也可按次序重排为数据向量。总之,实际数据集可参照(7.1)写为以下向量形式

$$\boldsymbol{d}_{\text{real}}=\boldsymbol{d}+\boldsymbol{n}_s+\boldsymbol{n}_c+\boldsymbol{n}_d+\boldsymbol{n}_m \tag{7.7}$$

此时$\boldsymbol{d}_{\text{real}}$变为随机向量,或者称为"时间序列"。因此,信息科学中的一个分支"时间序列分析"乃是地球物理数据处理的理论基础之一。

(二)地球物理信号模型

数据模型指的是用某种函数(包括待定系数)来逼近数据集,如果逼近误差很小,则只需要存储少量系数就可以恢复数据集,同时还可以任意内插或适当外推数据集。包含随机误差的统计数据模型将在下一段讨论,下面先讨论确定性的信号模型。

如果信号的频谱向高频的衰减足够快,则常用 m 阶多项式逼近信号。设信号 f 按等间距 Δ 取 $m+1$ 个样,即有$f_i, i=0,\cdots,m$ 用多项式逼近有

$$f(x)=\sum_{k=0}^{m} a_k x^k \tag{7.8}$$

离散化得线性方程组

$$\begin{bmatrix} f(x_0) \\ f(x_1) \\ \vdots \\ f(x_m) \end{bmatrix} = \begin{bmatrix} 1 & x_0 & \cdots & x_0^m \\ 1 & x_1 & \cdots & x_1^m \\ \vdots & \vdots & & \vdots \\ 1 & x_m & \cdots & x_m^m \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix} \tag{7.9}$$

式中的矩阵一般为非奇异的,可以用第二章所述的方法求解系数组$\{a_k\}$。

还可以用指数多项式逼近信号,即令

$$f(x)=\sum_{k=1}^{m} a_k e^{\lambda_k x} \tag{7.10}$$

离散取样时令 $x=i\Delta, i=0,1,\cdots,2m-1$(数据要多一倍),有$f_i \equiv f(i\Delta)$,求$a_k$和$\lambda_k$。

将取样长度记为 $T=2m\Delta$,设

$$z_k=e^{\lambda_k \Delta} \tag{7.11}$$

则有

$$f_i=\sum_{k=1}^{m} a_k e^{\lambda_k i\Delta}=\sum_{k=1}^{m} a_k z_k^i, i=0,1,\cdots,2m-1 \tag{7.12}$$

下面先用 Prony 方法求$\{z_k\}$。记 m 次多项式

$$P_m(z)=(z-z_1)(z-z_2)\cdots(z-z_m)=\sum_{i=0}^{m} \beta_i z^i \tag{7.13}$$

其中$\beta_m=1$。当$P_m=0$ 时(7.13)变为 m 次代数方程

$$z^m+\beta_{m-1}z^{m-1}+\cdots+\beta_1 z+\beta_0=0 \tag{7.14}$$

用数值方法求出它的根$(z_1,\cdots,z_m)$，再由(7.11)有

$$\lambda_k=\frac{1}{\Delta}\ln z_k, k=1,\cdots,m \tag{7.15}$$

代入(7.10)得求解$\{a_k\}$的线性方程组，便得到指数多项式逼近的数学模型。

如何求(7.14)式中的系数$\{\beta_k\}$呢？且看

$$\sum_{i=0}^{m} f_{k+i}\beta_i=\sum_{i=0}^{m}\Big(\sum_{j=1}^{m} a_j z_j^{k+i}\Big)\beta_i=\sum_{j=1}^{m} a_j z_j^k\Big(\sum_{i=0}^{m}\beta_i z_j^i\Big)=0 \tag{7.16}$$

这是由(7.13)式令$P_m=0$时得到的。由于$\beta_m=1$，(7.16)式移项得

$$\sum_{i=0}^{m-1} f_{k+i}\beta_i=-f_{k+m}, k=0,1,\cdots,m-1 \tag{7.17}$$

这便是求解系数$\{\beta_k\}$的线性方程组。因此，计算次序是解方程组(7.17)，然后解代数方程(7.14)，最后解方程组(7.10)。

由(7.10)式表示的指数多项式逼近是一种非线性信号模型。对一般的非线性数据模型，要拟合的函数表为

$$f=f(x;\boldsymbol{a}) \tag{7.18}$$

其中$\boldsymbol{a}$是向量，即$\{a_k, k=1,2,\cdots m\}$，给出某种形式的非线性函数f之后，可以用迭代方法求χ^2优越泛函

$$\chi^2(\boldsymbol{a})=\sum_{i=1}^{n}\left[\frac{f_i-f(x_i;\boldsymbol{a})}{\sigma_i}\right]^2 \tag{7.19}$$

取极小时的系数$\boldsymbol{a}$，其中n为数据个数，σ_i为第i个数据f_i的标准差(对信号$\sigma_i=1$)。步骤是先给出$\boldsymbol{a}$的初始向量，然后在迭代过程中不断修改，直至χ^2下降到接近零为止。与第三章第一节讨论的非线性问题线性化的步骤类似，迭代时要求雅可比矩阵，它的元为

$$\frac{\partial\chi^2}{\partial a_k}=-2\sum_{i=1}^{n}\frac{[f_i-f(x_i;\boldsymbol{a})]}{\sigma_i^2}\cdot\frac{\partial f(x_i;\boldsymbol{a})}{\partial a_k}, k=1,2,\cdots m \tag{7.20}$$

(三)地球物理数据模型

如(7.7)所示，实际数据带有误差，这种误差常常用平稳和各态遍历的时间序列来处理，因此这种数据集相当于平稳和各态遍态的随机过程，此时可用自相关和互相关函数来描述其特性。以地震记录而言，信号的统计性质是与$t=0$的原点有关的，因此不是严格平稳的。但是，如果取一段短的时窗，如在5s的记录中取150ms的记录段，它可以近似看成是平稳的。

对这种地震记录常可以用自回归(AR)模型逼近。此时，下一时刻数据的预测值(前一段波动的后果)可用当前及以前一段数据的线性组合表示为

$$u_m=b_1u_{m-1}+b_2u_{m-2}+\cdots+b_pu_{m-p}+\varepsilon_m \tag{7.21}$$

其中$\{b\}$为自回归系数集，即数据模型，ε为残差，应该是均值为零的白噪音。

令$m=p+1,p+2,\cdots,N$，代入(7.21)得矩阵方程

$$\begin{bmatrix}u_{p+1}\\u_{p+2}\\\vdots\\u_N\end{bmatrix}=\begin{bmatrix}u_p & u_{p-1} & \cdots & u_1\\u_{p+1} & u_p & \cdots & u_2\\\vdots & \vdots & \ddots & \vdots\\u_{N-1} & u_{N-2} & \cdots & u_p\end{bmatrix}\begin{bmatrix}b_1\\b_2\\\vdots\\b_p\end{bmatrix}+\begin{bmatrix}\varepsilon_1\\\varepsilon_2\\\vdots\\\varepsilon_p\end{bmatrix} \tag{7.22}$$

用最小二乘法求解(7.22)便可求得自回归系数。

由于大地的传递函数与源的褶积有时可用来表示某些地球物理信号，相应的数据模型称为滑动平均(MA)模型，即

$$g_i=\sum_{n=-n_l}^{n_r} c_n f_{i+n} \tag{7.23}$$

其中c_n为权系数，滑动模型是一个线性滤波器，它的输入为$\{f_i\}$，输出为$\{g_i\}$，n_l为第i个点向左滑移的点数，n_r为向右滑移的点数。当$n_r=0$时上式表示一个"因果"滤波器。

MA模型常被用来作光滑滤波。例如，令

$$c_n=1/(n_l+n_r+1)$$

用(7.23)便是进行线性窗口平均的光滑计算。这种线性光滑滤波处理常常使信号畸变。例如在图7.1(b)中，经16点光顺右部的信号幅度明显减小，信号变"胖"。

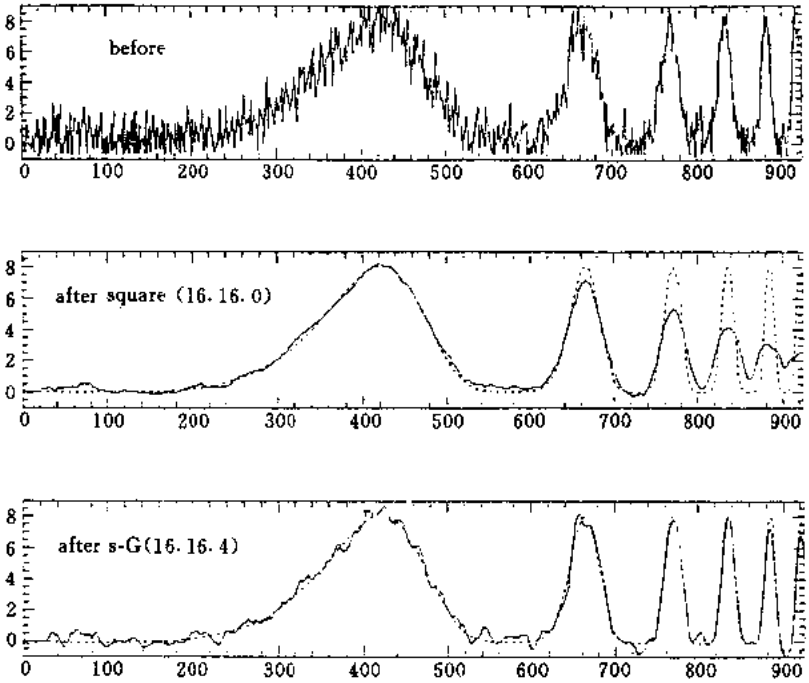


图7.1 光滑滤波的比较

(a)带噪音的信号；(b)线性光滑滤波的结果；(c)SG滤波的结果。图中虚线为信号

为了避免这种畸变，由Savitzky和Golay提出一种新的光滑滤波的方法(简称SG滤波)[30]，他们的主导思想是在移动窗口内用多项式拟合数据(见上一段)。对每个点f_i，作窗口内n_l+n_i+1个点的最小二乘多项式拟合，然后得到多项式在i点的值g_i，这个过程随窗口移动到f_{i+1}点而重复，重新用多项式拟合求系数$\{c_n\}$。

如果按这个想法重接作光滑计算，无疑是个笨办法，因为计算本身没有新的技巧。SG滤波的成功之处在于可以对假想的脉冲时序先做多项式拟合，然后仅用线性组合就可以求出对实际数据的拟合系数。于是，笨办法就变成了妙办法。

设多项式阶次为M，即$a_0+a_1 i+\cdots+a_m i^m$。注意g_0就是在$i=0$时的a_0。因为窗口内$i=-n_l\cdots,0\cdots,n_r$，相应最小二乘拟合问题的矩阵元为

$$A_{ij}=i^j, j=0,\cdots,M \tag{7.24}$$

用矩阵方程表示的最小二乘法方程为

$$(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})\cdot\boldsymbol{a}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{f} \text{ 或 } \boldsymbol{a}=(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{f} \tag{7.25}$$

代入(7.24)有

$$\{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\}_{ij}=\sum_{k=-n_l}^{n_r}A_{ki}A_{kj}=\sum_{k=-n_l}^{n_r}k^{i+j} \tag{7.26}$$

及

$$\{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{f}\}_j=\sum_{k=-n_l}^{n_r}A_{kj}f_k=\sum_{k=-n_l}^{n_r}k^j f_k \tag{7.27}$$

由于用单位向量e_n代替窗口内 f 数列时系数c_n就是a_0，因此有

$$c_n=\{(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{e}_n)\}_0=\sum_{m=0}^{M}\{(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\}_{0m}n^m \tag{7.28}$$

式中我们只需计算$(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}$的一行。用(7.28)计算系数后代入(7.23)作光滑，便完成了 SG 滤波。

在图 7.1 所示的例子中，用 4 阶多项式和 33 个点的滑动窗口取得的 SG 滤波结果示如(c)，可见畸变比 MA 模型小得多，但光滑程度略差。一般地说，在 SG 滤波中选用多项式阶次越高则信号畸变越小，而光滑程度越低。

二、最小二乘法数据处理

正如第三章第二节所述，最小二乘法是数据处理和反演中最常使用的数值计算方法。测量过程中不可避免带有噪音。只要信号与噪音的平方和(功率)可以准确地被测量出来。这时，以功率取极小为准则的最小二乘法就显示出它的独特优点。

(一)最小二乘法的导出

最小二乘法是 Hilbert 空间的线性逼近，用以下三种方法都可导出最小二乘解。

1. 投影法

设 H 为用一组规格化正交基$\{e_k\}$张开的 Hilbert 空间，x 为其中的一个元。设 $M=\mathrm{span}\{e_n\}$为 H 的一个子空间(n 为 k 的部分)，其中任意元记为 m，x 的其中的投影为 m'，则有

$$\|x-m'\|=\min\|x-m\|, m\in M \tag{7.29}$$

将 m 展成$\{e_k\}$的线性组合，有

$$\|x-\sum_{k=1}^{\infty}\alpha_k e_k\|=\min \tag{7.30}$$

同时，由投影定义知

$$\Big(x-\sum_{k=1}^{\infty}\alpha_k e_k\Big)\perp e_n, n=1,2,\cdots \tag{7.31}$$

因此，有

$$(x,e_n)=\Big(\sum_{k=1}^{\infty}\alpha_k e_k,e_n\Big)=\alpha_n, \tag{7.32}$$

其中(,)表示内积。式(7.32)表明，(7.30)取极小的充分和必要条件是系数$\{\alpha_k\}$等于规格化正交基的 Fourier 系数$\{c_k\}$

$$\alpha_k=(x,e_k)=c_k \tag{7.33}$$

2. 导数法

按最小二乘准则令目标泛函为

$$Q(\alpha_1,\alpha_2,\cdots)=\|x-\sum_{k=1}^{\infty}\alpha_k e_k\|^2 \tag{7.34}$$

展开后有

$$Q=(x-\Sigma\alpha_k e_k,x-\Sigma\alpha_n e_n)=\|x\|^2-2\Sigma\alpha_k(x,e_k)+\Sigma\alpha_k^2 \tag{7.35}$$

对 Q 取极小应使它对系数的偏导数为零

$$\frac{\partial Q}{\partial\alpha_n}=0,n=1,2,\cdots \tag{7.36}$$

代入(7.35)有

$$-2(x,e_n)+2\alpha_n=0,\text{或}\alpha_n=(x,e_n),n=1,2,\cdots \tag{7.37}$$

结果同(7.33)。

3. 匹配法

将(7.35)重写为

$$Q=\|x\|^2-\Sigma c_k^2+\Sigma c_k^2-2\Sigma\alpha_k c_k+\Sigma\alpha_k^2=\|x\|^2-\Sigma c_k^2+\Sigma(\alpha_k-c_k)^2 \tag{7.38}$$

此式右边前两项为模，取极小时必须使第三项为零，即

$$\alpha_k=c_k\equiv(x,e_k),k=1,2,\cdots \tag{7.39}$$

此式亦与(7.33)及(7.37)相同，同时，目标函数的极小应为

$$Q_{\min}=\|x\|-\sum_{k=1}^{\infty}\alpha_k^2 \tag{7.40}$$

4. 法方程

当 x 所在的空间的基$\{e_k\}$为有限而且不正交时，这时$\{\alpha_k\}$不是 Fourier 系数，但只要它们线性无关，就可找到一个元

$$x'\in M=\text{span}\{e_1,\cdots,e_n\};x\in H \tag{7.41}$$

使得

$$\|x-x'\|=\min\|x-m\|,m\in M \tag{7.42}$$

这等效于求系数组$\{\alpha_k\}$使得

$$Q(\alpha_1,\cdots,\alpha_n)=\|x-\sum_{k=1}^{n}\alpha_k e_k\|^2=\min \tag{7.43}$$

应用(7.33)所示的投影定理便有

$$\sum_{k=1}^{n}(e_k,e_l)\alpha_k=(x,e_l),l=1,\cdots,n \tag{7.44}$$

这个线性方程组称为最小二乘问题的法方程。正如在第三章中提到过，一般矩阵方程中系数矩阵的列向量不是相互正交的，因此或者用最小二乘法求法方程（7.44）的解（显式表示在（3.19）式），或者用奇异值分解把系数矩阵正交化，然后便得到类似于（7.33）中的投影解。

下面介绍最小二乘法在数据处理中的一些典型应用。

（二）系统辨识问题

如果一个线性系统有 n 个输入道$\{x\}$，但只有一个输出 y，它是$\{x\}$的线性组合

$$y=a_1x_1+a_2x_2+\cdots+a_nx_n \tag{7.45}$$

所谓系统辨识问题指由上述输入与输出确定系数集$(a_1,\cdots,a_n)$。

设对此系统测试 m 次，$k=1,2,\cdots,m$，则问题变为由下述方程组确定系数集：

$$a_1x_1(k)+a_2x_2(k)+\cdots+a_nx_n(k)=y(k),k=1,2,\cdots,m \tag{7.46}$$

记

$$\boldsymbol{a}=\begin{bmatrix}a_1\\ \vdots\\ a_n\end{bmatrix},\boldsymbol{X}=\begin{bmatrix}x_1(1) & \cdots & x_n(1)\\ \vdots & & \vdots\\ x_1(m) & \cdots & x_n(m)\end{bmatrix},\boldsymbol{y}=\begin{bmatrix}y(1)\\ \vdots\\ y(m)\end{bmatrix} \tag{7.47}$$

则(7.46)变为矩阵方程$\boldsymbol{x}_a=\boldsymbol{y}$，但测量带有误差。

如果 $m>n$，用最小二乘法得

$$\boldsymbol{a}=(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y} \tag{7.48}$$

当系统的输出不是一道而是多道时(记为 p)，则输入与输出的关系为

$$y_i=a_{1i}x_1+a_{2i}x_2+\cdots+a_{ni}x_n,i=1,\cdots,p \tag{7.49}$$

同样测试 m 次，记

$$\boldsymbol{A}=\begin{bmatrix}a_{11} & \cdots & a_{1p}\\ \vdots & & \vdots\\ a_{n1} & \cdots & a_{np}\end{bmatrix},\boldsymbol{Y}=\begin{bmatrix}y_1(1) & \cdots & y_p(1)\\ \vdots & & \vdots\\ y_1(m) & \cdots & y_p(m)\end{bmatrix} \tag{7.50}$$

则有矩阵方程$\boldsymbol{Y}=\boldsymbol{XA}$，系统由矩阵 $\boldsymbol{A}$ 来辨识。当它为负定矩阵时，我们可由目标泛函

$$Q(\boldsymbol{A})=(\boldsymbol{Y}-\boldsymbol{XA})^{\mathrm{T}}(\boldsymbol{Y}-\boldsymbol{XA})=\min \tag{7.51}$$

的极小化求矩阵 $\boldsymbol{A}$。在这种情况下导数法不能用，试用匹配法

$$Q=\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}-\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{XA}+\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{XA} \tag{7.52}$$

$$Q=\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}-\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}+[\boldsymbol{A}-(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}]^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})[\boldsymbol{A}-(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}] \tag{7.53}$$

显然，(7.53)取极小应满足

$$\boldsymbol{A}=(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y} \tag{7.54}$$

(三)载波信息恢复问题

在信息论发展早期，电子战的内容之一是由敌台发射的载波信息中检出低频信息。

以天线发射的高频窄带电台的电场分量可表为

$$E(0,t)=S(t)\mathrm{e}^{\mathrm{j}\omega t} \tag{7.55}$$

式中 $S(t)$为被载波的信号，j 为与天线有关的参数，ω 为角频率。设接收器矢径为 $\boldsymbol{r}$，其上电场为

$$E(\boldsymbol{r},t)\approx S(t)\mathrm{e}^{\mathrm{j}(\omega t-\boldsymbol{r}^{\mathrm{T}}\boldsymbol{k})} \tag{7.56}$$

式中波数矢量 $\boldsymbol{k}$ 在 x-y 平面上表为

$$\boldsymbol{k}=|\boldsymbol{k}|(\cos\theta\sin\theta)^{\mathrm{T}} \tag{7.57}$$

θ 为相对 x 轴波传播的方位角，$|\boldsymbol{k}|$为 $\boldsymbol{k}$ 的模，有 $k=|\boldsymbol{k}|=\omega/c=2\pi/\lambda$，$\lambda$ 为波长。设平面上第 l 个接收器矢径 $r_l=(x_l,y_l)^{\mathrm{T}}$，(7.57)代入(7.56)并考虑接收器的频率响应 $g(\theta)$有

$$x_l(t)=g_l(\theta)\mathrm{e}^{-\mathrm{j}k(x_l\cos\theta+y_l\sin\theta)}\cdot s(t)=a_l(\theta)s(t) \tag{7.58}$$

它为接收器上的接收信号。

如果在平面上有不同方位 θ 的 L 个接收台，而同时有 M 个敌对发射台载波信号，则有

矩阵

$$\boldsymbol{x}(t)=\boldsymbol{H}(\theta)\boldsymbol{S}(t)+\boldsymbol{V}(t) \tag{7.59}$$

式中

$$\boldsymbol{H}(\theta)=[a(\theta_1),\cdots,a(\theta_m)](L\times M)$$

$$\boldsymbol{S}(t)=[s_1(t),\cdots,s_M(t)]^{\mathrm{T}}$$

而 $\boldsymbol{V}(t)$ 表示干扰，由 $x(t)$ 确定敌台方位 $\{\theta_k\}$ 的问题称为定位问题，在文献[31]中有精辟的叙述。我们考虑在敌台方位已确定情况下如何用最小二乘法恢复信号 $\boldsymbol{S}(t)$。

Wiener 滤波器指的是设计一个线性滤波算子，它的输入为接收信号 $x(t)$，输出为 $s'(t)$，并使它与 $s(t)$ 的平方拟合差取极小

$$\varepsilon=\|s-s'\|=E\{(s-s')^{\mathrm{T}}(s-s')\}=\min \tag{7.60}$$

式中 E 为数学期望的符号。

设矩阵 A 为 Wiener 滤波器的线性算子，而 T 为某一正交变换算子(如传输网络)，此时滤波器的输出为

$$\boldsymbol{s}'=\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{y}=\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}\boldsymbol{x} \tag{7.61}$$

其中 $\boldsymbol{y}$ 是输入 $\boldsymbol{x}$ 的正交变换。(7.61)代入(7.60)有

$$\varepsilon=E\{(s-\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{y})^{\mathrm{T}}(s-\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{y})\}=E\{\boldsymbol{s}^{\mathrm{T}}\boldsymbol{s}-2\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{T}\boldsymbol{s}+\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y}\}=\min \tag{7.62}$$

由矩阵求导公式有

$$\Delta_A\varepsilon=-2E\{\boldsymbol{T}\boldsymbol{s}\boldsymbol{y}^{\mathrm{T}}\}+2E\{\boldsymbol{A}\boldsymbol{y}\boldsymbol{y}^{\mathrm{T}}\} \tag{7.63}$$

由(7.60)知，滤波器算子满足

$$AE\{\boldsymbol{y}\boldsymbol{y}^{\mathrm{T}}\}=TE\{\boldsymbol{s}\boldsymbol{y}^{\mathrm{T}}\}，或\ A-TE\{\boldsymbol{s}\boldsymbol{y}^{\mathrm{T}}\}E\{\boldsymbol{y}\boldsymbol{y}^{\mathrm{T}}\}^{-1} \tag{7.64}$$

如果信号与噪音相互独立而且均值为零，则有

$$E\{\boldsymbol{x}\}=0,E\{\boldsymbol{v}\}=0,E\{\boldsymbol{s}\boldsymbol{v}^{\mathrm{T}}\}=E\{\boldsymbol{v}\boldsymbol{s}^{\mathrm{T}}\}=0 \tag{7.65}$$

将信号与噪音的自相关矩阵记为

$$P=E\{\boldsymbol{s}\boldsymbol{s}^{\mathrm{T}}\},R=E\{\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}\} \tag{7.66}$$

则得

$$E\{\boldsymbol{s}\boldsymbol{y}^{\mathrm{T}}\}=E\{\boldsymbol{s}(\boldsymbol{H}\boldsymbol{s}+\boldsymbol{v})^{\mathrm{T}}\}\boldsymbol{T}^{\mathrm{T}}=\boldsymbol{P}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{T}^{\mathrm{T}}, \tag{7.67}$$

$$E\{\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\}=E\{(\boldsymbol{H}\boldsymbol{s}+\boldsymbol{v})(\boldsymbol{H}\boldsymbol{s}+\boldsymbol{v})^{\mathrm{T}}\}=\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^{\mathrm{T}}+\boldsymbol{R} \tag{7.68}$$

及

$$E\{\boldsymbol{y}\boldsymbol{y}^{\mathrm{T}}\}=\boldsymbol{T}(\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^{\mathrm{T}}+\boldsymbol{R})\boldsymbol{T}^{\mathrm{T}} \tag{7.69}$$

将(7.67～7.69)代入(7.64)得有 Wiener 滤波算子

$$\boldsymbol{A}=\boldsymbol{T}\boldsymbol{P}\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^{\mathrm{T}}+\boldsymbol{R})^{-1}\boldsymbol{T}^{-\mathrm{T}} \tag{7.70}$$

于是，(7.61)所示的滤波系统的输出 S' 与载波信号的最小平方拟合差取极小。

(四)最小平方数字滤波

由 Wiener 的最小二乘滤波思想发展了许多数字滤波器，除上述用于信号检测的滤波器外，还有许多线性滤波器，统称为最小平方滤波器，在地球物理中有广泛应用。

地球作为波传播的介质本身就是一个滤波器(大地滤波器)。它可用褶积表为 $s=h*r$，其中 s 为大地滤波器的输出，h 为大地滤波器的脉冲响应，r 为一次或二次源信号(如反射地震中的反射系数序列)。当信号 s 中含有误差时，可依最小二乘法则取目标泛函

$$Q(h_1,h_2,\cdots)=\|s-h*r\|^2=\sum_{n=0}^{\infty}(s_n-\sum_{k=0}^{\infty}r_{n-k}h_k)^2 \tag{7.71}$$

的极小，来求大地滤波器的脉冲响应。此时有

$$\frac{\partial Q}{\partial h_m}=2\sum_{n=0}^{\infty}\left(s_n-\sum_{k=0}^{\infty}r_{n-k}h_k\right)r_{n-m}=0 \tag{7.72}$$

重排(7.72)得

$$\sum_k\sum_n r_{n-k}r_{n-m}h_k=\sum_n s_n r_{n-m} \tag{7.73}$$

记源的自相关及它与接收信号的互相关为

$$a_\tau^{rr}=\sum_{n=0}^{\infty}r_{n+\tau}r_n,\text{和}a_\tau^{rs}=\sum_{n=0}^{\infty}s_{n+\tau}r_n,\tau=0,1,2,\cdots \tag{7.74}$$

则(7.73)式变为

$$\sum_{k=0}^{\infty}a_{m-k}^{rr}h_k=a_m^{rs},m=0,1,2,\cdots \tag{7.75}$$

对包含 N 个数据的时间序列，上式变为矩阵方程

$$\begin{bmatrix}a_0^{rr}&a_1^{rr}&\cdots&a_N^{rr}\\a_1^{rr}&a_0^{rr}&\cdots&a_{N-1}^{rr}\\\cdots&\cdots&\cdots&\cdots\\a_N^{rr}&a_{N-1}^{rr}&\cdots&a_0^{rr}\end{bmatrix}\begin{bmatrix}h_0\\h_1\\\vdots\\h_N\end{bmatrix}=\begin{bmatrix}a_0^{rs}\\a_1^{rs}\\\vdots\\a_N^{rs}\end{bmatrix} \tag{7.76}$$

由此可解得大地滤波器的脉冲响应。

将最小平方滤波器用于所谓的“预测问题”，指的是有一个时间序列$\{x_i\},i=0,\cdots,N$，要求此序列的预测值x_{N+p}，其中 $p>0$。设 $h=(h_0,h_1\cdots,h_N)^T$ 为线性预测滤波器，它的输入为$\{x_n\}$，输出为 $y=\{x_{n+p}\}$，有相关

$$a_m^{xy}=\sum_{n=0}^{\infty}x_{n+p+m}x_n=a_{p+m}^{xx},\text{其中输出 }y=\{x_{n+p}\} \tag{7.77}$$

由(7.75)式有

$$\sum_{k=0}^{\infty}a_{m-k}^{xx}h_k=a_{p+m}^{xx},m=0,1,2\cdots \tag{7.78}$$

这便是最小平方预测滤波器的法方程，写成矩阵形式有

$$\begin{bmatrix}a_0&a_1&\cdots&a_N\\a_1&a_0&\cdots&a_{N-1}\\\cdots&\cdots&\cdots&\cdots\\a_N&a_{N-1}&\cdots&a_0\end{bmatrix}\begin{bmatrix}h_0\\h_1\\\vdots\\h_N\end{bmatrix}=\begin{bmatrix}a_p\\a_{p+1}\\\vdots\\a_{p+n}\end{bmatrix} \tag{7.79}$$

解此方程得预测滤波器系数$\{h\}$，它的输出将是输入时间序列的预测值。

(五)叠前地震偏移波速确定问题

最小二乘思想在地球物理数据处理中的应用不胜枚举，下面举一个当前的热点问题叠前反射地震偏移中应用的一个例子。

在第六章第二节中曾谈过地震偏移问题，在哪里指的是平面波场向下入射时用有限差分法作波场偏移归位，以对反射源(地下构造)成像。实际地震记录多是共炮点记录，经过叠加后近似为平面波入射的波场，但叠加处理降低分辨率并产生波形畸变。因此，直接用共炮点记录(点源波场)作地震偏移具有诸多优越性。

早在20年代反射地震刚应用于油气勘探时就发现，反射波可以从任意方向返回，因此反射体位置不一定在观测点的正下方。将反射波能量正确地归位于地面观测点的正下方的处理叫地震偏移。地震偏移的计算的基础是波动方程式，但是要求其中位于系数项的地震波速度函数为已知。但是，准确的波速模型只有在反射能量归位之后才能求得。因此，地震学家便陷入这样的怪圈："他想解一个问题，但又非先有答案不可"。如何走出这个怪圈便成为近年来研究的焦点之一，而其中波速确定问题又是关键。

叠前的地震波场可写 $d(x_s,y_s,x_r,y_r,t)$，其中下标 s 表示点震源，r 为接收器，而波速模型为 $c(x,y,z)$。可以看出，地震数据是冗余的，其自由度比波速模型大得多。假设测区共有 N 个炮，则共炮点记录为

$$d_1,\cdots,d_N$$

作叠前偏移后由它们得到偏移道集相应为

$$R_1,\cdots,R_N$$

考虑到炮间距一般不大，而且对所有这些道集共用同一个横向变化缓慢的波速模型 C（或波阻抗模型 B），因此如果做偏移时使用了准确的波阻抗模型，则相邻炮的偏移道集最为相似。反之，如果波阻抗模型不准确，则相邻炮的偏移结果差异变大，相加后由于正负抵消，偏移道集的和的模将减小。根据最小二乘的思想，定义偏移道集的相似度为[32]

$$I(C)=\| R_1+\cdots+R_N \|^2_{WM} \tag{7.80}$$

式中 WM 意为加权范数，即对偏移道集加权后计算模。以(7.80)为目标泛函作极大化来确定波速模型 C，它是在网格点上的离散波速值。由于目标泛函是非线性的，计算可采用最速下降法等非线性迭代方法，在第九章中还要提及。

在极大化计算时要计算相似度的梯度

$$\nabla_c I=\left(\frac{\partial I(C)}{\partial C_m}\right),m=1,\cdots,M \tag{7.81}$$

式中 m 为网格点号，M 为总点号。实际上，此梯度为前传波场与反传波场的差加上它们的互相关，可结合偏移时波场的计算进行。

在下节中还要介绍二维最小二乘滤波器。

三、二维数字滤波

(一)二维 Fourier 滤波[33~34]

二维 Fourier 滤波在地球物理数据处理中有广泛应用，如地震处理中的 f-k 滤波，重磁数据处理中的波数域滤波等。

设 $f(x,y)$为区域位场，其 Fourier 变换表为

$$F(u,v)=\frac{1}{2\pi}\iint f(x,y)e^{-i(ux+vy)}dxdy \tag{7.82}$$

其中 u 和 v 为 x 或 y 方向的角波数。一个共中点地震道集为 $f(x,t)$，相应变换为

$$F(k,\omega)=\frac{1}{2\pi}\iint f(x,t)e^{-i(kx+\omega t)}dxdt \tag{7.83}$$

其中 k 为 x 方向波数，ω 为角频率。

在波数——频率域设计数字滤波器的传递函数 $H(u,v)$或 $H(k,\omega)$，用它乘以 $F(u,v)$或 $F(k,\omega)$，再逆 Fourier 变换回空间域或时空域，即达到滤波即消除噪音的目的。例如，消除表层干扰的位场低通滤波器为

$$H(u,v)=\begin{cases}1, |u|<k_x\ |v|<k_y\\ 0, |u|\geqslant k_x\ |v|\geqslant k_y\end{cases} \tag{7.84}$$

其中 k 为截断角波数。又如对地震道去除低速相干噪音的滤波器，

$$H(k,\omega)=\begin{cases}1, |k|<\omega/V\\ 0, |k|\geqslant\omega/V\end{cases} \tag{7.85}$$

其中 V 为信号的最小视速度。归纳各种二维 Fourier 滤波器，可以把它通用的传递函数写为

$$H(u,v)=\begin{cases}1, k_1<|u^2+v^2|\leqslant k_2\\ 0, \lambda_1 v<|u|\leqslant\lambda_2 v\end{cases} \tag{7.86}$$

其中 k_1 为低截波数(或频率)，k_2 为高截波数；λ_1 和 λ_2 为选定信号方向(倾角)的斜度范围。

由于空间域的褶积等于波数域的乘积，因此，二维 Fourier 滤波的输出为

$$f'(x,y)=\frac{1}{2\pi}\iint F(u,v)H(u,v)e^{i(ux+vy)}dudv \tag{7.87}$$

或

$$f'(x,t)=\iint F(k,\omega)H(k,\omega)e^{i(kx-\omega t)}dkd\omega \tag{7.88}$$

在实际应用中应增加去假频滤波和减小截断效应的有关技术。

(二)最佳二维线性滤波器设计[35]

上述 Fourier 变换滤波要求信号与噪音的"谱"有明显区别，因此可以用(7.86)这样的滤波器切除。如果信号与噪音的谱相互重叠，但我们有关于它们的谱的先验信息，可用王继伦先生提出的以下设计方法。

设二维数据包含信号与噪音两部分

$$d(x,y)=s(x,y)+n(x,y) \tag{7.89}$$

记 $h(x,y)$为二维滤波的脉冲响应函数，它的传递函数 $H(u,v)$，则滤波输出为 $d'=d*h$。记 $d'=s'+n'$，其中 s'为输出的信号成分，n'为输出的噪音成分。由上节讨论的最小二乘原理知，最小畸变滤波器应使

$$\iint[s(x,y)-s'(x,y)]^2dxdy=\min \tag{7.90}$$

对于压抑输出噪音的滤波器，应有

$$\iint[n'(x,y)]^2dxdy=\min \tag{7.91}$$

根据在波数域乘积等于空间域褶积的原理，对应于(7.90)的滤波器传递函数应使目标泛函

$$I_1=\iint[1-H(u,v)]^2|S(u,v)|^2dudv=\min \tag{7.92}$$

而对应于(7.91)的滤波器有

$$I_2=\iint H^2(u,v)|N(u,v)|^2dudv=\min \tag{7.93}$$

应用 Lagrange 乘子法(λ 为乘子)，同时要求畸变及噪音同时取极小的最佳线性滤波器

应使

$$I' = \iint\{[1 - H(u,v)]^2 S(u,v)|^2 + \lambda H^2(u,v)|N(u,v)|^2\}\mathrm{d}u\mathrm{d}v = \min \quad (7.94)$$

在离散情况下使 I' 对 $H(u,v)$ 的导数为零，最后可得此滤波的传递函数为

$$H(u,v) = \frac{|S(u,v)|^2}{|S(u,v)|^2 - \lambda|N(u,v)|^2} \quad (7.95)$$

(7.95)回代入(7.92)可求出拉格朗日乘子 λ。这种最佳滤波器是一种在最小二乘准则下的最佳滤波器。

(三)K-L 滤波器[36]

在模式识别研究中，有一种基于 K-L 变换的分类特征提取方法，它可以转换成一种基于 K-L 变换的滤波器，以提高地震记录的信噪比，也可以用来进行波类型的分类识别[37]。

K-L 变换是建立在线性正交展开原理上的数据处理方法之一，在模式识别、数据压缩等许多领域有广泛的应用。我们首先从线性正交展开的特例傅里叶展开出发，将它推广为 K L 展开，然后再进一步把它与奇异值分解联系起来。

某一周期性平稳随机过程的傅里叶展开为

$$x(t) = \sum_{n=-\infty}^{\infty} x_n \exp(\mathrm{i}n\omega_0 t) \quad (7.96)$$

其中，$\omega_0 = 2\pi/T$，T 为随机过程 $x(t)$ 的周期。由指数函数的正交性易知，傅里叶系数 x_m 和 x_n 是互不相关的，即它们的协方差矩阵为对角阵。

对于非周期性随机过程傅里叶展开已不适用，这就引出了 K-L 展开(或称 K-L 变换)，它表示为

$$x(t) = \sum_{n=1}^{\infty} \gamma_n x_n \varphi_n(t), a \leqslant t \leqslant b \quad (7.97)$$

