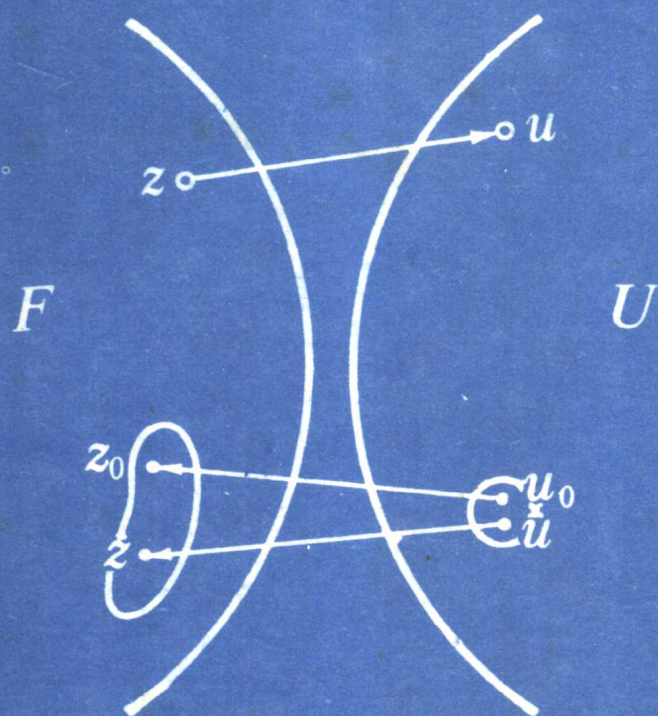


高等学校试用教材

重磁异常反演 理论与方法

刘天佑 编著



中国地质大学出版社

(鄂)新登字第 12 号

内 容 提 要

本书简明地介绍了重磁异常反演的一些理论和方法,如广义逆矩阵反演,Backus—Gilbert 反演理论,正则化方法,最小二乘参数估计,最大似然法参数估计,多解性及等效解,约束最优化,频率域最优化反演等问题,以及它们的实际应用效果。其中一部分经实践证明是行之有效的方法,另一部分则反映了苏联重磁反演理论与方法研究的某些动态。

本书可作为勘探地球物理专业的教学用书,也可供从事野外重磁资料解释的技术人员参考使用。

重磁异常反演理论与方法

刘天佑 编著

责任编辑 刘先洲

责任校对 陈爱玲

中国地质大学出版社出版

(武汉市 喻家山)

309 印刷厂印刷 (微机排版) 湖北省新华书店经销

*

开本 787×1092 1/16 印张 8.75 字数 224 千字

1992 年 1 月第 1 版 1992 年 1 月第 1 次印刷

印数 1—1500 册

ISBN 7-5625-0584-5/TP·15

定价: 2.75 元

前 言

反演问题是地球物理的核心问题。近 20 年来,由于电子计算机在地球物理勘探领域的广泛应用,地球物理反演的理论和方法得到了极大的发展。在西方有一些学者,如 Backus, G. E. 和 Gilbert, J. F. 从连续模型出发,建立了地球物理反演理论;Wiggins, R. A., Jackson, D. 和 Franklin, J. N. 从离散模型出发,提出了广义线性反演方法。他们把地球物理反演的理论抽象、提高和完善,使之成为一套系统严密的理论。加上广义逆矩阵理论,奇异值分解方法的应用,使得 Backus-Gilbert 反演理论和广义线性反演理论得到极为广泛的应用,成为近 20 年来勘探地球物理卓有成效的理论成果。在苏联,Тихонов, А. Н. 提出了解不适定问题的正则化方法;Страхов, В. Н., Цирульский, А. В. 从复变函数理论出发,成功地发展了重磁位场理论;Гольцман, Ф. М. 运用统计途径,建立了重磁异常的统计解释理论。这些从不同体系,不同角度进行研究所获得的成果,极大地丰富了地球物理反演的理论和方法。建立于古老经典的静位场理论基础之上的重磁勘探方法,由于正反演理论比其他的地球物理勘探方法(地震,电法等)成熟,所以,70 年代初以来在我国,它率先在电子计算机上实现了一系列的反演方法和技术。但是多年来对重磁反演的做法和认识仅仅停留在这样的水平上,即把原始观测数据当作是精确的,或者虽然含有误差,但是对反演结果来说是可以忽略的;对于反演结果的好坏也常常是用偏差平方和来衡量的,即把观测值与理论模型产生的理论值作对比,根据拟合程度如何来判断是否继续迭代下去,因为这些方法就是建立在经典最小二乘法的基础上的。同时,也忽视了重磁反问题非唯一性的严重程度,和缺乏对所求得解进行评价的方法和环节。Backus-Gilbert 反演理论和 Тихонов 正则化方法等的贡献在于,它们指出地球物理观测数据是有限的和包含误差的,对于这类不适定问题应当用正则化方法解决,或者说对解的估计应当在分辨与方差之间取折衷。对获得的解也应当进行评价。而不应当简单地认为模型理论值拟合了观测数据,就算反演完毕了。

近 20 年来,各种地球物理反演的理论和方法相继应用到重磁反演中,Backus-Gilbert 反演理论,广义逆矩阵理论,统计参数估计,最小二乘参数估计,最大似然法参数估计,正则化理论,约束最优化,动态规划等常常出现在文献资料中,使得我们经典的地球物理解释方法已无法与之适应。编写这一教学用书的目的在于,把地球物理反演的理论和方法(主要是重磁方面的)做一必要的介绍。书中的每一章都有作者自己的科研成果和教学方面的经验。其中许多方法在实践中证明是行之有效的。书中有较为详细的计算步骤和实例,便于理解和在计算机上实现。而另一些内容则是作者在苏联访问进修期间收集的新资料,它们基本上能反映苏联重磁反演理论与实践方面的某些动态。

全书共分七章。第一章为了顾及阅读理解的方便和连续性,首先简要回顾 70 年代中期以来我国重磁最优化反演的方法技术,然后,介绍了从统计学和概率论出发的统计参数估计,最大似然法估计,最小二乘估计的概念。用一定的篇幅介绍了从复变函数论出发的反问题多解性和等效解的概念。第二章介绍了正则化方法。正则化的概念 70 年代以来就为我国重磁界的同行所熟悉,但未见通俗系统的介绍,以及在重磁反演方面的应用效果。书中简明扼要地介绍了重磁反演中用到的正则化理论和方法,并以一定例子来加深对这一理论的理解。第三

章介绍了频率域最优化反演方法的实现,并对它作了较为全面的评价。除了实现振幅谱和相位谱反演外,还针对离散谱的特征提出了包络谱反演的途径,以及介绍了减少离散富里叶变换误差的方法。第四章讨论了最优化反演中的约束问题。目的是想说明在反演中加入先验信息的约束是减少多解性的重要措施。并从理论上和实践上对比了强制性的绝对约束方法与把约束条件加入到目标函数的约束最优化方法的差别。第五章内容较为丰富。较系统地讨论了广义逆矩阵的理论和方法,以及不同地球物理问题,采用广义逆矩阵方法的效果。它体现于,把广义逆矩阵方法加入到最优化反演计算中,突出地优越于我们熟悉的马奎特法和变尺度法,而且它还提供了资料分辨矩阵,信息密度矩阵和协方差矩阵等一系列辅助信息。第六章简要地介绍了 Backus-Gilbert 反演理论和从连续模型出发,在空间域用于求解重磁上下界面的问题。第七章系统讨论利用航磁异常计算地热场的方法原理和应用条件。介绍了一种取自 Mayhew 高空磁卫星资料反演思路的空间域调和级数模型法,以及在松辽盆地的应用效果。附录 1 粗略地介绍了数学模型的概念和物理地质模型的分类原则和建立的方法。

