计算物理丛书

最大熵方法

吴乃龙　袁素云　著

湖南科学技术出版社

ISBN 7-5357-0951-6

定价：6.50元

计算物理丛书

最大熵方法

吴乃龙　袁素云著

湖南科学技术出版社

湘新登字 004 号

内容提要

本书系统地介绍了最大熵方法的基本概念、基本原理及其在频谱分析、图像恢复、数学和物理等方面的应用,给出了详细的数学推导及大量的数值计算结果,还开列了许多有关的参考文献,并对最大熵方法研究的现状、疑难及展望作了中肯的评述.

本书既是最大熵方法的入门书,又包含许多新的研究成果,具有很好的可读性,可以作为物理和电子工程等有关专业的研究生教材,也可以作为有关专业的教师和科研人员的参考书.

计算物理丛书

最 大 熵 方 法

THE MAXIMUM ENTROPY METHOD

吴乃龙　袁素云著

责任编辑: 曾平安

*

湖南科学技术出版社出版发行

(长沙市展览馆路 3 号)

湖南省新华书店经销　国防科技大学印刷厂印刷

*

1991 年 12 月第 1 版第 1 次印刷

开本: 850×1168 毫米　1/32　印张: 11.375　字数:306 千字

印数:1–2000

ISBN 7–5357–0951–6

O · 93　　定价: 6.50 元

地科 86–041

总序言

计算物理学在中国的发展已有30年的历史. 这门学科首先为国防科研作出了重大贡献, 现已广泛应用于国民经济建设和基础科学研究. 我国的一大批计算物理科学家积累了丰富的实际经验, 急需及时地总结、提高、推广和交流.

继1982年成立中国核学会计算物理学会和1984年出版计算物理学报以来, 现在开始出版的计算物理学丛书是计算物理学界的又一重大里程碑.

这套丛书的对象是高年级大学生、研究生和广大的科技工作者, 内容首先反映中国人自己的成就, 并吸收国外的有用经验. 有理论, 有方法, 有算例: 既保证高水平的科学体系, 又注重解决实际问题, 以普及的形式表达专著的内容.

这套丛书的出版得到湖南科技出版社的大力支持. 对湖南科技出版社这种着眼于中华民族科技自立的崇高行为, 计算物理学工作者都高度赞赏, 并表示衷心的感谢.

对于参加这套丛书撰写的科学家、丛书编委会的编委、德高望重的名誉编委和编辑部的工作者所付出的辛勤劳动, 我代表中国核学会计算物理学会表示衷心的感谢.

祝计算物理学在为祖国的现代化作出贡献中茁壮成长!

秦元勋

1988年8月8日

前　言

本书是中国计算物理学会组织编写的《计算物理丛书》之中的一种. 写作本书主要根据下列材料: 作者的博士论文"The Maximum Entropy Method and Its Application in Radio Astronomy", 中国科学院、中国科学技术大学研究生院"最大熵方法"课程的讲稿, 以及作者在国内外发表的多篇论文和所作的多次学术报告.

从 1957 年 Jaynes 发表开创性论文算起, 最大熵方法经过了三十多年的发展, 已经成为一门理论基础坚实、应用广泛的学科. 在国外最大熵方法研究人员已经形成一支人数可观的队伍; 在国内也有不少人从事这方面的研究工作. 但更多的人, 仅对最大熵方法有所听闻, 或感兴趣而不甚了解, 想要学习而又苦于没有合适的教材. 在这种情况下, 写作一本最大熵方法的书, 就显得很有必要了.

本书的对象是物理和电子工程等专业的大学教师、研究生和科研人员, 并假定读者具有大学数学、物理和信号处理等方面的基础知识. 在写作本书时, 作者注意了系统性和可读性, 力图清楚地阐述最大熵方法的基本原理, 又尽量介绍新的研究成果. 本书提供大量的数值计算结果, 以帮助读者理解最大熵方法的基本理论. 本书开列了许多有关的参考文献, 为有志于对最大熵方法作进一步研究的读者提供方便.