核函数满足

$$\int_a^b \varphi_n(t)\varphi_m^*(t)\mathrm{d}t = \begin{cases} 1, n = m \\ 0, n \neq m \end{cases} \quad (7.98)$$

的正交性条件，而 γ_n 项是为满足协方差矩阵

$$E[x_n x_m^*] = 1, n = m \quad (7.99)$$

设置的，* 号表示共轭。容易证明，$x(t)$ 的相关函数

$$R(t,s) = E[x(t)x^*(s)] = \sum_n |\gamma_n|^2 \varphi_n(t)\varphi_n^*(s) \quad (7.100)$$

而

$$\int_a^b R(t,s)\varphi_k(s)\mathrm{d}s = |\gamma_k|^2 \varphi_k(t) \quad (7.101)$$

可见 $|\gamma_n|^2$ 为积分方程(7.101)的特征值，$\varphi_k(t)$ 为相应的特征函数；因此，K-L 变换式(7.97)中的系数及核函数可通过求解积分方程(7.101)而得到。

当 $x(t)$ 仅有 n 个离散取样的情况下，对应于 $x \in \mathrm{R}^n$，相应(7.97)式的 K-L 展开为

$$x = \sum_{j=1}^{n} c_j \boldsymbol{u}^j \quad (7.102)$$

其中 $\boldsymbol{u}^j$ 为正交矢量

$$(\boldsymbol{u}^i)^{\mathrm{T}}\boldsymbol{u}^j = \begin{cases} 1, j = i \\ 0, j \neq i \end{cases} \quad (7.103)$$

而系数

$$c_j = (\boldsymbol{u}^j)^{\mathrm{T}} x \tag{7.104}$$

对应于(7.100)式的相关函数变为 $n \times n$ 实对称矩阵

$$\boldsymbol{A} = E[xx^{\mathrm{T}}] \tag{7.105}$$

由(7.104)式知，系数的协方差

$$E[c_i c_j] = E[(\boldsymbol{u}^i)^{\mathrm{T}} x x^{\mathrm{T}} (\boldsymbol{u}^j)] = \lambda_i \delta_{ij} \tag{7.106}$$

此式与连续情况(7.99)式对应，其中 λ_i 为矩阵 $\boldsymbol{A}$ 的第 i 个特征值：

$$\boldsymbol{A}\boldsymbol{u}^j = \boldsymbol{A}\boldsymbol{\Lambda} \tag{7.107}$$

式中

$$\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \cdots, \lambda_n) \tag{7.108}$$

令特征向量组成为正交矩阵

$$\boldsymbol{U} = [\boldsymbol{u}^1, \boldsymbol{u}^2, \cdots, \boldsymbol{u}^n] \tag{7.109}$$

则有

$$\boldsymbol{U}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{U} = \boldsymbol{\Lambda} \tag{7.110}$$

即通过 K-L 变换(7.102)，把协方差矩阵 $\boldsymbol{A}$ 对角化。离散 K-L 变换的步骤是：由(7.105)式求 x 的协方差矩阵 $\boldsymbol{A}$，再求 $\boldsymbol{A}$ 的特征向量和特征值；由(7.104)式求系数组；最后根据数据压缩及子空间模式识别的需要，选定(7.102)式中级数展开的项数，即从 n 个特征矢量中取出其中 m 个构成新的变换矩阵，它为 $n \times m$ 阶矩阵，从而达到降维(如数据压缩)的目的。

对于地球物理来说，最感兴趣的应用是基于 K-L 变换的分类特征提取。由于自相关系数是协方差矩阵的对角线元素，由(7.107～7.108)式可知，大的特征值(即小的序号)对应过程的模式的相关性，而序号大的特征向量组成的变换矩阵则起到减小相关性，突出差异性的作用。因此变换矩阵起到分类特征提取的作用。以多道地震记录为例，反射信号一般具有较好的相干性，而噪音及绕射信号具有随机性，利用 K-L 变换分类特征提取可以构成提高信噪比的数字滤波器。

但是，由多道地震记录数据构成的矩阵 $\boldsymbol{A}$ 一般不是如(7.105)式所示的对称矩阵，它的分类特征分析不能通过典型的特征值分解求得，而必须通过奇异值分解(SVD)来求得。设 $\boldsymbol{A}$ 为秩等于 m 的 $n \times m$ 一般矩阵，可以用它构成以下实对称矩阵

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{O} & \boldsymbol{A} \\ \boldsymbol{A}^{\mathrm{T}} & \boldsymbol{O} \end{pmatrix} \tag{7.111}$$

它的维数为 $(n-m) \times (n+m)$，相应的谱展开写为

$$\boldsymbol{B}\boldsymbol{W} = \alpha \boldsymbol{W}, \text{其中 } \boldsymbol{W} = \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix} \tag{7.112}$$

其中 $\boldsymbol{u}$ 为 $(m \times 1)$，$\boldsymbol{v}$ 为 $(n \times 1)$。

将(7.111)代入(7.112)得

$$\boldsymbol{A}\boldsymbol{v} = \alpha \boldsymbol{u}, \boldsymbol{A}^{\mathrm{T}}\boldsymbol{u} = \alpha \boldsymbol{v} \tag{7.113}$$

及

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v} = \alpha^2 \boldsymbol{v}, \boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u} = \lambda^2 \boldsymbol{u} \tag{7.114}$$

注意(7.113)与奇异值分解定理证明中(2.79)式相同，而(7.114)与(2.91)及(2.92)式相同，

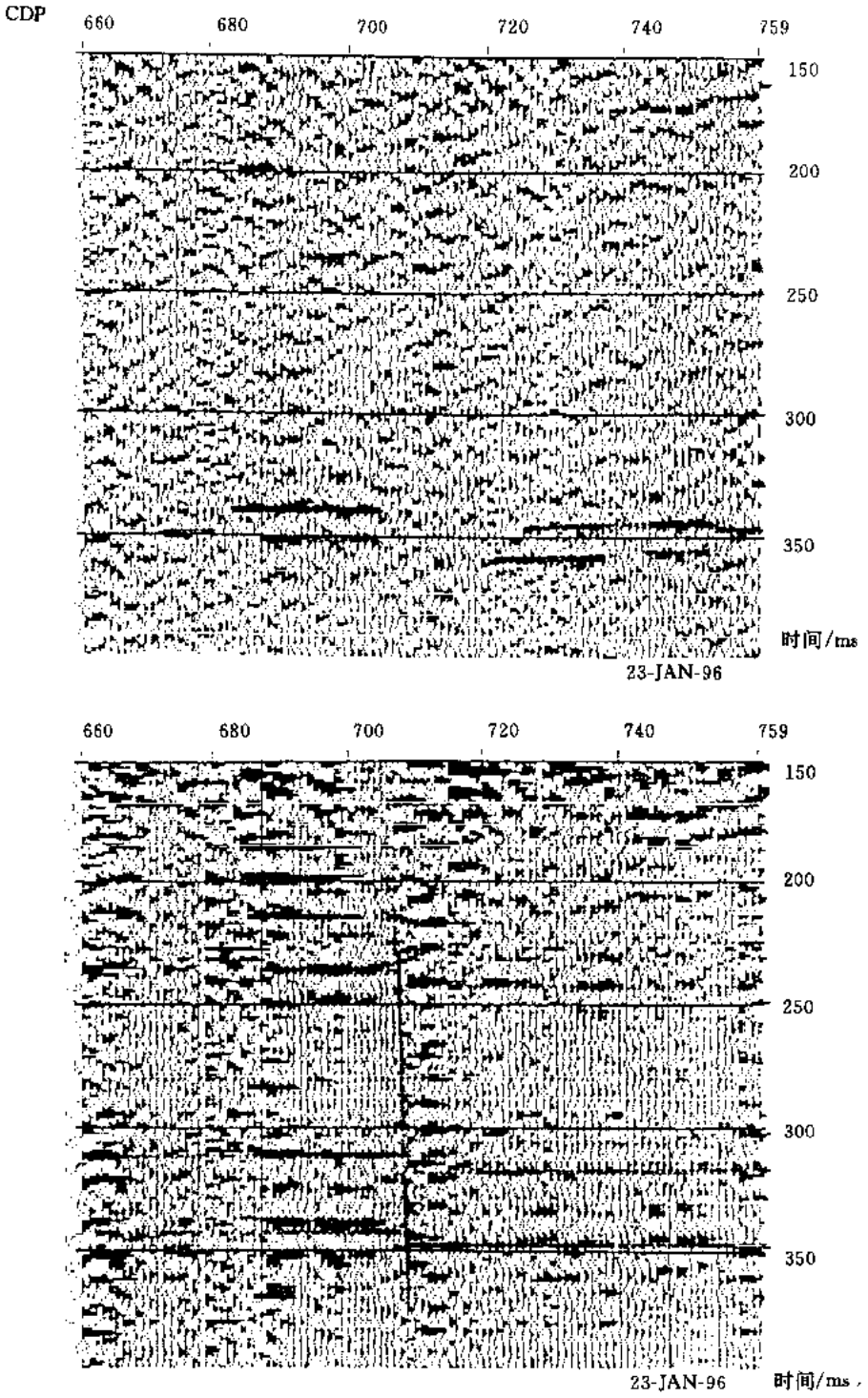


图 7.2　上海市北京东路反射地震叠后剖面段(上)及经非线性 K-L 滤波后的剖面段(下)
清晰显示出断层产状

因此按第二章第四节的证明不难将 K-L 变换推广到一般矩阵 A 的情况。

在应用时将二维数据写成矩阵形式，如在反射地震的叠前或叠后道集中提取一个数据窗。在图 7.2 中示出了在上海市区调查地震活动断裂带时在北京路取得的一段叠后剖面，时窗从 150ms 取到 400ms，从此数据窗的显示基岩面反射(350ms 左右)有错动，怀疑有断层存在，但是既不能肯定，也不能看出断层的产状。进行 K-L 滤波时将图 7.2(左)显示的数据 A 作奇异值分解并作非线性处理，然后根据奇异值总和的 70%作为信号的能量，另外 30%的干扰能量对应小奇异值并令其为零，因为

$$A = USV^{\mathrm{T}}, S = \mathrm{diag}(\alpha_1, \cdots, \alpha_m) \tag{7.115}$$

使

$$\sum_{i=1}^{q} \alpha_i^2 = 0.7 \sum_{i=1}^{m} \alpha_i^2 \tag{7.116}$$

确定 q，并计算

$$\tilde{A} = U\tilde{S}V^{\mathrm{T}}, \tilde{S} = \mathrm{diag}(\alpha_1, \cdots, \alpha_q) \tag{7.117}$$

作为滤波输出，取得图 7.2(右)的滤波结果，它清晰地显示了断层产状(详见：杨文采，上海大场—周浦断裂带市区物化探调查成果报告，地矿部物化探所，1996 年 2 月，廊坊)。在文献[33]中，还介绍了有关的非线性处理技术与其它应用实例。

二维小波变换作数字滤波是对二维 Fourier 滤波的重要改进，关于小波变换将在第五节讨论。

四、地球物理场的向下延拓

在第六章第二节中曾讨论过波场向下延拓问题，它是地震偏移处理的理论基础。我已故的朋友栾文贵先生对地球物理场的延拓理论曾作过全面深入的分析[38]，本节对此作一概括。

在基本方程(6.1)中，右端的源项在数学上代表着第一类的奇性，即在源发生的地方(奇点)场趋于无穷大。对于人工场源的观测方法，如人工地震或电磁法，接收器和场的计算必须绕过这些激发场源的奇点，因为近源点处场满足的方程要比(6.1)等偏微分方程复杂得多。对于某些被动场观测方法(如重磁场)，奇性出现在地下场源所在位置，它们正是被探测的对象。另一类奇性是由介质间断面[方程(6.1)系数项不连续]引起的奇性，相应的奇点反映地下构造(如反射面或不同电阻率岩层界面)。这种第二类的奇性在波场上表现为高阶奇性，而波场本身是连续的。当观测在地面进行时，波场向地下延拓的过程中在遇到奇性时会产生特殊的响应，因此有可能通过奇点的出现了解地下构造或异常源体。即使不可能直接圈定这些奇点，也可以提高分判地下构造的空间分辨率，因为地下某平面上的场更接近源体，这就是研究波场向下延拓的意义所在。

(一)位场向下延拓

令 (x, y, z) 为地下任一点，$z=0$ 表示地平面，$z=-H$ 表示地下某一平面(接近场源)，其中 H 为正实数，即此平面的深度。位场向下延拓问题乃是从地面上位场 $u(x, y, 0)=f(x, y)$ 或它的垂向导数 $g(x, y)$ 求地下位场 $u(x, y, -H)$。数学上为求解以下柯西问题

$$\begin{cases} \dfrac{\partial^2 u}{\partial z^2} + \left(\dfrac{\partial^2}{\partial x^2} + \dfrac{\partial^2}{\partial y^2}\right) u = 0, \quad -H \leqslant z < 0 \\ u|_{z=0} = f(x,y), \text{或} \dfrac{\partial u}{\partial z}\Big|_{z=0} = g(x,y) \\ u = 0 \text{ 如 } \sqrt{x^2 + y^2} \to \infty. \end{cases} \tag{7.118}$$

先来看向上延拓的情况(相当于正问题),即在(7.118)中 $0<z\leqslant H$ 的情况,用 Fourier 变换直接可得

$$\hat{g}(\xi,\eta) = -\sqrt{\xi^2+\eta^2}\,\hat{f}(\xi,\eta) \tag{7.119}$$

其中 ξ 和 η 为 x 及 y 方向的空间波数。对(7.119)作逆 Fourier 变换可求得波场向上延拓问题为解析解为

$$u(x,y,z) = \frac{-z}{2\pi}\int_{-\infty}^{\infty}\int \frac{f(x',y')}{[(x-x')^2+(y-y')^2+z^2]^{3/2}}\mathrm{d}x'\mathrm{d}y', z>0 \tag{7.120}$$

而对垂向导数的边界条件有

$$u(x,y,z) = -\frac{1}{2\pi}\int_{-\infty}^{\infty}\int \frac{g(x',y')}{[(x-x')^2+(y-y')^2+z^2]^{1/2}}\mathrm{d}x'\mathrm{d}y', z>0 \tag{7.121}$$

对于向下延拓问题(向上延拓的反问题),(7.120)式中 $u(x,y,z)$ 相当于地面的已知观测场,而积分号的 $f(x,y)$ 对应于下平面的未知场。因此,柯西问题(7.118)转变为求解下列第一类 Fredholm 积分方程

$$\int_{-\infty}^{\infty}\int K_z(x-x',y-y')u(x',y',z)\mathrm{d}x'\mathrm{d}y' = f(x,y), -H\leqslant z<0 \tag{7.122}$$

或

$$\int_{-\infty}^{\infty}\int L_z(x-x',y-y')u(x',y',z)\mathrm{d}x'\mathrm{d}y' = g(x,y) - H\leqslant z<0 \tag{7.123}$$

其中核函数表为

$$\begin{cases} K_z(x-x',y-y') = \dfrac{-z}{2\pi[(x-x')^2+(y-y')^2+z^2]^{3/2}} \\ L_z(x-x',y-y') = \dfrac{-1}{2\pi[(x-x')^2+(y-y')^2+z^2]^{1/2}} \end{cases} \tag{7.124}$$

如果下式成立

$$\int_{-\infty}^{\infty}\int |u(x,y,z)|^2\mathrm{d}x\mathrm{d}y < \infty \tag{7.125}$$

则方程(7.122)及(7.123)的解存在。条件(7.125)表明,位场向下延拓的最大深度为接近最浅的奇点,再深时解不存在。

不难导出由(7.124)表示的核函数的 Fourier 变换为

$$\begin{cases} \hat{K}_z(\xi,\eta) = \dfrac{1}{4\pi^2}\iint \dfrac{-z\mathrm{e}^{-\mathrm{i}(\xi\alpha+\eta\beta)}}{(\alpha^2+\beta^2+z^2)^{3/2}}\mathrm{d}\alpha\mathrm{d}\beta = \dfrac{1}{2\pi}\mathrm{e}^{z\sqrt{\xi^2+\eta^2}} \\ \hat{L}_z(\xi,\eta) = \dfrac{-1}{2\pi\sqrt{\xi^2+\eta^2}}\mathrm{e}^{z\sqrt{\xi^2+\eta^2}} \end{cases} \tag{7.126}$$

将(7.126)代入(7.122~7.123)的变换式并返回空间域有

$$u(x,y,z) = \frac{1}{\pi^2}\iint e^{-z\sqrt{\xi^2+\eta^2}}d\xi d\eta \iint f(x',y')\cos(\xi(x-x'))\cos(\eta(y-y'))dx'dy' \tag{7.127}$$

和

$$u(x,y,z) = \frac{-1}{\pi^2}\iint \sqrt{\xi^2+\eta^2}e^{-z\sqrt{\xi^2+\eta^2}}d\xi d\eta \iint g(x',y')\cos(\xi(x-x'))\cos(\eta(y-y'))dx'dy' \tag{7.128}$$

记着式中的 $z<0$，可知向下延拓的位场随波数指数增大，求解计算是不稳定的。

(二)电磁场向下延拓

针对低频电磁法而言，用线圈测量某个平面上的似稳电磁场强度并记为 $f(x,y)$，或它的垂向梯度 $g(x,y)$，延拓的基本方程是以下亥姆霍兹方程

$$\nabla^2 u + k^2 u = 0, k^2 = \omega^2\mu\varepsilon + i\omega\mu\sigma \tag{7.129}$$

其中 k 为对频率 ω 的角波数，σ 为电导率，ε 为介电系数，μ 为磁导率。在向下延拓时假定 k 为常数，则要求解以下边值问题

$$\begin{cases} \nabla^2 u + k^2 u = 0, -H \leqslant z < 0 \\ u|_{z=0} = f(x,y), \dfrac{\partial u}{\partial z}|_{z=0} = g(x,y) \\ u(x,y,z) = 0 \text{ 当 } \sqrt{x^2+y^2} \to \infty \end{cases} \tag{7.130}$$

由 Fourier 变换知场与其导数有以下关系

$$\hat{g}(\xi,\eta) = i\sqrt{k_a^2 - \xi^2 - \eta^2}\hat{f}(\xi,\eta) \tag{7.131}$$

其中 k_a 中的 a 表示空气中的场或介质参数。

应用 Fourier 变换(7.129)可改写为以下柯西问题

$$\begin{cases} \dfrac{d^2\hat{u}}{dz^2} + (k^2 - \xi^2 - \eta^2)\hat{u} = 0, -H \leqslant z < 0 \\ \hat{u}|_{z=0} = \hat{f}(\xi,\eta), \left.\dfrac{d\hat{u}}{dz}\right|_{z=0} = \hat{g}(\xi,\eta) \end{cases} \tag{7.132}$$

它的解为[36]

$$\hat{u}(\xi,\eta,z) = \hat{f}(\xi,\eta)\mathrm{ch}(iz\sqrt{k^2-\xi^2-\eta^2}) + \frac{\hat{g}(\xi,\eta)}{i\sqrt{k^2-\xi^2-\eta^2}}\mathrm{sh}(iz\sqrt{k^2-\xi^2-\eta^2}) \tag{7.133}$$

代入(7.131)有

$$\hat{u}(\xi,\eta,z) = \left[\mathrm{ch}(iz\sqrt{k^2-\xi^2-\eta^2}) + \sqrt{\frac{k_a^2-\xi^2-\eta^2}{k^2-\xi^2-\eta^2}}\mathrm{sh}(iz\sqrt{k^2-\xi^2-\eta^2})\right]\hat{f}(\xi,\eta) \tag{7.134}$$

因此，由逆 FT 可得电场向下延拓问题的解为

$$u(x,y,z) = \frac{1}{2\pi}\iint\left[\mathrm{ch}(iz\sqrt{k^2-\xi^2-\eta^2}) + \sqrt{\frac{k_a^2-\xi^2-\eta^2}{k^2-\xi^2-\eta^2}}\mathrm{sh}(iz\sqrt{k^2-\xi^2-\eta^2})\right]d\xi d\eta$$
$$\times \iint f(x',y')\cos(\xi(x-x'))\cos(\eta(y-y'))dxdy, -H \leqslant z < 0 \tag{7.135}$$

对不同频率的电磁场可得到多组向下延拓结果，综合它们的奇性反映将更准确的确定地下构造。

(三)声场向下延拓

令 $u(x,y,z,t)$ 为声场，向下延拓问题归结为解以下边值问题

$$\begin{cases}\dfrac{\partial^2 u}{\partial z^2}+\left(\dfrac{\partial^2}{\partial x^2}+\dfrac{\partial^2}{\partial y^2}+\dfrac{1}{c^2}\dfrac{\partial^2}{\partial t^2}\right)u=0,\ -H\leqslant z<0,\\ u\Big|_{z=0}=f(x,y,t),\dfrac{\partial u}{\partial z}\Big|_{z=0}=g(x,y,t).\end{cases}\tag{7.136}$$

其中，令 c 为常数声速(非常数声速问题见有关地震偏移文献，如〔7〕〔10〕)，为解此边值问题，必须规定

$$\xi^2+\eta^2<\omega^2/c^2\tag{7.137}$$

还要假设

$$\hat{m}(\xi,\eta,\omega)=\begin{cases}\hat{f}(\xi,\eta,\omega),\text{对 }\xi^2+\eta^2\geqslant\omega^2/c^2\\ 0,\text{其它}\end{cases}\tag{7.138}$$

及对相应的垂向导数

$$n(x,y,t)=\frac{-1}{(2\pi)^{3/2}}\int_0^\infty d\omega\iint\limits_{\xi^2+\eta^2\geqslant\omega^2/c^2}\sqrt{\xi^2+\eta^2-\omega^2/c^2}\hat{m}(\xi,\eta,\omega)e^{i(x\xi+y\eta+t\omega)}d\xi d\eta\tag{7.139}$$

因此，可用 m 和 n 代表(7.136)中的 f 和 g，求向下延拓问题的一种扩展了解空间的广义解。

用以上相同的 Fourier 变换方法，可得到与(7.122)类似的第一类 Fredholm 积分方程

$$\iiint K_z(x-x',y-y',t-t')u(z,x',y',t')dx'dy'dt'=m(x,y,t),\ -H\leqslant z<0\tag{7.140}$$

其中核函数为

$$K_z(x-x',y-y',t-t')=\frac{1}{8\pi^3}\int_0^\infty d\omega\iint\limits_{\xi^2+\eta^2\geqslant\omega^2/c^2}e^{z\sqrt{\xi^2+\eta^2-\omega^2/c^2}+i[\xi(x-x')+\eta(y-y')+\omega(t-t')]}d\xi d\eta\tag{7.141}$$

最后得到的声波场向下延拓的解析解为

$$u(x,y,z,t)=\frac{1}{\pi^3}\int_0^\infty d\omega\iint\limits_{\xi^2+\eta^2\geqslant\omega^2/c^2}e^{-z\sqrt{\xi^2+\eta^2-\omega^2/c^2}}d\xi d\eta$$
$$\times\iiint f(x',y',t')\cos(\xi(x-x'))\cos(\eta(y-y'))\cos(\omega(t-t'))dx'dy'dt'\tag{7.142}$$

(四)向下延拓问题的正则化[38~41]

在上述各种向下延拓问题的解析解中，地下场的计算都涉及到 $\exp(|z|\theta)$ 因子，说明如果数据 f 或 g 中含有的误差将会被指数放大，使计算变得十分不稳定。换句话说，向下延拓问题是所谓的不适定柯西边值问题。这时，不宜直接按有关公式(7.142)、(7.135)或(7.127)直接进行向下延拓计算，而要进行正则化处理。有关正则化的理论将在下章讨论，在此我们仅对向下延拓问题介绍如何进行正则化计算。

对不适定问题，我们考虑数据的不精确，不求它的精确解，而是在某一准则之下计算其

近似解。设带有误差的位场数据为

$$f^{\delta}(x,y) = f(x,y) + n(x,y), \quad -L < x,y < L \tag{7.143}$$

以最小能量准则和有限失拟能量为准则：

$$\begin{cases} J[u] \equiv \int_{-L}^{L}\int |u(x,y,z)|^2 \mathrm{d}x\mathrm{d}y = \min \\ I[u] \equiv \int_{-\infty}^{\infty}\int A(x,y;u) - f^{\delta}(x,y)|^2 \mathrm{d}x\mathrm{d}y = \delta^2, \quad -H \leqslant z < 0 \\ u(\pm L,y,z) = u(x,\pm L,z) = 0 \end{cases} \tag{7.144}$$

其中 I 与 J 为泛函，δ 为可忽略的小实数，而 A 满足第一类 Fredholm 积分方程(7.122)

$$A(x,y;u) = \int_{-L}^{L}\int K_z(x-x',y-y')u(x',y',z)\mathrm{d}x'\mathrm{d}y' \tag{7.145}$$

应用 Lagrange 最小乘子法并以 α 为乘子，取泛函

$$\begin{cases} M_\alpha[u] \equiv I[u] + \alpha J[u] = \min \\ u(\pm L,y,z) = u(x,\pm L,z) = 0 \end{cases} \quad \text{对} -H \leqslant z < 0 \tag{7.146}$$

从(2.146)中的 Eular 方程可导出

$$\int_{-\infty}^{\infty}\int K_z(x-x',y-y')u(x',y',z)\mathrm{d}x'\mathrm{d}y' + \alpha u(x,y,z) = f^{\delta}(x,y), \quad -H \leqslant z < 0 \tag{7.147}$$

它是第二类 Fredholm 积分方程，一定有稳定的解。对 x 和 y 进行 Fourier 变换上式变为

$$2\pi\hat{K}_z(\xi,\eta)\hat{u}(\xi,\eta,z) + \alpha\hat{u}(\xi,\eta,z) = \hat{f}^{\delta}(\xi,\eta), \quad -H \leqslant z < 0 \tag{7.148}$$

由(7.126)知，其中

$$\hat{K}_z(\xi,\eta) = \frac{1}{2\pi}\mathrm{e}^{z\sqrt{\xi^2+\eta^2}} \tag{7.149}$$

回代入(7.148)得

$$\hat{u}(\xi,\eta,z) = \frac{\exp(-z\sqrt{\xi^2+\eta^2})}{1+\alpha\exp(-z\sqrt{\xi^2+\eta^2})}\hat{f}^{\delta}(\xi,\eta), \quad -H \leqslant z < 0 \tag{7.150}$$

逆 Fourier 变换回空间域便得到稳定的近似解

$$\begin{aligned} u^{\delta}(x,y,z) = & \frac{1}{\pi^2}\iint \frac{\exp(-z\sqrt{\xi^2+\eta^2})}{1+\alpha\exp(-z\sqrt{\xi^2+\eta^2})}\mathrm{d}\xi\mathrm{d}\eta \\ & \times \iint f^{\delta}(x',y')\cos[\xi(x-x')]\cos[\eta(y-y')]\mathrm{d}x'\mathrm{d}y' \end{aligned} \tag{7.151}$$

其中正则化因子 α 与(7.144)中的允许失拟能量 δ 有关。这样处理之后，u^{δ} 便连续依赖于有误差的数据 f^{δ}，因为它与

$$\frac{\exp(|z|\theta)}{1+\alpha\exp(|z|\theta)} \tag{7.152}$$

的积分成正比。

用同样的过程对电磁场和声场作类似的正则化处理，便可得相应的向下延拓问题的近似解。如对电磁场，用(7.142)式计算不稳定，正则化后的近似解为

$$u^*(z,x,y,t)=\frac{1}{\pi^3}\int_{(}^{\infty}\mathrm{d}\omega\iint_{\xi^2+\eta^2\geqslant\omega^2/c^2}\frac{\exp(-z\sqrt{\xi^2+\eta^2+\omega^2/c^2})}{1+\alpha\exp(-z\sqrt{\xi^2-\eta^2+\omega^2/c^2})}\mathrm{d}\xi\mathrm{d}\eta\times$$
$$\iiint_{-\infty}^{\infty}f^*(x',y',t')\cos(\xi(x-x'))\cos(\eta(y-y'))\times$$
$$\cos(\omega(t-t'))\mathrm{d}x'\mathrm{d}y'\mathrm{d}t',\ -H\leqslant z<0 \tag{7.153}$$

其中正则化因子可将 u^* 代入下式计算

$$\iiint|A(x,y,z;u^*)-f^*(x,y,t)|^2\mathrm{d}x\mathrm{d}y\mathrm{d}z=\delta^2 \tag{7.154}$$

其中

$$A(x,y,t;u^*)=\iiint K_z(x-x',y-y',t-t')u^*(z,x',y',t')\mathrm{d}x'\mathrm{d}y'\mathrm{d}t' \tag{7.155}$$

式中核函数由(7.141)式定义。

五、小波变换与多尺度分析

自 1967 年快速 Fourier 变换算法(FFT)发明以来，如本章第三节所述的 Fourier 变换滤波和调和分析广泛应用于科学技术各个领域。然而，虽然正弦与余弦函数具有很好的正交性，但是它们都是无头无尾的函数，不具备局部性。换句话说，一个时间序列的谱可清晰地显示其频率成分，但不能指示某种频率成分出现在哪个时间段。为解决这个问题，Garbor 先生曾用一系列平移的窗口分划时间系列，然后对窗口内的时段进行 Fourier 变换(窗口 FT)，以便使调和分析具有局部性。然而 Garbor 变换找不到对应的正交系作为变换的基函数，因而无法精确地重建。这个问题直到 80 年代末期，发展了小波变换之后才得以解决。目前，小波变换已发展成一门专门的学问，有 500 页以上的多部巨著专门对小波变换进行论述，本书只能对小波变换的思路作简明的导引。

(一)连续小波变换[42~46]

在英语上，小波和地震子波使用同一个词 Wavelet，为区分它们中文把 Wavelet transform 译成小波变换。在历史上，1984 年 Grossman 曾发展了地震处理中的连续小波展开，并引入了 Wavelet 这个词，对小波变换理论的建立起了重要作用。

与 Fourier 变换类似，小波变换也要用一组函数 $h(x)$ 来逼近某一函数 $f(x)$。所不同的是，小波变换要求 $h(x)$ 同时具有以下三个特性：

(1)局部性，即 $h(x)$ 必须是支集紧的函数

$$h(x)=0\quad 对\ |x|\geqslant A \tag{7.156}$$

其中 A 是 x 轴的一个范围。

(2)与 Garbor 变换类似，它必须可以作平移变换，平移变换可记为

$$h_b(x)=h(x-b) \tag{7.157}$$

(3)除了平移外，这一组函数还可作伸缩变换，即

$$h_a(x)=h(ax) \tag{7.158}$$

由(7.156)知，如对 $x\geqslant A$ 有 $h(x)=0$，则对 $x\geqslant A/a$ 有 $h(ax)=0$，说明对 $a>1$，$h_a(x)$ 的支集更紧(也就是尺度更小)，这就是小波变换与多尺度分析紧密连系的根源。

结合以上三个特性，可以把这样的函数系写为

$$h_{a,b}(x) = |a|^{-1/2}h\left(\frac{x-b}{a}\right) \tag{7.159}$$

其中 a 称为伸缩因子，b 称为平移因子，它们都为实数。对 $a=1$ 和 $b=0$ 的 $h(x)$ 称为子波母函数，而给出一组 a 和 b 便得一组小波变换的基函数。

连续可积信号 $f(x)$ 的小波变换为它与基 $h_{a,b}(x)$ 在 L^2 空间的内积

$$W_{a,b}(f) = (f(x),h_{a,b}(x)) = \int_{-\infty}^{\infty} h_{a,b}(x)f(x)\mathrm{d}x; f(x) \in L^2(\mathrm{R}) \tag{7.160}$$

小波变换存在的条件为

$$W_h = \int_0^{\infty} \frac{|H(\omega)|^2}{|\omega|}\mathrm{d}\omega < \infty, \text{或} \int_0^{\infty} h(y)\mathrm{d}y = 0, \tag{7.161}$$

其中 $H(\omega)$ 为母函数 $h(x)$ 的 Fourier 谱。小波变换式(7.160)说明，用 2 个参数的基函数族可以把任意函数 $f \in L^2(\mathrm{R})$ 分解为线性组合。

要说明的是，小波基函数 $h_{a,b}(x)$ 并不一定是标准正交基。但是，由于它是具有伸缩和平移性质的仿射系，有可能把正交展开扩大到整个 $L^2(\mathrm{R})$ 空间，这个问题将在下一段中讨论。当然，在实际应用中，小波基函数还要求具有正交性和光滑性，才便于重建计算。

从(7.160)可以看到，对一维函数 $f(x)$ 的小波变换同时依赖于两个参数 a 和 b，也就是说，$W_{a,b}(f)$ 是一个二元函数，不像 Fourier 变换那么简单。相应地，小波逆变换(信号重建)定义为

$$f(x) = \frac{1}{W_h}\int_{-\infty}^{\infty}\int_0^{\infty} W_{a,b}(f)h_{a,b}(x)\frac{\mathrm{d}a\mathrm{d}b}{a^2} \tag{7.162}$$

其中 W_h 见(7.161)，取决于 h，它的唯一性仍需证明。相应地，对二维函数 $f(x,y)$ 的小波变换将取决于 4 个参数：a，b_x，b_y 和 θ，其中 θ 为坐标在平面上的方位角，b_x 和 b_y 为对应不同方向的平移因子。在 n 维欧几里德空间的情况下，连续信号表为 $f(x)$，小波基函数表为

$$h_{a,b,\theta}(x) = a^{-n/2}h\left[\Omega^{-1}(\theta)\frac{x-b}{a}\right] \tag{7.163}$$

其中 Ω 为在 R^n 空间旋转群的 Euler 角。相应的小波变换定义为

$$W_{a,b,\theta}(f) = \int_{\mathrm{R}^n} h^*_{a,b,\theta}(\boldsymbol{x})f(\boldsymbol{x})\mathrm{d}^n\boldsymbol{x} \tag{7.164}$$

其中 h^* 为 h 的复共轭。

连续小波变换具有能量守恒性质

$$\int_{-\infty}^{\infty} |f(x)|^2\mathrm{d}x = \frac{1}{W_h}\int_0^{\infty}\int_{-\infty}^{\infty} |W_{a,b}(x)||W^*_{a,b}(x)|\frac{\mathrm{d}a\mathrm{d}b}{a^2} \tag{7.165}$$

同时，函数微分的小波变换满足

$$W\left[\frac{\partial^n f}{\partial x^n}\right] = (-1)^n\int_{-\infty}^{\infty} f(x)\frac{\partial^n}{\partial x^n}[h^*_{a,b}(x)]\mathrm{d}x \tag{7.166}$$

(二)离散小波变换(DWT)与多尺度分解

在(7.160)式中首先对伸缩与平移参数作以下离散化

$$a = a_0^m, a_0 > 0, b = nb_0a_0^m \tag{7.167}$$

其中 m 和 n 为整数。由(7.159)有

$$h_{m,n}(x) = a_0^{-m/2}h(a_0^{-m}x - nb_0) \tag{7.168}$$

代入(7.160)式可把离散小波变换可写为

$$W_{m,n}(f) = [f(x), h_{m,n}(x)] = \int_{-\infty}^{\infty} f(x) h_{m,n}(x) \mathrm{d}x \tag{7.169}$$

而相应的信号重建公式为

$$f(x) = \sum_{m} \sum_{n} [f(x), h_{m,n}(x) h_{m,n}(x)] = \sum_{n=-\infty}^{\infty} W_{m,n}(f) * h_{m,n}(-x) \tag{7.170}$$

在离散情况下，由于 m 和 n 实际上都只能是有限的，很难找出真正正交的子波集，来完成函数的分析与重建。为解决此问题，Daubechies 提出用迭代方法求出以下形式的近似正交系

$$\psi_{j,i}(x) = 2^{j/2} \psi(2^j x - i) \tag{7.171}$$

其中 i 和 j 为整数，作为拟正交子波集，使 $\psi(x)$ 正交于步长 $b=2^{-j}i$ 的平移和 $a=2^j$ 的伸缩。因而，对支集较紧的实函数 $f \in L^2$，可进行以下小波分析

$$f(x) = \sum_{j=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} W_{j,i}(f) \psi_{j,i}(x) \tag{7.172}$$

详细算法见本节第三段。

Mallat(1989)提出了多尺度分析的概念，他注意到在离散情况下拟正交小波集只是 L^2 空间的一个子空间集，还有信息落在它的补空间，即(7.172)式还有遗失信息。他定义了以下"尺度函数"中，满足

$$\int_{-\infty}^{\infty} \phi(x) \mathrm{d}x = 1 \tag{7.173}$$

及

$$\phi_{ji} = 2^{j/2} \phi(2^j x - i) \tag{7.174}$$

当 $j' > j$ 和 $i' > i$ 时，ϕ_{ji} 正交于 $\phi_{j'i'}$。因此，尺度函数 ϕ_{ji} 产生了一个子空间集

$$V_0 \subset V_1 \subset \cdots \subset V_j \subset V_{j+1},$$

它伴随拟正交小波集 ψ_{ji}，并张在它的正交补空间 $W_0, W_1, \cdots, W_j, W_{j+1}$，使

$$V_{j+1} = V_j \oplus W_j。 \tag{7.175}$$

相应地，小波分解可写为

$$f(x) = f_0(x) + \sum_{j=0}^{\infty} f_j(x) \tag{7.176}$$

式中右边第一项为对应尺度函数 ϕ 展开的连续光滑逼近，而 f_j 为与 ψ_{ji} 对应的第 j 阶小波细节，表示不同尺度的分解信号成分。其中，$f_1(x)$ 为一阶小波细节，为尺度最小的信号成分，随 j 增大 $f_j(x)$ 的尺度加大。于是，信号 $f(x)$ 被分解为不同尺度的小波细节 $f_j (j=1,2,\cdots)$ 和光滑逼近 f_0，这个过程称为信号的多尺度分析，在地球物理中有广泛应用。