实际上,重磁反演的理论和方法是极为广泛的,本教学用书所涉及的内容只不过是沧海一粟,无法包罗万象。一些在实践中证明是极其有效的方法,如人机对话反演就没有收编进来。作者认为,随着计算机软件的完善,采用正演拟合结合其他快速反演方法选初值的人机对话反演解释系统是很有前景的,它充分地发挥了解释人员的经验和先验信息的作用,以及计算机方便快速两方面的优势,是克服多解性,提高推断解释质量的一种重要手段,不应当忽视这种看起来简单的方法。此外,在航磁资料解释中,为了减轻人工切线法的繁重工作量及人为的主观性,采用自动反演成图的解释系统,如 Werner 反褶积法加上马奎特反演法,也是很有前途的。

由于本教学用书所讨论的问题涉及的面较广泛,并且一些方法理论还在不断地完善之中;同时,也由于作者的水平和所接触的领域极其有限,所介绍的内容未免有片面性和谬误之处,承蒙读者批评指正。但是,与其等到这些理论与方法发展成熟了再来总结,还不如现在就把这一方面的信息、经验和实践介绍给读者,它也许能起到抛砖引玉的作用,帮助读者去思索和探索这一问题。

作者衷心地感谢管志宁教授审阅了 1989 年的初稿并提出了极为宝贵的意见,促使作者做了内容上较大的变动。也衷心地感谢苏联科学院乌拉尔分院地球物理研究所 Цирульский 教授和列宁格勒大学 Гольдман 教授以及全苏地球物理勘探研究所 Калинин, Т. Б. 所给予的帮助;最后,感谢中国地质大学出版社对出版该书给予的热情支持。

作者 1990 年 11 月于武汉

目 录

第一章 重磁异常反演的一些理论问题	(1)
1.1 重磁异常定量解释方法的回顾	(1)
1.1.1 高斯法与广义最小二乘法	(2)
1.1.2 最速下降法	(3)
1.1.3 阻尼最小二乘法	(5)
1.2 地球物理反问题的统计提法	(6)
1.2.1 统计估计的一般原则	(7)
1.2.2 估计量的好坏标准	(9)
1.3 地球物理反问题的最大似然法参数估计	(9)
1.4 地球物理反问题的最小二乘参数估计	(10)
1.5 复变函数论与重磁异常反问题的多解性及等效解	(11)
1.5.1 柯西型积分、重力位的柯西型积分表达式	(12)
1.5.2 反问题解的等效性定理	(15)
1.5.3 场源边界及密度不同组合的等效族的例子	(16)
第二章 正则化方法	(19)
2.1 不适定问题的概念	(19)
2.2 正则算子的概念	(21)
2.3 第一类积分方程的正则解	(22)
2.4 线性方程组的正则解	(23)
2.5 选择正则参数 α	(25)
2.5.1 根据误差值 δ 法	(26)
2.5.2 牛顿切线法	(26)
2.5.3 准最优准则法	(29)
第三章 频率域最优化反演	(31)
3.1 理论谱与离散富里叶变换谱精度的比较	(31)
3.2 二维板状体包络谱的性质, 提取及反演方法	(33)
3.3 二维板状体相位谱的性质、计算及反演方法	(37)
3.4 频率域反演方法的初步评价	(39)
3.5 提高位场离散富里叶变换精度的移样法和等效源续尾叠样法	(42)
3.5.1 抽样定理的推广	(43)
3.5.2 离散富里叶变换 DFT 误差方程	(45)
3.5.3 移样法	(45)
3.5.4 等效源续尾叠样法	(47)
3.6 利用灰色系统方法把空间域剖面外推以减少有限截断的误差	(47)
第四章 最优化反演中的约束问题	(51)

4.1 约束矩阵法	(51)
4.2 罚函数法	(54)
第五章 广义逆矩阵反演	(59)
5.1 利用广义逆矩阵的选择法	(59)
5.1.1 广义逆矩阵反演方法	(59)
5.1.2 用奇异值截断法解反问题	(62)
5.1.3 用加阻尼的广义逆法解反问题	(64)
5.1.4 广义逆矩阵方法的辅助信息及算法的评价	(65)
5.1.5 实例	(68)
5.2 重磁界面流动窗口系数法反演	(69)
5.2.1 核函数的性质及方程的病态	(69)
5.2.2 奇异值分解与正则化方法	(73)
5.2.3 利用广义逆矩阵的流动窗口系数法反演	(74)
第六章 巴克斯-吉尔伯特 (Backus-Gilbert) 线性反演	(81)
6.1 Backus-Gilbert 反演理论	(81)
6.1.1 模型构制	(81)
6.1.2 Backus-Gilbert 非唯一性评价	(83)
6.1.3 非线性问题和 Frechet 导数	(87)
6.2 二维磁性下界面反演	(87)
6.2.1 二维磁性界面模型的正演和 Frechet 导数	(87)
6.2.2 获得稳定解的两种途径	(88)
第七章 利用磁测资料估算居里等温面及其应用	(95)
7.1 影响居里面深度的因素和场的分离问题	(95)
7.1.1 影响居里面深度的因素	(95)
7.1.2 磁性下界面场的提取和分离	(97)
7.2 空间域连续模型的磁性体下界面的反演方法	(97)
7.2.1 方法原理	(99)
7.2.2 模型试验结果	(104)
7.3 用地热学方法估算居里深度点和地温梯度	(106)
7.4 利用航磁资料估算地温场方法及在松辽盆地的应用	(108)
附录 I 建立物理-地质模型	(114)
附录 I 有关章节中涉及的一些数学概念	(119)
参考文献	(128)
英文目录	(131)

第一章 重磁异常反演的一些理论问题

1.1 重磁异常定量解释方法的回顾

地球物理勘探的最终目的是要根据实测异常的分布特点,作出对于地质构造和矿产分布的有关结论,通常称为地球物理异常的推断解释。研究引起异常的地质体的空间位置、大小、形状及产状这样的问题称为定量解释。通常把已知场的分布求地质体模型的参数称为反问题,反问题是异常定量解释的基础。在重磁勘探中,反演计算的方法可以分为特征点法,任意点法和选择法三大类。而70年代以来又把重磁勘探方法用于区域密度界面和磁性界面的解释,在这类问题中反演计算的方法又可分为直接法和迭代法,这些方法通常只能在计算机上实现。

特征点法是利用异常曲线上的特征点如极值点,半极值点,零值点以及拐点等来计算地质体产状要素的方法,常用的有特征点解析法,切线法。由于它只利用少数特殊点进行计算,所以其抗干扰能力较差,一般只适用于单个矿体引起的规则光滑异常的计算。

任意点法是利用异常曲线上的任意点来计算地质体的产状。它的特点是能够利用较多的观测值参加计算,从而提高了计算结果的精度和抗干扰能力,并且可以将多点的计算结果进行比较,来评价计算结果的可靠程度。但是任意点法一般计算公式比较复杂,所以手算不便,应用不多。特别在斜磁化条件下的许多形体目前还没有任意点法的计算公式。

不管是特征点法还是任意点法,对于形状复杂的地质体,或多个地质体引起的复杂叠加异常来说,定量计算的效果往往都不好。

选择法,如解释二度体磁异常的扇形量板法,其实质是将实测异常曲线与一系列的已知形状模型体所产生的理论异常曲线进行比较,当实测曲线与某一条理论曲线相符合时,我们就将该理论曲线对应的模型体作为实际地质体的解释。选择法的优点是利用整条观测曲线进行解释,当个别点受到歪曲时对解释结果影响不大。它较适合于复杂异常的解释。但是解释的工作量相当大,并且没有对三度体异常进行计算的量板。

除了这三大类方法外,还有一些方法如积分法,矢量法,复梯度法,希尔伯特变换法等,这些方法大多属于异常变换法,都是为了压制某种成分的干扰,突出有意义的异常而进行的。

70年代以来,国内外开始应用电子计算机解地球物理反问题,这些方法可以分为两类:一类针对线性问题的(或非线性问题作了线性化处理的),常采用直接解法和迭代法;另一类是针对非线性问题的,即在最小二乘意义下最优化的迭代法,这种方法的实质就是选择法。下面我们简单回顾一下这类方法的思路及发展过程。而其他的直接解法和迭代法将在其他章节介绍。

1.1.1 高斯法与广义最小二乘法

在重磁资料的解释中,一般都是多参数的非线性最优估计问题,理论模型异常

$$\Delta Z = f(\underline{x}, b_1, b_2, \dots, b_n) \quad (1.1.1)$$

是参数 b_i ($i=1, 2, \dots, n$) 的非线性函数。其中 $\underline{x} = (x, y, z)$ 为变量。 ΔZ_i ($k=1, 2, \dots, m$) 为 m 个观测点 $\underline{x}_i$ ($k=1, 2, \dots, m$) 上的实测异常,现在要求与观测异常相对应的地质体参数值 b_i ($i=1, 2, \dots, n$),使理论模型异常值与实测异常值偏差的平方和

$$\Phi = \sum_{k=1}^m [\Delta Z_k - f(\underline{x}_k, b_1, b_2, \dots, b_n)]^2 \quad (1.1.2)$$

为最小。设 b_i 的初始值为 $b_i^{(0)}$, 则

$$b_i = b_i^{(0)} + \delta_i \quad (i=1, 2, \dots, n) \quad (1.1.3)$$

问题在于如何求出校正值 δ_i , 下面介绍高斯提出的方法。

因为 f 是 b_i 的非线性函数,直接求 b_i 非常困难,所以应当把 f 在 $b^{(0)}$ 附近线性化。我们在 $b^{(0)}$ 附近将 f 写为泰勒展开式,并略去 δ_i 的二次及二次以上的项,则有

$$\begin{aligned} f(\underline{x}_k, b_1, b_2, \dots, b_n) &= f(\underline{x}_k, b_1^{(0)} + \delta_1, b_2^{(0)} + \delta_2, \dots, b_n^{(0)} + \delta_n) \\ &= f^{(0)}(k) + \frac{\partial f^{(0)}(k)}{\partial b_1} \delta_1 + \frac{\partial f^{(0)}(k)}{\partial b_2} \delta_2 + \dots + \frac{\partial f^{(0)}(k)}{\partial b_n} \delta_n \end{aligned} \quad (1.1.4)$$

式中 $f^{(0)}(k) = f(\underline{x}_k, b_1^{(0)}, b_2^{(0)}, \dots, b_n^{(0)})$

这里 $f^{(0)}(k)$ 和 $\frac{\partial f^{(0)}(k)}{\partial b_i}$ 当 $b_i^{(0)}$ 给定时都是 $\underline{x}$ 的函数,可以直接计算得到。

将 (1.1.4) 代入 (1.1.2) 式可得:

$$\Phi \doteq \hat{\Phi} = \sum_{k=1}^m [\Delta Z_k - (f^{(0)}(k) + \frac{\partial f^{(0)}(k)}{\partial b_1} \delta_1 + \dots + \frac{\partial f^{(0)}(k)}{\partial b_n} \delta_n)]^2 \quad (1.1.5)$$

现在要求 δ_i , 使 $b_i = b_i^{(0)} + \delta_i$ 为参数的模型体理论曲线与实测曲线尽可能地符合,即使得 $\hat{\Phi}$ 值为极小,也就是说应当满足关系:

$$\frac{\partial \hat{\Phi}}{\partial \delta_i} = \frac{\partial \hat{\Phi}}{\partial b_i} = 0 \quad (i=1, 2, \dots, n)$$

由此可得:

$$\begin{aligned} 0 &= \frac{\partial \hat{\Phi}}{\partial b_i} = \frac{\partial \hat{\Phi}}{\partial \delta_i} \\ &= 2 \sum_{k=1}^m [\Delta Z_k - (f^{(0)}(k) + \frac{\partial f^{(0)}(k)}{\partial b_1} \delta_1 + \dots + \frac{\partial f^{(0)}(k)}{\partial b_n} \delta_n)] (-\frac{\partial f^{(0)}(k)}{\partial b_i}) \\ &= 2 [\delta_1 \sum_{k=1}^m \frac{\partial f^{(0)}(k)}{\partial b_1} \cdot \frac{\partial f^{(0)}(k)}{\partial b_i} + \dots + \delta_n \sum_{k=1}^m \frac{\partial f^{(0)}(k)}{\partial b_n} \cdot \frac{\partial f^{(0)}(k)}{\partial b_i} \\ &\quad - \sum_{k=1}^m \frac{\partial f^{(0)}(k)}{\partial b_i} (\Delta Z_k - f^{(0)}(k))] \end{aligned} \quad (1.1.6)$$

整理可得:

$$\sum_{i=1}^n \sum_{j=1}^n [(\frac{\partial f^{(0)}(k)}{\partial b_i} \cdot \frac{\partial f^{(0)}(k)}{\partial b_j}) \delta_i] = \sum_{i=1}^n [\frac{\partial f^{(0)}(k)}{\partial b_j} (\Delta Z_k - f^{(0)}(k))] \quad (1.1.7)$$

将 (1.1.6) 式写成方程组形式:

$$\begin{cases} a_{11}\delta_1 + a_{12}\delta_2 + \dots + a_{1n}\delta_n = g_1 \\ a_{21}\delta_1 + a_{22}\delta_2 + \dots + a_{2n}\delta_n = g_2 \\ \dots\dots\dots \\ a_{n1}\delta_1 + a_{n2}\delta_2 + \dots + a_{nn}\delta_n = g_n \end{cases} \quad (1.1.8)$$

$$\text{式中 } a_{ij} = \sum_{k=1}^m \frac{\partial f^{(0)}(k)}{\partial b_i} \frac{\partial f^{(0)}(k)}{\partial b_j}, \quad g_i = \sum_{k=1}^m \frac{\partial f^{(0)}(k)}{\partial b_i} (\Delta Z_k - f^{(0)}(k))$$