下面再讨论离散小波变换的算法。

(三)Daubechies 小波变换系数[43,44]

从(7.169)可知，离散小波变换可视为一种线性滤波器。如果滤波器系数矩阵 $\boldsymbol{H}$ 是一个正交矩阵，则很容易由(7.172)式重建信号，比利时数学家 Daubechies(1988)按(7.171)的思想首次实现了一组拟正交子波集的迭代逼近，并给出了一组小波变换系数矩阵，称为 Daubechies 小波滤波器系数。下面就以其中最局部化的系数 DAUB4 来介绍此系数组的计算方法。

局部性最好的系数矩阵每行只有 4 个非线元素(c_0,c_1,c_2,c_3)，隔行改变 c_2 和 c_0 的符号，如下式所示：

$$\begin{bmatrix} c_0 & c_1 & c_2 & c_3 & & & & & & & & \\ c_3 & -c_2 & c_1 & -c_0 & & & & & & & & \\ & & c_0 & c_1 & c_2 & c_3 & & & & & & \\ & & c_3 & -c_2 & c_1 & -c_0 & & & & & & \\ & & & & & & \ddots & & & & & \\ & & & & & & & c_0 & c_1 & c_2 & c_3 & \\ & & & & & & & c_3 & -c_2 & c_1 & -c_0 & \\ c_2 & c_3 & & & & & & & & c_0 & c_1 & \\ c_1 & -c_0 & & & & & & & & c_3 & -c_2 & \end{bmatrix} \tag{7.177}$$

每两项右移二个单元表示出平移，而伸缩反映在每行元素的增加上，要作 z^l 的变化，对 DAUB4 来说是固定的。如果偶数行不改变系数符号并只右移一个单元，(7.176)蜕化为通常的褶积而且可用 FFT 方法作滤波，现在奇数行与偶数行有符号反转的区别，而且不平移，看上去好像把两套褶积混合在一起。这种设计思想体现了(7.175～7.176)式的多尺度分析。矩阵中奇数行类似于一个滑动平均的四点圆滑滤波，对应着尺度函数 ϕ。随着符号的改变，偶数行不再是圆滑滤波，而与提取小波细节的小波函数 ψ 相关，将偶数行系数的各阶矩取为零就可以实现这一点。这样组合的滤波器也称为二次镜像数字滤波器。

设信号 f 离散取样后有 N 个数据，它的重建应包含 $N/2$ 个圆滑的成分和 $N/2$ 细节成分。要使矩阵(7.177)正交，重建用的逆矩阵应是它的转置：

$$\begin{bmatrix} c_0 & c_3 & & & & & & & c_2 & c_1 \\ c_1 & -c_2 & & & & & & & c_3 & -c_0 \\ c_2 & c_1 & c_0 & c_3 & & & & & & \\ c_3 & -c_0 & c_1 & -c_2 & & & & & & \\ & & & & \ddots & & & & & \\ & & & & c_2 & c_1 & c_0 & c_3 & & \\ & & & & c_3 & -c_0 & c_1 & -c_2 & & \\ & & & & & & c_2 & c_1 & c_0 & c_3 \\ & & & & & & c_3 & -c_0 & c_1 & -c_2 \end{bmatrix} \tag{7.178}$$

它是 DAUB4 离散小波逆变换矩阵。

由于(7.178)是(7.177)的逆矩阵，应有

$$c_0^2 + c_1^2 + c_2^2 + c_3^2 = 1,\ c_2c_0 + c_3c_1 = 0 \tag{7.179}$$

为决定 4 个系数我们还需两个方程，这可以由二阶矩为零的附加条件取得：

$$c_3 - c_2 + c_1 - c_0 = 0,\ 0 \cdot c_3 - 1 \cdot c_2 + 2 \cdot c_1 - 3c_0 = 0 \tag{7.180}$$

解由(7.179)和(7.180)组成的方程组便求出 DAUB4 系数组

$$c_0 = (1 + \sqrt{3})/4\sqrt{2}, c_1 = (3 + \sqrt{3})/4\sqrt{2},$$
$$c_2 = (3 - \sqrt{3})/4\sqrt{2}, c_3 = (1 - \sqrt{3})/4\sqrt{2}. \tag{7.181}$$

对于局部性放宽的其它系数组可用类似的步骤求得，如对 DAUB6，系数有 6 个，此时要

加上三阶矩为零的附加条件导出 6 个方程。最后解出 DAUB6 系数组为

$$c_0=(1+\sqrt{10}+\sqrt{5-2\sqrt{10}})/16\sqrt{2},\ c_1=(5+\sqrt{10}+3\sqrt{5+2\sqrt{10}})/16\sqrt{2},$$

$$c_2=(10-2\sqrt{10}+2\sqrt{5+2\sqrt{10}})/16\sqrt{2},\ c_3=(10-2\sqrt{10}-2\sqrt{5+2\sqrt{10}})/16\sqrt{2},$$

$$c_4=(5+\sqrt{10}-3\sqrt{5+2\sqrt{10}})/16\sqrt{2},\ c_5=(1+\sqrt{10}-\sqrt{5+2\sqrt{10}})/16\sqrt{2}. \tag{7.182}$$

在 Daubechies(1988)的文章中已算出了各种高阶矩的系数组[43]。因此类似于(7.177)和(7.178)的小波分析和重建矩阵都已被计算出来。

小波是什么样子呢？用一个单位向量 $\boldsymbol{e}$ 乘以(7.178)作 DWT 重建，就可得到小波函数。图 7.3 给出 DAUB4 和 DAUB20 小波函数，所用单位向量的长度为 1024。对 DAUB4 小波，单位向量第 5 个元为 1，其它为零，对 DAUB20 小波，第 24 个元素为 1，其余为零。这些小波都有高阶连续导数。但是其中 DAUB4 小波在某些点能向左求导，而不能向右求导。

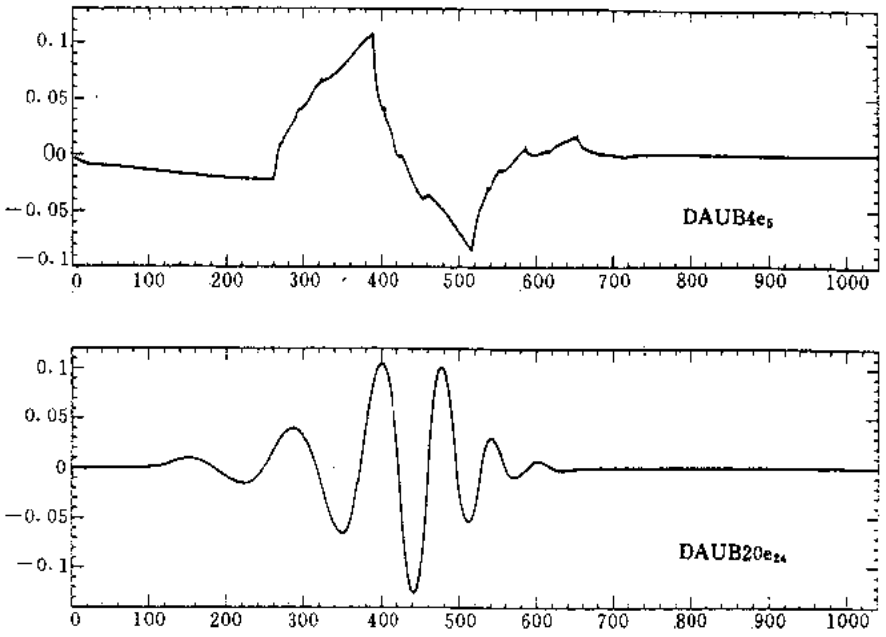


图 7.3　小波家族中的二个小波函数

DAUB4 和 DAUB20，e_5 为单位向量(在第 5 个元为 1)，e_{24} 为单位向量(第 24 个元为 1)

(四)离散小波变换的金字塔算法

设数据为 $\{y_i, i=1, N\}$，其中 N 为 2 的整数幂。由 Mallat(1989)等提出的金字塔式算法首先将它乘以小波变换滤波矩阵，如(7.177)等，按奇偶号重排得到长为 $N/2$ 的圆滑向量和另一半元构成的高阶小波细节。将此圆滑向量再乘以 $N/2\times N/2$ 的小波矩阵，又得到长为 $N/4$ 的再圆滑向量和低一阶的小波细节，$N=16$ 时如下式所示。

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \\ y_{10} \\ y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{15} \\ y_{16} \end{bmatrix} \overset{\times M_{16}}{\Rightarrow} \begin{bmatrix} s_1 \\ d_1 \\ s_2 \\ d_2 \\ s_3 \\ d_3 \\ s_4 \\ d_4 \\ s_5 \\ d_5 \\ s_6 \\ d_6 \\ s_7 \\ d_7 \\ s_8 \\ d_8 \end{bmatrix} \xrightarrow{\text{permute}} \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \\ s_6 \\ s_7 \\ s_8 \\ d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \\ d_6 \\ d_7 \\ d_8 \end{bmatrix} \overset{\times M_{8\times 8}}{\Rightarrow} \begin{bmatrix} S_1 \\ D_1 \\ S_2 \\ D_2 \\ S_3 \\ D_3 \\ S_4 \\ D_4 \\ d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \\ d_6 \\ d_7 \\ d_8 \end{bmatrix} \xrightarrow{\text{permute}_{8\times 8}} \begin{bmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ D_1 \\ D_2 \\ D_3 \\ D_4 \\ d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \\ d_6 \\ d_7 \\ d_8 \end{bmatrix} \overset{etc.}{\Rightarrow} \begin{bmatrix} S_1 \\ S_2 \\ D_1 \\ D_2 \\ D_1 \\ D_2 \\ D_3 \\ D_4 \\ d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \\ d_6 \\ d_7 \\ d_8 \end{bmatrix}, \tag{7.183}$$

式中右边第一列为数据，它乘上如(6.177)的16×16的小波变换滤波矩阵之后示如第二列，其中奇数行 s_j 为圆滑滤波的元素，偶数行为小波细节数据 d_j。把 s_j 依序排在前8个元，而 d_j 重排到后8个元，它相当于一阶小波细节，要在迭代中保留起来。然后对前8个经首次光滑的元再做小波变换，即乘以8×8的相应小波系数矩阵，然后再进行重排，结果取得二次光滑成分 s_{1-4} 和二阶小波细节 D_{1-4}，它与一阶小波细节 d_{1-4} 一起被保存起来。最后，对 s_{1-4} 再做小波分析，即乘以4×4的系数矩阵，取得的 s_1 和 s_2 为信号 y_i 的最光滑(线性)逼近。当然，实际数据的取样数一般较大，即 $N \gg 16$，相应小波系数矩阵虽然很大，但由于系数相同只是符号及位置变化，不必存储整个矩阵。相应迭代过程也是类似的，只不过迭代次数多一些罢了。具体需要迭代多少次，取决于数据的数量和数据处理的目的，不一定要做到数据的线性逼近。

由于整个过程是由线性正交算子的乘法组合而成，这样进行的离散小波分析是正交的，而逆变换(Inverse DWT)只需将(7.183)的过程从右向左倒转进行就可以了。当然，使用的矩阵是如(7.178)的正交逆矩阵，即小波系数矩阵的转置。

从(7.177)和(7.178)两个矩阵中可以看出，由于数据有限，在边界上对数据向量作了折叠，即最后几个小波系数会同时受数据向量的头和尾的影响。对短的数据这种影响是要考虑的。

小波变换在地球物理中的应用还刚刚开始，目前已用于异常分解及去噪、数据压缩及提高分辨处理，以及奇性检测和信号增强上[47]。随着时间的推移，它在地球物理上的应用将逐渐拓开，尤其是对数据采集和存储，有可能使引起数字化以来的一次新的革命。

第八章　地球物理中的偏微分方程反演

经过前两章的讨论之后，我们将继续第3～5章的讨论，对地球物理反演问题作进一步的深入剖析。

傅承义先生曾说过："地震像一盏灯，它发生时可以照亮地球内部"。要想通过地震波看到地球内部，还必须戴上一付"魔镜"，这就是反演。反演把波场记录转化为地球内部的地震波速结构，使我们看到碰撞造山带造成的地壳加厚和缩短，裂谷或某些沉积盆地下方的地幔的上隆和流体的活动，以及大型韧性剪切带和推复体的存在。然而，反演的多解性又使人们迷惑不解，为什么同样的数据集合得出不同的反演结果？因此，我们要先讨论反演的一些理论问题，明确反演的内涵和局限性。

一、广义解与正则化

（一）反问题的广义解

在第四章中，我们曾对正问题和反问题作过讨论。所谓正问题，指给定所有可能的地球参数的集合（即地球模型）和第六章中讲述的拟微分算子，计算合成的地球物理场数据集合。与此对应，反问题指的是由观测数据和正问题的拟微分算子求地球模型。我们知道，对于正问题而言，拟微分算子是由物理上的本构方程和运动方程推导出来的确定性算子，不存在争议的余地。但是对于反问题而言，我们并不知道对应正问题的逆算子是否存在，是否唯一，以及是否具有使计算发生困难的奇异性质。因此，即使对于无限精确的合成数据，也必须研究反问题的解的存在性、唯一性以及构造逆算子的方法。这方面的研究称为精确反演，主要是数学家感兴趣。

按照数学家的逻辑思维，他们可以按正问题算子的不同和理想的初始-边界条件对反问题分成许多类，如双曲型二阶偏微分方程反问题、抛物型或椭圆型二阶偏微分方程反问题等等，然后逐一证明反问题解的存在性、唯一性和下面要谈到的适定性等等。这些研究导致一些能起指导作用的定理或引理。例如，得出这样的定理："如果正问题的映射为模型空间和数据空间之间的双向同胚映射，则对应反问题是适定的"。然后，你得去证明有关算子是否双向同胚算子映射。

地球物理学家的思维与数学家有很大区别，他们很清楚数据的观测是需要大量资金和时间的，仪器精度再高数据中也必定有观测误差和环境噪音的影响。因此，他们想的是：现在观测到了有限和不精确的数据集，如何从数据中提取尽可能多的关于地球内部的信息？他们坚信，即使有关的反问题的精确解不存在或不唯一，数据中仍然含有地球信息，而且可以被提取出来。

经过多年的研究，数学家与地球物理学家在以下这一点上已经取得了共识：实际反问题的求解同时不可避免地制造一些人造假象，而控制这些人造假象应该是反问题研究的一个

主要内容。至于人造假象的来源，数学家归结为反问题的不适定性，而物理学家归结为信息的不足和沿时间轴正演与反演的非对称性（见第一章）。因此，作为指导反问题研究的理论，反演的内涵已扩大为：从数据中提取尽可能多信息的同时，控制求解过程中产生的人造假象[48]。

根据对上述反演内涵的共识，许多数学家和地球物理已不再把重点放在反问题的精确解上，而在考虑对实际数据寻找近似或等效的解估计（即广义解）上。

反问题的广义解可分为以下三大类。

(1)拟解(Quasisolution)

拟解是由吉洪诺夫等人最早提出的一种广义解[49]，它考虑到实际数据含有误差，用合成数据与观测数据失拟最小来定义。

设 $\tilde{d}$ 为一实际数据集，$d=A(m)$ 为对应的理想数据集，其中 A 为正问题的映射算子。在数据空间中使 $\tilde{d}$ 和 d 之间的距离最小的广义解称为拟解。

记 $\mathrm{dis}(x,y)$ 为 Hilbert 空间中两个元 x 和 y 之间的距离，它可定义为

a. 上确界距离　$\mathrm{dis}(x,y)=\sup|x(t)-y(t)|$

b. 二次距离　$\mathrm{dis}(x,y)=\left\{\int_a^b|x(t)-y(t)|^2\mathrm{d}t\right\}^{1/2}$

c. 加权二次距离　$\mathrm{dis}(x,y)=\left\{\int_a^b|x(t)-y(t)|^2W(t)\mathrm{d}t\right\}^{1/2}$

d. 矩权二次距离　在(c)中令 $W(t)=|t-t_0|^k$，$k=1,2,\cdots$

理论研究得出有关拟解的以下两个定理，对应用具有指导意义。

定理 1：设模型空间 M 为一紧集，A 为一非线性连续映射，p 为数据空间 D 向 $A(M)$ 的投影。如果对任何精确数据 $d\in A(M)$ 方程 $Ax=d$ 有一个解，则对有误差实际数据 $\tilde{d}$ 方程 $Ax=\tilde{d}$ 有一个拟解

$$x=A^{-1}[p(\tilde{d})], \tag{8.1}$$

它连续地依赖于 $\tilde{d}$。

这个关于拟解存在性的定理，对于线性反问题，定理 1 有更精确的表示。

定理 2：假设模型空间 M 和数据空间 D 都是向量赋范空间的一部分，M 是凸紧集，A 是连续的线性映射，而且 $Ax=0$ 只有零解，在 D 中任何超球都是凸的，则对于任何实际数据 $\tilde{d}$，方程 $Ax=\tilde{d}$ 有一个连续依赖于 $\tilde{d}$ 的拟解，它定义为

$$x=A^{-1}[p(\tilde{d})] \tag{8.2}$$

这个定理说清了对实际资料线性反问题拟解存在的条件。

以数据拟合为最佳准则求出的广义解都是拟解，如对线性反问题的最小二乘解估计(3.19)就是例子。在第三章和第四章中我们已讨论过，这样的拟解对应分辨率高的解估计，但对方差等人造因素没有任何控制。当加入对方差的控制项时得到的解估计只能是近似的，这就是第二类广义解，即近似解。

(2)近似解(Approximate solutions)

凡是用某种确定性泛函的极小化为准则，并结合数据拟合质量定义的广义解，都属于近似解。

用某种距离来定量地表示数据拟合质量与拟解的定义相同。设 $\mathrm{dis}_D(\tilde{d},Ax)$ 为数据空间的某种距离（如加权二次距离），$\mathrm{dis}_M(x,x')$ 为模型空间中的某种距离，λ 为 Lagrange 乘子，

可选目标泛函为

$$\phi(x) = \mathrm{dis}_D(\tilde{d}, Ax) + \lambda \mathrm{dis}_M(x, x'), 0 \leqslant \lambda < 1 \tag{8.3}$$

它的极小化便是一种近似解。当(8.3)中取二次距离并令 $x=x'$ 便有线性反问题的最小范数解(3.24)。当取 $\lambda=0$ 时由(8.3)为准则取得的近似解蜕化为拟解。例如，以脉冲度准则取得的广义解就属于拟解的范畴(见第四章第四节)。

(3)等效解(Equivalent solutions)

在讨论向下延拓问题的(7.144)式中，我们曾令失拟泛函等于一个小数 δ^2(有限失拟)。因此，我们可以用以下形式扩展解空间，即令目标泛函

$$\phi(x) \leqslant \delta^2 \tag{8.4}$$

定义一种广义解，它即是一种等效的解估计。在近似解中，我们假定 $\phi(x)$ 的全局极小是存在的，因此近似解是唯一的；而等效解是非唯一的，因为(8.4)式已把 x 的解空间扩展了。在实际数据无论在数量还是在质量上都很糟的情况下，等效解也可以提取出其中的一些信息，但比近似解的可信度差一些。

地球物理反演中广泛应用的是第二种广义解——近似解，它可用不适定问题的正则化方法导出来。在第七章第四节曾提到不适定问题及正则化，下面要对此作进一步讨论。

(二)不适定问题的正则化

数学上的不适定问题是 J. Hadamard 在研究偏微分方程解的稳定性时提出来的。对于给定的偏微分方程(组)和定解条件，定解问题中解的存在性、唯一性和稳定性是三个理论上的基本问题。定解问题解的稳定性指对于某个定解问题如果存在唯一的解，而且在定解条件中原始资料作微小变化时解也仅作微小的变化。因此，偏微分方程(组)的合理的定解问题应当满足解的存在、唯一和稳定三个要求，并将存在性、唯一性和稳定性统称为定解问题的适定性。凡是给定的条件不满足上述三个适定条件中的任一个的定解问题称为不适定问题(ill-posed problem)，即数学上提法不合适的问题，否则称为适定问题(well-posed problem)(以上关于解的稳定性的概念并不能作为一个数学定义，因为还没有对所谓的"微小变化"作定量规定。稳定性的概念可以进一步用拓扑理论来说明，即定解问题的稳定性依赖于原始资料空间和解空间中拓扑的选择。在这里拓扑指的是拓扑空间的某种极限关系。为了避免涉及数学上拓扑学的内容，我们不再深入讨论稳定性的拓扑解释，而只要求注意到拓扑学是研究抽象空间中几何物体附贴性质而不是研究空间的距离特性就可以了)。

例如，在数学物理问题中 Laplace 方程的 Direchlet 边值问题

$$\begin{cases} \nabla^2 u(x,y) = 0, (x,y) \in \Omega(\mathrm{R}^2) \\ u(x,y) = d(x,y), (x,y) \in S \end{cases} \tag{8.5}$$

是适定的，其中平面区域 Ω 的边界 S 是封闭的。这个边值问题解的存在性和唯一性已经得到了证明。由极值原理知，如果调和函数 u 在某一区域内具有极值，则此极值必定出现在边界上。因此，当数据有微扰 δd 时，必定有 $|\delta u_{\max}| \leqslant |\delta d_{\max}|$，说明解连续依赖于数据。

椭圆型方程的柯西问题是最早由 Hadamard 研究的不适定问题，表示为

$$\begin{cases} \nabla^2 u(x,y) = 0 \\ u(x,0) = 0 \\ \dfrac{\partial u(x,0)}{\partial y} = \dfrac{1}{n}\sin(nx) \end{cases} \tag{8.6}$$

这个问题的解存在，可表为

$$u(x,y)=\frac{1}{n^2}\sin(nx)\mathrm{sh}(ny) \tag{8.7}$$

但是，当 $n\to\infty$时，$\sin(nx)/n^2\to 0$，而 $\mathrm{sh}(ny)\to\infty$。因此数据项 $\sin(nx)/n$ 的微小变化可能导致 $u(x,y)$的很大变化，此问题是不适定的。以上例子说明正问题也可能不适定，反问题不适定的情况就更多了。

我们还是用距离空间为背景来研究反问题的稳定性。通常正则方法不必象 BG 理论那样把模型及数据限制在希尔伯特空间(即可以不必规定内积)，而只要求它们处在某种 Banach 空间就可以了。如果和第四章第二节中一样，把反问题归纳为以下第一类 Fredholm 积分方程

$$\int_a^b G(x,y)m(y)\mathrm{d}y=d(x),e\leqslant x\leqslant f \tag{8.8}$$

其中 m 和 d 属于某个 Banach 空间(由于上式即为内积形式，实际上可规定 m 和 d 处于希尔伯特空间)。不难证明，方程(8.8)的求解是不稳定的，即当 $d(x)$有微小的误差时解得 $m(y)$的误差可能非常大。实际上，假定 α 是一个任意的数，由 Riemann 引理可知当 $n\to\infty$时有

$$\alpha\int_0^\pi G(x,y)\sin(ny)\mathrm{d}y\to 0 \tag{8.9}$$

将(8.8)和(8.9)相加，有

$$\int_0^\pi G(x,y)[m(y)+\alpha\sin(ny)]\mathrm{d}y\to d(x),n\to\infty \tag{8.10}$$

当 $\alpha\to\infty$时方程(8.10)的解为

$$m(y)+\alpha\sin ny\to\infty$$

而此时由(8.2)式知数据项

$$\tilde{d}(x)=d(x)+\alpha\int_0^\pi G(x,y)\sin ny\mathrm{d}y\to d(x) \tag{8.11}$$

说明数据的微小变化对应方程(8.8)的解的变化可能为无穷大。因此，第一类 Fredholm 积分方程的求解属于不适定问题。

从数学考虑，问题的不适定性是由算子的奇异性引起的。根据某种合理的准则修改算子以压制奇性，并使计算稳定的理论称为正则化。

稳定性指初始数据微小的变化对应广义解也作微小变化，或者说解连续地依赖于数据。栾文贵把这种连续依赖的性质分为三种类型[38]：

1. 适定型，设数据扰动为 δ，解的扰动为 δ_m，δ_m 正比于 δ，即 $\delta_m\propto\delta$。

2. Hölder 型：$\delta_m\propto\delta^\alpha$，$0<\alpha<1$

3. 对数型，$\delta_m\propto(\ln\frac{1}{\delta})^{-\alpha}$

其中 Hölder 型的依赖关系使计算处于临界状态，对数型计算不稳定。对不具稳定性的不适定问题，可用以下正则化方法使计算稳定。

(三)Miller 正则化方法

设数据的误差为 δ 并具有可加性，因此(8.8)可用 Fredholm 算子 G 写为以下形式

$$Gm=d+\delta \tag{8.12}$$

由于 m 和 δ 都是未知函数，我们不来求方程 $Gm=d$ 的精确解，而把问题改写为：设 ε 为正比

于误差的范数$\|\delta\|$的给定实数，求 m 使它在数据空间 $\mathscr{B}_d$ 满足

$$\|Gm - d\| \leqslant \varepsilon \tag{8.13}$$

同时在模型空间 $\mathscr{B}_m$ 满足

$$\|Bm\| \leqslant E \tag{8.14}$$

其中 $G:\mathscr{B}_m \to \mathscr{B}_d$ 为连续算子，$B:\mathscr{B}_m \to \mathscr{B}_m$ 为一线性算子，它具有连续有界的逆算子 B^{-1}，E 为给定实数。如果 B 为单位算子则(8.14)式就是有限能量约束的一种形式。由于(8.12)的解是不适定的，允许解集相当于非紧集 $\mathscr{H}$，而(8.13)及(8.14)的条件规定了子空间 $\mathscr{P}$，如果 ε 和 E 选择得适当，则 $\mathscr{H}(G)$ 和 $\mathscr{P}$ 的交 $\mathscr{K}$ 可化为紧集，问题(8.12)转化为(8.13)和(8.14)的条件适定问题。

在 Miller 正则化方法中，把(8.13)和(8.14)的求解转化为求以下泛函极小

$$\phi(\alpha, m) = \|Gm - d\|^2 + \alpha\|Bm\|^2 \tag{8.15}$$

式中 α 为某一正实数。由于 B^{-1} 为有界线性算子，ϕ 的极小化问题有唯一解，可记为 $\hat{m}_\alpha$，令

$$\varepsilon_\alpha = \|G\hat{m}_\alpha - d\|^2 \tag{8.16}$$

$$E_\alpha = \|B\hat{m}_\alpha\|^2 \tag{8.17}$$

可以证明当 α 从0变到∞时 ε_α 连续地增大，而 E_α 连续地减小，因此可以适当地选择 α 使(8.13)和(8.14)同时得到满足。当选 $\alpha<1$时相当于把(8.17)式的第二项作为约束，而在 $\alpha>1$时把第一项作为约束。

为了选择合适的 ε 和 E，我们可以对紧集 $\mathscr{K}$ 的形状取为简单的对称流形。如果令 $\alpha=(\varepsilon/E)^2$，则(8.15)式变为

$$\phi(m) = \|Gm - d\|^2 - (\varepsilon/E)^2\|Bm\|^2 \tag{8.18}$$

于是，当 $\phi(m)\leqslant\varepsilon^2$时，(8.18)式变为

$$\frac{\|Am - d\|^2}{\varepsilon^2} + \frac{\|Bm\|^2}{E^2} \leqslant 1$$

说明解 $\hat{m}$ 是位于 $\mathscr{K}$ 内部的一个无限维超椭球，记为 $\mathscr{K}_0 \subset \mathscr{K}$。类似地当 $\phi(m)\leqslant 2\varepsilon^2$ 时有超椭球集 $\mathscr{K}_1$，它含了紧集 $\mathscr{K}$(见图8.1)。只要 $\phi(\hat{m})\leqslant\varepsilon^2$则 $\hat{m}$ 就位于 $\mathscr{K}$ 之内。

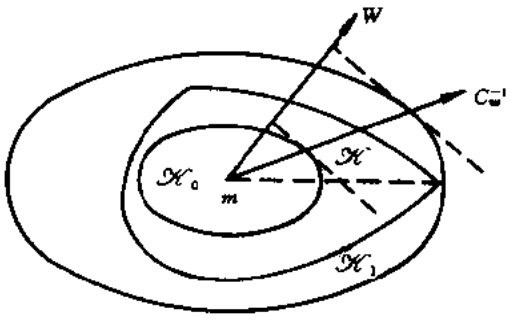


图 8.1 集 $\mathscr{K}$，$\mathscr{K}_0$和 $\mathscr{K}_1$的示意图，$\mathscr{K}_0$和 $\mathscr{K}_1$表示似超椭球

由于集 $\mathscr{K}_0$和 $\mathscr{K}_1$具有简单的几何结构，我们来考虑以下新算子

$$C = C^*G + (\varepsilon/E)^2B^*B \tag{8.19}$$

其中 G^* 为 G 的伴随算子，B^* 为 B 的伴随算子。由线性算子的定义可知，算子 $C:\mathscr{B}_m \to \mathscr{B}_m$ 是正定的，即对任意 m 有$(Cm,m)>0$；C 也是自伴随的，即 $C^*=C$。此外，C 具有连续的逆算子 C^{-1}(或者说 C 是正则化算子)，并可把解估计表为

$$\hat{m} = C^{-1}G^*d \tag{8.20}$$

实际上，利用算子 C 的性质和(8.19)及(8.20)式可以把(8.18)式用内积写为

$$\phi(m) = (C[m - \hat{m}], (m - \hat{m})) + \|d\|^2 - (d, G\hat{m}), \tag{8.21}$$

此式说明解估计(8.20)实际上是使以下泛函

$$\phi(\hat{m}) = \|d^2\| - (d, G\hat{m}) \tag{8.22}$$

极小化的结果，由(8.15)可知它相当于最小二乘最小范数解估计的一个变种。在利用 Miller 正则化求(8.20)的解估计时，先由先验知识给出(8.14)式中算子 B 和常数 E 并给出允许的数据拟合差的范数 ε，然后由(8.19)式确定正定自伴随算子 C 及它的逆 C^{-1}，最后由(8.20)式确定解估计 $\hat{m}$。由于在 $\phi(\hat{m}) \leqslant \varepsilon^2$ 时 $\hat{m}$ 位于紧集 $\mathscr{K}$ 之内，$\hat{m}$ 与数据 d 有连续的依赖关系，算法必定是稳定的。比较 BG 方法的(4.121)和(8.18)可知，Miller 正则化中的比值 $\alpha = (\varepsilon/E)^2$ 相当于折衷参数，可以通过改变 α 的方法得到一组解估计。当 α 越小时解估计的总体分辨率提高，即解估计越接近于连续地球模型的最小二乘或脉冲度准则解估计，而 α 加大时分辨率和误差都降低，趋向于最小范数解估计。要注意 BG 理论中解估计的分辨率和高斯分布的方差是逐点计算的，而正则化方法中只用 ε 来限制误差的范数，并未提出对解估计定量评价的方法。

(四)Tikhonov 正则化原理

这一类正则化方法的思想是对模型 m 规定一种稳定化的泛函 $\Omega(m)$，它对于任意正实数 E^2 满足

$$\Omega(m) \leqslant E^2 \tag{8.23}$$

对比(8.17)可知，正则化泛函 $\Omega(m)$ 规定了第一类 Fredholm 积分方程(8.8)中解 m 的极限条件，即把允许解集限制在紧集之中，在求解计算中起到稳定作用。实际上，对方程(8.8) Tikhonov 选取的正则化泛函为

$$\Omega(m) = \int_a^b (p(y)|m'(y)|^2 + q(y)|m(y)|^2)\mathrm{d}y \tag{8.24}$$

式中 $p(y)$ 和 $q(y)$ 为权函数，只取正值，而 $m'(y)$ 为 m 的导数。由于(8.23)和(8.24)中包括了对解的导数的范数(光滑度)的限制，比单纯使用最小范数作限制起到了更强的稳定作用。再者，由上两式规定的解集只是 Banach 空间中的某些特殊的函数类型，称为正则化函数族。

现在我们要来找同时满足方程(8.8)和(8.23)的解估计。类似于拉格朗日乘子法的思想，首先构造出以下光滑泛函

$$\phi_\alpha(m) = \frac{1}{2}\int_c^f \left\{\int_a^b G(x,y)m(y)\mathrm{d}y - d(x)\right\}^2 \mathrm{d}x + \frac{\alpha}{2}\Omega(m) \tag{8.25}$$

其中任意常数 $\alpha > 0$ 称为正则化因子。然后，用变分法求泛函(8.25)达到极小的解估计 $\hat{m}(y)$。由于解估计依赖于正则化因子 α，也可以把它写为 $\hat{m}_\alpha(y)$。泛函 ϕ 取极小的必要条件是它的一次变分为零，即

$$\delta\phi_\alpha(m) = 0$$

代入(8.25)及(8.24)可得

$$\begin{aligned}\delta\phi_\alpha(m) = {} & \alpha\int_a^b [Lm(y)]\delta m(y)\mathrm{d}y \\ & + \int_a^b \left[\int_a^b G^{\mathrm{T}}G(y,\tau)m(\tau)\mathrm{d}\tau\right]\delta m(y)\mathrm{d}y \\ & - \int_a^b \left[\int_c^f d(x)G(x,y)\mathrm{d}x\right]\delta m(y)\mathrm{d}y \\ & + p(y)m'(y)\delta m(y)\Big|_a^b\end{aligned} \tag{8.26}$$

其中 L 为以下微分算子

$$Lm(y) \equiv \quad (p(y)m'(y))' + q(y)m(y)$$

而

$$G^{\mathrm{T}}G(y,\tau) = \int_c^d G(x,y)G(x,\tau)\mathrm{d}x \tag{8.27}$$

令(8.26)的变分为零，我们得到以下柯西边值问题的微-积分方程式

$$\alpha Lm(y) + \int_a^b G^{\mathrm{T}}G(y,\tau)m(\tau)\mathrm{d}\tau = \int_c^d d(x)G(x,y)\mathrm{d}x \tag{8.28}$$

和相应的边值条件

$$pm'(y)|_a = pm'(y)|_b = 0 \tag{8.29}$$

用数值方法(如差分和数值积分)求解上述柯西边值问题，便可以获得(8.8)的正则化解估计 $\hat{m}_\alpha(y)$。

从以上讨论可知，Tikhonov 正则化方法的特点是考虑了解满足的光滑条件。在第三章和本章讨论的其它方法都没有运用这种条件，仅仅 BG 展布准则是个例外。在 BG 展布中用平均核的二阶矩作为极小化的泛函，相当于对解估计起到光滑作用。Tikhonov 正则化考虑的是总体的光滑性。如果在(8.24)中令 $p=1$和 $q=0$，则正则化泛函变为

$$\Omega(m) = \int_a^b \left(\frac{\mathrm{d}m(y)}{\mathrm{d}y}\right)^2 \mathrm{d}y \tag{8.30}$$

它表示 $m(y)$的总体光滑度。当 $\alpha<1$时，由(8.25)和(8.30)可知这时正则化解估计相当于以总体光滑度为约束的最小二乘解估计。从应用的观点来说，我们还希望了解上述正则化方法的收敛性，即解估计 $\hat{m}_\alpha(y)$在什么条件下收敛于解 $m(y)$以及收敛的速度如何，这些问题可以在有关参考文献中找到[38]。Tikhonov 正则化方法最早由他本人应用于地球物理中位场资料的延拓问题[57]，采用了如同(8.30)型式的正则化泛函。

(五)小结

我们已讨论了实际资料反问题属于不适定问题，必须求助于广义解，而其中的近似解是目前应用最普遍的广义解。不适定问题来自算子自身固有的奇异性，而正则化思想是合理地修改算子使其奇异性得到改善。在一般情况下，反问题正则化是将

$$\Phi + \lambda\Omega \quad \text{极小化} \tag{8.31}$$

其中$0<\lambda<\infty$称为正则化因子或折衷因子。Φ 是度量数据拟合质量的泛函，它与映射的脉冲度(第四章第四节)或广义解的分辨率关系密切，当 Φ 被极小化时脉冲度提高，但解估计不稳定可能包含被指数放大了的数据误差的影响。Ω 称为稳定化泛函或正则化算子，它起到圆滑(压制奇性)的作用，当 然也带来降低分辨率与畸变等副作用。$\Phi+\lambda\Omega$ 是经过了正则化了的泛函，它包含了上篇中各反演方法的例子，如

最小范数广义解　$\Omega=x^{\mathrm{T}}x$

奇异值分解和 BG 最佳折衷广义解　$\Omega=\mathrm{Var}[x]$

其中 var 为方差。下面还要再讨论一种思路不同的广义解，其中

$$\Omega = H[x]$$

这里 H 为负熵。

二、最大熵法[50～52]

数据误差的随机性质，是实际数据反演时必须认真考虑和分析的，以最大熵为准则的反演方法具有独特之处。

(一)χ^2拟解

我们仍然以线性反问题开始来讨论数据为随机向量时拟合质量的问题。将(8.8)式改写为

$$\int G_i(y)m(y)\mathrm{d}y = d_i + \delta_i, \quad i = 1,2,\cdots,N \tag{8.32}$$