将 (1.1.8) 式写成矩阵形式:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_n \end{pmatrix} = \begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_n \end{pmatrix} \quad (1.1.9)$$

或记为

$$A\delta = g \quad (1.1.10)$$

式中

$$A^{[n \times n]} = P^T P$$

$$P^{[m \times n]} = \left(\frac{\partial f^{(0)}(k)}{\partial b_i} \right)$$

$$g^{[n \times 1]} = P^T (\Delta Z - f^{(0)})$$

在实际计算中,当给定了观测值及模型体初值 $b_i^{(0)}$ 后, a_{ij} 和 g_j 都可以计算出来,求解上面方程组可得改正量 δ ,进而可求出 $b^{(1)} = b^{(0)} + \delta$,以 $b^{(1)}$ 作为新的初值再重复以上步骤计算,直至 $|\delta_i|$ ($i=1, 2, \dots, n$) 小于允许误差为止。

前面说过,非线性问题需要反复迭代和修正,关键在于泰勒展开式 (1.1.4) 是近似的,因此求得的 b_i 也是近似的,其近似程度依赖于 $|\delta_i|$ 的大小,若 $|\delta_i|$ 较大,则 δ_i 的二次项及二次以上的项不能忽略。逐次迭代的结果, b_i 逐步逼近于真值,而 $|\delta_i|$ 也逐渐减小,当小到二次项可以忽略时, (1.1.4) 式便成为精确等式。当然,如果二次和二次以上的项其偏导数均为零时,不论 δ_i 有多大, (1.1.4) 式也总是精确的成立,这就是线性模型的情况,所以线性问题用不着迭代,而非线性问题则需要迭代。

以上方法通常称为高斯法。高斯法在应用时,要求矩阵 A 是正定的和非奇异的。否则便不能区分极大、极小和鞍点,或者干脆不能求解。高斯法在应用时还有一个困难,就是迭代过程不稳定,当 δ 过大,使泰勒展开的高次项不容忽略时,就会发生这样的情况,以参数 $b^{(l+1)}$ 计算的 Φ 值比以 $b^{(l)}$ 计算的 Φ 值还要大 (这里 l 为迭代次数的指标),即是说高斯法不能保证平稳地收敛于 Φ 的极小值。

后来,根据多次应用的实际经验,对高斯法提出一个改进方案,这时不是把 δ 作为校正值,而是将 δ 作为校正方向,记为 δ_i ; 而将 δ_i 乘上一个常数 α 作为校正值。问题变为要求 α 值,使

$$\Phi(b^{(l)} + \alpha\delta_i^{(l)}) \quad (l=1, 2, \dots) \quad (1.1.11)$$

为极小,这就是单变量求极小的问题,然后取

$$b^{(l+1)} = b^{(l)} + \alpha\delta_i^{(l)} \quad (1.1.12)$$

作为新的初值,通常 α 取在 $[0, 1]$ 之间。

这种技术称为广义最小二乘法,它使迭代的稳定性有所改善。但是,即使这样,当初值取得不合适时,这种方法也经常不收敛。

1.1.2 最速下降法

一般说来, $\Phi(b)$ 是一个相当复杂的函数,但是在它的极小值附近不大的范围内,可以视为二次型。假如我们在最优值附近给定了一个初值 $b^{(0)}$, 设最优值为 b , 初值到最优值的改正量为 δ , 即 $b = b^{(0)} + \delta$ 。若该改正量不大,我们可以将函数 Φ 展成 n 维泰勒级数并忽略高次项:

$$\Phi(b_i) = \Phi(b_i^{(0)} + \delta_i)$$

$$= \phi(b^{(0)}) + \sum_{i=1}^n \frac{\partial \phi(b^{(0)})}{\partial b_i} \delta_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 \phi(b^{(0)})}{\partial b_i \partial b_j} \delta_i \delta_j$$

表示为矩阵形式有

$$\phi = \phi_0 + g^T \underline{\delta} + \frac{1}{2} \underline{\delta}^T G \underline{\delta} \quad (1.1.13)$$

式中

$$g = \text{grad} \phi = \begin{bmatrix} \frac{\partial \phi}{\partial b_1} \\ \frac{\partial \phi}{\partial b_2} \\ \vdots \\ \frac{\partial \phi}{\partial b_n} \end{bmatrix} \quad G = \begin{bmatrix} \frac{\partial^2 \phi}{\partial b_1 \partial b_1} & \frac{\partial^2 \phi}{\partial b_1 \partial b_2} & \cdots & \frac{\partial^2 \phi}{\partial b_1 \partial b_n} \\ \frac{\partial^2 \phi}{\partial b_2 \partial b_1} & \frac{\partial^2 \phi}{\partial b_2 \partial b_2} & \cdots & \frac{\partial^2 \phi}{\partial b_2 \partial b_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 \phi}{\partial b_n \partial b_1} & \frac{\partial^2 \phi}{\partial b_n \partial b_2} & \cdots & \frac{\partial^2 \phi}{\partial b_n \partial b_n} \end{bmatrix}$$

(1.1.13) 式是二次型，这就说明了函数 ϕ 在最优值附近总是可以用二次型来近似。在最优值附近的 $(b_1, b_2, \dots, b_n)$ n 维空间中

$$\phi(\underline{b}) = \phi(\underline{b}^{(0)})$$

确定一个 n 维椭球面。当 $n=2$ 时其图形可示意表示在图 1.1.1 中，图中所示的等值线比较接近实际的曲面等值线，不是理想的椭球面。

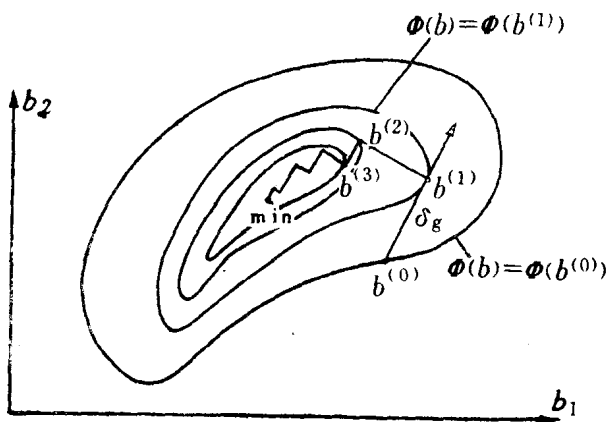


图 1.1.1

最速下降法的思想是这样：要求 $\phi(\underline{b})$ 的极小值，从 $\underline{b}^{(0)}$ 出发，应该沿着函数 ϕ 减小得最快的方向去找。

从 $\underline{b}^{(0)}$ 出发，函数 $\phi(\underline{b})$ 在 $\underline{b}^{(0)}$ 点减小得最快的方向是什么方向呢？这应当是函数 ϕ 在 $\underline{b}^{(0)}$ 点的梯度方向的反方向。我们知道在 $\underline{b}^{(0)}$ 点函数 ϕ 的梯度矢量为：

$$\text{grad} \phi = \left(\frac{\partial \phi}{\partial b_1}, \frac{\partial \phi}{\partial b_2}, \dots, \frac{\partial \phi}{\partial b_n} \right)_{\underline{b}=\underline{b}^{(0)}}$$

若另取一个方向 $\underline{c}$ ，在 $\underline{b}^{(0)}$ 点 $\phi(\underline{b})$ 在 $\underline{c}$ 方向的变化速率为

$$\left[\frac{d}{d\alpha} \phi(\underline{b}^{(0)} + \alpha \underline{c}) \right]_{\alpha=0} = \left(\frac{\partial \phi}{\partial b_1} \right)_{\underline{b}^{(0)}} c_1 + \left(\frac{\partial \phi}{\partial b_2} \right)_{\underline{b}^{(0)}} c_2 + \cdots + \left(\frac{\partial \phi}{\partial b_n} \right)_{\underline{b}^{(0)}} c_n,$$

由上式可见函数 Φ 在 $\underline{b}^{(0)}$ 点沿 $\underline{c}$ 方向的变化速率就是在 $\underline{b}^{(0)}$ 点函数 Φ 的梯度矢量在 $\underline{c}$ 方向上的投影, 显然 $\underline{c}$ 与 $\text{grad}\Phi$ 重合时此值最大, 所以 $\Phi(\underline{b})$ 在 $\underline{b}^{(0)}$ 点增加最快的方向就是 $\text{grad}\Phi$ 方向, 而其反方向就是函数 Φ 减小最快的方向, 我们将该方向记为 $\underline{\delta}_g$ 。所以最速下降法有时也称为梯度法。

我们沿着最速下降方向 $\underline{\delta}_g$ 寻找校正量 $\underline{\delta}$, 这时 (1.1.3) 式可以写成:

$$\underline{b} = \underline{b}^{(0)} + \lambda \underline{\delta}_g$$

式中 $\underline{\delta}_g = -\left(\frac{\partial\Phi}{\partial b_1}, \frac{\partial\Phi}{\partial b_2}, \dots, \frac{\partial\Phi}{\partial b_n}\right)^T$

问题在于如何求得 λ , 使

$$\Phi(\lambda) = \Phi(\underline{b}) = \Phi(b_1^{(0)} - \lambda \frac{\partial\Phi}{\partial b_1}, b_2^{(0)} - \lambda \frac{\partial\Phi}{\partial b_2}, \dots, b_n^{(0)} - \lambda \frac{\partial\Phi}{\partial b_n})$$

取极小。

因为函数 Φ 为非负, 如果 $\lambda \underline{\delta}_g$ 较小, 将 $\Phi(\underline{b}^{(0)} + \lambda \underline{\delta}_g)$ 泰勒展开到一次项, 可写出:

$$\Phi(\underline{b}^{(0)} + \lambda \underline{\delta}_g) \doteq \Phi_0 - \lambda \underline{\delta}_g^T \underline{\delta}_g = \Phi_0 - \lambda |\text{grad}\Phi_0|^2$$

使 Φ 取极小, 我们取等式右边为零, 由此可得:

$$\lambda = \frac{\Phi}{\left(\frac{\partial\Phi}{\partial b_1}\right)^2 + \left(\frac{\partial\Phi}{\partial b_2}\right)^2 + \dots + \left(\frac{\partial\Phi}{\partial b_n}\right)^2} \bigg|_{\underline{b}=\underline{b}^{(0)}} = \frac{\Phi_0}{|\text{grad}\Phi_0|^2}$$

将 λ 代入可求出第一次迭代的修正值 $\underline{b}^{(1)}$:

$$\underline{b}^{(1)} = \underline{b}^{(0)} - \frac{\text{grad}\Phi}{|\text{grad}\Phi|^2} \Phi$$

以 $\underline{b}^{(1)}$ 作为新的初值, 继续沿 $\underline{b}^{(1)}$ 点函数的 $\underline{\delta}_g$ 方向进行计算, 求出 λ 和第二次迭代改正值, 如此继续下去, 即可最后逼近 Φ 的极小值。其迭代过程如图 1.1.1 所示。

最速下降法一般能保证收敛, 但收敛速度常常很慢, 特别在函数 Φ 的 n 维椭球面是拉长椭球面的情况更是如此。

1.1.3 阻尼最小二乘法

阻尼最小二乘法也称马奎特法, 1963 年, 马奎特对高斯法作了有效的改正, 这种方法在实际中应用效果较好。

由前面叙述我们知道, 用高斯法进行迭代时, 校正矢量的步长较大, 如果初值给得较好, 能很快收敛, 但它的收敛性不稳定, 当初值不好时则容易发散。而最速下降法的情况正好相反, 它一般能保证收敛, 但步长很小, 计算很费时间。

另外, 高斯法的校正矢量的方向 ($\underline{\delta}_g$) 和最速下降法的方向 ($\underline{\delta}_g$) 也往往不一致, 特别是当 $\Phi(\underline{b})$ 的等值面是拉长椭球的情况, 两者方向之间的夹角 γ 可达 $80-90^\circ$, 如图 1.1.2 所示, 例如对于二次型, 高斯法的方向是直接指向 Φ 的极小点, 而最速下降法方向 $\underline{\delta}_g$ 是负梯度方向, 它是垂直通过 $\underline{b}^{(0)}$ 点函数的等值面的。正因为这样, 对于非线性函数当等值面拉得很长时, $\underline{\delta}_g$ 远远偏离负梯度方向。这是造成高斯法经常失败的主要原因。

马奎特法是在高斯法和最速下降法之间取某种插值, 它力图以最大步长前进, 同时又能紧靠负梯度方向, 这样既能保证收敛, 又加快收敛速度。它的基本思想是: 在迭代过程中的每一步, 尽量使用接近高斯改正量的方向, 以使校正步长尽可能地增大, 如当这种情况不能收敛时, 则改用接近最速下降的方向 $\underline{\delta}_g$, 同时缩小步长, 以保证收敛。马奎特方法的校正矢量我们以 $\underline{\delta}_0$ 表示。

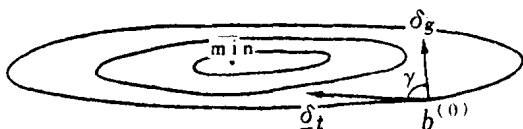


图 1.1.2

实现这种思想的数学方法，只要将方程

$$A\delta_t = g$$

改变为

$$(A + \lambda I)\delta_0 = g \quad (\lambda > 0) \quad (1.1.14)$$

就可以了。式中 δ_0 是我们要求得的对参数的校正矢量，即马奎特校正矢量， I 为单位矩阵， λ 为由我们选取来控制校正方向和步长的任意正数。校正矢量 δ_0 是 λ 的函数，即

$$\delta_0 = \delta_0(\lambda)$$

由方程 (1.1.10) 变成方程 (1.1.14)，改善了收敛性这是很明显的，如果矩阵 A 是奇异的，当 $\lambda > 0$ 时 $(A + \lambda I)$ 则为正定的。