其中$\{d_i\}$为有限的理论数据，$\{\delta_i\}$为误差随机向量，它不为零向量时(8.32)式反演出的$\tilde{m}(y)$必定也有误差。那么，$\tilde{m}(y)$的误差是怎么分布的？

一般认为地球物理数据的测量误差属于高斯分布。经反演计算之后，地球模型的误差除了与数据的随机性质有关之外，还与算子的性质有关。从(4.119)知，对有限数据集反演的解估计的方差

$$\mathrm{Var}[\tilde{m}(y_0)] = \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i(y_0)\alpha_j(y_0)e_{ij} \tag{4.120}$$

其中$\{\alpha_i\}$为与算子有关的待定系数组，e_{ij}为数据$\{\tilde{d}_i\}$的协方差矩阵的元，大致可用d_i和d_j的乘积估计。(4.120)式提示我们，作为一种简化的统计模型，用χ^2分布来描述地球模型的随机性质也许是可以接受的。

我们知道，对一组服从高斯正态分布的随机取样(均值为$\bar{x}$，方差为σ^2)

$$x_1, x_2, \cdots, x_N$$

量χ^2定义为

$$\chi^2 = \frac{1}{\sigma^2}[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_N - \bar{x})^2] \tag{8.33}$$

而量χ^2服从以分布函数

$$F(\chi^2) = \frac{1}{\Gamma(\lambda)2^\lambda}\int_0^{\chi^2} u^{\lambda-1}\mathrm{e}^{-u/2}\mathrm{d}u \tag{8.34}$$

其中Γ为伽马函数，$\lambda = N/2$为自由度。所谓的置信度

$$W(\chi^2) = 1 - F(\chi^2) \tag{8.35}$$

将线性反问题(8.32)作离散化，将$\tilde{m}(x)$的离散取样记为随机向量

$$(m_1, m_2, \cdots, m_L)^{\mathrm{T}}$$

则(8.32)写为

$$\tilde{d}_i = \sum_{k=1}^{L} G_{ik}m(y_k) + \delta_i, \quad i = 1,\cdots,N \tag{8.36}$$

其中

$$G_{ik} = G_i(m_k)\Delta r = G_i(m_k)\frac{m_{k+1} - m_k}{2} \tag{8.37}$$

根据χ^2的定义，对于解估计$\{\tilde{m}_k\}$来说，量

$$\chi^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}\left[\tilde{d}_i - \sum_{k=1}^{L}G_{ik}\tilde{m}(y_k)\right]e_{ij}^{-1}\left[\tilde{d}_j - \sum_{k=1}^{L}G_{jk}\tilde{m}(y_k)\right] \tag{8.38}$$

当数据服从正态分布时，记 σ_i 为第 i 个数据的标准差，(8.38)退化为

$$\chi^2 = \sum_{i=1}^{N} \left[\frac{\tilde{d}_i - \sum_{k=1}^{L} G_{ik} \tilde{m}(y_k)}{\sigma_i} \right]^2 \tag{8.39}$$

如果将量 χ^2 取极小我们将得到一种广义解，即 χ^2 拟解。当(8.39)式中的 $\sigma_i \equiv 1$ 时，χ^2 拟解退化为一般的最小二乘解估计。在(8.39)中的权 $(1/\sigma_i)$ 的作用是：对误差大的数据在目标泛函中给予较小的权，以减少它们在优化计算中起的作用。

(二)解估计的概率

涉及到有误差的数据 $\{\tilde{d}_i\}$，我们再问一个问题：由它们反演取得的解估计 $\{\tilde{m}_j\}$，与精确数据 $\{d_i\}$ 反演出的模型 $\{m_j\}$ 相比，$\{m_j = m_j\}$ 的概率 $\mathrm{Prob}(\boldsymbol{m}/\tilde{\boldsymbol{d}})$ 是多少？显然，$\mathrm{Prob}(\boldsymbol{m}/\tilde{\boldsymbol{d}})$ 越大越好，它的极大化也许能提供一条实际反演的新准则。

回忆关于两个事件 A 和 B 发生的条件概率的 Bayes 定理

$$\mathrm{Prob}(A|B) = \mathrm{Prob}(A) \frac{\mathrm{Prob}(B|A)}{\mathrm{Prob}(B)} \tag{8.40}$$

其中 $\mathrm{Prob}(A|B)$ 为给定事件 B 发生的条件下事件 A 发生的概率，$\mathrm{Prob}(A)$ 和 $\mathrm{Prob}(B)$ 为事件 A 或 B 发生的无条件概率。

广义地看，事件 A 或 B 可视为某种假设，或某种推理，这时 $\mathrm{Prob}(A|B)$ 就与推理的似真度(Plausibility)有关。令 I 为某些采集的背景信息(事件 B)，H 为由其推理出的某种假说(事件 A)，则 $\mathrm{Prob}(H/I)$ 称为贝叶斯先验(Bayesian Prior)。当一组数据(如用于反演的地球物理数据)出现时，由 Bayes 定理(8.40)有

$$\mathrm{Prob}(H|D_a I) = \mathrm{Prob}(H|I) \frac{\mathrm{Prob}(D_a|HI)}{\mathrm{Prob}(D_a|I)} \tag{8.41}$$

此式右边分子项 $\mathrm{Prob}(D_a\ HI)$ 指对背景信息和假说，预测数据出现的概率，是可以计算的。右边分母项是一规格化常数，称为“先验预测概率”。

如果进行做实验的检测假说，取得另一组数据 D_c，则有

$$\mathrm{Prob}(H|D_c D_a I) = \mathrm{Prob}(H|D_a I) \frac{\mathrm{Prob}(D_c|HD_a I)}{\mathrm{Prob}(D_c|D_a I)} \tag{8.42}$$

根据概率乘积公式

$$\mathrm{Prob}(AB|C) = \mathrm{Prob}(A|C)\mathrm{Prob}(B|AC) \tag{8.43}$$

可将(8.42)写为

$$\mathrm{Prob}(H|D_c D_a I) = \mathrm{Prob}(H|I) \frac{\mathrm{Prob}(D_c D_a|HI)}{\mathrm{Prob}(D_c D_a|I)} \tag{8.44}$$

现在把反问题求解看成是一个推理过程，反演的解估计 $\tilde{m}$ 在数据 $\tilde{d}$ 有误差时被假设为可能的解，其中数据 $\tilde{d}$ 包含有多组数据(即 D_a, D_c 等等)。按(8.44)式有

$$\mathrm{Prob}(\boldsymbol{m}|\tilde{\boldsymbol{d}}I) = \mathrm{Prob}(\tilde{\boldsymbol{d}}|\boldsymbol{m}I) \frac{\mathrm{Prob}(\boldsymbol{m}/I)}{\mathrm{Prob}(\tilde{\boldsymbol{d}}/I)} \tag{8.45}$$

它代表了给定数据与先验知识发生时解估计接近精确(数据)的解的概率，式中 $\mathrm{Prob}(\tilde{\boldsymbol{d}}/\boldsymbol{m}I)$ 为给定模型时正问题解出现的概率，可用最大似然法计算；$\mathrm{Prob}(\boldsymbol{m}/I)$ 取在没有数据出现时模型出现的概率，与反演系统的熵有关，见下面讨论；分母 $\mathrm{Prob}(\tilde{\boldsymbol{d}}/I)$ 应为接近于1的常数，可以忽略。

（三）最大熵法（MEM）

假如某种过程产生一系列可能的事件集合，它们的出现概率记为

$$p_1, p_2, \cdots, p_n$$

能否找到一种测度，来度量这个集合中有多少“选择的可能性”？这是 Shannon 信息论早期就提出的问题。

显然，这种测度应是$\{p_i\}$的连续函数$E(p_1, \cdots, p_n)$。再者，如果所有 p_i 都相同，则有 $p_i = 1/n$，那么 E 将是 n 的单调上升函数。也就是说，集合中有更多的事件时就有更多的选择可能性。

此外，如果选择划分为相继的两个步骤，则原先的测度 E 应等于后继各个选择 E 值的加权和。例如，设 $n=3, p_1=\frac{1}{2}, p_2=\frac{1}{3}, p_3=\frac{1}{6}$，则测度为

$$E(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}) \tag{8.46}$$

将此选择分为两步，先作两种可能性选择

$$E'(\frac{1}{2}, \frac{1}{2}) \tag{8.47}$$

然后再将后种可能性划分为两种选择

$$E''(\frac{2}{3}, \frac{1}{3}) \tag{8.48}$$

则最后还是实现了 $n=3$的选择，此时应有

$$E(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}) = E(\frac{1}{2}, \frac{1}{2}) + \frac{1}{2}E(\frac{2}{3}, \frac{1}{3}) \tag{8.49}$$

满足上述三个条件的测度函数为

$$E = -k\sum_{i=1}^{n} p_i \lg p_i \tag{8.50}$$

式中 k 为取决于度量单位的常数，可规格化为1。式(8.50)与统计力学中熵的定义式一样，因此称为概率集$(p_1, \cdots, p_n)$的熵。由于事件选择的可能性越大，确定性就越差，因此熵是不确定性的度量。相反，负熵 $H=-E$(negentropy)就是确定性的度量。与线性反演问题联系起来，可知在没有任何数据发生时，设 m 为各态遍历的随机向量，(8.45)式分子中 Prob(m/I) 正比于 $\exp(E)=\exp(-H)$：

$$\text{Prob}(\boldsymbol{m}/I) \propto \exp(-H) \tag{8.51}$$

在展示反演结果时，我们常把 $\boldsymbol{m}$ 的数值量化成 L 个灰度级（整数），记为$\{m_k, k=1, \cdots, L\}$。于是，整个图像的灰度总量为

$$M = \sum_{k=1}^{L} |m_k| \tag{8.52}$$

那么，由组合公式可知，$\boldsymbol{m}$ 总共有的可能选择为

$$\text{Prob}(\boldsymbol{m}/I) = \frac{M!}{m_1! m_2! \cdots m_L!} \propto \exp\Big(-\sum_{k=1}^{L} m_k \ln\Big(\frac{m_k}{M}\Big) + \frac{1}{2}(\ln M - \sum_{k=1}^{L} \ln m_k)\Big) \tag{8.53}$$

比较(8.51)可知

$$H(\boldsymbol{m}) \approx \sum_{k=1}^{L} m_k \ln(m_k/M) \tag{8.54}$$

其实，由(8.50)的定义也可导出(8.54)式。

根据上节关于正则化的讨论(8.31)，对实际数据反演，我们可以用

$$\chi^2(\tilde{m}) + \lambda H(\tilde{m}) = m:n \tag{8.55}$$

为准则，求 χ^2 最佳拟合的同时，熵 E 最大(负熵最小)的广义解。由(8.51)代入(8.45)可知，这种广义解也就是 Prob($\boldsymbol{m}|\tilde{\boldsymbol{d}}I$)最大的广义解，即解估计接近精确解可能性最大的解估计，称为最大熵近似解。将(8.39)与(8.54)代入(8.55)有

$$Q(\boldsymbol{m}) = \sum_{i=1}^{N} \left[\frac{\tilde{d}_i - \sum_{k=1}^{L} G_{ik} m_k}{\sigma_i} \right]^2 + \lambda \sum_{k=1}^{L} m_k \ln m_k = m:n \tag{8.56}$$

式中要求 $m_k \geqslant 0$(灰度)。

注意最大熵准则与最小范数准则仅差在对后一项 m_k 是否取对数上。由于取了对数，泛函 Q 受模型的少量强峰值(即"亮点")的影响比最小范数正则化小得多。还要注意在(8.56)中，Q 已不再是线性泛函，因此计算困难要大一些。

将(8.56)式简写为

$$Q(\boldsymbol{m}) = \|\boldsymbol{d} - \boldsymbol{G}\boldsymbol{m}\|_2 + \lambda \|f(m_k)\|_1 = Q_1 + \lambda Q_2 \tag{8.57}$$

其中 d 为带误差的数据向量，$f(m_k) = m_k \ln m_k$，可求出以下导数向量和二阶导数矩阵(Hassian 矩阵)

$$\left.\begin{aligned} &\nabla Q_1 = 2(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{m} - \boldsymbol{G}^{\mathrm{T}}\boldsymbol{d}) \\ &\frac{\partial^2 Q_1}{\partial m_k \partial m_l} = [2\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}]_{kl} \\ &[\nabla Q_2]_k = f'(m_k) \\ &\frac{\partial^2 Q_2}{\partial m_k \partial m_l} = \delta_{kl} f''(m_k) \quad \text{(对角线矩阵)} \end{aligned}\right\} \tag{8.58}$$

如将泛函 $Q(\boldsymbol{m})$ 在 $\boldsymbol{m}_0$ 处作台劳展开并取前两项，有

$$Q(\boldsymbol{m}) \approx Q(\boldsymbol{m}_0) + \sum_k \frac{\partial Q}{\partial x_k}(m_k - m_0) + \frac{1}{2}\sum_k \sum_l \frac{\partial^2 Q}{\partial m_k \partial m_l}(m_k - m_0)(m_l - m_0) \tag{8.59}$$

将(8.58)代入(8.59)可以快速计算 Q。因此，第九章要讨论的仿真淬火与遗传算法等非线性优化方法，可以用来求出使 Q 取极小的广义解 $\boldsymbol{m}$。

有一种取巧的 MEM 算法，充分利用了(8.58)中二阶导数矩阵$[\nabla^2 Q_2]$为对角阵的特点和迭代计算的逼近思想。由于$[\nabla^2 \phi_2]$为对角阵，如果$[\nabla^2 Q_1]$也能用对角阵近似，则(8.59)简化为

$$Q(\boldsymbol{m}) \approx Q(\boldsymbol{m}_0) + [\nabla Q](\boldsymbol{m} - \boldsymbol{m}_0) + \frac{1}{2}(\boldsymbol{m} - \boldsymbol{m}_0)[\nabla^2 Q](\boldsymbol{m} - \boldsymbol{m}_0) \tag{8.60}$$

为$2Q/\partial \boldsymbol{m}^{\mathrm{T}} = 0$求极小，有

$$[\nabla Q] + \frac{1}{2}[\nabla^2 Q](\boldsymbol{m} - \boldsymbol{m}_0) = 0 \tag{8.61}$$

设

$$\Lambda_k - \left[\sum_i (G_{ik})^2\right] \approx [\nabla^2 Q_1]/2 \tag{8.62}$$

为 Hassian 矩阵$[\nabla^2 Q_1]$的近似对角阵，将(8.57)、(8.58)及(8.62)代入(8.61)得迭代格式

$$-\Delta m_k \approx \frac{\nabla Q_1 - \lambda \nabla Q_2}{\Lambda_k + \frac{\lambda}{2} f''(m_k)} \tag{8.63}$$

因此在迭代的第 k 步，可以由 m_k 近似估计修改增量 Δm_k，以大大提高最大熵法的计算速度。

三、地球物理线性反演[53~56]

现在考虑与基本方程(6.1)有关的反演问题，分成各种地球物理数据的情况，分别用 Born 近似等方法构成线性反演的方程式。非线性反演问题将在以后讨论。

(一)位场反问题

对重力、磁法等位场反问题，基本方程(6.1)退化为 Poisson 方程及地面 $z=0$上的边界条件：

$$\begin{cases} u_{xx} + u_{yy} + u_{zz} = \rho(x,y,z), \quad -\infty < x,y,z < \infty \\ u|_{z=0} = v(x,y) \end{cases} \tag{8.64}$$

反问题是给定地面的位场观测数据 $v(x,y)$求场源函数 $\rho(x,y,z)$。由位论知，(8.64)的正问题有解

$$u(x,y,z) = \frac{1}{4\pi}\iint \mathrm{d}x'\mathrm{d}y' \int_0^\infty \frac{\rho(x',y',z')}{\sqrt{(x-x')^2 + (y-y')^2 + (z-z')^2}}\mathrm{d}z' \tag{8.65}$$

代入地面数据有

$$v(x,y) = \frac{1}{4\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_0^\infty \frac{\rho(x',y',z')}{\sqrt{(x-x')^2 + (y-y')^2 + z'^2}}\mathrm{d}z'\mathrm{d}x'\mathrm{d}y' \tag{8.66}$$

这是带奇异核的第一类 Fredholm 积分方程。

由(8.66)可知，位场反问题本身就是线性的。由于数据 $v(x,y)$的自由度小于未知函数 $\rho(x,y,z)$的自由度，位场反问题的解具有高度的非唯一性。如果没有辅助的信息，由(8.66)求出任意一个解估计是没有意义的。通常的做法是把(8.66)式的积分区间局限在一个比较小的域内，如给定以下可谓的"硬边界"约束

$$a < x' < b, \quad f < y' < g, \quad p' < z' < q; \tag{8.67}$$

然后在此区域内对 $\rho(x',y',z')$离散化。

对于重力测量来说，通常测量的数据是垂直方向的重力加速度，相应(8.64)式中的边界条件变为

$$\partial u/\partial z|_{z=0} = g_z(x,y) \tag{8.68}$$

Poission 方程对此边值问题的解为

$$g_z(x,y) = \frac{f}{4\pi}\int_a^b\int_f^g\int_p^q \frac{z'\rho(x',y',z')}{\sqrt{(x-x')^2 + (y-y')^2 + z'^2}}\mathrm{d}z'\mathrm{d}y'\mathrm{d}x' \tag{8.69}$$

式中 f 为万有引力常数，ρ 为密度，(8.67)所示的硬边界已放入(8.69)式中。

如果测量的数据是航空磁测的 $\Delta T(x,y,z=H)$，H 为飞行高度，则反演比较复杂。与重力场不同，地磁场是由可视为磁偶极子的磁性元引起的，因此单位体积中磁性元的磁位为

$$dU_m = \frac{1}{4\pi}\frac{1}{r^3}(\boldsymbol{r}\cdot\boldsymbol{M}dV) \tag{8.70}$$

其中矢径

$$\left.\begin{aligned}\boldsymbol{r} &= \boldsymbol{i}(x-x') + \boldsymbol{j}(y\quad y') + \boldsymbol{k}(H-z')\\ r &= |\boldsymbol{r}| = [(x-x')^2 + (y-y')^2 + (H-z')^2]^{1/2}\end{aligned}\right\} \tag{8.71}$$

下半空间 z' 为负。因此磁位为

$$U_m = \frac{1}{4\pi}\iiint_V \frac{\boldsymbol{r}\cdot\boldsymbol{M}}{r^3}dV = -\frac{1}{4\pi}\iiint_V \left\{\boldsymbol{M}\cdot\nabla\left(\frac{1}{r}\right)\right\}dV \tag{8.72}$$

其中 V 为(8.67)规定的体积，梯度∇是对测点(x,y,H)而言。航空磁测的 ΔT 指的是由地壳上部磁性物质引起的磁异常在地球正常场上的投影，而地球正常磁场是由地核物质运动引起的。设地球正常磁场的磁倾角为 I，磁偏角(与 x 轴夹角)为 D，则 ΔT 可表为

$$\Delta T = -\frac{\partial U_m}{\partial x}\cos I\cos D - \frac{\partial U_m}{\partial y}\cos I\sin D - \frac{\partial U_m}{\partial z}\sin I \tag{8.73}$$

用(8.72)式求导并代入(8.73)可以得到一个比较复杂的积分方程，它将 $\Delta T(x,y,H)$表为磁化强度向量 $\boldsymbol{M}=(M_x,M_y,M_z)$的线性积分。

根据(8.73)与(8.72)可知，仅根据 $\Delta T(x,y)$反演磁化强度向量 $\boldsymbol{M}(x',y',z')$几乎是不可能的。在(8.73)中令磁倾角 $I=90°$，则有

$$\Delta T = -\partial U_m/\partial x = \Delta Z \tag{8.74}$$

这相当于在磁极附近航磁场只有垂直分量 ΔZ，而水平分量可忽略不计。在一般情况下，由于 ΔT 与 ΔZ 之间存在线性关系，在波数域很容易互相转换[53~56]，因此可先用 Fourier 二维滤波器将 ΔT 转化为 ΔZ，这个计算过程称为“化向磁极”。于是，磁测数据反演化为解线性方程组

$$\Delta Z(x,y,H) = -\frac{1}{4\pi}\frac{\partial}{\partial z}\iiint_V \frac{\boldsymbol{r}\cdot\boldsymbol{M}}{r^3}dV \tag{8.75}$$

如果认为磁源以感应磁化为主，而剩余磁性可以忽略，即有 $M_x=M_y=0$，(8.75)进而简化为

$$\Delta Z(x,y,H) = -\frac{1}{4\pi}\iiint_V \frac{3(z'-H)^2 - r^2}{r^5}M_z(x',y',z')dx'dy'dz' \tag{8.76}$$

这时线性反问题已与重力反问题类似，所不同的只是(8.76)中的核函数。

关于直流电法中电阻率成像问题，将在本章第五节专门讨论。

(二)地热反问题

将表6.1中有关系数代入(6.1)式，地温测量的正问题可表为

$$\begin{cases}\partial u/\partial t = \alpha\nabla^2 u + f(\boldsymbol{x},t), & -\infty < |\boldsymbol{x}| < \infty\\ u(\boldsymbol{x},t)|_{t=0} = F(\boldsymbol{x}), & 0 < t < \infty\end{cases} \tag{8.77}$$

式中 u 为地下温度场，α 为岩石热扩散系数，f 为热源密度，F 为地球初始温度。正问题的解可表为[57]

$$u(\boldsymbol{r},t) = \frac{1}{8(\pi\alpha t)^{3/2}}\int_{-\infty}^{\infty} F(\boldsymbol{x})\exp\left(-\frac{|\boldsymbol{x}-\boldsymbol{r}|^2}{4\alpha t}\right)d\boldsymbol{x} + \frac{1}{8(\pi\alpha)^{3/2}}\int_0^t \frac{d\tau}{(t-\tau)^{3/2}}$$
$$\int_{-\infty}^{\infty} f(\boldsymbol{x},\tau)\exp\left(-\frac{|\boldsymbol{x}-\boldsymbol{r}|^2}{4\alpha(t-\tau)}\right)d\boldsymbol{x} \tag{8.78}$$

其中 $\boldsymbol{r}$ 为观测点，$\boldsymbol{x}$ 为地下任一点。

假定地壳中热源(放射性物质)蜕变非常缓慢,因此可认为与 t 无关,则(8.78)式简化为

$$u(\boldsymbol{r},t)=\frac{1}{4\pi\alpha}\int\frac{f(\boldsymbol{x})}{|\boldsymbol{x}-\boldsymbol{r}|}\left\{1-\Phi\left(\frac{|\boldsymbol{x}-\boldsymbol{r}|}{2\sqrt{\alpha t}}\right)\right\}\mathrm{d}\boldsymbol{x}+\frac{1}{8(\pi\alpha t)^{3/2}}\int F(\boldsymbol{x})\exp\left(-\frac{|\boldsymbol{x}-\boldsymbol{r}|^2}{4\alpha t}\right)\mathrm{d}\boldsymbol{x} \tag{8.79}$$

其中

$$\Phi(a)=\frac{2}{\sqrt{\pi}}\int_0^a \mathrm{e}^{-w^2}\mathrm{d}w,\quad a=\frac{|\boldsymbol{x}-\boldsymbol{r}|}{2\sqrt{\alpha t}} \tag{8.80}$$

取决于观测数据和研究对象。反问题可有多种提法。首先,可由观测的地温场 $u(x,y,o,t)$ 反演热源密度 $f(x,y,z)$,此时要假定初始温度 $F(x,y,z)$ 及 $\alpha(x,y,z)$ 为已知。这时(8.80)化为第Ⅱ类 Fredholm 积分方程,它的求解是适定问题。

另一种可能的反演问题是由多年来测定的地温场 $u(x,y,o,t)$ 反推初始的地温场 $F(x,y,z)$。这个反问题属于初始条件反问题。当然,这时仍需假定地球内部没有热源,因此由(8.80)有

$$u(\boldsymbol{r},t)=\frac{1}{8(\pi\alpha t)^{3/2}}\int_{-\infty}^{\infty}F(\boldsymbol{x})\exp\left(-\frac{|\boldsymbol{x}-\boldsymbol{r}|^2}{4\alpha t}\right)\mathrm{d}\boldsymbol{x} \tag{8.81}$$

这一反问题为第一类 Fredholm 积分方程的反问题,以前已经讨论了很多。

如果给定 $u(x,y,o,t)$,而要求 $\alpha(x,y,z)$,则导致非线性反问题,尚未有深入研究。

(三)电磁法勘探反问题[56~60]

设人工场源为 $f(x,y,z,t)$,它的似稳电磁场可由基本方程(6.1)表为

$$u_{xx}+u_{yy}+u_{zz}-gu_t=f(x,y,z,t),g=\mu\sigma,u_t=\partial u/\partial t \tag{8.82}$$

将地球电结构视为背景 g_0 和扰动 α 两部分。

$$g(x,y,z)=g_0(x,y,z)-\alpha(x,y,z) \tag{8.83}$$

相应的标量场视为正常场和异常场两部分

$$u(x,y,z,t)=u_0(x,y,z,t)+u_a(x,y,z,t) \tag{8.84}$$

其中

$\alpha=\mu\cdot\delta\sigma$,为电导率的扰动(设磁导率 μ 为常数);$g_0=\mu\sigma_0$,σ_0 为地球背景电导率。已知 u_0 为入射电磁场,由 f 入射 g_0 引起;u_a 为二次感应电磁场,由 α 扰动引起。将(8.83~8.84)代入(8.82)得

$$\nabla^2[u_0+u_a]-(g_0-\alpha)\frac{\partial[u_0+u_a]}{\partial t}=f \tag{8.85}$$

对似稳场上式两边作 Fourier 变换,由于 u_a 仅与 α 有关,有

$$\nabla^2u_0+\omega g_0u_0=f(x,y,z,\omega) \tag{8.86}$$

此方程可用有限元或有限差分求解 u_0。对于二次场

$$\nabla^2u_a+\omega g_0u_0=\omega\alpha[u_0+u_a] \tag{8.87}$$

由于(8.87)是非线性,在 α 很小时假定 $u_a\ll u_0$,则依 Born 近似有

$$\nabla^2u_a+\omega g_0u_0\approx\omega\alpha u_0 \tag{8.88}$$

令源为点源,位于点 s,而接收线圈位于 r。应用 Green 函数法(8.88)的解可写为

$$u_a(s,r,\omega)\approx\omega\int_{-\infty}^{\infty}\int_{0}^{\infty}\alpha(x)G(x,s,\omega)G(x,r,\omega)\mathrm{d}x \tag{8.89}$$

其中 Green 函数满足

$$\nabla^2 G(\boldsymbol{x},\boldsymbol{y},\omega) + \omega g_0 G(\boldsymbol{x},\boldsymbol{y},\omega) = -\delta(\boldsymbol{x}-\boldsymbol{y}) \tag{8.90}$$

由于(8.90)与(8.86)很相似，G 可用有限差分法或有限元法在求 u_0同时计算出。

由实测多频电磁信号 $u(\boldsymbol{s},\boldsymbol{r},\omega)$中减去入射场 $u_0(\boldsymbol{s},\boldsymbol{r},\omega)$可求出 $u_s(\boldsymbol{s},\boldsymbol{r},\omega)$。今点源 $f=-\delta(\boldsymbol{s}-\boldsymbol{r})$，则有

$$G(\boldsymbol{x},\boldsymbol{s},\omega) = u_0(\boldsymbol{x},\boldsymbol{s},\omega), G(\boldsymbol{x},\boldsymbol{r},\omega) = u'_0(\boldsymbol{x},\boldsymbol{r},\omega) \tag{8.91}$$

其中 u'_0为位于接收点位置的“虚场源”在地下产生的伪“入射场”，可用(8.86)计算。将(8.91)代入(8.89)得

$$u_s(\boldsymbol{s},\boldsymbol{r},\omega) \approx \omega\int \alpha(\boldsymbol{x}) u_0(\boldsymbol{x},\boldsymbol{s},\omega) u'_0(\boldsymbol{x},\boldsymbol{r},\omega)\mathrm{d}\boldsymbol{x} \tag{8.92}$$

这是第一类 Fredholm 积分方程，用于反演地下电导率的扰动 $\delta_\sigma=\alpha/\mu$。对低频电磁法而言，式中 ω 一般有3～6组。由(8.92)可见，观测频率的增加可以弥补空间测点的不足。当 $g_0=$ const 或对层状介质 $g_0=g(z)$，u_0及 u'_0可找到相应的解析公式计算[58～60]，而不必用有限元等数值方法。

(四)散射地震线性反演

在第五章中，我们已经详细地讨论过地震层析成像问题，这里仅扼要回顾一下其要点，以便将它与其他地球物理线性反演作一比较，以看出所有地球物理反演的共性。

对散射地震的标量波场，由(6.1)有

$$\nabla^2 u - \frac{1}{c^2(\boldsymbol{x})}\frac{\partial^2 u}{\partial t^2} = -\delta(\boldsymbol{x}-\boldsymbol{s})\delta(t), \tag{8.93}$$

其中 $c(\boldsymbol{x})$为波速，$\boldsymbol{s}$ 为点震源位置。根据逆散射原理将平方波慢函数分解为两部分

$$\frac{1}{c^2(\boldsymbol{x})} = \frac{1+\alpha(\boldsymbol{x})}{c_0^2(\boldsymbol{x})} \tag{8.94}$$

其中 α 为平方波慢扰动，并假设为

$$-1 \leqslant \alpha(\boldsymbol{x}) \leqslant 1 \text{ 和 } c(\boldsymbol{x}) = c_0(\boldsymbol{x}), \text{如果 } |\boldsymbol{x}| \to \infty \tag{8.95}$$

$1/c_0^2$为平方背景波慢。相应地，波场也分解为入射波场 u_0和散射波场 u_s 两部分

$$u(\boldsymbol{x},t) = u_0(\boldsymbol{x},t) + u_s(\boldsymbol{x},t) \tag{8.96}$$

将(8.94～8.96)代入(8.93)并转换到频率域，得

$$\left[\nabla^2 + \frac{\omega^2}{c_0^2(\boldsymbol{x})}\right] u_0(\boldsymbol{x},\boldsymbol{s},\omega) = -\delta(\boldsymbol{x}-\boldsymbol{s}) \tag{8.97}$$

和

$$\left[\nabla^2 + \frac{\omega^2}{c_0^2(\boldsymbol{x})}\right] u_s(\boldsymbol{x},\boldsymbol{s},\omega) = -\alpha(\boldsymbol{x})\frac{\omega^2}{c_0^2(\boldsymbol{x})} u(\boldsymbol{x},\boldsymbol{s},\omega) \tag{8.98}$$

应用 Born 近似于(8.98)式右边，有

$$\left[\nabla^2 + \frac{\omega^2}{c_0^2(\boldsymbol{x})}\right] u_s(\boldsymbol{x},\boldsymbol{s},\omega) = -\alpha(\boldsymbol{x})\frac{\omega^2}{c_0^2(\boldsymbol{x})} u_0(\boldsymbol{x},\boldsymbol{s},\omega), \text{如果 } u_s \ll u_0 \tag{8.99}$$

设接收器位于 $\boldsymbol{r}$，则(8.99)的 Born 近似解为

$$u_s(\boldsymbol{r},\boldsymbol{s},\omega) \approx \omega^2 \int \frac{G(\boldsymbol{x},\boldsymbol{s},\omega)G(\boldsymbol{x},\boldsymbol{r},\omega)}{c_0^2(\boldsymbol{x})}\alpha(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \tag{8.100}$$

其中 Green 函数满足

$$\nabla^2 G(x,y,\omega)+\frac{\omega^2}{c_0^2(x)}G(x,y,\omega)=-\delta(x-y) \tag{8.101}$$

此式为 Fredholm 第一类积分方程，当 c_0 为已知时代入散射波场数据可求出平方波慢扰动 α。当 c_0 为常数时，有

$$G(x,s,\omega)=u_0(x,s,\omega)=\frac{\exp(ik|x-s|)}{4\pi|x-s|},k=\omega/c_0 \tag{8.102}$$

及

$$G(x,r,\omega)=\tilde{u}_0(x,r,\omega)=\frac{\exp(ik|x-r|)}{4\pi|x-r|} \tag{8.103}$$

又当接收器非常接近点源时，令 $s=-r$，

由(8.101～103)有

$$u_s(r,\omega)\approx k^2\int_{z>0}\alpha(x)\frac{\exp(2ik|x-r|)}{(4\pi)^2|x-r|^2}dx \tag{8.104}$$

由本节可知，与基本方程(6.1)有关的反问题可以分为两类，一类是(6.1)右边源项的反问题，本身就是线性问题。另一类是(6.1)左边系数项的反问题，它一般是非线性的，可以通过 Born 近似粗略地转化为线性反问题，以估计其系数项的空间扰动。

四、非线性迭代地震反演

在上节及第五章第七节中已讨论了波动方程的逆散射问题。方程(8.100)的 Born 近似反演要求平方波慢扰动很小，这个条件在实际上常不能满足。用广义 Radon 变换求逆得出的反演公式(5.221)也是在 Born 近似的假定上取得的，只不过用了比较复杂的背景波速函数 C_0，它还要通过其它反演方法得来。至此，我们可用两种不同的途径来实现地震反演：一是通过尽量接近实际的背景波速函数 C_0 来使 Born 近似的条件得到满足，除(5.221)所述的方法，黄联捷和作者还推导了背景波速为垂向线性变化时的 Born 近似反演公式[61,62]。另一途径是扔掉 Born 近似，直接去解非线性波动方程(8.98)，这就与迭代过程联系了起来。本节详细讨论后一种途径[63~68]。

(一)波动方程散射问题的积分解

设 x 为地球介质中任意点，x_s 为点震源位置，由声波方程知标量波场满足

$$\nabla^2 u(x,t)-\frac{1}{c^2(x)}\frac{\partial^2 u(x,t)}{\partial t^2}=-\delta(x-x_s)\delta(t) \tag{8.105}$$

和零初始条件，式中 u 为声场，c 为介质波速。假定 c_0 为一均匀的参考波速，$\alpha(x)$ 为波速扰动(无量纲)，并定义

$$V(x)\equiv\frac{\alpha(x)}{c_0^2}=\frac{1}{c_0^2}-\frac{1}{c^2(x)} \tag{8.106}$$

代入(8.105)式得

$$\nabla^2 u(x,t)-\frac{1}{c_0^2}\frac{\partial^2 u(x,t)}{\partial t^2}+V(x)\frac{\partial^2 u(x,t)}{\partial t^2}=-\delta(x-x_s) \tag{8.107}$$

由于波场为入射波场 u_i 和散射波场 u_s 的叠加，即

$$u(x,t)=u_i(x,t)+u_s(x,t) \tag{8.108}$$

而入射波场与散射源 $V(\boldsymbol{x})$ 无关，应满足

$$\nabla^2 u_i(\boldsymbol{x},t) - \frac{1}{c_0^2}\frac{\partial^2 u_i(\boldsymbol{x},t)}{\partial t^2} = -\delta(\boldsymbol{x}-\boldsymbol{x}_s)\delta(t) \tag{8.109}$$

代入(8.107)知散射波满足

$$\nabla^2 u_s(\boldsymbol{x},t) - \frac{1}{c_0^2}\frac{\partial^2 u_s(\boldsymbol{x},t)}{\partial t^2} = -V(\boldsymbol{x})\frac{\partial^2 u(\boldsymbol{x},t)}{\partial t^2} \tag{8.110}$$

此式的解为

$$u_s(\boldsymbol{x},t) = \iint \mathrm{d}x'\,\mathrm{d}t G(\boldsymbol{x},t|\boldsymbol{x}'t')V(x')u(\boldsymbol{x}',t') \tag{8.111}$$

式中 Green 函数满足

$$\nabla^2 G - \frac{1}{c_0^2}\frac{\partial^2 G}{\partial t^2} = -\frac{\partial^2}{\partial t^2}\delta(t-t')\delta(\boldsymbol{x}-\boldsymbol{x}') \tag{8.112}$$

如果参考波速 c_0 为常数，(8.112)式的解为

$$G(\boldsymbol{x},t|\boldsymbol{x}',t') = \frac{1}{4\pi|\boldsymbol{x}-\boldsymbol{x}'|}\frac{\partial^2}{\partial t^2}\delta\left(\frac{|\boldsymbol{x}-\boldsymbol{x}'|}{c_0} \quad (t \quad t')\right) \tag{8.113}$$

而(8.109)式的解为

$$u_i(\boldsymbol{x},\boldsymbol{x}_s;t) = \frac{1}{4\pi|\boldsymbol{x}-\boldsymbol{x}_s|}\delta\left(\frac{|\boldsymbol{x}-\boldsymbol{x}_s|}{c_0} - t\right) \tag{8.114}$$

注意(8.113)式中 Green 函数是表示分布的广义函数，具有旋转对称性，即

$$G(\boldsymbol{x},t|\boldsymbol{x}',t') = G(\boldsymbol{x}',t'|\boldsymbol{x},t) \tag{8.115}$$

如果 c_0 不是常数，则可以用数值方法(如有限差分)求满足(8.109)及(8.112)式的入射波场及格林函数。

将(8.109)与(8.111)式代入(8.108)式得总声场

$$u(\boldsymbol{x},t) = u_i(\boldsymbol{x},t) + \iint \mathrm{d}x'\,\mathrm{d}t' G(\boldsymbol{x},t|\boldsymbol{x}',t')V(\boldsymbol{x}')u(\boldsymbol{x}',t') \tag{8.116}$$