由 (1.1.14) 式，当 $\lambda = 0$ 时， $\delta_0 = \delta_t$ ，马奎特法退化为高斯法，而当 $\lambda \rightarrow \infty$ 时， $\delta_0 \doteq \frac{1}{\lambda}g$ ， g 是最速下降法的方向，这时马奎特校正方向就是最速下降法的方向。由此可见：

(1) 当 λ 越来越大时， δ_0 的长度越来越小，且

$$\lim_{\lambda \rightarrow \infty} \|\delta_0\| = 0$$

这里 $\|\delta_0\|$ 表示矢量 δ_0 的长度。

(2) 当 λ 由零逐渐增大时， δ_0 的方向逐渐由高斯法方向 δ_t 转向最速下降法的方向 δ_g ， λ 越大 δ_0 越接近 δ_g 。

(3) 对 $\lambda > 0$ 的任意实数， δ_0 方向使 Φ 在半径为 $\|\delta_0\|$ 的球面上取极小。

上述 (2) 说明 $\delta_0(\lambda)$ 当 λ 增大时是越来越接近 Φ 的负梯度方向，即总是指向 Φ 的下降方向，又根据 (1)，当 $\lambda \rightarrow \infty$ 时， $\|\delta_0\| \rightarrow 0$ ，因此，对于任意第 l 次迭代值 $\underline{b}^{(l)}$ ，我们总可以找到充分大的 λ ，使

$$\Phi(\underline{b}^{(l)} + \delta_0(\lambda)) < \Phi^{(l)} = \Phi(\underline{b}^{(l)})$$

如果我们取 $\underline{b}^{(l+1)} = \underline{b}^{(l)} + \delta_0^{(l)}$ ，就保证了 Φ 在每次迭代中改善。所以上述 (1)，(2) 是保证下面迭代过程得以稳定进行的理论根据。而 (3) 则说明在跨出同样步长的情况下，以 $\delta_0(\lambda)$ 的方向为最好，这便从理论上说明了这种方法的优点。

1.2 地球物理反问题的统计提法

上面我们简单回顾的高斯法，广义最小二乘法，最速下降法及阻尼最小二乘法等非线性反演的最优化方法都没有涉及到观测资料的误差问题。实际上，用于求解反问题的地球物理观测资料总是包含有随机误差的，观测资料本身就是具有某种特征的随机值。

反演理论的基本前提是观测数据和模型参数之间存在着一定的关系,任何解反问题的方法,即根据观测数据估计一个模型的参数,它们都会把数据中的误差映射到估计的模型参数中去。因此,模型参数的估计值本身就是随机变量。从概率统计的观点出发,应当把确定模型参数的问题归结为统计估计问题,用概率论和数理统计方法加以解决。

1.2.1 统计估计的一般原则

为了便于理解地球物理反问题统计提法,我们先简单回顾一下涉及到概率论与数理统计的一些概念。

随机变量 在一定条件下,受随机因素的影响而变动的数量我们称之为随机变量。例如在一个测点上用重力仪或磁力仪进行多次观测,由于它受地下,地表以及仪器内部随机因素影响,它们的观测值不可能完全相同。随机变量的变化是有一定规律的,它可以由随机事件的概率来刻画,可以通过考察随机事件的频率的方法来认识这种规律。我们在相同条件下做大量试验,用列表法及图示法来表示随机变量的频率分布,并由此得到概率分布。例如对于连续随机变量的频率分布可以用直方图来表示。当观测数据不断增加时,划分区间的间隔不断缩小,那么直方图上边的折线将接近于一条连续的曲线,这条曲线 $P(x)$ 就是连续随机变量 ξ 的概率密度。如果对任意 $x_1, x_2 (x_1 \leq x_2)$, 则有

$$P(x_1 \leq \xi < x_2) = \int_{x_1}^{x_2} P(x) dx \quad (1.2.1)$$

其中 $P(x_1 \leq \xi < x_2)$ 表示事件 “ $x_1 \leq \xi < x_2$ ” 的概率。

关于概率密度 $P(x)$, 我们指出两个性质:

$$(1) \quad P(x) \geq 0,$$

$$(2) \quad \int_{-\infty}^{\infty} P(x) dx = 1.$$

关于第一个性质,是由于概率不为负的缘故。关于第二个性质,由于事件 $-\infty < \xi < +\infty$ 是必然事件,因此,它的概率必为 1。

分布函数与概率密度 对于任何一个实数 x , “ $\xi < x$ ” 是一个随机事件,它有概率 $P(\xi < x)$, 我们记为

$$F(x) = P(\xi < x) \quad (1.2.2)$$

我们称 $F(x)$ 为 ξ 的分布函数。显然,分布函数是用来描述随机变量概率规律的。分布函数当 $x \rightarrow +\infty$ 时, $\lim_{x \rightarrow +\infty} F(x) = 1$ 。连续随机变量的分布函数 $F(x)$ 与概率密度 $P(x)$ 的关系由 (1.2.1), (1.2.2) 可得

$$F(x) = \int_{-\infty}^x P(x) dx$$

因此

$$P(x) = \frac{dF(x)}{dx} \quad (1.2.3)$$

即概率密度 $P(x)$ 是分布函数 $F(x)$ 的微商。

正态分布 (高斯分布) 连续随机变量 ξ 的概率密度为

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}} \quad (\sigma > 0) \quad (1.2.4)$$

则称 ξ 服从正态分布并记为 $\xi \sim N(a, \sigma)$, 当 $a=0, \sigma=1$ 时的正态分布称为标准正态分布。(1.2.3), (1.2.4) 式是对一维随机变量而言, 对于 n 维随机变量则有

$$P(x_1, x_2, \dots, x_n) = \frac{\partial F(x_1, x_2, \dots, x_n)}{\partial x_1 \partial x_2 \dots \partial x_n} \quad (1.2.5)$$

数学期望与方差 在应用中我们常常需要知道描述随机变量一些特征的数,例如,它可能取值的平均数是什么,取值的分散程度如何等等,随机变量的这些数字特征,不仅在一定程度上简单、形象地描绘了随机变量的特征,而且在实际问题中这些数字特征用数理统计的方法也比较容易估计。

随机变量取值的平均数,我们用数学期望来描述。设离散随机变量 ξ 可能取的值 $x_1, x_2, \dots$, 若相应概率为 $P(\xi=x_i)=P_i$, 我们称

$$E\xi = \sum_{i=1}^{\infty} x_i P_i \quad (1.2.6)$$

为 ξ 的数学期望。

对于随机变量取值的分散程度,我们用方差来描述。即用 $|\xi - E\xi|$ 来描述 ξ 的取值的摆动情况,由于绝对值在数学上的处理极不方便,因此,我们用 $(\xi - E\xi)^2$ 来描述 ξ 的摆动情况, ξ 是个离散随机变量,可能取值为 $0, 1, \dots, n$, 因此 $(\xi - E\xi)^2$ 可能取值为 $(0 - E\xi)^2, (1 - E\xi)^2, \dots, (n - E\xi)^2$, 而取这些值的概率分别为 $P_0, P_1, \dots, P_n$, 因此 $(\xi - E\xi)^2$ 平均摆动大小为 $\sum_{i=0}^n (k - E\xi)^2 P_i$, 这是一个平均方差和,我们称之为方差。