此式即为声散射问题的积分解。对于波动方程的逆散射问题，指的是已知部分波场 $u(\boldsymbol{x},t)$ 的观测值求散射源 $V(\boldsymbol{x})$。由于(8.116)式右边积分号内有未知函数 $u(\boldsymbol{x}',t)$，因此(8.116)式是非线性积分方程。如果散射波很弱，对任何 $\boldsymbol{x}$ 和 t 满足 $u_s \ll u_i$，则可以用 u_i 代替(8.116)式积分号内的 u，使(8.116)式变为第一类 Fredholm 积分方程来求解，这就是上节讨论过的 Born 近似。本节要讨论的是地震勘探的情况下散射波场不可忽略的情况，即非线性积分方程(8.116)的求解问题。

如上所述，逆散射问题已经归结到对 $V(\boldsymbol{x})$ 的成像，那么 $V(\boldsymbol{x})$ 在地震勘探中的物理意义是什么?这个问题一直困扰着不少人，使得逆散射方法迟迟不能推广。将(8.106)式写为

$$V(\boldsymbol{x}) = \frac{c^2(\boldsymbol{x}) - c_0^2}{c_0^2 c^2(\boldsymbol{x})} = r(\boldsymbol{x})\frac{[c_0 + c(\boldsymbol{x})]^2}{c_0^2 c^2(\boldsymbol{x})} = r(\boldsymbol{x})[s(\boldsymbol{x}) + s_0]^2 \tag{8.117}$$

其中 s_0 为参考介质波慢，$s(\boldsymbol{x})$ 为 $\boldsymbol{x}$ 点波慢，

$$r(\boldsymbol{x}) = \frac{c(\boldsymbol{x}) - c_0}{c(\boldsymbol{x}) + c_0} \tag{8.118}$$

即常密度介质的垂向反射系数，可知 $V(\boldsymbol{x})$ 与反射系数成正比，与波慢和的平方也成正比，其单位为 $\mathrm{s}^2/\mathrm{m}^2$，即波慢的平方单位。因此，新参数 $V(\boldsymbol{x})$ 可称为平方波慢扰动，它对波速低的介质变化尤其敏感，同时也反映了介质波速的突变。由此可知，对 $V(\boldsymbol{x})$ 的逆散射成像不同于偏

移，对圈定和揭示低速薄层或透镜体尤其有利。

还要注意到，作为散射源的 $V(x)$ 现在还不等价于反射源，而只是均匀背景介质中点散射的集合。例如，对由 c_0 和 c_1 组成的两层介质，如取 c_0 为参考介质，由(8.117)式求出的 $V(x)$ 在 c_1 介质内部并不等于零，而我们知道这里并不发生反射。为使逆散射方法可以用于反射地震的情况，还必须将 c_0 推广到具有无穷阶导数的变化介质的情况，即令 $c_0=\bar{c}(x)$，其中 $\bar{c}(x)$ 为单位球内 $c(x)$ 的平均值，并属于 C^∞ 函数族

$$\bar{c}(x)=\frac{1}{4\pi}\int_0^{2\pi}\mathrm{d}\theta\int_0^{\pi}\mathrm{d}\phi c(x,\phi,\theta)\sin\phi \tag{8.119}$$

其中 x 为球心位置，θ 为水平角，ϕ 为垂直角。于是，(8.117)式变为

$$V(x)=\frac{c(x)-\bar{c}(x)}{c(x)+\bar{c}(x)}\left(\frac{1}{c(x)}+\frac{1}{c(x)}\right)^2=r(x)\left(\frac{1}{c(x)}+\frac{1}{c(x)-\Delta c(x)}\right)^2 \tag{8.120}$$

其中

$$\Delta c(x)=c(x)-\bar{c}(x);\quad r(x)=\frac{\Delta c(x)}{2c(x)-\Delta c(x)} \tag{8.121}$$

于是，$V(x)$ 便只与波速的变化有关，可同时反映反射界面和薄层散射体。然而，在反演时波速 $c(x)$ 是未知的，这时仍然无法由(8.119)式求出参考背景 $\bar{c}(x)$ 的空间变化。从实用角度出发，在只考虑纵波反射时可以把(8.119)式中平均的范围缩小到 x 点正上方±45°的球锥范围内，即令

$$\bar{c}(x)=\frac{2}{\pi}\int_{-\pi/4}^{\pi/4}\mathrm{d}\theta\int_{-\pi/4}^{\pi/4}\mathrm{d}\phi c(x;\theta,\phi)\sin\phi \tag{8.122}$$

而 $c(x)$ 可从上向下逐渐由上层的成像结果推算出来，使逆散射反演时 $\bar{c}(x)$ 可以实时计算之。

经过这样的推广，便可以将(8.111)式中的散射波视为反射信号。对地面某一测点 x_r，有

$$u_s(x_r,x_s,t;V)=\int\mathrm{d}t'\int\mathrm{d}yG(x_r,t|y,t')V(y)u(y,x_s,t';V) \tag{8.123}$$

令 B 为以 $G(x_r,t|y,t')u(y,x_s,t')$ 为核的非线性积分算子(在总波场 u 之中包含有散射地 u_s)，则对应 (x_r,x_s,t) 的第 i 个反射地震数据可用泛函表示为

$$u_{s_i}=B_iV,\quad i=1,2,\cdots,I \tag{8.124}$$

其中 I 为地震数据的总个数，取决于炮点数、接收点数和每道记录样点数。式(8.124)构成了一个非线性泛函方程组，逆散射反演问题化为此泛函方程组求解问题。

(二)迭代线性化反演

由(8.124)式可知，地震反演是由已知的反射地震数据 $u_{s_i}(i=1,\cdots,I)$ 求取平方波慢扰动 $V(x)$。从(8.120)式可见，$V(x)$ 本质上为一广义函数，从有限个数据不可能求出 $V(x)$ 唯一的解，只能求在某种准则之下的 $V(x)$ 的一种逼近，即解估计。由于 B 为非线性算子，求解估计的过程可分解为一系列线性逼近的过程，即迭代计算过程，而在每一迭代步都用线性化方法将(8.124)式化为线性泛函求解，实行线性化的基础是泛函的微分。

设 δV 为 $V(x)$ 的微小变化，则根据泛函强导数定义有

$$B_i(V+\delta V)-B_iV=\int F_i(y)\delta V(y)\mathrm{d}y \tag{8.125}$$

式中 F_i 为算子 B_i 对模型 V 的 Fréchet 导数，它是一个线性泛函，计算方法见下段。令

$$\delta u_{s_i} = B_i(V + \delta V) - B_i V \tag{8.126}$$

它为两种相近的散射模型产生的数据差。如果把(8.125)及(8.126)式当作约束，求$\|\delta V\|$取极小的解估计，则可令目标函数为

$$\phi = (\delta V, \delta V) - \sum_{i=1}^{I} \alpha_i [(F_i, \delta V) - \delta u_{si}] \tag{8.127}$$

式中括号(,)表示内积，如(8.125)式所示，α_i为拉格朗日乘子。(8.127)式对δV求极小得

$$\delta V = \sum_{i=1}^{I} \alpha_i F_i \tag{8.128}$$

此式回代入(8.125)式得

$$\delta u_{s_i} = \sum_{k=1}^{I} (F_i, F_k)\alpha_i, \quad i = 1, \cdots, I \tag{8.129}$$

将δu_s视为反射地震数据与合成数据之差，由此式可求出系数$\{\alpha_i\}$，还可由(8.127)式计算$\delta V(x)$，即平方波慢扰动的迭代修改增量。

由以上分析便得出了以下迭代线性化格式

$$V_{q+1} = V_q + \mu_q \sum_{i=1}^{I} \alpha_{i,q} F_{i,q}, \quad q = 0, 1, 2, \cdots \tag{8.130}$$

式中q注记迭代次数，μ_q为第q次迭代时的松驰因子，在0到1之间取值。初始模式V_0为一光滑函数，只要按下节所述的方法计算出每次迭代对应的算子B的Fréchet导数并由(8.129)解出系数组$\{\alpha_i\}$，便可通过迭代修改式(8.130)不断修正模型$V_q(x)$，最后求出在$\|\delta V\|$取极小准则下平方波慢扰动的解估计。

上面的讨论是针对$V(x)$为连续函数的情况下导出的迭代线性化格式。如果事先将地球模型离散化为一组向量

$$\boldsymbol{V}^{\mathrm{T}} = (V_1, \cdots, V_J) \tag{8.131}$$

则(8.124)式可写为

$$u_{s_i} = B_i(V_1, \cdots, V_j)^{\mathrm{T}} \tag{8.132}$$

由于模型离散化后已局限在有限维欧几里德空间，即此时$\boldsymbol{V} \in \mathrm{R}^J$，因此对应的Fréchet导数已退化为雅可比矩阵

$$F_{ij} = \left(\frac{\partial B_i}{\partial V_j}\right)_{IXJ} \tag{8.133}$$

由(8.125)式有

$$\delta u_s = \boldsymbol{F}\delta \boldsymbol{V} \tag{8.134}$$

这是线性方程组，可以用阻尼最小二乘法写出它的解为

$$\delta \boldsymbol{V} = (\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F} + \lambda I)^{-1}\boldsymbol{F}^{\mathrm{T}}\delta \boldsymbol{u}_s \tag{8.135}$$

写成逐次线性化迭代格式为

$$\boldsymbol{V}_{q+1} = \boldsymbol{V}_q + \mu_q(\boldsymbol{F}_q^{\mathrm{T}}\boldsymbol{F}_q + \lambda_q \boldsymbol{I})^{-1}\boldsymbol{F}_q^{\mathrm{T}}(\boldsymbol{u}_s - \boldsymbol{u}_{s,q}) \tag{8.136}$$

式中，$q=0,1,2\cdots$为迭代次数，$\boldsymbol{F}_q$为第q次迭代计算的雅可比矩阵，λ_q为第q次迭代时选取的阻尼因子，$\boldsymbol{I}$为单位矩阵，u_s为实测地震数据向量，$u_{s,q}$为用$\boldsymbol{V}_{q-1}$合成的地震数据。

(三)Fréchet导数的计算

在(8.125)中泛函Fréchet的计算是迭代算法的关键之一，其推导要求出模型参数变化与数据的关系。设$V_1(x)$和$V_2(x)$为两处相近的地球平方波慢扰动模型，它们产生的数据差

可由(8.123)得

$$\Delta u_s = u_s(x_r, x_s, t; V_1) - u_s(x_r, x_s, t; V_2)$$
$$= \int dt' \int dy [V_1(y) u_1(y, t') - V_2(y) u_2(y, t')] G(x_r, t | y, t') \tag{8.137}$$

式中 u_1 和 u_2 分别为在 V_1 和 V_2 介质中点源 x_s 的总波场，G 为点 y 处散射源向测点 x_r 的回传算子，背景波速为 c_0。

构造一个新函数(相当于源在 x_r 处的假想波场)

$$\tilde{u}(x, x_r, t; V) = G(x_r, t | x, t') + \int dt' \int dy \tilde{G}(x, t | y, t') V(y) \tilde{u}(y, x_r, t; V) \tag{8.138}$$

式中 $\tilde{G}$ 为 G 的共轭 Green 函数，即

$$\tilde{G}(x, t | y, t') = G(x, -t | y, -t') \tag{8.139}$$

为书写简练，以下我们将时间变量 t 纳入空间变量中，记 $y=(y_1, y_2, y_3, t)$，其中 (y_1, y_2, y_3) 为点 y 的空间坐标，同时，dy 也包含了对时间的导数。注意 u 和 $\tilde{u}$ 都是可以通过正演合成求出的总场，计算方法几乎是一样的，引入假想波场 $\tilde{u}$ 是为了导出泛函的变化率。

令 $\tilde{u}_1$ 为介质 $V_1(x)$ 中的假想波场代入(8.138)式移项，再代入(8.137)中的 Green 函数有

$$\Delta u_s = \int dy V_1(y) u_1(y) G(x_r | y) - \int dy V_2(y) u_2(y) \tilde{u}_1(y)$$
$$- \int dy V_2(y) u_2(y) \int dy' \tilde{G}(y | y') V_1(y') \tilde{u}_1(y') \tag{8.140}$$

式中 y 与 y' 为介质中的两个散射点。将上式最后一项改变积分次序并调换 y 与 y' 的记号，并应用(8.139)式有

$$\Delta u_s = \int dy V_1(y) u_1(y) G(x_r | y) - \int dy V_2(y) u_2(y) \tilde{u}_1(y)$$
$$- \int dy V_1(y) \tilde{u}_1(y) \int dy' G(y | y') V_2(y') u_2(y') \tag{8.141}$$

由(8.108)和(8.110)式知，上式最后一个积分等于 $u_2 - u_i$，其中 u_2 为点源在 V_2 介质中的总波场，u_i 为点源(x_s 处)在参考介质 c_0 中的波场，与介质的波慢扰动无关，因此

$$\Delta u_s = \int dy [V_1(y) u_1(y) G(x_r | y) - V_2(y) u_2(y) \tilde{u}_1(y) + V_1(y) \tilde{u}_1(y) (u_2(y) - u_i)] \tag{8.142}$$

上式第二及第三项合并有

$$\Delta u_s = \int dy [V_1(y) - V_2(y)] u_2(y) \tilde{u}_1(y) +$$
$$\int dy [V_1(y) u_1(y) G(x_r | y) - V_1(y) \tilde{u}_1(y) u_i(y)] \tag{8.143}$$

由于

$$G(x_r | y) = \tilde{u}_1(y, V_1) - \int dy \tilde{G}(x_r | y) V_1(y) \tilde{u}_1(y, V_1) \tag{8.144}$$

及

$$u_i(y) = u_1(y, V_1) - \int dy G(x_s, y) V_1(y) u(y, V_1) \tag{8.145}$$

代入(8.143)式得

$$\Delta u_s = \int \mathrm{d}\boldsymbol{y}[V_1(\boldsymbol{y}) - V_2(\boldsymbol{y})]u_2(\boldsymbol{y})\tilde{u}_1(\boldsymbol{y}) + \int \mathrm{d}\boldsymbol{y}[V_1(\boldsymbol{y})\tilde{u}_1(\boldsymbol{y})u_{1s}(\boldsymbol{y}) - V_1(\boldsymbol{y})u_1(\boldsymbol{y})\tilde{u}_{1s}(\boldsymbol{y})] \tag{8.146}$$

式中 u_{1s} 为源位于 $\boldsymbol{x}_s$ 的二次散射波场，即

$$u_{1s} = \int \mathrm{d}\boldsymbol{y}' G(\boldsymbol{x}_s|\boldsymbol{y}')V_1(\boldsymbol{y}')u_1(\boldsymbol{y}') \tag{8.147}$$

及

$$\tilde{u}_{1s} = \int \mathrm{d}\boldsymbol{y}' \tilde{G}(\boldsymbol{x}_r|\boldsymbol{y}')V_1(\boldsymbol{y}')\tilde{u}_1(\boldsymbol{y}') \tag{8.148}$$

在(8.146)式右边积分的第二项为二次散射波场耦合的差异。对于深层反射信号而言，如果 $\boldsymbol{x}_r$ 与 $\boldsymbol{x}_s$ 之间的距离远小于它们到散射体的距离，则由于 u_1 大 u_{1s} 也大，$\tilde{u}_1$ 大时 $\tilde{u}_{1s}$ 也大，它们可以大部分抵消，因此有

$$u_s(\boldsymbol{x}_r,\boldsymbol{x}_s,V_1) - u_s(\boldsymbol{x}_r,\boldsymbol{x}_s,V_2) \backsimeq \int \mathrm{d}\boldsymbol{y}[V_1(\boldsymbol{y}) - V_2(\boldsymbol{y})]u_2(\boldsymbol{y})\tilde{u}_1(\boldsymbol{y}) \tag{8.149}$$

令 $V_1 = V_2 + \delta V$ 并注意到 Fréchet 导数的定义式(8.125)，知 Fréchet 导数可用下式计算

$$F(\boldsymbol{y},\boldsymbol{x}_r,\boldsymbol{x}_s,V) \backsimeq \tilde{u}_r(\boldsymbol{y},\boldsymbol{x}_r,V)u_2(\boldsymbol{y},\boldsymbol{x}_s;V) \tag{8.150}$$

推广到迭代线性化的场合，可用(8.150)式代入(8.130)式进行迭代，其中

$$F_{i,q} \equiv F(\boldsymbol{y},\boldsymbol{x}_r,\boldsymbol{x}_s,V_q) = \tilde{u}_1(\boldsymbol{y},\boldsymbol{x}_r,V_q)u_2(\boldsymbol{y},\boldsymbol{x}_s,V_q) \tag{8.151}$$

V_q 为第 q 次迭代的平方波慢扰动，$\tilde{u}_1$ 和 u_2 为用 V_q 合成的假想或实际波场。

(四)非线性迭代的最速下降算法与共轭梯度法

将(8.136)的迭代格式写为以下函数形式

$$V_q(\boldsymbol{x}) = V_{q-1}(\boldsymbol{x}) + \mu_q \delta V_q(\boldsymbol{x}),\ q = 1,2,\cdots \tag{8.152}$$

式中 V_q 为第 q 次迭代计算的平方波慢扰动，δV_q 为 q 次迭代的扰动修改增量，μ_q 为松驰因子。为考虑数据的拟合差，由(8.137)可以看出，对应(8.152)式反射数据也有以下关系

$$u_s(\boldsymbol{x}_r,\boldsymbol{x}_s,t;V_q) \backsimeq u_s(\boldsymbol{x}_r,\boldsymbol{x}_s,t;V_{q-1}) + u_q\Delta q(\boldsymbol{x}_r,\boldsymbol{x}_s,t) \tag{8.153}$$

式中 $\boldsymbol{x}_r$ 为测点，$\boldsymbol{x}_s$ 为炮点，Δq 为两次迭代之间反射波场的差异，它可表为

$$\Delta q(\boldsymbol{x}_r,\boldsymbol{x}_s) = \int \mathrm{d}\boldsymbol{y}\delta V_q(\boldsymbol{y})\tilde{u}(\boldsymbol{y},\boldsymbol{x}_r;V_{q-1})u(\boldsymbol{y},\boldsymbol{x}_s;V_{q-1}) \tag{8.154}$$

式中 $\tilde{u}$ 为源放在测点时的假想总波场，它与总波场 u 都可以用上一次迭代的地球模型 V_{q-1} 计算出来。注意(8.153)式是近似式，它忽略了高次项。

令 $V(x)$ 为真实的波慢平方扰动，它产生的理想反射数据可用函数 $u_s(\boldsymbol{x}_r,\boldsymbol{x}_s;V)$ 描述(为简明起见，以后推导不再写时间变量 t，应意识到它在波场中的存在)。于是，反射数据的逐点拟合差为

$$e(\boldsymbol{x}_r,\boldsymbol{x}_s;V_q) = u_s(\boldsymbol{x}_r,\boldsymbol{x}_s;V_q) - u_s(\boldsymbol{x}_r,\boldsymbol{x}_s;V) \tag{8.155}$$

代入(8.153)式有

$$\begin{aligned} e(\boldsymbol{x}_r,\boldsymbol{x}_s;V_q) &= u_s(\boldsymbol{x}_r;\boldsymbol{x}_s;V_{q-1}) + \mu_q\Delta q(\boldsymbol{x}_r,\boldsymbol{x}_s) - u_s(\boldsymbol{x}_r,\boldsymbol{x}_s;V) \\ &= e(\boldsymbol{x}_r,\boldsymbol{x}_s;V_{q-1}) + \mu_q\Delta q(\boldsymbol{x}_r,\boldsymbol{x}_s) \end{aligned} \tag{8.156}$$

根据定义，数据拟合均方差为

$$E(V_q) = \iiiint \mathrm{d}\boldsymbol{x}_s\mathrm{d}\boldsymbol{x}_r\mathrm{d}\boldsymbol{x}'_s\mathrm{d}\boldsymbol{x}'_r e^*(\boldsymbol{x}_r,\boldsymbol{x}_s;V_q)W(\boldsymbol{x}_r,\boldsymbol{x}_s;\boldsymbol{x}'_r,\boldsymbol{x}'_s)e(\boldsymbol{x}'_r,\boldsymbol{x}'_s;V_q) \tag{8.157}$$

式中上角 * 表示共轭，而 W 为数据的协方差矩阵。假定数据服从高斯分布和各自独立，则有

$$W(\boldsymbol{x}_r,\boldsymbol{x}_s;\boldsymbol{x}'_r,\boldsymbol{x}'_s)=W(\boldsymbol{x}_r,\boldsymbol{x}_s)\delta(\boldsymbol{x}_r-\boldsymbol{x}'_r)\delta(\boldsymbol{x}_s-\boldsymbol{x}'_s)$$

于是(8.157)式变为

$$\begin{aligned}E(V_q)&=\iint \mathrm{d}\boldsymbol{x}_s\mathrm{d}\boldsymbol{x}_r W(\boldsymbol{x}_r,\boldsymbol{x}_s)\,|e(\boldsymbol{x}_r,\boldsymbol{x}_s;V_q)|^2\\&=\iint \mathrm{d}\boldsymbol{x}_s\mathrm{d}\boldsymbol{x}_r W(\boldsymbol{x}_r,\boldsymbol{x}_s)\,|e(\boldsymbol{x}_r,\boldsymbol{x}_s;V_{q-1})+\mu q\Delta q(\boldsymbol{x}_r,\boldsymbol{x}_s)|^2\\&=E(V_{q-1})-2\mu_q p_q+\mu_q^2 Q_q\end{aligned}\tag{8.158}$$

其中

$$p_q=-\operatorname{Re}\iint \mathrm{d}\boldsymbol{x}_s\mathrm{d}\boldsymbol{x}_r W(\boldsymbol{x}_r,\boldsymbol{x}_s)e(\boldsymbol{x}_r,\boldsymbol{x}_s;V_{q-1})\Delta^* q(\boldsymbol{x}_r,\boldsymbol{x}_s)\tag{8.159}$$

$$Q_q=\iint \mathrm{d}\boldsymbol{x}_s\mathrm{d}\boldsymbol{x}_r W(\boldsymbol{x}_r,\boldsymbol{x}_s)\,|\Delta^* q(\boldsymbol{x}_r,\boldsymbol{x}_s)|^2\tag{8.160}$$

代入(8.154)式它们都是可以计算的。

为使数据拟合均方差取极小，首先可将(8.158)式对 μ_q 求导。由 $\mathrm{d}E/\mathrm{d}\mu q=0$ 得

$$\mu_q=p_q/Q_q\tag{8.161}$$

代回(8.158)式有

$$E(V_q)=E(V_{q-1})-p_q^2/Q_q\tag{8.162}$$

由此可见，为使第 q 次迭代的拟合均方差取极小，要求 p_q 取极大和 Q_q 取极小。

为使拟合均方差极小化求地球模型 $V(\boldsymbol{x})$，要求泛函相对于函数 V 的偏导数。用泛函的 Gateanx 微分定义(称为泛函的弱导数，而上文讲的泛函的 Fréchet 导数称为强导数)，有

$$\begin{aligned}\mathrm{d}E(V)&=\lim_{\alpha\to 0}\frac{E(V+\alpha\delta V)-E(V)}{\alpha}\\&=2\operatorname{Re}\iint \mathrm{d}\boldsymbol{x}_s\mathrm{d}\boldsymbol{x}_r W(\boldsymbol{x}_r,\boldsymbol{x}_s)e(\boldsymbol{x}_r,\boldsymbol{x}_s,V)\Delta^*(\boldsymbol{x}_r,\boldsymbol{x}_s;\delta V)\\&=2\int \mathrm{d}\boldsymbol{x}\delta V(\boldsymbol{x})g(\boldsymbol{x};V)\end{aligned}\tag{8.163}$$

此式由代入(8.160)等式导出，其中

$$g(\boldsymbol{x};V)=\operatorname{Re}\iint \mathrm{d}\boldsymbol{x}_s\mathrm{d}\boldsymbol{x}_r W(\boldsymbol{x}_r,\boldsymbol{x}_s)e(\boldsymbol{x}_r,\boldsymbol{x}_s;V)\tilde{u}^*(\boldsymbol{x},\boldsymbol{x}_r;V)u^*(\boldsymbol{x},\boldsymbol{x}_s;V)\tag{8.164}$$

对于第 q 次迭代，自然有

$$q(\boldsymbol{x};V_q)=\operatorname{Re}\iint \mathrm{d}\boldsymbol{x}_s\mathrm{d}\boldsymbol{x}_r W(\boldsymbol{x}_r,\boldsymbol{x}_s)e(\boldsymbol{x}_r,\boldsymbol{x}_s;V_q)\tilde{u}^*(\boldsymbol{x},\boldsymbol{x}_r;V_q)u^*(\boldsymbol{x},\boldsymbol{x}_s;V_q)\tag{8.165}$$

将(8.155)式及(8.165)式代入(8.159)式知

$$p_q=-\int \mathrm{d}\boldsymbol{y}\delta V_q(\boldsymbol{y})g(\boldsymbol{y},V_{q-1})\tag{8.166}$$

由(8.162)式知要使 $E(V_q)$ 极小化要使 p_q 取极大，由(8.166)式知，使积分 P_g 取极大必须使

$$\delta V_q(\boldsymbol{x})=g_{q-1}(\boldsymbol{x})=g(\boldsymbol{x},V_{q-1})\tag{8.167}$$

因此有

$$p_q=\int \mathrm{d}\boldsymbol{x}g_{q-1}^2(\boldsymbol{x})\tag{8.168}$$

因此，在给定 Q_q 后使拟合均方差取极小的迭代线性化公式由(8.161)、(8.167)代入(8.152)式有

$$V_q(x) = V_{q-1}(x) - \frac{p_q}{Q_q} g_{q-1}(x) \tag{8.169}$$

这是非线性迭代的"最速下降法"。

同时考虑 p_q 取极大和 Q_q 取极小而使拟合均方差取极小的迭代线性化公式可用类似方法导出，其结果为

$$\delta V_q(x) = -g_{q-1}(x) + \frac{p_q}{p_{q-1}} \delta V_{q-1}(x), q > 1 \tag{8.170}$$

而

$$V_q(x) = V_{q-1}(x) + \frac{p_q}{Q_q}\left[-g_{q-1}(x) + \frac{p_q}{p_{q-1}}\delta V_{q-1}(x)\right] \tag{8.171}$$

这种迭代格式不仅要用本次迭代的修改增量 $\delta V_q(x) = -g_{q-1}(x)$，而且还要用到上次迭代的修改增量 $\delta V_{q-1}(x) = -g_{q-2}(x)$ 来求当前迭代的地球模型，使解空间的搜索具有较好的连续性，它在计算数学中被称为"共轭梯度下降法"。

由上述讨论可以把拟合均方差取极小的迭代方法的计算过程总结如下。

1. 给出背景平均波速及平方波慢扰动的初始猜测 $V_0(x)$，计算数据拟合差 $e_0(x_r, x_s, t)$，波场 u_0 及虚拟波场 $\tilde{u}_0$，然后求 $g_0(x)$ 及拟合均方差 $E(V_0)$。

2. 按以下次序迭代计算 $p_q, \delta V_q, \Delta q, Q_q, V_q, e_q, g_q$ 及 $E(V_q)$，公式参见以上讨论。

3. 返回2进行 $q+1$ 次迭代或根据 $E(V_q) < \varepsilon$ 的准则停止迭代。

五、电导率成像反演

在第三节中讨论过位场和电磁场反问题的 Born 线性反演。本节围绕电导率 σ 的成像问题，讨论非线性反演的一般方法。

(一)直流电法层析成像[69~71]

在图5.1中，用供电电极代替震源，用接收电极代替检波器接收电位，就得到了用电法进行层析成像的数据采集装置，可用来对地下物体的电导率进行层析成像。由于电测仪器便宜和测量快速，电法层析成像成本低，很容易推广。

二维电导率成像的正问题为以下 Poisson 方程的边值问题

$$\left.\begin{aligned} &\frac{\partial}{\partial x}\left(\sigma \frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial y}\left(\sigma \frac{\partial u}{\partial y}\right) = 0 \\ &u|_{\partial\Omega} = f(x, y) \end{aligned}\right\} \tag{8.172}$$

式中 u 为待求电位，Ω 为二维研究区域，$\partial\Omega$ 为其边界，$f(x, y)$ 为给定边界的电位。如果给定边界上的电流密度 J_0，则(8.172)中边值条件改为

$$\sigma \frac{\partial u}{\partial n}\bigg|_{\partial\Omega} = J_0 \tag{8.173}$$

其中 n 为 $\partial\Omega$ 的外法线方向。

根据场论知

$$\nabla \cdot (u\sigma\nabla u) = \nabla u \cdot \sigma\nabla u + u\nabla \cdot (\sigma\nabla u) = \sigma|\nabla u|^2 \tag{8.174}$$

其中代入了(8.171)中的 Poisson 方程 $\nabla \cdot (\sigma\nabla u) = 0$。将此式在 Ω 上作积分并化为界面积分有

$$\iint_{\Omega}\sigma_i|\nabla u|^2\mathrm{d}x\mathrm{d}y=\iint_{\Omega}\nabla\cdot(u\sigma\nabla u)\mathrm{d}x\mathrm{d}y$$
$$=\int_{\partial\Omega}\sigma u(\nabla u\cdot n)\mathrm{d}S=\int_{\partial\Omega}V_0I_0\mathrm{d}S \tag{8.175}$$

表明在 Ω 内全部耗散能量等于边界电能的供入。

电导率成像时在边界 $\partial\Omega$ 上逐点供电，记第 i 个供电点的电流及电位为 I_{0i} 和 V_{0i}，在边界 $\partial\Omega$ 上逐点测量电位 $u_i(x',y')$，其中 $(x',y')\in\partial\Omega$，由第六章第三节所示的有限元法可唯一地确定区域 Ω 内每个点电位 $u_i(x,y)$，并可用计算出 $\nabla u_i(x,y)$。将区域 Ω 按有限元剖分并将其节点记为 $j=1,2,\cdots,J$，则有(8.175)有

$$\sum_{j=1}^{J}\sigma_j|\nabla u_{ji}|^2A_j=V_{0i}I_{0i},\quad i=1,2,\cdots,I \tag{8.176}$$

式(8.176)为线性方程组，A_j 为第 j 个元的面积。从原理上电导率成像只需求解方程组(8.176)，但实际上常因供电点及测点不够密而且不能完全覆盖边界 $\partial\Omega$，这种方法实现有一些困难。

将边界 $\partial\Omega$ 分为两段，$\partial\Omega_1$ 为有电位测量的边界，而 $\partial\Omega_2$ 为无法测量的边界(如图5.1中的下边界)，缺乏 $u_i|_{\partial\Omega_2}$ 使我们无法用有限元法计算域内每个节点的电位 u_{ji}。这时，可以用迭代法逼近边界上的电位值。设第 k 次迭代时已有电导率模型 $\sigma^{(k)}$，由(8.172)知

$$\left.\begin{aligned}&\nabla\cdot(\sigma^{(k)}\nabla u)=0\\&\sigma^{(k)}\frac{\partial u}{\partial n}\Big|_{\partial\Omega_1}=J_i\end{aligned}\right\} \tag{8.177}$$

则它可计算

$$u_i^{(k)}|_{\partial\Omega_2}=g(\sigma^{(k)}) \tag{8.178}$$

同时用有限元法计算域内 Ω 各节点的 $u_{ji}^{(k)}$，接下来用各正正则化方法解方程(8.176)求 $(\sigma_j^{(k+1)},j=1,\cdots,J)$。当然，收敛性与给定的初始模型 $\sigma^{(0)}$ 及 $\partial\Omega_2$ 的大小有关。

(二)一维地球大地电磁法反演[71~73]

在(3.101)中曾讨论过大地电磁反演电磁反射系数的问题。本节大地电磁法反演地下电阻率，先从水平层状介质地球模型讨论起。设这种介质共有 N 层，每层的电导率为 σ_n，厚度为 h_n，它们是要反演的地球参数。

大地电磁场源来自无穷远处的天空，垂直入射到地面上，因此场方程中场源项为零。选取平面入射波波前沿 xy 平面，有

$$\frac{\partial\boldsymbol{E}}{\partial x}=\frac{\partial\boldsymbol{E}}{\partial y}=0,\quad\frac{\partial\boldsymbol{H}}{\partial x}=\frac{\partial\boldsymbol{H}}{\partial y}=0 \tag{8.179}$$

用下标记电场或磁场沿某方向的分量，由于大地电磁场频率低(<1000Hz)，可忽略高阶无穷小 $\partial^2/\partial t^2$ 项，则(6.36)及(6.37)可写为

$$-\frac{\partial E_y}{\partial z}=\mathrm{i}\omega\mu H_x,\quad\frac{\partial E_x}{\partial z}=\mathrm{i}\omega\mu H_y,\quad H_z=0 \tag{8.180}$$

$$-\frac{\partial H_y}{\partial z}=\sigma E_x,\quad\frac{\partial H_x}{\partial z}=\sigma E_y,\quad E_z=0 \tag{8.181}$$

可见 E_y 只与 H_x 有关，H_y 只与 E_x 有关，我们只需测量其中的任一组，即可求出平面电磁波的波阻抗

$$Z_{xy}=\frac{E_x}{H_y}=\frac{E}{H}\text{ 或 }Z_{yx}=\frac{E_y}{H_x}=-Z_{xy} \tag{8.182}$$

这样测量出的波阻抗 $Z=E/H$ 称为标量波阻抗。如果地下介质不是水平层状介质而是水平不均匀介质，则 $Z_{xy}\neq -Z_{yx}$，必须同时测量 H_x、H_y 及 E_x 和 E_y，来计算所谓的张量波阻抗(见下一段)。

方程(8.180)及(8.181)为二阶常微分方程，可求出它的通解为

$$\left.\begin{aligned}E_x(z)&=A_n e^{-k_n z}+B_n e^{k_n z}\\ H_y(z)&=-\frac{k_n}{i\omega\mu}(A_n e^{-k_n z}-B_n e^{k_n z})\end{aligned}\right\} \tag{8.183}$$

其中 n 为深度 z 对应的层号，$k_n^2=i\omega\mu\sigma_n$，A_n 和 B_n 为第 n 层的积分常数。设 z_n 和 z_{n+1} 分别为第 n 层的顶面和底面深度，代入(8.182)有波阻抗

$$\left.\begin{aligned}Z_n&=Z_{0n}\frac{e^{-k_n Z_n}+B_n e^{k_n Z_n}/A_n}{e^{-k_n Z_n}-B_n e^{k_n Z_n}/A_n}\\ Z_{n+1}&=Z_{0n}\frac{e^{-k_n Z_{n+1}}+B_n e^{k_n Z_{n+1}}/A_n}{e^{-k_n Z_{n+1}}-B_n e^{k_n Z_{n+1}}/A_n}\end{aligned}\right\} \tag{8.184}$$

式中 $Z_{on}=i\omega\mu/k_n$，为第 n 层的特征阻抗。联立(8.184)并消去公共项得

$$Z_n=Z_{0n}\frac{Z_{0n}(1-e^{-2k_n h_n})+Z_{n+1}(1+e^{-2k_n h_n})}{Z_{0n}(1+e^{-2k_n h_n})+Z_{n+1}(1-e^{-2k_n h_n})} \tag{8.185}$$

此式给出自下而上递推计算地面 $z=0$ 处波阻抗 $Z(0)$ 的正演公式。一般来说周期为 T 的大场电磁波阻抗可表为

$$Z(0)=f(\sigma_1,\sigma_2,\cdots,\sigma_N;h_1,h_2,\cdots,h_{N-1};T) \tag{8.186}$$

f 为非线性函数，反问题为由不同周期 T 的地面阻抗 $Z(0)$ 观测数据求 $\{\sigma_n,\rho_n\}$。

注意到波阻抗 $Z(0)$ 为复数，它分为振幅 $|Z|$ 与相位 θ 两个参数。早期的反演只用到其振幅，即令

$$\rho_s=\frac{1}{\omega\mu}|Z|^2 \tag{8.187}$$

并称其为视电阻率。把不同周期 T 的 ρ_s 数据写为向量

$$\boldsymbol{\rho}_s=(\rho_{s1},\cdots,\rho_{sm})^{\mathrm{T}} \tag{8.188}$$

把地层参数 $\{\sigma_n,h_n\}$ 依(8.186)改写为向量

$$\boldsymbol{x}=(x_1,\cdots,x_{2N-1})^{\mathrm{T}} \tag{8.189}$$

用最小范数广义解可令目标泛函

$$Q(\boldsymbol{x})=\sum_{m=1}^{M}(\rho_{sm}-\rho_{cm})^2+\lambda\sum_{n=1}^{2N+1}\Delta x_n^2=m:n \tag{8.190}$$

求地层参数的广义解的修改增量 $\{\Delta x_n\}$，从而通过迭代不断逼近这种广义解，式中 ρ_{cm} 为用递推公式(8.185)和(8.187)式计算出的视电阻率。令 $\Delta\rho_s$ 为数据拟合差向量，由(8.190)可得线性方程组

$$[\boldsymbol{A}_0^{\mathrm{T}}\boldsymbol{A}_0+\lambda\boldsymbol{I}]\Delta\boldsymbol{x}=\boldsymbol{A}_0^{\mathrm{T}}\Delta\boldsymbol{\rho}_s=0 \tag{8.191}$$

其中 $\boldsymbol{A}_0$ 为 Jacobian 矩阵，计算很有讲究，见文献[72]。当然初始模型给定也很有讲究，尤其是层数 N 与电导率变化类型要给对，如未知时可先用下章第五节人工神经网络的方法预先将地层分类。