一般,设 ξ 可能取值 $x_1, x_2, \dots$, 相应概率为 $P(\xi=x_i)=P_i$, 我们称

$$D\xi = \sum_{i=1}^{\infty} (x_i - E\xi)^2 P_i \quad (1.2.7)$$

为 ξ 的方差,称 $\sqrt{D\xi}$ 为 ξ 的标准差,若把 (1.2.7) 展开,则有

$$D\xi = \sum_{i=1}^{\infty} x_i^2 P_i - (E\xi)^2$$

类似的有连续随机变量 ξ 的数学期望 (1.2.8) 和方差 (1.2.9)

$$E\xi = \int_{-\infty}^{\infty} x P(x) dx \quad (1.2.8)$$

$$D\xi = \int_{-\infty}^{\infty} (x - E\xi)^2 P(x) dx \quad (1.2.9)$$

对于正态分布,则有 ξ 的数学期望和方差

$$\begin{cases} E\xi = a \\ D\xi = \sigma^2 \end{cases} \quad (1.2.10)$$

参数估计. 最大似然法 我们知道,每一个样本 x_i 都是随机变量,它的概率分布时常是依赖于一些参数的。例如, x_i 服从正态分布时,概率密度就要依赖于参数 a 和 σ 。设 x_i 的分布为 $f(x_i, \theta)$, 其中 θ 是参数。当 x_i 为离散随机变量时, $f(x_i, \theta)$ 代表概率分布;当 x_i 为连续随机变量时, $f(x_i, \theta)$ 代表概率密度。在许多情况下,我们知道了 $f(x_i, \theta)$ 的一般形式,但是不能预先确定参数 θ 的值。最大似然法就是通过若干实验来决定这个参数的方法。

因为 $x_1, x_2, \dots, x_n$ 是相互独立的,所以 n 维随机变量 $(x_1, x_2, \dots, x_n)$ 的分布函数为 $\prod_{i=1}^n f(x_i, \theta)$ 。在做了 n 次试验后得到样本值 $(x_1, x_2, \dots, x_n)$ 。为了确定 θ 的值,自然提出: θ 取什么值时,最有可能得到这一批样本值? 换句话说,当 θ 取什么值时,概率分布 $\prod_{i=1}^n f(x_i, \theta)$ 对于这一批样本值 $(x_1, x_2, \dots, x_n)$ 达到最大值。这样,就把确定 θ 的问题化为函数求极值的问题。这种确定 θ 的方法就叫做最大似然法。

我们称

$$L = \prod_{i=1}^n f(x_i, \theta) \quad (1.2.11)$$

为似然函数。

要使似然函数取最大值，必然满足条件

$$\frac{\partial L}{\partial \theta} = 0 \quad (1.2.12)$$

由于 L 和 $\ln L$ 同时达到最大值，但有时 $\ln L$ 便于计算，所以上面方程也可改写为

$$\frac{\partial \ln L}{\partial \theta} = 0 \quad (1.2.13)$$

由方程 (1.2.12) 或 (1.2.13) 解出的 θ 是 $x_1, x_2, \dots, x_n$ 的函数，把它记作 $\hat{\theta}(x_1, x_2, \dots, x_n)$ ，我们称 $\hat{\theta}(x_1, x_2, \dots, x_n)$ 为总体参数 θ 的最大似然估计量。

1.2.2 估计量的好坏标准

设参数为 θ ，不管用什么方法我们得到估计量 $\hat{\theta}(x_1, x_2, \dots, x_n)$ 。现在给出判别或比较估计量好坏的几个标准。

1. 无偏性 因为估计量 $\hat{\theta}(x_1, x_2, \dots, x_n)$ 是随机变量 $x_1, x_2, \dots, x_n$ 的函数，因此 $\hat{\theta}$ 也是一个随机变量，随机变量是在它的数学期望上下摆动取值的，既然用 $\hat{\theta}$ 来估计参数 θ ，因此，自然希望 $\hat{\theta}$ 的数学期望就是 θ 。一般定义，如果估计量 $\hat{\theta}(x_1, x_2, \dots, x_n)$ 满足 $E\hat{\theta} = \theta$ ，则称 $\hat{\theta}(x_1, x_2, \dots, x_n)$ 是无偏估计量。

2. 有效性 方差刻画了一个随机变量取值的分散程度，方差越小，说明随机变量取值越集中。因此我们给出有效性的标准。通常定义，如果 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 都是参数 θ 的无偏估计量，且

$$E(\hat{\theta}_1 - \theta)^2 \leq E(\hat{\theta}_2 - \theta)^2,$$

则称无偏估计量 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 有效。

3. 一致性 由 n 次试验结果 $x_1, x_2, \dots, x_n$ ，对总体的参数 θ 做出估计量 $\hat{\theta}_1(x_1, x_2, \dots, x_n)$ 。我们自然希望试验的次数越多，估计量 $\hat{\theta}(x_1, x_2, \dots, x_n)$ 接近参数 θ 的效果越好。于是便引出一致性准则。我们定义，如果参数 θ 的估计量 $\hat{\theta}(x_1, x_2, \dots, x_n)$ 满足：对任意 $\varepsilon > 0$ 有

$$\lim P(|\hat{\theta}(x_1, x_2, \dots, x_n) - \theta| > \varepsilon) = 0$$

则称 $\hat{\theta}(x_1, x_2, \dots, x_n)$ 为 θ 的一致估计量。实际上，一致估计量就是要求 $\hat{\theta}(x_1, x_2, \dots, x_n)$ 依概率收敛到 θ 。

1.3 地球物理反问题的最大似然法参数估计

我们指出，应当把地球物理反问题归结为参数分布的统计估计问题。也就是根据尺度为 m ，概率密度为 $P(u; z)$ 的观测空间给出的随机变量 $u = (u_1, u_2, \dots, u_m)$ 来估计参数 z 和确定估计的误差。

我们常常把重磁性反问题用观测数据 u 和模型参数 z 的关系表示为 $Az = u$ ，一般认为重磁观测数据 u 符合多变量高斯分布，数据不相关，而把 Az 看作随机向量 u 的数学期望。对于随机变量 u_i ，其概率密度

$$P(u_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(u_i - Az)^2}{2\sigma^2}}$$

那么 m 维随机向量 $u = (u_1, u_2, \dots, u_m)$ 的分布函数为

$$L = \prod_{i=1}^m P(u_i) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^m e^{-\frac{1}{2\sigma^2} \sum_{i=1}^m (u_i - Az)^2}$$

取对数

$$\ln L = -m \ln \sigma - m \ln \sqrt{2\pi} - \frac{1}{2\sigma^2} (u - Az)^T (u - Az) \quad (1.3.1)$$