在(8.187)的视电阻率表达式中,只应用了波阻抗的振幅信息,而没有考虑复波阻抗的相位,而相位信息也应该用于提高反演的分辨率,记 θ_c 为复波阻抗的相位向量,(8.188)改写为

$$\boldsymbol{d} = (\rho_{a1}, \cdots, \rho_{am}; \theta_{c1}, \cdots, \theta_{cm})^{\mathrm{T}} \tag{8.192}$$

相应 Jacobian 矩阵写为

$$\boldsymbol{A} = \begin{Bmatrix} a_{ij} \\ b_{ij} \end{Bmatrix}, i = 1, M;\ j = 1, M \tag{8.193}$$

其中

$$a_{ij} = \frac{1}{|\rho_{ci}|}\frac{\partial|\rho_{ci}|}{\partial x_j}x_j;\ b_{ij} = \frac{\partial\theta_{ci}}{\partial x_j}x_j \tag{8.194}$$

于是,将(8.191)改写为

$$(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + \lambda\boldsymbol{I})\Delta\boldsymbol{x} - \boldsymbol{A}^{\mathrm{T}}\Delta\boldsymbol{d} = 0 \tag{8.195}$$

当此方程用于非线性迭代求地球模型修改增量 $\Delta\boldsymbol{x}$ 时,要按第三章第五节所述联合反演的方法,把 ρ_s 和 θ_c 数据及相应矩阵归一化,因为它们的量纲是不同的。

(三)二维地球大地电磁法反演

设地球介质沿 x 方向(走向)不变化,则大地电磁场对 x 方向的偏导数为零,由(6.36～6.37)有

$$\frac{\partial E_x}{\partial y} - \frac{\partial E_y}{\partial z} = \mathrm{i}\omega\mu H_x, \frac{\partial H_x}{\partial z} = \sigma E_y, \frac{\partial H_x}{\partial y} = \sigma E_z \tag{8.196}$$

及

$$\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} = \sigma E_x, \frac{\partial E_x}{\partial z} = \mathrm{i}\omega\mu H_y, \frac{\partial E_x}{\partial y} = -\mathrm{i}\omega\mu H_z \tag{8.197}$$

(8.196)仅与磁场分量 H_x 有关,因为磁场平行构造走向 x;而(8.197)式仅与电场分量 E_x 有关,因为电场平行构造走向;这说明垂直反射的电磁波场 x 方向可分解为两组独立的线性偏振波。

在测量电场垂直分量时要求垂直钻孔,一般不测,仅测量磁场及电场的水平分量。再者,测线与构造走向 x 可能有夹角 ϕ,这时地表面波阻抗应该用如下二阶复张量表示,它与 ϕ 角有关,与测量场分量的关系是

$$\begin{Bmatrix} E_x \\ E_y \end{Bmatrix} = Z\begin{Bmatrix} H_x \\ H_y \end{Bmatrix},\ Z = \begin{Bmatrix} Z_{xx} & Z_{xy} \\ Z_{yx} & Z_{yy} \end{Bmatrix} \tag{8.198}$$

在电性主轴方向观测时,有 $Z_{xx} = Z_{yy} = 0$,通过坐标旋转等办法可使它们为零,这时 $Z_{xy} \neq Z_{yx}$,对应有

$$\rho_s = \rho_{xy} = 0.2T|Z_{xy}|^2;\ \rho_s' = \rho_{yx} = 0.2T|Z_{yx}|^2 \tag{8.199}$$

将(8.198)改写为

$$Z = \frac{\begin{Bmatrix} (E_{x1}H_{y2} - E_{x2}H_{y1})(E_{x2}H_{x1} - E_{x1}H_{x2}) \\ (E_{y1}H_{y2} - E_{y2}H_{y1})(E_{y2}H_{x1} - E_{y1}H_{x2}) \end{Bmatrix}}{H_{x1}H_{y2} - H_{x2}H_{y1}} \tag{8.200}$$

式中下脚标1表示由方程(8.196)计算的对应 H_x 磁水平极化的电磁场,脚标2表示由方程(8.197)计算的对应 E_x 电水平极化的电磁场。电磁场分量一般用有限元法正演计算,所用的

变分泛函与(6.65)式相似。例如对 H_x，使泛函

$$F(H_x)=\iint_{\Omega}\left\{\frac{1}{\sigma}\left(\frac{\partial H_x}{\partial y}\right)^2+\frac{1}{\sigma}\left(\frac{\partial H_x}{\partial x}\right)^2-i\omega\mu H_x^2\right\}d\Omega=m:n \tag{8.201}$$

利用第六章第三节的剖分方法并代入边界条件，可得到求解各节点电磁场分量的线性方程组。

反演方法与一维情况相似，不过这时数据量已增加为(M 为总测点数)

$$\boldsymbol{d}\approx(\rho_{a1},\cdots,\rho_{am},\rho'_{a1},\cdots,\rho'_{am};\theta_{c1},\cdots,\theta_{cm},\theta'_{c1},\cdots,\theta'_{cm})^{\mathrm{T}} \tag{8.202}$$

而地球模型为(y,z)平面上每个剖分网格内的平均电导率$\{\sigma_j\}$。迭代反演时求模型修改增量 Δx 的线性方程组同(8.195)，其中矩阵 $\boldsymbol{A}$ 的含义同(8.193)。与一维反演最大的区别在于，二维及三维大地电磁场反演时地球参数大为增加，迭代反演要求有较接近的先验初始模型 x_0，否则迭代极易失败。在算法上，最近提出的共轭梯度法看来优点较多，详见文献[74]。

总的说，在这两节中我们讨论的非线性迭代反演算法都还存在依赖于初始模型选择和迭代能否收敛等问题，这些问题与非线性过程深层次的规律有关，将在下章深入讨论。

第九章　非线性反演方法

上一章所述的迭代线性化方法思路简单，但不一定能成功。比如我们能造过江的小木船，但任务却是过海。把许多小木船拼合起来，能完成过海的任务吗？

那就要看运气了。如果目的地只是陆沿的小岛，航行时没有大风浪，也许就成功了。

但是，科学不能靠运气。要想过海，就要研究海洋和航海的规律。线性反演比如渡江，非线性反演则如过海，线性迭代反演就像组合出来的"小木船"。

一、迭代、分形和浑沌

地球物理场能量很小，除天然地震震源物理研究外，场正演问题都归结为线性偏微分方程。但是，反问题都是非线性的。

非线性问题有什么共同规律？分形与浑沌。

(一)牛顿迭代与分形

如第三章及第八章后两节所述，非线性迭代的最基本方法是牛顿迭代法。也就是说，将函数展成台劳级数，略去高次项，从一次项中提出修改增量和 Jacobian 矩阵，构成线性方程组。牛顿迭代法收敛很快，但是收敛取决于初始猜测。

1988 年，Peitgen 与 Saupe 的论文集中发表了一个有趣的试验结果[75]，他考虑以下简单的非线性方程

$$z^3-1=0 \tag{9.1}$$

此方程的一个实根为 $z=1$，两个复根为

$$z=\exp(\pm 2\pi i/3) \tag{9.2}$$

用牛顿迭代格式

$$z_{j+1}=z_j-\frac{z_j^3-1}{3z_j^2} \tag{9.3}$$

来逼近，得到的是实根还是那一个复根？

当然，初值 z_0 可以是复平面 $z=x+iy$ 中的任一点。可以猜测，z_0 在复平面上可分为若干个区域，z_0 在某个区域用(9.3)作迭代后收敛为实根 $z=1$，在另外区域收敛于复根。习惯于线性思维的人会认为这些区域是有清晰边界分开的几块，如 z_0 等于 1 的邻域牛顿迭代将收敛于实根 $z=1$，它的面积大约占 z 平面的 1/3 左右，而其它区域收敛收敛于复根。事实并非如此，初值 z_0 的收敛区域是分形的，如图 9.1 所示。

从图 9.1 可见，黑色区域的面积的确是选初值区域($-2\leqslant x\leqslant 2$，$-2\leqslant y\leqslant 2$)的 1/3，但它的边界是分形的，即含有所有的尺度，彼此自相似。为什么像(9.1)式那么简单的方程和像(9.3)那么简单的迭代格式会导致这么复杂的分形图象？为什么初值在这种边界上的微小变化会使迭代收敛到完全不同的根？

问题归结为方程(9.1)的非线性,而非线性是系统走向浑沌的必要条件。对于非线性系统,初值的微小变化会使系统状态在几个"吸引子"之间回弹,其几何表现就是分形。

(二)分形地球模型

本书中的大部分内容,都把地球参数看成是实函数集,即 Hilbert 空间的元,这是确定性模型。

在上一章第二节关于最大熵法的讨论中,我们把实际反演的地球模型看成是随机的,并研究它的概率,这是地球的随机模型。确定性模型隐含着地球物质有序分布的假定,而随机模型隐含着地球物质随机分布的假定。我们现在进一步假定地球物质分布是自相似或自仿射的,具有多尺度的层次结构,这就导致地球的分形模型。

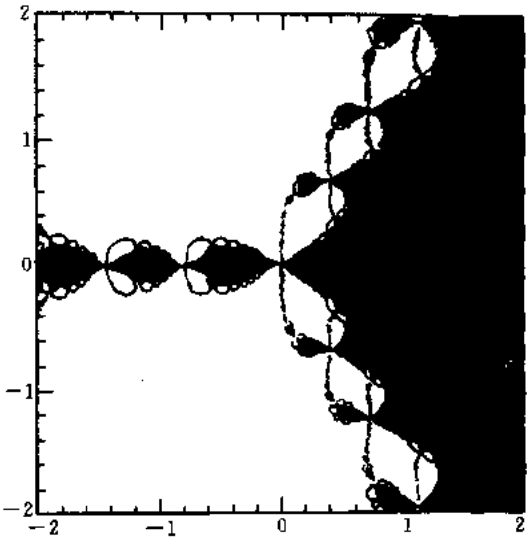


图 9.1 实虚轴在(−2,2)范围内的复平面 z
黑色区域经牛顿迭代后收敛于实根 $z=1$ 初值区,
白色为收敛于复根(9.2)的区域

从分形的观点描述地球的根据是:地球是无标度的复杂对象,其尺度可由几毫米的微裂隙到上万公里的地球直径,而不同尺度之间的现象具有相似性。

人有特征尺度,即人的平均身高,在 1.6m 或 5ft 左右。因此,人造的东西也有特征尺度,如火车的高度在 2m 上下,轮船和高楼平均为几十米,这种特征尺度称为标度。

自然现象一般具有多尺度的特征,没有特征尺度。分形几何学把不同尺度的现象用标度律联系起来

$$p(\lambda t)=\lambda^{\alpha}p(t),0<\alpha<1 \tag{9.4}$$

式中 $p(t)$ 为某种层次的尺度,$p(\lambda t)$ 为它放大 λ 倍之后的尺度,α 为标度指数。而

$$D_0=2-\alpha \tag{9.5}$$

等于 Mandelbrot 分维数。

维数指的是一个几何对象中一个点所置的独立坐标的个数,如地球表面的一个点用经纬度表示,它的维数是 2。在分形几何学中,维数可以为分数,分数的维数称为分维数。

对二维情况,一个正方形每边都放大三倍(尺度放大),则变为 9 个原正方形,有

$$2=\ln 9/\ln 3$$

对整数维为 d 的几何对象,每个方向都放大 L 倍,结果得到 N 个原来的对象,有

$$d=\ln N/\ln L$$

每个方向放大 L 倍等效于此方向测量尺度(或度量的单位)缩小为原来的 $\varepsilon=1/L$ 倍。因此,在一般情况下,用很小的度量单位 ε 研究对象的尺度变化时,可定义

$$D_0=\lim_{\varepsilon\to 0}\frac{\ln N(\varepsilon)}{\ln(1/\varepsilon)} \tag{9.6}$$

这就是 Mandelbrot 分形维。

1992 年 Korvin 编了一本名为"地学中的分形模型"的书[76],书中列举了与地球科学有关的许多分形模型。其中谈到,1984 年美国地调所出动数十辆消防车对内华达岩石出露区

进行冲洗，然后对其裂隙作详细填图，得出该区裂隙系统的平均分维数为 1.744。用大尺度的区域断裂构造图计算此区断裂系统的分维数为 1.773，证实了不同层次的地球断裂系统之间具有自相似性。陈颙与特科特等人的专著对此也有精彩的描述[77~79]。

关于分形几何学与其它分维数（如相关维 D_2、信息维 D_1 等）的讨论详见有关专著（如[80]）。以下只介绍对时间系列计算分形维 D_0 的方法。传统的介绍 D_0 分维数的方法多用时间系列的功率谱计算。由于地球物理资料的功率谱在高频段含有大量噪音，这种计算方法几乎不能用。我们研究出以下算法，在反射地震资料处理上取得良好效果。

对平面曲线，其总长度为

$$L=\varepsilon^{D_0}(N+f/\varepsilon) \tag{9.7}$$

式中 ε 为度量单位（尺子），N 为量得的尺数，f 为尺子量完后的剩余长度（$f<\varepsilon$），D_0 为 Mandelbrot 分形维数。将(9.7)两边取对数，有

$$\ln(N+f/\varepsilon)=-D_0\ln y+\ln L \tag{9.8}$$

设时间序列为 $\{s_1,s_2,\cdots,s_m\}$，取样率为 Δt，则用 $\varepsilon_1=\Delta t$ 为尺子量出它对应的曲线长度为

$$N_1+f_1/\varepsilon_1=\frac{1}{\Delta t}\sum_{i=1}^{m-1}[(s_{i+1}-s_i)^2+\Delta t^2]^{1/2} \tag{9.9}$$

在令 $\varepsilon_2=2\Delta t$ 为尺子时量出

$$N_2+f_2/\varepsilon_2=\frac{1}{2\Delta t}\sum_{i=1}^{(m-1)/2}[(s_{i+2}-s_i)^2+4\Delta t^2]^{1/2} \tag{9.10}$$

取 $\varepsilon_3=4\Delta t$，有

$$N_3+f_3/\varepsilon_3=\frac{1}{4\Delta t}\sum_{i=1}^{(m-1)/4}[(s_{i+4}-s_i)^2+16\Delta t^2]^{1/2} \tag{9.11}$$

将(9.9～9.11)代入(9.8)有方程

$$\ln(N_j+f_j/\varepsilon_j)=-D_0\ln\varepsilon_j+\ln L,\ j=1,2,3 \tag{9.12}$$

用最小二乘法易求出方程组(9.12)中的两个未知数 D_0 及 L。当然，还可取 $\varepsilon_4=8\Delta t$ 等，以提高求分形维 D_0 的准确度。下节中还要提到，反演迭代输出序列的分形维是指示迭代状态的一种有用参数。

（三）非线性迭代与浑沌

设 x_n 为第 x_n 步的迭代输出，x_{n+1} 为再下一步的迭代输出，二次方程

$$x_{n+1}=rx_n(1-x_n) \tag{9.13}$$

虽然很简单，但其迭代过程（演化）却是很复杂的。这个方程称为 May 生态方程。将 x_{n+1} 及 x_n 视为若干年后池塘中大鱼的产量，由于 x_n 越大繁殖越多，所以 x_{n+1} 与它成正比；又因大鱼越多吃的小鱼也越多，x_{n+1} 又与 $(1-x_n)$ 成正比。这就是生态方程的含义，系数 r 与饲料总量有关。

将 x_n 及 x_{n+1} 视为若干年后你的一笔银行存款总值，当年存款 x_n 越多次年本利就越多，所以 x_{n+1} 与 x_n 成比例。但是，存款越多银行的利率下降越多，x_{n+1} 又与 $(1-x_n)$ 成比例。系数 r 为控制参数，与银行存款总量有关。可见，生态方程反映许多自然与人文发展的规律。

将(9.13)式中 x_{n-1} 视为常数，则它是一个关于 x_n 的二次方程，有两个根。这意味着演化问题存在两种选择（线性问题只有一种选择）。x_n 有两种选择将造成迭代输出不稳定，在两种选择中跳来跳去。例如，池塘鱼的产量和水果产量常出现大年与小年的区别，这种演化称

为二齿分叉(Pitchfork bifurcation)。

取决于控制系数r,二齿分叉可能不断进行下去,即由两叉变四叉,四叉变八叉。具体地说,随r从很小变到$r=r_1=1.0$时,开始第一次分叉。当$r=r_2=3$时,再次分四叉,等等。此后,迭代变得非常不稳定,并很快变得没有规律和不可预测(即浑沌)。

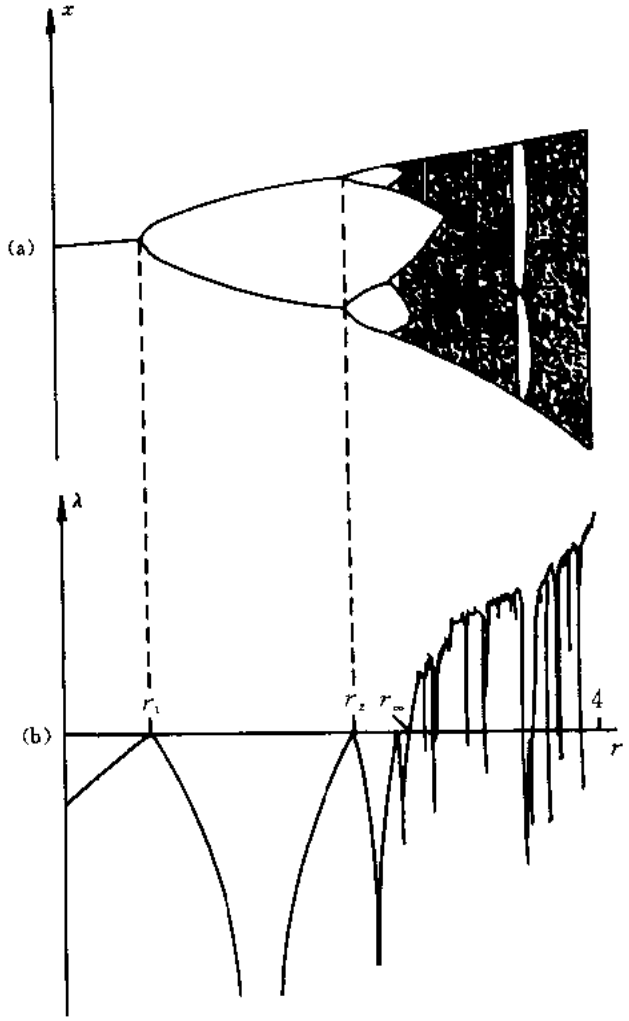


图 9.2　二次映射(9.13)的迭代输出x_n
随r的变化,黑色区表示浑沌区(上),以及 Lyapunov 指数的变化(下)

图 9.2 示出二次映射的迭代输出随控制系数的分叉过程[81],以及相应的 Lyapunov 指数。由图可见,二次映射迭代随外部控制参数r的增大导致有规律性的分叉,直至走向浑沌。

在非线性动力学中,浑沌指的是非线性系统演化的一种不确定和无规则状态。分叉、间歇、突变(如相变)都是典型的不规则状态。在地球科学中,火山爆发是典型的间歇,地震发生是能量的突然释放,其形成的断裂裂隙具有分形结构[75~79]。

浑沌发生的必要条件是系统为非线性。多层次的复杂非线性系统(如人类社会)由于其自组织的困难,较易演化为混沌运动(如战争)。开放的耗散(Dissipative)系统由于固有的非线性性质,也经常出现浑沌。但是,非线性只是浑沌运动发生的必要条件,而不是充分条件。

浑沌运动的特征包括:

(1)不可预测性,指初始条件有微小的差别将导致最终结果迥然不同。

设迭代映射方程为 $x_{n+1}=f(x_n)$，例如当 f 为二次函数时，它变成(9.13)的 May 生态方程。f 在一般情况下指任何导致浑沌结果的函数。如果初始条件 x_0 带有微小误差 ε。经过 N 次迭代后其误差被指数放大，记 $f_N(x_0)$ 为不带误差的迭代结果，$f_N(x_0+\varepsilon)$ 为带误差的迭代输出，有

$$|f_N(x_0+\varepsilon)-f_N(x_0)|=\varepsilon e^{N\lambda(x_0)} \tag{9.14}$$

因此定义

$$\lambda(x_0)=\lim_{N\to\infty}\lim_{\varepsilon\to 0}\frac{1}{N}\ln\left|\frac{f_N(x_0+\varepsilon)-f_N(x_0)}{\varepsilon}\right| \tag{9.15}$$

为 Lyapunov 指数。还可将(9.15)写为

$$\lambda(x_0)=\lim_{N\to\infty}\frac{1}{N}\ln\left|\frac{\mathrm{d}f_N(x_0)}{\mathrm{d}x_0}\right|\sim\frac{1}{N}\ln\frac{\varepsilon_N}{\varepsilon} \tag{9.16}$$

可见 Lyapunov 指数表示经 N 次迭代后系统演化轨道加速偏离的指数。设 $|\Delta I|$ 为经过一次迭代后系统信息的平均损失，有

$$\lambda(x_0)=\ln 2|\Delta I| \tag{9.17}$$

说明 λ 与 $|\Delta I|$ 成正比。根据 Shannon 信息论，系统信息量等于对该系统作完备描述编码所需的最少 bit 数目。当 $\lambda>0$ 时，每次迭代的信息损失都大于零，系统的熵不断增大以导致浑沌的发生。图 6.2 下方示出了二次迭代的 λ 随 r 的变化并将它与系统的分叉和浑沌作对比。由图可见，$\lambda<0$ 时对应系统的稳定，在 $\lambda=0$ 的点系统发生分岔，而 $\lambda>0$ 的点对应浑沌。因此，Lyapunov 是指示状态的重要标量参数。

(2)整体行为的有规律性。虽然系统在未来的具体状态具有不确定性和不可预测，但是"表面上看起来疯狂杂乱，其实自有规矩"(莎士比亚)。所有系统演化的轨迹形成的相空间的图形中，存在若干个吸引轨迹的若干个很小的空间(称为吸引子)，使轨迹不断收缩到其中，或者突跳到另一个吸引子附近。这种现象表示整体行为仍具有规律性。

整体行为的规律性还表现在不同层次的运动的相似性(分形)上。Feigenbaum 证明，无论是那种形如 $x_{n+1}=f(x_n)$ 的混沌运动，其转化为浑沌的尺度特征都由两个普适常数控制，更说明浑沌理论具有整体规律性[82]。

(3)形式周期性，浑沌状态的发生有时会重复出现，但这种重复是不确定的。例如，大地震的发生时多时少，既包括高频度的复复出现，又没有准确的周期[77~79]。

非线性科学研究的全面展开，还是 90 年代的事。19 世纪建立了线性科学的理论框架，它在 20 世纪发展为完整的体系。但是非线性科学理论框架的建立，将是 21 世纪的事。对正问题的研究尚且如此，对非线性反问题的研究更加零星。接下来将介绍我怎样根据浑沌理论的基本思想，指导非线性反演研究取得的一些成果。

二、地震道非线性浑沌反演[83~86]

(一)间断型折衷曲线与线性反演的失败

回忆第四章的第六节中讨论的 *BG* 折衷准则时，我们曾用图 4.6 展示了折衷曲线，并试图用改变折衷因子 θ 来找到在分辨率与方差之间取最佳折衷的广义解。在第八章的第一节讨论正则化理论时知，一般情况下反演的正则化因子都起到折衷因子的作用，即在提高广义

解的质量(拟合优度)与压抑人造假象之间取折衷。

我的研究发现,上述理论隐含着一个前提:即折衷曲线是连续可微的,只有此时才存在最佳折衷的参数 θ_{opt}。但实际情况并非总是如此,至少有一部分反演问题,它的折衷曲线是间断型的(如图 9.3 所示),这里找不到最佳折衷的 θ_{opt}。

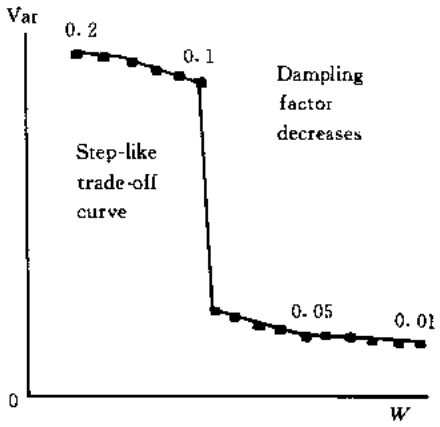


图 9.3 阶梯型折衷曲线示意图
纵坐标为方差 σ^2,横坐标为分辨宽度 W
(见第四章第六节)

以最简单的一个线性问题为例,让我们做一个地震道脉冲反褶积的试验。假设地震道为子波 w 与反射系统序列 r 的褶积

$$s_t = \sum_{\tau=0}^{m} w_{t-\tau} r_\tau \qquad (9.18)$$

由地震道 s 反演 r 称为脉冲反褶积。用最小二乘法求(9.18)的广义解为

$$r=(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}+\lambda\boldsymbol{I})^{-1}\boldsymbol{W}\boldsymbol{s} \qquad (9.19)$$

其中 $\boldsymbol{W}$ 为不同延时的子波 $w_{t-\tau}$ 构成的矩阵,λ 为阻尼系数,即正则化因子。方程(9.19)可用第三章(3.82)式介绍的自然逆法求解。

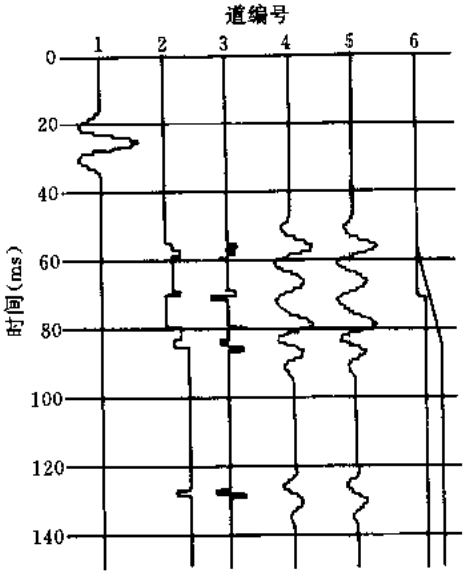


图 9.4 数值试验的资料
道号 1 为子波,2 为波阻抗模型,3 为反射系数序列,4 为用褶积合成的地震道,5 为加微小误差的合成地震道,6 为非线性迭代的初始波阻抗模型

在图 9.4 的前 4 道中分别示出了子波、地层波阻抗、反射系数序列与合成地震道。在合成地震道中加入微小的高斯噪音,其均值为零,方差为地震道最大幅度 s_{max} 的 2.5%,示如图 9.4 的第 5 道。线性反演时,选取 $\lambda_k=0.5/5^{k-1}$,$k=1,2,\cdots,10$;即 $\lambda_1=0.5$,$\lambda_2=0.1$,$\lambda_3=0.02$,…。反演结果示于图 9.5。

由图可见,当 $k<4$(即 $\lambda_k>0.04$)时,反射系数序列的反演结果是有序的,但是分辨率很低;而在 $k>5$ 之后反演的结果方差突然增大并把信号全部掩盖,无法找到分辨率较高而方差又不明显的最佳折衷解估计。试验表明,对应线性反演的折衷曲线是间断型时(如图 9.3)所示,此时无法找到符合实际要求的广义解。

(二)非线性地震道反演走向浑沌

上述线性反演失败的一个原因是,地震道数据 s 只含有限带的信息,而反射序列 r 却要求具有类似脉冲函数的全频带信息。换句话说,反演要求的一部分信息在输入信号中不存在。

在第三章第七节中已介绍了由地震道反演地层波阻抗的广义线性反演方法。由于波阻抗序列的频谱随频率的增加而减少,与地震道频谱的高频段相似,因此估计作波阻抗时反演算子性状会有所改善。但这时问题变为非线性反问题,要进行迭代,迭代公式为

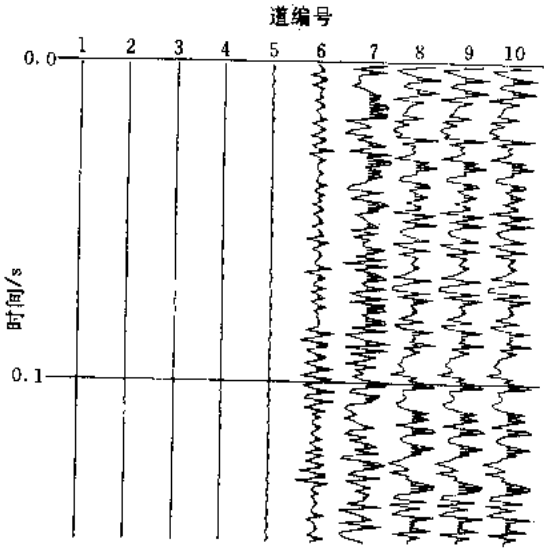


图 9.5　带微小误差的地震道的线性反演结果

道编号 k 对应几何级数递减的阻尼系数，$k>5$ 之后方差突然爆发，预示折衷曲线属于图 9.3 的间断型

$$z_k=z_{k-1}+\mu_k\Delta m_k \tag{9.20}$$

式中 k 为迭代次数，z_k 为第 k 次迭代输出的波阻抗向量，Δm_k 为 k 次迭代波阻抗的修改增量，由下述方程解得　　$\Delta m_k=\left(A_{k-1}^{\mathrm{T}}A_{k-1}-\lambda_k I\right)^{-1}A_{k-1}^{\mathrm{T}}(s-s_{k-1})$　　(9.21)

式中 A_{k-1} 为前次迭代求出的 Jacobian 矩阵，由(3.121)式计算，s_{k-1} 为前次迭代模型 z_{k-1} 的合成地震道。

由上节分析知，对非线性迭代系统，控制参数是极其重要的。在(9.20～9.21)式中，有两个控制参数：正则化因子 λ_k 与弛张因子 μ_k。

为了求得高分辨率的解估计，阻尼因子 λ_k 必须随迭代逐渐减小。例如，可用几何级数序列

$$\lambda_k=0.92\lambda_{k-1},\lambda_0=0.1\sim0.2 \tag{9.22}$$

表示 λ_k 随 k 的减小，与下节仿真淬火算法中温度下降的公式相似。

弛张因子 μ_k 要考虑到抑制反演中不可避免的人造因素。这就是说，如果(9.21)中的 Δm_k 功率越大，它包含的人造假象成分也就越多。因此，可令

$$\mu_k=1-\frac{\|\Delta m_k\|}{\|z_{k-1}\|} \tag{9.23}$$

将(9.23)代入(9.20)有

$$z_k=z_{k-1}+\left(1-\frac{\|\Delta m_k\|}{\|z_{k-1}\|}\right)\Delta m_k \tag{9.24}$$

这个公式就是 May 生态演化方程(9.13)的变种。根据上节讨论的关于浑沌理论的知识可以预测，地震道反演迭代的过程具有浑沌性质，当数据有微小误差时，随 k 增大到一定程度($k\to\infty$)，反演系统将输出无序的波阻抗序列$\{z_k\}$。

试验的结果正是如此。在图 9.6 上图示出当数据不含误差时的波阻抗反演输出序列，分辨率随迭代提高。下图示出带微小误差的反演输出序列，当 $k>43$ 时，z_k 突然变为无序，而在 $k=65$ 时，迭代的无序影响到计算机软件系统而引起死机。

我在 1989 年进行的这次试验使我确信，反问题的非线性与混沌理论研究的正问题的非

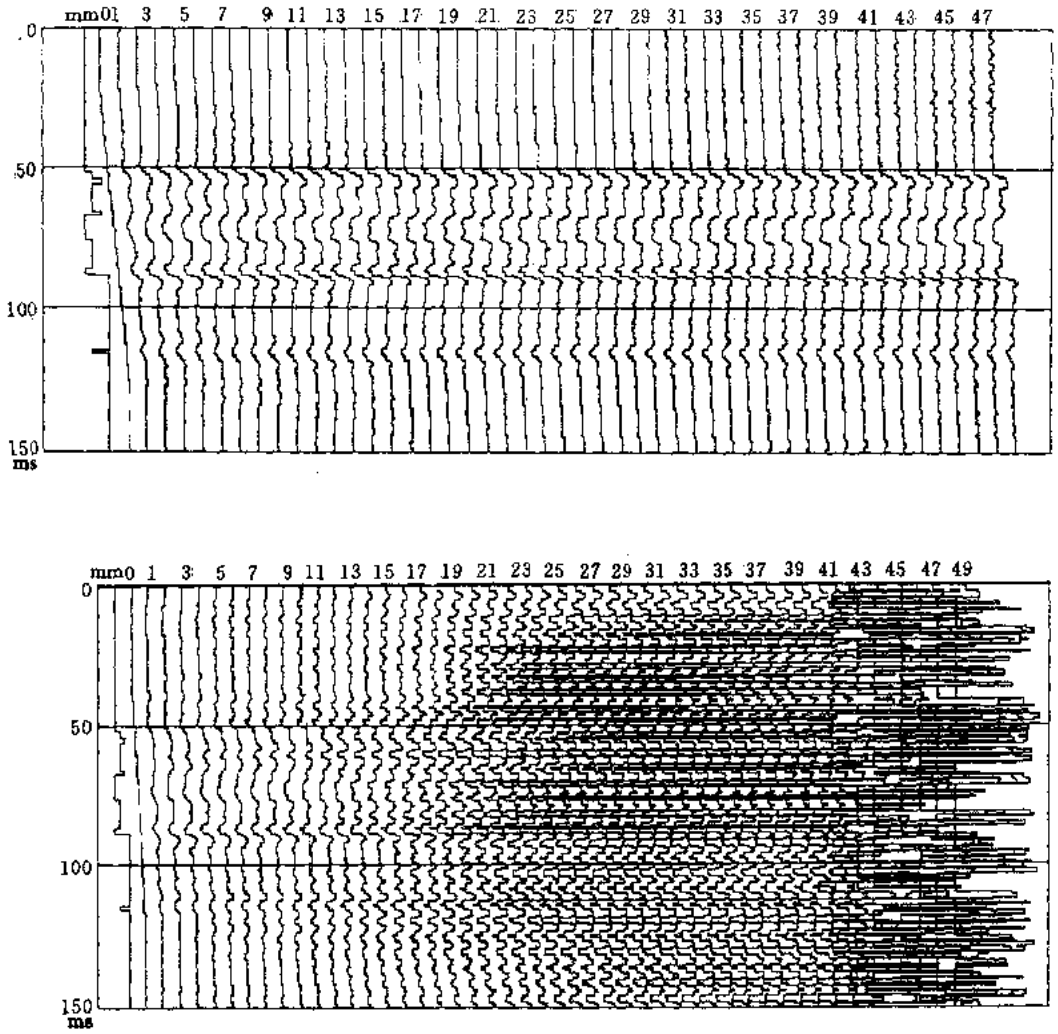


图 9.6 地震道非线性反演输出序列

(上)为精确的输入数据的反演;(下)为带微小误差的输入数据的反演

线性之间遵循某些共同的演化规律。即非线性迭代过程包含着若干个以跃变连接的阶段(相当于相变),而最终的阶段为浑沌状态,这是与线性化迭代的经典提法不同之处。经典的迭代理论只有两种模式:收敛与发散,其依据是迭代过程是连续渐进的。

(三)地震道非线性混沌反演

应用混沌理论的思想研究地震反问题时,要看到两者在非线性领域的共性,也要注意正问题与反问题的不同点。

浑沌理论研究非线性常微分方程的演化,而地震反演研究的是与线性偏微分方程求广义逆有关的逐次逼近。

浑沌理论研究初始条件微小变化使系统状态走向无序的规律,地震反演研究初始与边界条件两者的变化使系统走向无序的规律。

浑沌理论中系统演化特征取决于 Poincare 映射,如(9.13)式所示的 May 生态方程。在地震反演中,系统的 Poincare 方程复杂一些,如写成标量形式,可有

$$z_{i,k} \simeq z_{i,k-1} + \left(1 - \frac{\Delta m_{i,k}}{\sum_i z_{i,k-1}}\right)\Delta m_{i,k} \tag{9.25}$$

其对应的系统方程是

$$\frac{\mathrm{d}z_{i,k}}{\mathrm{d}k} = \left(1 - \frac{\Delta m_{i,k}}{\sum_i z_{i,k-1}}\right)\frac{\mathrm{d}(\Delta m_{i,k})}{\mathrm{d}k} - \frac{\Delta m_{i,k}}{\sum_i z_{i,k-1}} \cdot \frac{\mathrm{d}(\Delta m_{i,k})}{\mathrm{d}k}$$

$$= \left(1 - \frac{2\Delta m_{i,k}(\lambda_k)}{\sum_i z_{i,k-1}}\right)\frac{\mathrm{d}[\Delta m_{i,k}(\lambda_k)]}{\mathrm{d}k} \tag{9.26}$$

在浑沌理论中，Lyapunov 指数(李指数)是指示非线性动力学系统特征的重要参数，由(9.15)式定义。对地震反演也可以应用 Lyapunov 指数，但是其定义要有所改变。

对一个非线性动力系统而言，混沌运动来源于对初始条件极其敏感的依赖性。对一个非线性反演系统而言，系统的状态和输出同时依赖于数据的误差和初始模型两方面，因此李指数的定义也可以不止一种。

考虑到地震道反演的状态参数及输出是波阻抗向量(或函数)，因此要用模来代替(9.15)式中的量。于是，可以把依赖于初始模型的李指数定义为

$$L_1(Z_0) = \lim_{k\to\infty}\lim_{\delta\to 0}\frac{1}{k}\lg\frac{\|Z - Z_k\|}{\delta} \tag{9.27}$$