当观测数据的概率取得极大值时,即似然函数取得最大值时,得到的模型参数 z 的估计值是最佳的。不难看出, (1.3.1) 式取极大相当于

$$(u - Az)^T (u - Az) \quad (1.3.2)$$

取极小。所以模型参数的最大似然估计等价于最小二乘解估计。这两种不同观点出发得出的参数估计是一致的。

例 球形矿体埋深 D , 在 P_i 点实测重力异常值为 Δg_i ($i=1, 2, \dots, m$), 我们熟知, 球体的重力异常值 Δg 为

$$\Delta g(P_i, M, D) = f_i(M, D) = \frac{f \cdot M \cdot D}{(x_i^2 + y_i^2 + D^2)^{3/2}}$$

这里 f 是万有引力常数, M 是球形矿体的剩余质量。如果在 m 个测点上观测值与计算值之差为随机误差 ε_i , 它具有正态分布, 零均值和方差 σ^2 , 那么随机变量 Δg_i 的分布也是正态的, 均值为 $f_i(M, D)$ 和方差 σ^2 :

$$P(\Delta g_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2} (\Delta g_i - f_i(M, D))^2\right]$$

观测值向量 $\Delta g = (\Delta g_1, \Delta g_2, \dots, \Delta g_m)$ 将具有多维正态分布

$$P(\Delta g; M, D) = \prod_{i=1}^m P(\Delta g_i) = \frac{1}{(\sqrt{2\pi}\sigma)^m} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^m (\Delta g_i - f_i(M, D))^2\right]$$

问题归结为求未知参数 M, D, σ 的分布, 第一、二个参数即剩余质量, 埋深, 第三个是误差特征。注意到这里 M, D 与 $f_i(M, D)$ 是非线性关系, 实际计算时还要通过高斯线性化方法来实现。

1.4 地球物理反问题的最小二乘参数估计

最小二乘法是高斯 (1795 年) 为了用天文观测数据来估算行星运动轨道而提出的, 迄今已近 200 年, 科学技术的发展给最小二乘法的理论和应用都带来了深刻的影响。这是一种在实际中最为广泛应用的方法。在 1.1 中我们叙述的高斯法就是利用最小二乘法求解地球物理反问题的一种方法, 它通过线性化的方法来求解一个非线性的重磁反演问题, 这是我们所熟知的。但是在观测数据中考虑误差影响, 用统计观点来讨论它还不多。我们这里所说的地球物理反问题的最小二乘参数估计, 就是以统计观点, 用最小二乘法来求解模型参数问题。

最小二乘法是采用这样的方法来估计参数 z , 即极小化残差 ε_i 的平方和:

$$\rho = \sum_{i=1}^m \varepsilon_i^2 = \varepsilon^T \varepsilon = (u - Az)^T (u - Az) \quad (1.4.1)$$

满足 (1.4.1) 条件的 $\hat{z}$ 称为最小二乘估计。 ρ 的极小化等价于解方程组

$$\frac{\partial \rho}{\partial z_i} = 0, \quad (i = 1, 2, \dots, n)$$

或

$$A^T A z = A^T u \quad (1.4.2)$$

我们称 (1.4.2) 为正规方程组, 若矩阵 $A^T A$ 是非退化的, 则可以记正规方程组的解为

$$\hat{z} = (A^T A)^{-1} A^T u \quad (1.4.3)$$

最小二乘估计具有如下的重要特征, 它与原始观测的误差分布无关:

1. 最小二乘估计是无偏的, 即

$$E(\hat{z}) = z \quad (1.4.4)$$

实际上, 我们注意到残差 ε 的期望值为零, 即 $E(\varepsilon) = 0$, 把它代入 (1.4.3) 式右端可得:

$$E(\hat{z}) = E[(A^T A)^{-1} A^T (Az + \varepsilon)] = z$$

2. 最小二乘估计的协方差矩阵:

$$\text{Cov}(\hat{z}) = E[(\hat{z} - z)(\hat{z} - z)^T]$$

把 (1.4.3) 代入 (1.4.4) 并考虑到 $E(\varepsilon) = 0$, 那么有

$$\begin{aligned} \text{Cov}(\hat{z}) &= E\{[(A^T A)^{-1} A^T (Az + \varepsilon) - z][(A^T A)^{-1} A^T (Az + \varepsilon) - z]^T\} \\ &= (A^T A)^{-1} A^T E(\varepsilon \varepsilon^T) A (A^T A)^{-1} \end{aligned}$$

由于我们认为残差 $\varepsilon_i = y_i - \varphi_i(m)$ 是独立具有零均值和同一方差 σ^2 , 它不取决于参数 z , 即

$$E(\varepsilon \varepsilon^T) = \sigma^2 I$$

因此有

$$\text{Cov}(\hat{z}) = \sigma^2 (A^T A)^{-1} \quad (1.4.5)$$

即最小二乘估计具有最小的协方差。

3. 如果观测资料的方差 σ^2 不知道, 那么可以按下式估计

$$\hat{\sigma}^2 = \frac{1}{m-n} (u - A\hat{z})^T (u - A\hat{z})$$

差值 $(u - A\hat{z})$ 称为残差矢量, $\hat{\sigma}^2$ 称为残余方差。这种估计是无偏的, 把它与 σ^2 一样地代入 (1.4.5) 式, 我们可以得到最小二乘估计的协方差矩阵的无偏估计。

在 1.3 中我们已经说过, 如果残差 ε 具有正态分布, 那么最小二乘估计与最大似然法估计是一致的。在多数地球物理问题中, 关于 ε 正态分布的假设可以认为是似然的, 这是广泛利用最小二乘法解地球物理反问题的基础。最小二乘法在这种情况下具有最大似然估计的特征: 可靠性和渐近有效性。

1.5 复变函数论与重磁异常反问题的多解性及等效解

在重磁勘探中, 通常把已知场源参数 (几何形状, 赋存的空间位置, 密度或磁化率等), 求位场的分布称为正问题; 而把给定了位场分布求场源的几何形状, 赋存的空间位置和密度或磁化率称为反问题, 正问题的解是唯一的, 而反问题在一般情况下没有单一解, 即理论上可以找到不同的场源分布建立相同的外部位场的例子。例如, 埋深相同, 密度和体积不一样, 但总质量一样的同心球; 磁化率和体积不一样, 但总磁矩一样的同心球, 它们都可以产生完全相同的重力异常或磁异常。此外, 还有象总质量或磁矩相同的共焦椭球, 同心水平圆柱体等等。形状稍为复杂一些的例子如图 1.5.1 所示, 这是上顶点埋深相同的一组不同三角形截面水平棱柱体, 在有限长观测剖面上, 它们可以建立完全相同的磁场。

理论上已经证明, 几何形状为凸形的星状体, 反问题的解是唯一的。而对于非凸形的或层状的模型, 反问题解的非唯一性就十分严重。许多具有实际意义的场源体, 反问题的解往往是不唯一的。尤其是观测值中包含有各种误差时 (仪器测量误差, 基点联测, 测点测量误差, 主要磁场或正常地磁场的计算误差等等), 反问题的解常常是不稳定的。即外部场有一个很微小的变化, 它都可以造成反问题解的极大波动。因此, 研究反问题解的非唯一性和寻找有实际意义的解是十分重要的。下面我们主要讨论反问题解的非唯一性和等效性。