式中 Z 为真实(或理论)的波阻抗模型，Z_k 为第 k 次迭代计算的波阻抗输出，$\delta = \|Z - Z_0\|$ 为初始偏差的模。显然，L_1 应显示初始模型的偏差随反演迭代的演化。

对于数据中误差的仿李指数，可参照(9.15)式定义为

$$L_2(s) = \lim_{k\to\infty}\lim_{\|\delta\|\to 0}\frac{1}{k}\ln\frac{\|s - s_k\|}{\|\delta\|} \tag{9.28}$$

式中 s 为地震道数据，s_k 为第 k 次迭代时的合成地震数据，δ 为实测数据中的误差，L_2 表示数据误差在非线性反演迭代过程中使轨道偏离的指数。如果综合考虑初始模型和数据误差的影响，还可以定义

$$L_3 = \alpha L_1 + \beta L_2, \alpha, \beta \geqslant 0, \alpha + \beta = 1 \tag{9.29}$$

式中 α 和 β 相应为对初始模型和数据误差的作用加权时用的权系数。在混沌理论中，李指数为负是动力学系统有序的指示，在地震反演中对应方差很小的稳定状态。但如果其绝对值很大，则表示解估计的分辨率很低。在动力学系统中，李指数由负变正往往是分岔或相变的标志，在地震反演中它们变为大的正数，应该指示所谓的“发散”，即输出无序的波阻抗序列。总之，由(9.27)和(9.29)式规定的仿李指数应该成为反演迭代系统状态的一种重要表征。可惜的是，(9.27)和(9.29)式定义的仿李指数只能用于数值试验，而不能直接用于实际地震道反演，因为此时两式中的分母项是未知的。当然，可有多种近似计算李指数的方法，使得它们可用于实际资料的反演。由于具体的近似计算方法纯属技术问题，不必详述。这里只指出当 k $<$10 时，(9.27)及(9.29)式中有取极限项，很难算准，因此这时近似计算出的李指数会有很大差别。用大写字母 L_1 及 L_2 表示由(9.27)及(9.29)式计算出的精确仿李指数，用小写字母 l_1 及 l_2 表示参照此两式近似计算出的精确仿李指数，以示它们的区别。试验表明，如果我们的目的是找出反映迭代系统状态的表征，它们之间有无区别并不重要。

图 9.7 表示对图 9.4 有误差数据反演时(对应图 9.6 下)李指数(图 9.7 上)及近似计算

的仿李指数(图 9.7 下),相应阻尼因子 λ_k 是随几何级数减少的。如图可见,李指数变化有如下规律:

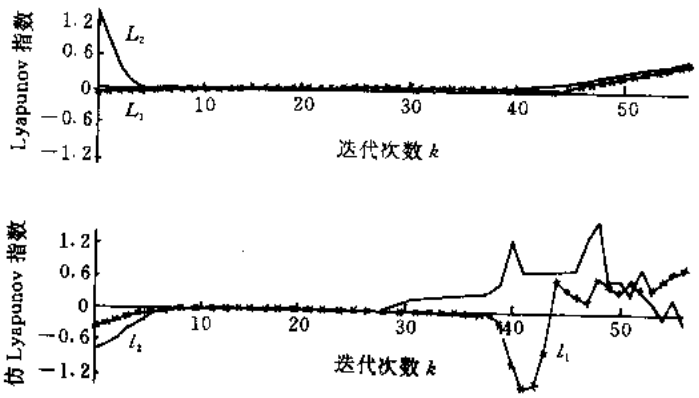


图 9.7 李指数(上)与近似计算的仿李指数(下)

此图对应的迭代输出序列为图 9.6(下),阻尼系数 λ_k 在迭代时呈几何级数递减

①$k=1\sim6$,李指数趋向于零,输出波阻抗序列方差很小,但是分辨率提高。

②$k=7\sim22$,李指数为小于零但接近于零的小数,输出序列分辨率继续提高,方差也增大。

③$k=23\sim41$,李指数为大于零的正数,输出序列稳定,同时具有高分辨率及大方差。

④$k>41$,李指数突然增大,输出序列变为无序。

输出序列的这种阶段性与浑沌理论中的相变有一定的相似性。

从应用的观点看,在第二阶段结束前,即两个李指数同时由负变正时停止反演的迭代过程.将会取得分辨率尽可能高的解估计。这时虽然方差不是最小,但仍然远小于信号幅度。

陈志德的试验表明,输出序列分维数 D_0 的变化率最大值对应李指数由负变正,也可以用作判断迭代是否中止的准则。

图 9.7 还表明,在李指数由负变正之前及之后的几个迭代步,近似计算的仿李指数 l_1 和 l_2 与准确的 L_1 和 L_2 基本相同,可以代替反演过程中的 L_1 和 L_2 计算(实际反演迭代在第

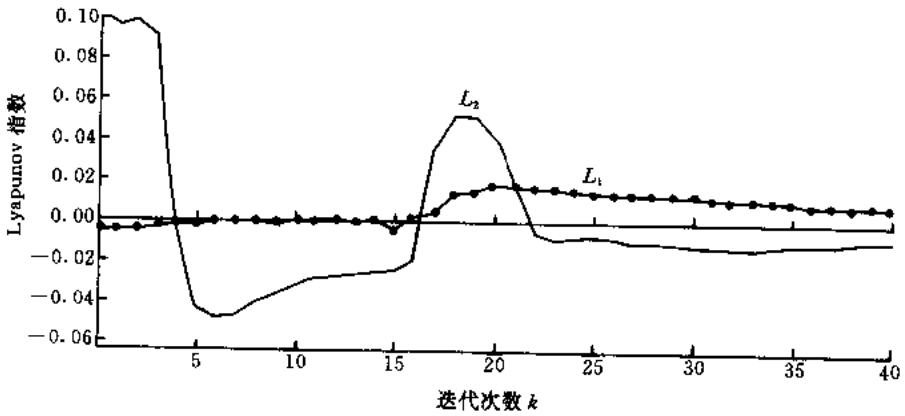


图 9.8 阻尼系数 λ_k 呈算术级数递减时的李指数变化,

L_2 呈多峰状显示迭代的复杂性

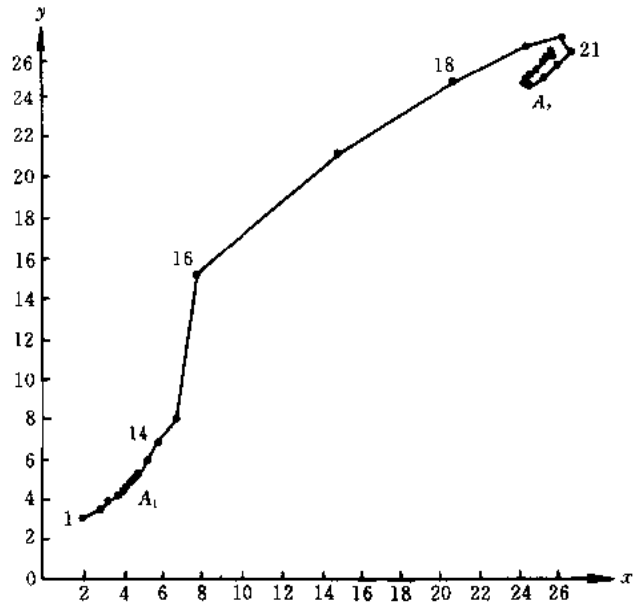


图 9.9　对应图 9.8 波阻抗输出序列的相空间重建

三阶段前中止)。

当然，非线性地震反演的控制参数 λ_k 也可以随其它级数形式减少。例如，呈算术级数递减

$$\lambda_k=\begin{cases}\lambda_{k-1}-\Delta\lambda, & \text{当 } \lambda_{k-1}>\Delta\lambda\\ \Delta\lambda, & \text{当 } \lambda_{k-1}\leqslant\Delta\lambda\end{cases} \tag{9.30}$$

这时，地震道反演仍然有走向无序的最终迭代阶段，但阶段性更为复杂。图 9.8 显示了 λ_k 呈算术级数递减时李指数的变化，其中 L_2 呈多峰状，表示在 $k=16\sim22$ 时输出有一个准无序状态，然后又回到 $L_2<0$ 的亚稳定阶段。一般来说，阶段与相空间吸引子对应(见图 9.9)，而奇异吸引子的出现可使迭代过程复杂化[84]。

地震道浑沌反演方法已经过国际验证，同时在大庆试验了几百公里的地震剖面，取得了良好效果。但是，与非线性科学的宏大和复杂性相比，这项研究只不过是肤浅的起步。

我希望看到的不仅是非线性迭代分阶段变化以及最终走向浑沌的规律性，这可以说是预料之中的事。我更希望看到非线性迭代过程具有自组织的功能，使解空间的搜索保持有序。

三 仿真焠火反演[65,87,88]

从第八章第一节的讨论知，反演问题大多转化为泛函极小问题。如果(6.31)式中泛函是非线性的，可以通过求 Fréchet 导数线性化，也可以不求 Fréchet 导数，直接在解空间搜索泛函极小对应的广义解。在后一类方法中，仿真焠火(Simulated Annealing)是目前最接近实际应用的一种。

(一)仿真焠火与系统能量最小状态

焠火是众所周知的一种金属加工工艺，指金属从高温逐渐冷却的一种热处理。高温时，分子相对运动的热活动性大，在冷却时热活动性逐渐消失，最后形成原子有序排列，处于热动力系统的最小能量状态。快速的突然冷却常不能达到最小能量状态，只有在缓慢冷却的条件下，在原子失去热活动性的同时有足够的时间让它们重新分布，才能达到最小能量状态。

焠火这一物理过程的计算机仿真就成为一种求取泛函极小的最优化算法。与焠火过程类似，这种算法要求温度缓慢下降，这是与降低计算成本相矛盾的。

仿真焠火算法将非线性优化问题与统计力学中的热平衡问题进行类比，找出了优化问题求解与固体焠火过程之间的相似性，从而另辟优化问题数值求解的新途径。

对于非线性优化问题，设非负目标函数为 $\phi(v)$，要求

$$\phi(v^*)=\min\phi(v_i),v_i\in V \tag{9.31}$$

其中 v_i 为某一个可能的解，所有 v_i 的集合记为 $V=\{v_i\}$，$i=1,2,\cdots,N$；v^* 为 V 中使 ϕ 取极小的一个解估计。在与淬火过程类比时，ϕ 视为系统的内能，要使它由高到低，最后取极小而终止。v_i 为解空间的一个点，对应为固体内的一种粒子状态，它的扰动范围随内能的减小而逐渐减小。温度 T 用控制参数 $T'=kT$ 类比，其中 k 为波尔兹曼系数。仿真淬火算法在求解优化问题时先取很大的 T'，T' 可称为仿真温度，然后逐渐降低 T'，计算时不同的迭代步对应着时间与温度 T' 的变化。当系统温度高内能大时粒子的随机扰动大，相应于系统状态的 v_i 扰动的自由度也大。随迭代进行温度降低，当前解 v_i 不断更新而形成序列，构成了对解空间的搜索。

按照统计力学中关于淬火过程的理论，仿真淬火算法遵循的基本规律是：某种状态 v_i 在当前出现的概率服从波尔兹曼分布

$$f=\frac{e^{-\phi(v_i)/kT}}{\sum_{j=1}^{N}e^{-\phi(v_j)/kT}} \tag{9.32}$$

由(9.32)式可见，当仿真温度 $T'=kT\to 0$ 和目标函数 ϕ 取极小时($v_i=v^*$)，系统处于内能最低的基态，这时概率 f^* 取极大而接近于 1，说明系统最稳定，熵最小。因此，仿真温度降至某一阈值以下，而且系统的熵达到最小可作为停止迭代的准则。

仿真淬火算法的关键在于：在给定仿真温度 T'_i 和当前状态 v_i 之下，如何确定下一个状态 v_{i+1}，即如何构造迭代修改序列以便使解空间中的搜索合理地进行，而最后找到使 ϕ 取全局极小的 v^*。仿真淬火算法的创始人 N. Metropolis 于 1953 年提出了一种抽样算法。他说，假定当前的状态为 v，利用随机扰动产生一种新状态 v'，其对应系统内能的变化为 $\Delta\phi=\phi(v')-\phi(v)$。根据(9.32)式，这种新状态可以被接受的概率为

$$p=\exp(-\Delta\phi/T') \tag{9.33}$$

如果 $\phi(v')$ 比 $\phi(v)$ 减小或相同，即 $\Delta\phi\leqslant 0$，则有概率 $p\geqslant 1$，这种新状态毫无疑问要被接受。如果 $\Delta\phi>0$，由(9.33)式计算 $p<1$，但也不一定为零。这时新状态虽然使目标函数增大，但为保证解空间的搜索是全局性的，不至于落入局部极小而无机会逃出，给予 $p>0$ 的机会是正确的，它给予从局部极小逃出的机会。根据(9.33)式建立了 Metropolis 内循环抽样算法，将在下面详述。

根据仿真淬火的算法原理，可以把算法的实现步骤大致归结如下。

第一步，选取 $i=0$ 时的初始模型及仿真温度。随机选取初始地球模型 v_0；并依照经验选取初始温度 T'_0。也可通过实验选取 T'_0，例如，随机抽样一组模型$\{v_j\}$，计算对应的目标函数值$\{\phi_j\}$，取$\{\phi_j\}$的方差或最大 $\Delta\phi_j$ 的若干倍为 T'_0。

第二步，在给定 v_i 和仿真温度 T'_i 之下用 Motropolis 抽样对解空间进行迭代搜索，具体由以下内循环迭代过程构成。

(1)令 $k=0$ 时当前解估计 $v(0)=v_i$，参数 $T=T'_i$；

(2)按某一半随机准则在当前解 $v(k)$ 附近产生一个邻近子集 $V_n(v(k))\subset V$，它不能为空集，与 $v(k)$ 的距离(或半径)与参数 T 成比例。在子集中随机地取出一个新的扰动模型 v' 作为下一步迭代的候选解，计算

$$\Delta\phi=\phi(v')-\phi(v(k)) \tag{9.34}$$

(3)如 $\Delta\phi\leqslant 0$，则接收 v' 为下步迭代的当前解，即令 $v(k+1)=v'$。如若 $\Delta\phi>0$，则按(9.33)式计算概率 p'，并计算一个在(0,1)之间的随机数 p^*。如果 $p'>p^*$，则接受 v' 为下一步迭代的当前解。否则令 $v(k+1)=v(k)$，即保留原来的当前解估计。

(4)令 $k=k+1$，如果 $k+1$ 不超过某个规定的次数返回(9.32)继续抽样搜索，否则用当前解估计 $v(k+1)$ 跳出内循环，转入下述主程序的第三步。

第三步，按一定准则降温。此时令 $i=i+1$，$T_{i+1}=\lambda_i T_i$，$\lambda_i<1$。例如，λ_i 可在(0.2,0.99)之间变化，但随 i 增大而增大。

第四步，判断是否停止外循环迭代，如果不是，则返回第二步，以当前 $v_{i+1}=v_{(k+1)}$ 及 T_{i+1} 重复 Metropolis 抽样搜索。外循环停止计算的准则是 T' 小于某一规定阈值和系统的熵达到极小；或者由(9.33)式计算当前解的概率 $f\to 1$ 等。

仿真焠火算法的收敛性可以用马尔可夫链理论来研究，因为非平稳可数马尔可夫链可恰当地描述解空间的搜索序列。总的来说，在仿真焠火算法中，如果仿真温度 T' 的变化等参数满足一定的条件收敛是有保证的，但收敛速度较慢。如果 T'_0 足够高而且下降充分慢，以及对每一次下降，Metropolis 抽样都很稳定，则能达到 $\phi(v_i)$ 的全局极小值。这些条件正好要求加大计算成本，这是仿真焠火固有的缺点。

(二)仿真焠火地震反演

进行地震反演时，目标泛函可选类似(6.31)的任何形式。例如，对不考虑正则化的情况，以拟解选

$$\phi_k(r)=\sum_{j=1}^{J}[s_{(j)}-s_{k(j)}]^2 \tag{9.35}$$

为计算第 k 次迭代时的目标泛函值，其中 r 为地震道反射系数序列，s 为地震道数据，s_k 为第 k 次迭代时用 $(k-1)$ 次迭代结果计算的合成地震道。

初始温度 T_0 可选为 10000～20000。

每一次迭代，温度下降为 $T_{k+1}=0.92T_k$，这与浑沌反演的阻尼 λ_k 下降一致，几何级数递减。

每一次迭代随机试验拟解的修改最大次数约为 3000。设当前第 j 个模型参数为 r_j，随机扰动的公式为

$$r'_j=r_j+\mathrm{Rand}(I)\cdot\beta(T_k) \tag{9.36}$$

式中 Rand(I) 为一个[−0.5,05]之间取值的随机数，β 为与当前迭代温度有关的权。温度 T_k 越低时，权 β 应越小才能提高成功率。

按(9.36)修改模型参数后作地震合成并代入(9.35)式求 $\phi_k(r')$，如果 ϕ_k 减小则修改成功，否则按 Metropolis 算法确定是否用 r' 代替 r。接下去再作扰动试验。

对每个温度级(即每个 k)，规定最大扰动成功次数(如 500)，在此级成功修改多次后即降低温度进入 $(k+1)$ 次迭代。

迭代停止的准则是：

(1)在 k 个温度级修改 r3000 次，成功率为 0；

(2)温度已下降到室温；

(3)$|\phi_{k+1}-\phi_k|<\varepsilon$，$\varepsilon$ 为规定的小数。

按以上参数准则，我用仿真焠火进行了与混沌反演相似的地震道反演试验，对无误差数

据结果示如图 9.10，子波为 11 点零相位 Ricker 子波，频率为 30Hz。

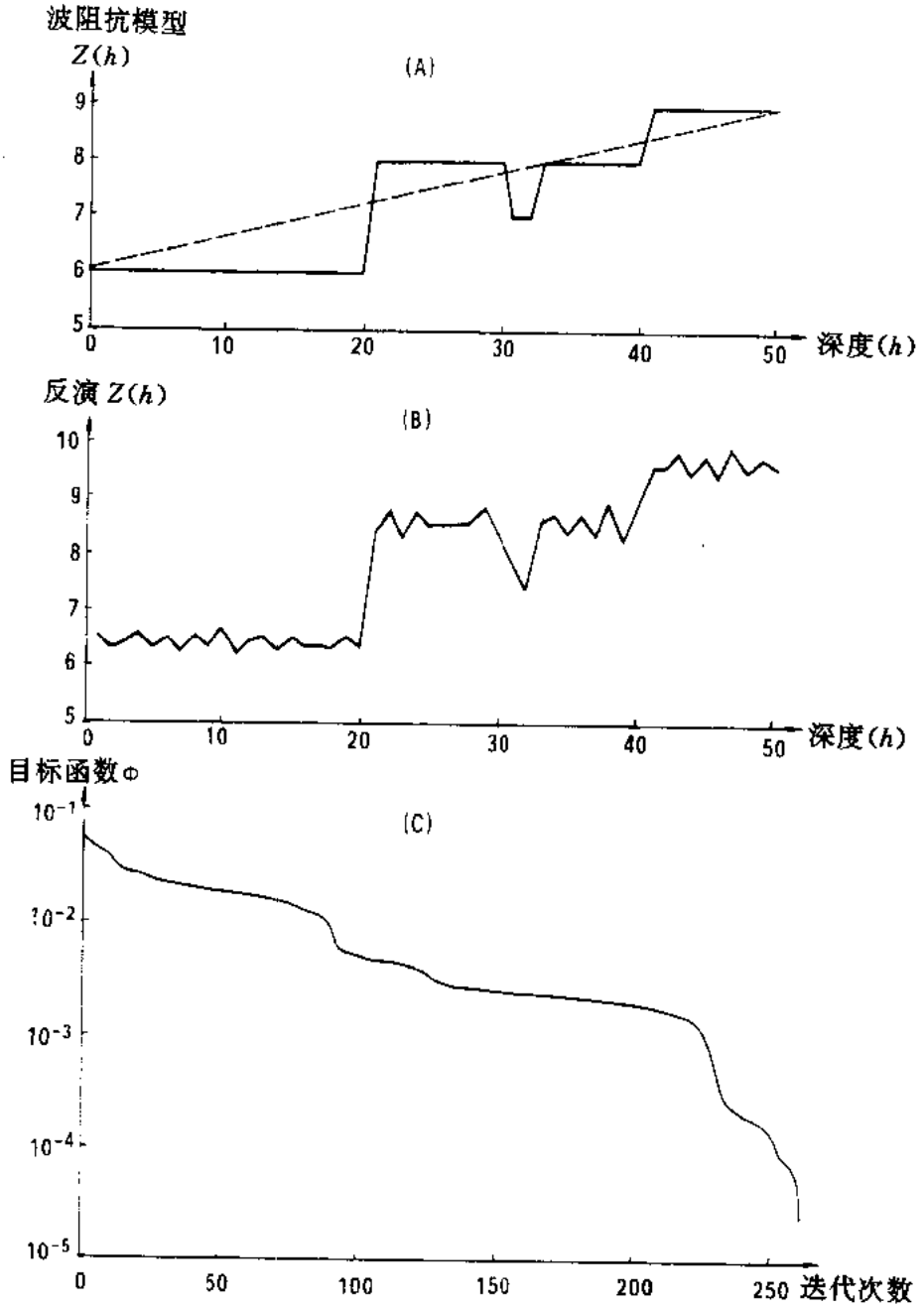


图 9.10　仿真淬火地震反演试验(精确数据)

(A)地层波阻抗模型，虚线为初始模型；(B)反演最终取得的波阻抗序列；(C)目标泛函随迭代的下降曲线

由图可见，仿真淬火反演的分辨率是不错的，如中段低阻抗薄层有清晰显示；但是，方差较大，即抖动厉害。

从目标泛函的下降曲线看，它呈现阶梯形状，与浑沌反演的阶段性类似。在迭代 230 次以后 ϕ_k 已接近全局极小，最终在 $k=265$ 时达到全局极小"山谷"，$\phi_{min}=4.92\times10^{-5}$。

以上是用双精度计算的反演结果。当用单精度计算时，迭代次数和 ϕ 曲线下降台阶减

少，迭代在 238 次停止，$\phi_{min}=4.88\times10^{-4}$。

如果在数据中加入微小高斯噪音，则迭代 226 次，$\phi_{min}=1.59\times10^{-3}$，没有明显的全局极小"山谷"出现。此时计算时仍比浑沌反演多一倍。

上述试验没有考虑用正则化压制人造假象的情况。用最小范数及最大熵准则(8.56)代替(9.35)建立目标泛函，然后用上述仿真淬火算法反演的波阻抗序列，取得反演结果方差减小但分辨率降低，未能达到明显改善解估计的目的。

总之，仿真淬火是一个很稳定的非线性反演方法，只要参数选择适宜，就能给出分辨率高的反演结果，但是与浑沌反演相比，仿真淬火的优点还不突出。

从生物工程仿真而产生的遗传算法，也是解非线性优化问题的一个姣姣者，与仿真淬火享有同等的地位。

四、遗传算法[89~91]

(一)遗传工程与最优化

生命科学向来与地球物理无缘，这种情况在进入非线性科学的新时期被打破了。在(9.13)的生态方程中，已经看到生命科学与数理科学的结缘。

非线性科学从自然界存在的非线性系统中提取各种非线性过程的概念和规律，找出其共性，并通过计算机仿真而转用其它领域。遗传算法(Genetic Algorithm)就是这样发生的。下一节讨论的人工神经网络也是这样。

生物通过遗传的演化是一种最优的自然择优演化系统。在这种系统中，优秀种属通过适应环境的生存竞争不断向高级演化，而劣等种属不断被淘汰。在遗传法则中，适应性强的染色体优生生存，使带有随机性的繁殖过程带有适者生存的总体趋向性。这种演化过程的计算机仿真，就是在随机选择中根据目标泛函取值的增大趋势进行合理取舍。

(二)非线性优化与模型编码

假定有一组未知参量

$$x_i(i=1,2,\cdots,M)$$

构成模型向量 $\boldsymbol{m}$，它的非线性目标函数为 $\Phi(\boldsymbol{m})$。根据先验知识，对每个未知量都有上下界 a_i 及 b_i，即

$$a_i\leqslant x_i\leqslant b_i$$

同时可用间隔 d_i 把它离散化，使

$$d_i=(b_i-a_i)/N \tag{9.37}$$

于是，所有允许的模型 $\boldsymbol{m}$ 将被限制在集

$$x_i=a_i+jd_i \quad (j=0,1,\cdots,N) \tag{9.38}$$

之内。

通常目标泛函(如经济学中的成本函数)表示观测函数与某种期望模型的失拟，因此非线性优化问题即为在上述限制的模型中求使 $\Phi(\boldsymbol{m})$ 极小的一个模型。对少数要求拟合最佳的问题，求目标函数的极大与失拟函数求极小是一致的。对于地球物理问题，通常要进行双重离散化。首先，地球模型一般用连续函数表示，反演时要离散化为参数集才能用于计算。有时，也将未知函数展开成已知基函数的集，用其系数作为离散化的参数集 x_i，第二次离散化

的需要是因为每一个未知参数在其变化范围内再次被离散化，以使离散模型空间最终包含着有限个非线性优化可选择的模型，其个数为

$$K = \prod_{i=1}^{M} N_i \tag{9.39}$$

其中 M 为未知参数 x_i 的个数。由此式可见，K 决定于每个参数离散化的间隔 d_i 及其变化范围 (a_i, b_i)，在大多数情况下它们只能靠先验知识来选择。

一般而言，优化问题非线性化的程度越高，逐次线性化的方法越不稳定，而对蒙特卡洛法却没有影响，因为 C 此法从有限模型空间中随机地挑选新模型并计算其目标函数 $\Phi(m)$。遗传算法与此不同的是同时计算一组模型（开始时是随机地选择的），然后把它进行二进制编码，并通过繁殖、杂交和变异产生一组新模型进一步对有限的模型空间搜索。编码的方法可有多种，下面举最简单的例说明之，对于有符号的地球参数反演时的编码方式一般要更复杂些。

假设地球为有三个水平层的层状模型，含层底界面深度 $h_j(j=1,2,3)$ 及层速度 $v_j(j=1,2,3)$ 这两组参数。如某个模型的参数值为（十进制）

$h_1=6, h_2=18, h_3=28,$　　单位为 10m

$v_1=16, v_2=24, v_3=30,$　　单位为 hm/s

按正常的二进制编制法它们可分别用以下字符串表为

和

0	0	1	1	0	1	0	0	1	0	1	1	1	0	0
1	0	0	0	0	1	1	0	0	0	1	1	1	1	0

为了减少字节，这种编码方法改变了惯用的单位制，只是按精度要求（深度为 10m，波速为 hm/s）来规定参数的码值，同时也意味着模型空间离散化间距 d_i 都规格化为一个单位（即 10m 或 hm/s）。当然，在此编码的基础上，还可以写出多种新的编码字符串。例如，三参数值的对应字节顺序重排，就可组成以下新的二进制码串

0	1	1	0	0	1	1	0	1	1	1	0	0	0	0
1	1	1	0	1	1	0	0	1	0	0	1	0	0	0

模型参数的二进制编码是一种数学上的抽象，通过编码把具体的非线性优化问题和生物演化过程联系了起来，因为这时形成的编码字符串就相当于一组遗传基因的密码。不仅是二进制编码，十进制编码也可直接用于遗传算法。根据生物系统传代过程的规律，这些基因信息将在繁殖中传到下一代，而下一代将按照“适者生存”的原则决定种属的发展或消亡，而优化准则或目标函数就起到了决定“适者生存”的作用，即保留失拟较小的新模型，而放弃失拟大的模型。在传代过程中用编码表示的基因部分地交合和变异，即字符串中的一些子串被保留，有的改变，以使传代的过程向优化的目标演化。总的来说，遗传算法可分为三步，即繁

殖、杂交及变异。

(三)遗传算法的三步骤

1. 繁殖(配对)

一开始随机地选取 Q 个字符串作为初始模型并计算目标函数 $\Phi(m_k)$，$k=1,2,\cdots,Q$。按 Q 个父母传代的规律，来产生新的一组模型(字符串)，产生和似然选择新模型组由目标函数导出的某些测度(如概率)来确定。记选择第 k 个字符串的概率为 $P_r(m_k)$，它的计算方法常用的有以下两种

(1)P_r 为 $\Phi(m_k)$的线性函数

$$P_r(m_k)=a-b\Phi(m_k) \tag{9.40}$$

其中

$$b=Q^{-1}(\Phi_{max}-\Phi_{avg})^{-1} \quad a\geqslant b\Phi_{max} \tag{9.41}$$

Φ_{max}和 Φ_{avg}分别为 Q 个模型计算的目标函数的极大值和平均值。

(2)P_r 为 $\Phi(m_k)$的指数函数

$$P_r(m_k)=A\exp[-B\Phi(m_k)] \tag{9.42}$$

其中

$$B=(\Phi_\sigma)^{-1} \quad A=\left[\sum_j \exp(-B\Phi_j)\right]^{-1} \tag{9.43}$$

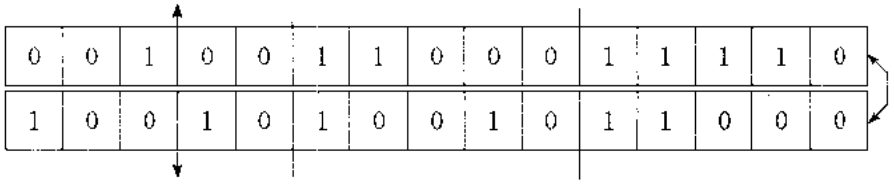
Φ_σ 为 Φ 值分布的标准差。

将 Q 个模型编好码并计算出目标函数 $\Phi(m_k)$来之后，就可以根据(9.40)或(9.42)式计算选择它们的概率，以决定那些父母能继续传种。对非线性地球物理反问题，用指数函数或许好一些，但计算量大。接下去就用杂交来产生新模型集。

2. 杂交

将 Q 个字符串随机地配对成 $Q/2$ 对"父母"。然后发生一个 0 到 1 之间的随机数，用它来决定这对父母是否可以杂交繁殖。设 P_c 为杂交概率，如果随机数大于 P_c，则这对"父母"不能用来杂交，而直接进入下一轮传代。如果随机数小于 P_c，则随机在二进制字符串中选中一个位置把字符串截成两段，然后交换其中的一段而生成两个新的杂交字符串。新的字符串的十进制内容可能变化极大，要检查一下是否超出了先验决定的范围。

作为上面举过的字符串的例子，如果必须进行杂交而且位置随机地选为第四位与第三位之间，则杂交结果的十进制变为(4,24,30)与(18,18,28)，即



注意，为节省篇幅，这只是说明字符串杂交的例子，没有涉及地球模型中波速与深度的含义。

3. 变异

变异是生物演化中产生新种属的重要步骤，即遗传基因不完全来源于父母，而随偶然因素或环境发生少量的改变。对本文讨论的问题而言，二进制字符串的随机地选择的一个或几个位的极性发生变化称为变异。设变异概率为 P_m，例如选 $P_m\leqslant 1/l$，l 为字符串长度或二进制总位数，它控制极性改变的似然度。经过变异之后，由 Q 个二进制字符串组成的下一代

"崽"就产生了，它又作为新的输入再产生后代。

在遗传算法中三个步骤各具其特殊的功能。繁殖步决定哪些上代能生存和传种，其依据是目标函数的拟合度。杂交步决定了模型中包含的信息(基因)的分享和混合，而变异步使传代的轨迹散开，便于模型空间搜索的展开。实际上，$P_m=0$(即无变异)的遗传是没有活力的，对优化问题相当于陷入局部极值而跳不出来，P_m 很大的遗传就更类似于蒙特卡洛法完全随机的搜索。

总结以上所述，在用遗传算法解优化问题之前，要先考虑好以下因素

编码的格式与方法；

初始的染色体集团；

繁殖概率的特性(线性、指数或其它)；

参加繁殖的规模，即 Q 组字符串的 Q 大小；

杂交概率 P_c；

变异概率 P_m。

这些因素要根据实际问题具体分析选择。当然，这些因素及参数在传代过程中也可以变化，不一定保持不变，不能修改。

(四)遗传算法的示例

为了使读者对遗传算法有一个具体的印象，这里举一例说明。设 $0\leqslant x\leqslant 127$，求方程 $\Phi=x^2$ 的极大值。由于 $x<128$，因此可用 7 位二进字符串($l=7$)来编码，并给定参数 $Q=4$，$P_c=1.0$，$P_m=0.0$(只要杂交，不要变异)。繁殖概率可用(9.40)及(9.41)的线性型，如

$$P_r(x_k)=\frac{\Phi(x_k)}{Q\Phi_{avg}} \tag{9.44}$$

其中平均拟合度 Φ_{avg} 也就是拷贝的期望值，意味着小于它的模型不传代，而大于它的模型向下传代。上述 P_r 的计算方法与指数型计算结果也相差不大。随机选择的 4 个初始模型及第一代的遗传见下表。

i	x	二进串	$\Phi(x)$	$P_r(x)$	父母	杂交点	产仔	x_i
1	15	0001111	225	0.012	1	000—1111	0000101	5
2	70	1000110	4900	0.265	3	110—0101	1101111	111
3	101	1100101	10201	0.553	2	1—000110	1111000	120
4	56	0111000	3136	0.170	4	0—111000	0000110	6
Φ_{avg}			4616					6696

注意在初始模型中有的繁殖概率很小(0.012)，但对第一代没有取舍，只是在配对时有意识地将它与 P_r 最大的模型($i=3$)配成一对再杂交，同时把繁殖概率为中等的两个模型配成一对。对于第二代，有两个 P_r 很小的模型($i=1$ 和 4)便剔除了，因此便提高了拟合度，同时平均拟合度也得以提高。

i	x	二进串	$\Phi(x)$	$P_r(x)$	父母	杂交点	产仔	x_i
1	5	0000101	25	0.0009	2	110111—1	1101110	110
2	111	1101111	12321	0.46	3	111100—0	1111001	121
3	120	1111000	14400	0.5378	2	1101—111	1101000	68
4	6	0000110	36	0.0013	3	1111—000	1111111	127
Φ_{avg}			6696					11874

第二代产生的 $x_i=127$ 便得到了最优化问题的解，比蒙特卡洛方法快一个级次，其中繁殖与杂交的作用是非常明显的。此外，由于本问题($\Phi=x^2$)不存在局部极小，因此变异可以省略，这就是选 $P_m=0.0$ 的原因。

目前虽然有理论研究给出的遗传算法基础定理，但对具体的优化问题，还不知道如何肯定传代到什么时候才能取得问题的解。因此，上述标准的遗传算法总带有半定量和直觉性质，要用搜索轨道等图示来帮助算法的执行和控制。当然，发展一些技巧或者将它与其它算法配合使用都能进一步克服它存在的缺点。

(五)遗传算法在地震反演中的应用

以地震走时反演为例，根据最小二乘准则使合成数据与实测数据的拟合差取极小，目标函数可取为

$$\Phi_t=\frac{1}{N_A}\sum_{i=1}^{N_A}\frac{(T_{i,0}-T_{i,s})^2}{(\Delta T)^2} \tag{9.45}$$

式中 $T_{i,0}$为观测资料中提取出的地震走时；$T_{i,s}$为合成地震或射线追踪算出的地震走时；ΔT为所有合成地震走时的平均值；N_A为合成地震数据的个数，它可以少于实测 $T_{i,0}$的个数，因为在射线追踪时有阴影区存在，不一定能算出合成数据 $T_{j,0}$。利用射线追踪计算走时的方法很多，可参阅文献[90]和[91]。对于少数几个波速为常数的水平层，走时反演的参数编码方法可参照第二节中介绍的分别对深度和波速编码方法，二进制码的字符串位数 l 不会太大。要注意的是由深度定出的字符串应符合数值由浅到深增大的规律，这一约束条件不应在杂交和传代过程中破坏。这种不等式的约束($h_1<h_2<h_3\cdots$)在遗传算法中是容易实现的。

对于波场反演，较方便的做法是将地球介质作等间距的划分。例如，将水平层状介质细分为 100 个等厚度的水平层。在上地壳可假定波速小于 6400m/s(相当于解空间的硬约束)，而波速间距为 100m/s，则可将波速用 100m/s 为单位，每层用 6 位二进制子字符串表示波速，地层模型总共用 600 位二进制字符串表示($l=600$)。初始模型可随机地选取 24～192 个，然后通过繁殖杂交与变异。杂交概率在 0.5～1.0 之间，变异概率可小于 0.01。目标函数(即失拟方程)在频率域可表示为

$$\Phi_w=\frac{\sum_j\sum_k[A(\omega_k)P_0(\omega_k,v_j)-CP_s(\omega_k,v_j)]^2}{\sum_j\sum_k[A(\omega_k)P_0(\omega_k,v_j)]^2} \tag{9.46}$$

式中 $P_0(\omega_k,v_j)$为实测地震道的频谱；ω_k 为角频率；v_j 为第 j 层的波速；$P_s(\omega_k,v_j)$为相应的合成地震道；$A(\omega_k)$为地震仪及检波器的频率滤波器，例如，可取

$$A(\omega)=\mathrm{sinc}^4(\omega/\omega_N) \tag{9.47}$$

式中 ω_N 为 Nyquist 频率，即 $\omega_N=\pi/\Delta t$，Δt 为时间取样率。参数 C 为振幅拟合因子，它起到合成与观测记录之间幅度上匹配的作用。C 的计算常用地震道的包络函数的平均比值。例如，设 $E[]$为波动信号的包络函数，可令

$$C=\frac{1}{J}\sum_{j=1}^{J}\frac{E[P_0(t_{max},v_j)]}{E[P_s(t_{max},v_j)]} \tag{9.48}$$

式中 t_{max}为包络极大值的对应时间，J 为总层数。包括函数可通过复数道的模取得。

用遗传算法作波速反演时失拟最小的模型将一直保存到传代停止。什么时候停止传代

还没有理论上可计算的好办法，一般要显示解空间的搜索范围及局部密度，以此来判断是否可以停止传代。值得指出的是，由(9.45)和(9.46)式给出的目标函数对于有误差的数据是有问题的，反演的目标不是追求对有误差数据的完美拟合，而是要求出准确而且分辨率最高的解估计。

遗传算法在执行中可能出现两类问题。其一称为"早熟"问题，即在传代之初就随机地选中了比较好的模型，它在传代中起主导作用，而使其后的计算因散不开而白白浪费。通常，增大 Q 值可以改善这种情况。另一类问题正相反，即传相当多代之后仍然找不到一个特别好的解估计，即可能有几百个算出的目标函数值都大同小异。这时，最好修改目标函数的比例因子(即(9.46)式的分母)，以使繁殖概率 P_r 的变化范围加大。

对于高维地震模型的反演，由于参数太多，相应的模型字符串太长，目前用遗传算法作反演的计算成本还嫌太高。实际上，为了加快计算，不仅要改进反演技巧和传代的控制技术，而且还要大幅度提高正演计算的速度，在为对遗传算法大量的计算花费在正演合成上。

遗传算法的最大优点将在并行计算上体现出来。

五、人工神经网络的应用[92~94]

(一)人工神经网络与人工智能

前面所有关于地球物理反演的讨论，都只是把计算机当成人类的计算工具，是地球物理学家用计算机来进行计算。

但是，人类也可以使计算机具有"智能"，教计算机进行反演或数据处理。这时，计算机必须有能进行思维的人工神经元和神经网络。

人工神经网络是生物神经网络的计算机仿真。这时，人工神经网络被安排成生物神经网络相似的结构模式，以加权等形式将神经元连接起来。这些权系数在"学习"过程中被自适应地估定和修正，然后用于模式识别等智能活动。

人工神经网络种类很多，发展很快。本节只能通过反传网络和反馈网络两个典型介绍人工神经网络的基本原理。

生物神经网络可分为感觉神经层、分析层和动作层三个层次。感觉神经元组成了第一个层次，它接受色、香、味、触等外来信号并将它输入到大脑，人工智能中感觉神经层相当于计算机的输入层。隐蔽在大脑中的神经细胞根据以往学习到的经验对输入层传来的信息进行分析判断，如看到和闻到的食物是否能吃，要不要马上吃。然后把分析判断的结果以信息的形式传给动作神经元，以便分泌唾液，启动口腔和蠕动肠胃。在人工智能研究中，动作神经层称为输出层，而进行分析判断的脑神经层称为隐层。人工智能的进展已建立了神经网络的多种数学模型和算法，并通过计算机实现了部分人类的智能活动(如下棋，识字等)。

(二)神经网络的设计

神经元是人类智能活动的基本单位，也是神经网络中的一个节点。神经元的数学模型是它的传递函数 f，它把元的输入信号转变为输出信号。最常用的神经元传递函数有以下三种。

1. 阈值型

$$f(x)=\begin{cases}1, & x>0\\ 0, & x\leqslant 0\end{cases} \tag{9.49}$$

2. 分段线性型

$$f(x)=\begin{cases}1, & x>a\\ x/a, & 0<x\leqslant a\\ 0, & x\leqslant 0\end{cases} \tag{9.50}$$

3. S 型

$$f(x)=[1+e^{-x}]^{-1} \tag{9.51}$$

或用双曲正切函数

$$f(x)=\frac{1}{2}\left\{1+\text{th}\left(\frac{x}{b}\right)\right\} \tag{9.52}$$

式中，b 为常数。

根据实际问题的需要，把许多神经元连成网络，并训练它去解决问题，称为神经网络的设计。设计的步骤一般可分为以下五步。

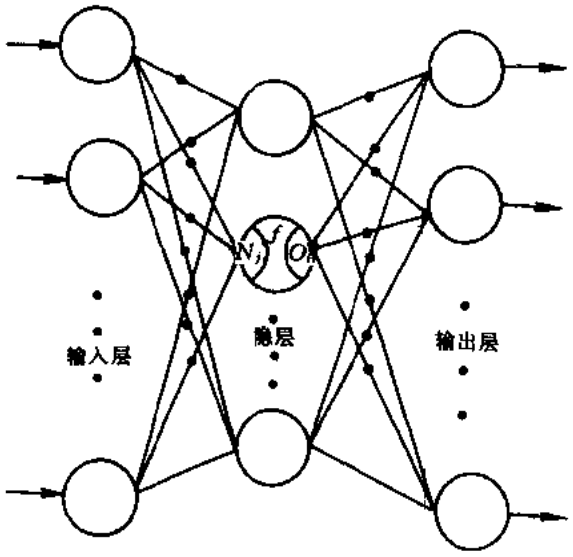


图 9.11 反传神经网络

圈圈表示神经元，其传递函数如(9.51)式所示；

黑点表示权 W_{ij}。反传指训练时由输出误差反求 W_{ij}

1. 结构设计。确定输入层和输出层中神经元(节点)的数目和类型，隐层的层数、类型和内部连接方式，它与输入输出层的连接方式。目前，神经网络的结构设计根据要解决的问题不同而异，主要凭设计者的经验。如果没有经验，可套用神经网络技术中比较成熟的反传神经网络，它分为输入层、输出层和隐层三部分，如图 9.11 所示。

2. 训练。在反传神经网络中，每一个节点的传递函数和与其它节点的连接方式都是设计者规定的，但是连接任两外点之间的权系数 W_{ij} 是未知的。所谓训练就是根据已知的输入和期望输出权系数。求权系数的方法即反传神经网络算法，将在下节详细讨论。

3. 用非理想的已知数据和期望输出检查网络的可信度。将含误差的数据送入神经网络，检查网络的实际输出与期望输出之间的偏离，以估计可信度。

4. 在专家监视下用大量实际数据作输入，估算神经网络的成功率。换句话说，由于工程问题的复杂性，任何设计好的神经网络的应用成功率都难以达到百分之百。因此，在试用期间要有专家监视。

5. 推广应用。

(三)反传神经网络算法

现在来讨论反传神经网络的训练，即根据已知的期望输出 $\{y_k\}$ 和实际输出 $\{\hat{y}_k\}$ 求权系

数$\{W_{ij}\}$，其中k表示输出层的顺序号；j为当前层号；i为前一层的节点号。神经元的传递函数选用S型，用(9.51)式表示。对网络中的任一个节点j，其输入来自前层的所有节点i，总输入为

$$N_j=\sum_i W_{ij}O_i \tag{9.53}$$

其中，O_i为前层第i个节点的输出。通过节点j之后的输出为

$$O_j=f(N_j)=[1+\exp(-N_i)]^{-1} \tag{9.54}$$

此式是代入(9.51)式的结果。信号由输入层通过隐层到输出层，网络的实际输出

$$\hat{y}_k=O_k,\qquad k=1,2,\cdots,K \tag{9.55}$$

其中，K为输出层的节点数。

根据最小平方原理，定义期望输出与实际输出之间的误差函数为

$$E=\frac{1}{2}\sum_k |y_k-\hat{y}_k|^2 \tag{9.56}$$

求系数组$\{W_{ij}\}$可通过反传和逐渐迭代来解决。首先进行初始化，即随机地赋予网络一组权系数$W_{ij}(t=0)$，这里t为迭代次数。由正传($j=1,2,\cdots,k$)求每一个节点的输入(N_j)和输出(O_j)，然后反传($j=k,k-1,\cdots,1$)修改得权系数$W_{ij}(t=1)$。每一个权系数的修改增量要通过误差函数E对权W_{ij}的偏导数计算

$$\frac{\partial E}{\partial W_{ij}}=\frac{\partial E}{\partial N_j}\frac{\partial N_j}{\partial W_{ij}}=\delta_j O_i \tag{9.57}$$

式中

$$\delta_k=\frac{\partial E}{\partial N_j} \tag{9.58}$$

为误差函数对节点j输入的偏导数；O_i为节点j前层的输出，它是利用(9.53)式两边求导取得的。

首先从输出层k开始计算δ_k，因为这里期望输出与实际输出是已知的，由(9.58)式有

$$\delta_k=\frac{\partial E}{\partial \hat{y}_k}\frac{\partial \hat{y}_k}{\partial N_k}=-(y_k-\hat{y}_k)f'(N_k) \tag{9.59}$$

其中，f由(9.51)式定义，f'可由(9.51)式求出。(9.59)式代入(9.57)式得输出层与隐层之间的权系数修改增量

$$\Delta W_{ik}(t)=-\frac{\partial E(t)}{\partial W_{ik}}=(y_k-\hat{y}_k)f'(N_k)O_i \tag{9.60}$$

式中，O_i为隐层各节点的输出。然后即可对这组权系数进行修改(对第t次迭代)

$$W_{ik}(t+1)=W_{ij}(t)+\mu\Delta W_{ik}(t) \tag{9.61}$$

式中，μ为松弛因子(也称为学习率)，$\mu>0$。

对隐层内部，我们知道的只是其后层已修改的系数组$W_{jm}(t+1)$。这里m表示当前层j的后层，用它来修改前层的系数组W_{ij}称为反传。这时由(9.58)式有

$$\delta_j=\frac{\partial E}{\partial N_j}=\frac{\partial E}{\partial O_j}\frac{\partial O_j}{\partial N_j}=\frac{\partial E}{\partial O_j}f'(N_j) \tag{9.62}$$

由于后层有许多节点(记为$m=1,2,\cdots$)与j节点连接，因此

$$\frac{\partial E}{\partial O_j}=\sum_m \frac{\partial E}{\partial N_m}\frac{\partial N_m}{\partial O_j}=\sum_m \delta_m W_{jm} \tag{9.63}$$

此式推导应用了(9.58)式和(9.53)式。由于反传计算时后层 m 的偏导数 δ_m 及系数 W_{jm} 已经修改，因此(9.63)式不难计算出来。代(9.63)到(9.62)和(9.57)式中便得隐层 j 节点的第 t 次迭代修改增量

$$\Delta W_{ij}(t+1)=-\frac{\partial E(t)}{\partial W_{ij}}=-\sum_m \delta_m W_{jm} f'(N_j) O_i \tag{9.64}$$

再用(9.61)式进行迭代修改(框图见图 9.12)，迭代一般是收敛的。

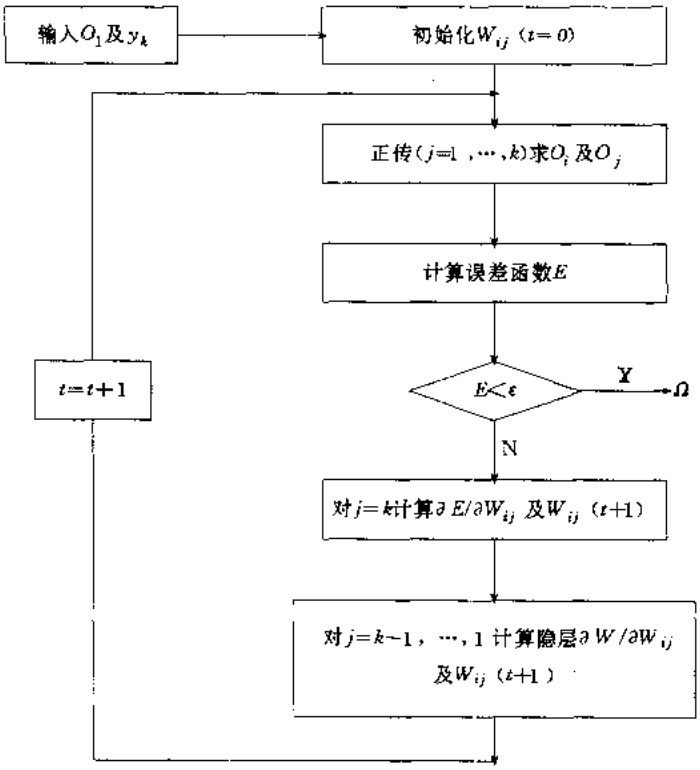


图 9.12 反传神经网络算法程序框图

由上述可知，BP 神经网络算法可以应用于非常微妙的非线性关系的分析，把已知的经验推广到未知数据的场合。它在数据检查，地震废道剔除，测井层位对比等问题上的应用取得了一定的成功，对非线性地球物理反演的应用也有一定潜力。

在简单情况下，可以用神经网络直接进行非线性反演。例如，单界面二维重力反问题

$$\Delta g(x)=\int_a^b \Delta\delta \lg \frac{(x-y)^2+H^2}{(x-y)^2+[H-z(y)]^2}\mathrm{d}y \tag{9.65}$$

由 Δg 求 $z(x)$，就是一个典型的非线性反问题。式中，$\Delta\delta$ 为上下界面的平均密度差；H 为界面平均埋深；z 为界面起伏。给出多种 $z(x)$ 的模型并由(9.65)式计算 $\Delta g(x)$，并用 $\Delta g(x)$ 作神经网络的输入，$z(x)$ 作期望输出，则可以确定反传神经网络的系数组 $\{W_{ij}\}$。然后，用这种神经网络作为实际资料反演的工具，推广到复杂界面形状反演的情况。

(四)反馈神经网络

将输出信息经过一步时移(Z 变换，记为 Z^{-1})并反接到输入层，称为反馈。在图 9.13 中，给出了人工反馈网络流程图。

Hopfield 网络是反馈网络中较成功的一种，它具有极小化的功能，在反演中有应用前

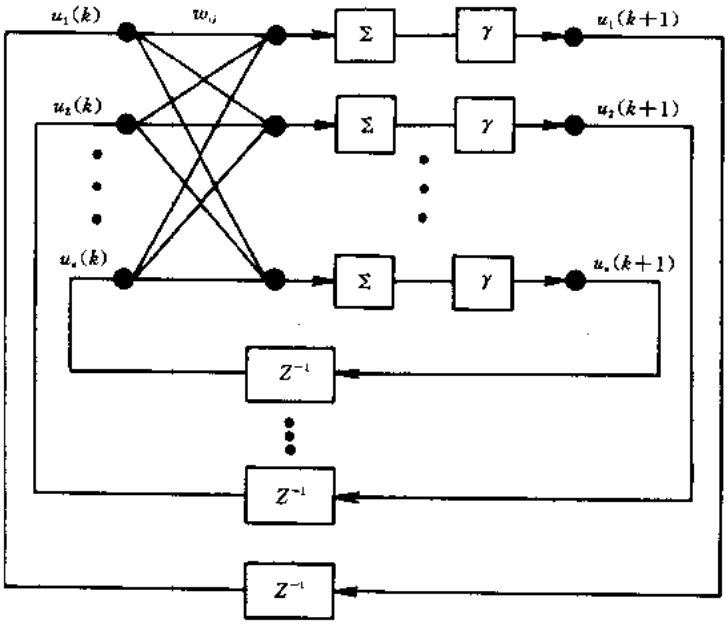


图 9.13　人工反馈网络流程图

γ 为(9.51)式所示的 S 型神经元，Z^{-1}表示时移

景。

这类网络可用非线性微分方程描述

$$\frac{\mathrm{d}u}{\mathrm{d}t}=-\frac{u}{R}+Wv+\theta$$

式中 R 为常数，$v=[r_1(u_1),r_2(u_2),\cdots,r_n(u_n)]^{\mathrm{T}}$ 组成的向量，r_i 为(9.51)式定义的神经元函数。θ 为常数向量。

对有限节点的网络，上式写成

$$c_i\frac{\mathrm{d}u_i}{\mathrm{d}t}=\sum_{j=1}^{n}W_{ij}v_j-\frac{u_i}{R_i}+\theta_i \tag{9.66}$$

其中 W_{ij}为$(n\times n)$对称矩阵，为前层第 i 个节点与后层第 j 个节点连接的权。u_i 为网络的输入变量，v_i 为网络的状态变量。c_i 及 R_i 都是正常数，θ_i 为偏移量。

写成向量形式，(9.66)式变为

$$\boldsymbol{C}\dot{\boldsymbol{u}}=\boldsymbol{W}\boldsymbol{p}(\boldsymbol{u})-\boldsymbol{A}\boldsymbol{u}+\boldsymbol{\theta} \tag{9.67}$$

式中

$$\boldsymbol{C}=\mathrm{diag}\{c_1,\cdots,c_n\} \tag{9.68}$$

$$\boldsymbol{A}=\mathrm{diag}\{R_1^{-1},\cdots,R_n^{-1}\} \tag{9.69}$$

$$\boldsymbol{p}=[r_1(u_1),\cdots,r_n(u_n)]^{\mathrm{T}} \tag{9.70}$$

它等价于关于 v 的微分方程

$$\dot{v}=\boldsymbol{D}(v)[\boldsymbol{W}\boldsymbol{v}-\boldsymbol{p}(v)+\boldsymbol{\theta}] \tag{9.71}$$

其中

$$\boldsymbol{D}(v)=\mathrm{diag}\{[c_1u'_1(v_1)]^{-1},\cdots,[c_nu'_n(v_n)]^{-1}\} \tag{9.72}$$

为正定，而

$$p(v)=[R_1^{-1}u_1(v_1),\cdots,R_n^{-1}u_n(v_n)]^{\mathrm{T}} \tag{9.73}$$

注意对S形神经元，v_i 及 r_i 的域值为$(-1,1)$。对于任一初始条件

$$v(0)=v_0\in(-1,1)^n\subset\mathbf{R}^n \tag{9.74}$$

方程(9.71)的解存在而且唯一。

负反馈网络的特点在于，输出信号的反馈信息对系统起到调节抑制作用，使系统逐渐趋于平衡状态，这是一些能量处于极小的状态。利用反馈网络使系统处于稳定的平衡状态的过程，也就是目标泛函趋向极小的过程。

设常微分方程为

$$\partial x/\partial t=f(x) \tag{9.75}$$

其中 $x(t)$ 为 n 维向量，则在 f 连续可微时对任意初始条件

$$x(0)=\bar{x}_0 \tag{9.76}$$

存在唯一解。如果 $\bar{x}_0$ 满足 $f(\bar{x}_0)=0$，则称 $\bar{x}_0$ 为该方程的平衡点(与初始条件无关的点)。如果它在其邻域内是唯一的，则又称为孤立平衡点。

根据平衡点的特性及矩阵 $D(v)$ 的正定性，易见下列三个命题成立

(1)由(9.71)知向量 $V_e\in(-1,1)^n$ 是此网络平衡点的充要条件是

$$Wv_e+\theta=p(v_e) \tag{9.77}$$

(2)此反馈网络的平衡点不超过 2^n 个，且当 $t\to\infty$ 时网络的每个非平衡解趋向于某个平衡点。

(3)如果 v_e 是网络的平衡点，则它也是能量函数

$$E(v)=-\frac{1}{2}v^{\mathrm{T}}Wv+\sum_{i=1}^{n}\frac{1}{R_i}\int_0^{v_i}r_i^{-1}(x)\mathrm{d}x\quad v^{\mathrm{T}}\theta \tag{9.78}$$

的极小点。

根据上述命题可知，随着时间推移，此类反馈网络系统会自动搜索其能量函数的极小点。一旦找到某个极小点，该系统便处于稳定的平衡状态，搜索也就停止，这时的状态变量 v_e 便是反演要求的模型参数。

方程(9.77)给出了设计Hopfield反馈网络的基本方程式。具体地说，设计时给出一组向量

$$\{v_{e1},\cdots,v_{em}\}\subset(-1,1)^n \tag{9.79}$$

为网络的渐近平衡点，求网络的权 W 和向量 θ。由(9.77)有

$$Wv_{ei}+\theta=p(v_{ei}),i=1,\cdots,m \tag{9.80}$$

这是关于求 W 和 θ 的方程组，含有 $m\times n$ 个方程。由于 W 是对称的，有 $n(n+3)/2$ 个未知数。

由于

$$\theta=p(v_{em})-Wv_{em} \tag{9.81}$$

由(9.80)消去 θ 有

$$W(v_{ei}-v_{em})=p(v_{ei})-p(v_{em}) \tag{9.82}$$

由此线性方程组先解出矩阵 W 的元，然后再由式(9.81)求出 θ。至于网络的其它常数，如 c_i 及 R_i 都是预先根据需要选定的。

对于地球物理反演问题，将地球模型参数记为向量 v，数据记为 d(含误差)，对于线性反问题

$$Gv=d \tag{9.83}$$

由(3.18)或(3.24)式知，应用最小二乘法有目标泛函

$$E(v)=v^{\mathrm{T}}G^{\mathrm{T}}Gv \quad 2v^{\mathrm{T}}G^{\mathrm{T}}d + d^{\mathrm{T}}d=\min \tag{9.84}$$

由于 $d^{\mathrm{T}}d$ 为常数，(9.84)等效于

$$E(v)=\frac{1}{2}v^{\mathrm{T}}Wv-\theta^{\mathrm{T}}v \tag{9.85}$$

比较(9.78)与(9.85)式可知，只要略加修改可把线性反问题转化为 Hopfield 网络搜索极小点的问题，当然，这样做是意义不大的，因为其它线性反演的许多方法都可以取得更好的结果。对非线性反演，要把最优化反演的目标泛函写成(9.78)的形式，然后使网络趋于平衡点。这种非线性优化的方法也许不如仿真淬火那样趋向全局极小，但给我们提供了非线性反演系统自组织的思路。

到目前为止，人工神经网络系统在地球物理上的应用还限于模式识别上，如对大地电磁曲线类型的自动识别、地震道的分类特征识别等，直接应用于反演尚未取得令人信服的结果。

神经网络在自适应反演上也有应用前景。目前反演方法众多，不同的反演方法基于不同的假设条件，有不同的特点和应用范围。将数据先输入神经网络进行分析，然后由神经网络决定采用哪种或哪几种反演方法进行反演，充分地利用了神经网络的优点，使反演方法与数据自相适应，近期不难在计算机上实现。

神经网络应用的另一个方向在反演的自组织上。对非线性反演，尤其是各种地球物理资料的联合反演，将不同的反演方法和迭代过程组织成一个系统，用神经网络去指挥这个系统，使之输出分辨率最优而方差最小的地球模型，乃是下一阶段非线性反演要研究的一个重要方向。

结 束 语

一、这本书是用方程堆起来的，除了方程及其解释，所剩无几。

所有方程又归结为基本方程(6.1)，反演指的是求它的系数项或小项(极个别情况下为初始条件)。方程解得好，也不一定能解决好实际问题。

实际问题可能是石油赋存在哪里？矿的储量多大？地基下的裂隙影响大坝稳定性吗？地壳哪部分在局部融熔？

反演是解决实际问题的一种很有用的方法，研究反演的理论与方法，可使人在解决实际问题时做得更好更准确。

真正的反演专家要做到理论与应用的珠联璧合。

二、地球物理反演是在一个三角架上的人造水晶，这个三角架的三条腿分别是地球科学、数理科学和信息科学。当前其特点可用"显著，广泛，有限"六个字来概括。

30 年来从无到有，从零散到系统，建立起其科学系统和方法体系，成果可谓显著。

从全球壳幔结构到北京机场立交桥墩质量检测，从塔里木油气资源评价到三峡楚墓的发掘，应用可谓广泛。

但是，60 年代后期开创的地球物理反演理论，还没有跳出 50 年代吉洪诺夫提出的正则化思想，其进展也实在有限。

三、有 40 多亿年历史的地球，给人类提出数不清的谜。21 世纪的地球物理，可能面临一些极其古怪的难题。例如，八千万年以前，中国地幔是否强烈上隆？

21 世纪的计算机，其计算能力又将提高几个级次。21 世纪地球物理数据的积累，也将提高几个级。应该发展哪些新反演理论与方法，解决地球科学面临的古怪难题？

写完这本书，我仍然不能回答这个问题。但是，肯定要跳出正则化的折衷主义思维，要发展新的非线性反演理论与方法，尤其是自组织的非线性反演理论与方法。最重要的是，要和三条腿一起生长。

收笔有诗一首，为付梓作兴：

人生过秋方知秋，
流连书海又白头。
呈书反演圆旧梦，
阵列方程填壑丘。
学识无限虚怀谷，
未知天地泛扁舟。
莫谈文章千古恨，
只望后生造琼楼。

参考文献(下篇)

〔1〕杨文采,地球物理勘查通用的场方程,地球与空间科学观测技术进展(陈运泰主编),p.447～454,地震出版社,1995。

〔2〕杨文采,岩石的粘弹性谐振Q模型,地球物理学报,30,399～411,1987。

〔3〕杨文采,地球物理反演与地震层析成像,地质出版社,1987。

〔4〕Christenson, R. M., Theory of Viscoelasticity. Academic Press, 1982.

〔5〕Ben-Menafem, A, and Sighn, S. J., Seismic waves and sources, Springer-Verlag. 1989.

〔6〕Nabighian, M. N. (ed), Electromagnetic Methods in Applied Geophysics. SEG Publication, 1985.

〔7〕Berkout, A. J., Seismic migration, Imaging of acoustic energy by wavefield extrapolation. Elsevier, 1985.

〔8〕杨文采,李幼铭,应用地震层析成像,地质出版社,1992。

〔9〕吴如山,Araujo, F. V., Multi-frequency backscattering tomography: the Principle and the reconstruction methods; 6hh SEG International Meeting, Expanded Abstract, 967～970, 1991。

〔10〕马在田,地震成像技术——有限差分法偏移,石油工业出版社,1989。

〔11〕Riley, K. F., Mathematical Methods for the Physical Sciences, Cambridge Univ. Press, 1974.

〔12〕Silvester, P. P., Finite Elemen: for Electrical Engineers, Cambridge Univ. Press, 1983.

〔13〕徐世浙,地球物理中的有限单元法,科学出版社,1994。

〔14〕Sheriff, R. E., Encyclopedic dictionary of exploration geophysics, SEG Publication, 1984.

〔15〕Sheriff, R. E. and Geldart, L. P., Exploration Seismology, Cambridge Univ. Press, 1983.

〔16〕Johnston, D. H. and Toksoz, M. N., Seismic Wave Attenuation, SEG reprint series No. 2, 1981.

〔17〕Futterman, W. I., Dispersive body waves, J. Geophys. Res., 67, 5279～5291, 1962.

〔18〕Biot, M. A., Theory of propagation of elastic waves in a fluidsaturated solid, J. Acoustic Soc. Am., 28: 168～191, 1956.

〔19〕Biot, M. A., Mechanics of propagation of elastic waves in a fluid-saturated solid, J. Appl. Physics, 33: 1482～1498, 1962.

〔20〕Rabbel, W, Seismic anisotropy at the Continental Deep drilling site (KTB), Tectonophysics, 232, 329～341, 1994.

〔21〕Zaback, M, D. and Emmermann, R., Scientific Rationale of an ICDP, ICDP Report, Potsdam, Germany, 1993.

〔22〕Morig, R., Untersuchungen zum zusammenhang von seismischen und lithologischen parametern sedimentarer festgesteine, Ph. D. Thesis, TU Berlin, 1994.

〔23〕Murphy, W. F., Winkler, K. W. and Kleinberg, R. L., Acoustic relaxation in sedimentary rocks: Dependence on grain contacts and fluid saturation, Geophysics, 51, 757～766, 1986.

〔24〕Strack, K. M., Exploration with Deep Transient Electromagnetics, Elsevier Science Publ. Co., Int. 1992.

〔25〕Telford, W. M., Geldart, L. P., Sheriff, R. E. and Keys, D. A., Applied Geophysics (2nd Ed.), Cambridge Univ. Press, 1990.

〔26〕White, J. E. and Sengbush, R. L., Production Seismology, SEG Publication 1987.

〔27〕Ward S. H. and Hohmann, G. W., Electromagnetic Theory for geophysical applications, in Nabighian, M. N. (ED.) "Electromagnetic Methods in Applied Geophysics, I: theory", 131～311, SEG publication, 1985.

〔28〕黄联捷,Microscopic approaches to wave propagation in complex media. Ph.D these, University Paris Ⅵ, 1994。

〔29〕李幼铭、胡建行,细胞自动机在地震波传播研究中的应用,地球物理学报,38,651～661,1995。

〔30〕Savitzky, A. and Golay, M. J. E., Analytical Chemistry, 36, 1627～1639, 1964.

〔31〕Krim, H. and Viberg, M, Two decades of array signal Processing research. IEEE Signal Processing, 13: 4, 67～94, 1996.

〔32〕Chaveut, G. and Jacewitz, C. A., Determination of background velocities by multiple migration fitting. Geophysics, 60: 2, 476～490, 1995.

〔33〕杨文采、郭爱缨等,重磁异常频率域解释的理论与方法,物化探研究报道,2,134～178,地球物理探矿研究所,

1978。

〔34〕Yilmaz, D., Seismic Data Processing, Soc. Expl. Geophys. 1987.

〔35〕王继伦，二维最佳线性滤波器设计原理，地球物理学报，28，157～168，1985。

〔36〕杨文采，非线 K-L 滤波及其在反射地震资料处理中的应用，石油物探，35:2，17～26，1996。

〔37〕Oja, E., Subspace Method of Pattern Recongnition, Research Studies Press, 1983.

〔38〕栾文贵，地球物理中的反问题，科学出版社，1989。

〔39〕Tikhonov, A. N., Continuation of potential fields in a layered medium by the regularization method. Earth Physics, No. 12, 1968.

〔40〕Garnir, H. C. (ed.), Singularities in Boundary Value Problems, D. Reidel Pub. Co., 1980.

〔41〕Marozov, V. A., Methods for Solving Incorrectly Posed Problems, Springer-Verlag, 1984.

〔42〕Beylkin, G., Coifman, R. and Rokhlin, V., Fast wavelet transforms and numerical algorithms I, Comm. Pure Applied Math. XLIV, 2, 141～184, 1991.

〔43〕Daubechies, I., Orthogonal bases of compactly supported wavelets, Comm. Pure Applied Math., 909～996, 1988.

〔44〕Daubechies, I., Ten Lectures on Wavelets, Capital city Press, Montpelier, Verment, 1992.

〔45〕Mallat, S. G., A theory for multiresolution signal decomposition: the wavelet representation, IEEE Tran. Pattern Analysis and machine intelligence, 11, 7, 674～693, 1989.

〔46〕Mallet, S. G., Hwang, W. L., Singularity detection and processing with wavelets, IEEE trans. Infor. Theory: 38, 617～643, 1992.

〔47〕侯遵泽、杨文采，中国重力异常的小波变换与多尺度分析，地球物理学报，40，122～132，1997。

〔48〕Sabatier, P. C. (ed), Basic Methods of Tomography and Inverse Problems, Adam Hilger, 1985.

〔49〕Tikhonov, A. N. and Arsenin, V. Y., Solutions of Ill-Posed Problems. Wiley, 1977.

〔50〕Teukolsky, *et al.*, Numerical Recipes in FORTRAN: the Art of Scientific Computing. William H. Press, 1992.

〔51〕Skilling, J. (ed), Maximum Entropy and Bayesian Methods, Kluwer, 1989.

〔52〕Smith, C. R. and Crandy, W. T. (ed), Maximum-Entropy and Bayesian Methods in Inverse Problems, Reidel, 1985.

〔53〕Bott, M. H., Inverse metheds in the interpretation of magnetic and gravily anomalies, in "Methods in computional Physics", Vol. 13, 133～162, 1973.

〔54〕王家林、王一新、万明浩，石油重磁解释，石油工业出版社，1991.

〔55〕管志宁、张昌达、程方道、申宁华，磁法勘探重要问题理论分析与应用，地质出版社，1993。

〔56〕管志宁、安玉林，区域磁异常定量解释，地质出版社，1991。

〔57〕Budak, B. M., Samarskii, A. A. and Tikhonov, A. N., A collection of Problems on Mathematical Physics, Pergamon Press, 1964.

〔58〕Tarantola, A., Inverse Problem Theory, Elsevier, 1987.

〔59〕傅良魁，应用地球物理学电学原理，地质出版社，1989。

〔60〕Kautman, A. A. and Killer, G. V., Frequency and Transient Soundings, Elsevier, 1983.

〔61〕黄联捷、杨文采，参考波速线性变化时声波方程逆散射反演，地球物理学报，34:1，80～98，1991。

〔62〕黄联捷、杨文采，声波方程逆散射反演的近似方法，地球物理学报，34:5，626～634，1991。

〔63〕杨文采，地震波场反演的 BG-逆散射方法，地球物理学报，38:5，358～361，1995。

〔64〕杨文采，波动方程的逆散射与迭代线性化反演，石油物探，34:3，114～121，1995。

〔65〕杨文采，非线性地震反演方法的补充及比较，石油物探，34:4，110～116，1995。

〔66〕Norton, S. J., Iterative seismic inversions, Geophys. J., 94, 457～468, 1988.

〔67〕Fawcett, J., Two-dimensional velocity inversion of the acoustic wave equation, Wave Motion, 7, 503～513, 1985.

〔68〕Tarantola, A., Inversion of seismic reflection data in the acoustic approximation, Geophys., 49, 1259～1266, 1984.

〔69〕Yorkey T. J. *et al.*, Comparing Reconstruction algorithms for electrical impedance tomography, IEEE Trans. on Biomedical Eng., Vol BME-34. 843～852, 1987.

〔70〕Webster, J. G. (1ed) Electrical Impedance Tomography. Adam Hilger, 1990.

〔71〕柳重堪，信号处理的数学方法，东南大学出版社，1992。

〔72〕陈乐寿、刘任、王天生，大地电磁测探资料处理与解释，石油工业出版社，1989。

〔73〕Madden, T. R., Inversion of low-frequency electromagnetic data, in Oceanographic and Geophysical Tomography, 377～408, Elsevier, 1990.

〔74〕Mackie, B. L. and Madden, T. R., Three-dimensional magnetotelluric inversion using conjugate gradients, Geophys. J. Int., 115, 215～229, 1993.

〔75〕Peitgen, H. O. and Saupe, D. (eds), The Science of Fractal Images, Sprirger Verlag, 1988.

〔76〕Korvin, G., Fractal Models in Earth Sciences, Elsevier, 1992.

〔77〕陈颙，分形与浑沌在地球科学中的应用，学术期刊出版社，1989。

〔78〕D. L. 特科特（陈颙等译），分形与浑沌——在地质学与地球物理学中的应用，地震出版社，1993。

〔79〕国家地震局地球物理研究所，地震科学中的非线性问题，地震出版社，1993。

〔80〕Feder, J., Fractaes, Plenum Press, 1988.

〔81〕Schuster, H. G., Deterministic Chaos, VCH, 1987.

〔82〕Feigenbaum, M. J., Universal Behavior in Nonlinear System. Los Alamos Science, 1980.

〔83〕杨文采，非线性地震道的混沌反演（Ⅰ）：理论与数值试验，地球物理学报，36，222～232，1993。

〔84〕杨文采，非线性地震道的混沌反演（Ⅰ）：Lyapunov 指数和吸引子，地球物理学报，36，376～387，1993。

〔85〕陈志德、杨文采、吴永刚，用于储层追踪的地震波阻抗混沌反演，石油地球物理勘探，30，1～10，1995。

〔86〕牛彦良、杨文采、吴永刚，跨孔地震 CT 中的逐次线性化方法，地球物理学报，38，378～386，1995。

〔87〕王山山、李青仁、管叶君，约束模拟退火反演，石油地球物理勘探，30，27～35，1995。

〔88〕Rothman, D. H., Nonlinear inversion, Statistical mechanics, and residual statistics estimation, Geophysics, 50, 2797～2807, 1985.

〔89〕杨文采，地球物理反演的遗传算法，石油物探，341，116～122，1995。

〔90〕Sambrige, M. and Drijkoningen, G., Genetic algorithms in seismic waveform inversion, Geophys. J. Int., 109, 323～342, 1992.

〔91〕Goldberg, D. E., Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley, 1989.

〔92〕杨文采，神经网络算法在地球物理反演中的应用，石油物探，34：2，116～121，1995。

〔93〕Khanna, T., Foundation of Neural Networks, Addison-Welsley, 1990.

〔94〕王伟，人工神经网络原理，北京航空航天大学出版社，1995。

〔95〕Aki, K. and Richards, P. G., Quantitative Seismology, W. H. Freeman and Company, 1980（中译："定量地震学"，地震出版社，1986）。

〔96〕傅承义、陈运泰、祁贵仲，地球物理学基础，科学出版社，1985。

〔97〕曾融生，固体地球物理学导论，科学出版社，1990。

〔98〕顾功叙，地球物理勘探基础，地质出版社，1990。

〔99〕刘光鼎主编，中国海区及邻域地质地球物理，科学出版社，1992。

〔100〕刘光鼎、曾融生、朱志文等，地球物理学，科学出版社，1994。

〔101〕White, J. E. and Sengbsu, R. L. 开发地震学（王一民等译），地震出版社，1992。

〔102〕杨文采、李幼铭等，应用地震层析成象，地质出版社，1993。

〔103〕李世雄、刘家琦，小波变换和反演数学基础，地质出版社，1994。

〔104〕Hunter, J. H. Jr, and Wilson, R. E., Waves in Astrophysics, The New York Academy of Sciences, Vol. 773, 1995.

〔105〕刘光鼎、陈运泰、陈颙、李幼铭（主编），八十年代地球物理进展论文集，科学文献出版社，1989。