最大熵方法是作者的博士后研究课题, 得到博士后

科学基金的资助，也是中国科学院理论物理研究所的开放课题，得到理论物理研究所的资助．世界实验室中国高等科学技术中心 VAX-8550 计算机房和中国科学院北京天文台 VAX-11／780 计算机房为数值计算和作图提供了许多方便．中国科学院理论物理研究所郝柏林教授对本书的写作给予了宝贵的支持和鼓励．大连理工大学王宏禹教授认真审阅了书稿并提出了许多有益的建议．作者在此一并表示感谢．

作　者

1990 年 6 月于北京

《计算物理丛书》编辑委员会

目　录

第一章　最大熵方法概论 ………………………………………… (1)
§1.1　什么是最大熵方法 ………………………………………… (1)
§1.2　熵的定义 ………………………………………… (4)
§1.3　为什么要用最大熵方法 ………………………………………… (6)
§1.4　最大熵方法研究现状和展望 ………………………………………… (9)
第二章　MEM1 及其在频谱分析中的应用 ………………………… (16)
§2.1　熵 H1 的定义和表达式 ………………………………………… (16)
§2.2　公式化和解 ………………………………………… (26)
§2.3　等价方法和信号模型 ………………………………………… (38)
§2.4　频谱分析算法和算例 (已知相关函数) ……………… (58)
§2.5　频谱分析算法和算例 (已知时间序列) ……………… (80)
§2.6　定阶准则 ………………………………………… (111)
第三章　MEM2 及其在图像恢复中的应用 ………………………… (129)
§3.1　熵 H2 的定义和表达式 ………………………………………… (129)
§3.2　公式化和隐式解 ………………………………………… (136)
§3.3　显式解 ………………………………………… (146)
§3.4　等价方法和信号模型 ………………………………………… (160)
§3.5　R-λ 算法 ………………………………………… (165)
§3.6　图像恢复算法和应用举例 (Ⅰ) ………………………… (174)
§3.7　图像恢复算法和应用举例 (Ⅱ) ………………………… (191)
附录 3.A　倒谱分析 ………………………………………… (208)
附录 3.B　图像恢复 ………………………………………… (215)

第四章　最大熵方法分析与比较 ………………………… (220)
§4.1　GMEM ……………………………………………… (220)
§4.2　熵的表达式 ………………………………………… (227)
§4.3　解的存在性、唯一性、一致性和统计性质 …… (231)
§4.4　分辨率增加和数据外推（实验结果） ………… (242)
§4.5　分辨率增加和数据外推（理论分析） ………… (256)
§4.6　谱峰位置和相对功率估计（实验结果） ……… (264)
§4.7　谱峰位置和相对功率估计（理论分析） ……… (273)
§4.8　对 MEM 三个学派的评述 ……………………… (283)
第五章　最大熵方法在数学和物理中的应用 ……………… (286)
§5.1　解矩问题 …………………………………………… (287)
§5.2　解积分问题 ………………………………………… (299)
§5.3　解偏微分方程 ……………………………………… (311)
§5.4　预测统计力学 ……………………………………… (321)
§5.5　粒子数按能级的统计分布 ………………………… (326)
§5.6　经典统计系综 ……………………………………… (333)
§5.7　量子统计系综 ……………………………………… (339)
结束语 …………………………………………………… (345)
参考文献 ………………………………………………… (346)

第一章　最大熵方法概论

本章描述最大熵方法的全貌，介绍基本原理而不涉及具体细节. 通过阅读这一章，可以首先了解最大熵方法的各个方面. 这是很有益处的. 在掌握基本原理之后，读者可以根据自己的需要，重点阅读后继几章中感兴趣的部分，解决自己的实际问题. 也只有这样，才能悟出最大熵的真谛而不仅仅是停留在一般性了解之上.

§1.1　什么是最大熵方法

在数学、物理、工程技术和其它领域中，常常要根据所测量的数据、所给的条件或所作的假设(以下统称为数据)求解问题. 对于解，通常关心三个问题：存在性、唯一性和稳定性. 如果这三个要求中至少有一个不满足，则该问题称为不适定问题(ill-posed problem 或 ill-conditioned problem). 造成这类问题的原因是数据不完全或有噪声，或既不完全又有噪声. (在这里，"噪声"是广义的. 测量数据与"真实"数据的任何偏差都称为噪声).

对付这类不适定问题，有各种各样的方法. 最大熵方法(the Maximum Entropy Method，缩写 MEM)只是其中的一种. 最大熵方法是说：在所有的可行(可能)解中，应该选择其熵最大的一个. 最大熵方法的表述，就是这么的简单. 这里需要说明几点. (1) 在对解的三个要求中，稳定性被暂时撇在一边. 如果要对它加以考虑，也只在用最大熵方法把解求出来之后才进行. (2) 如果解本来不存在，"选择其熵最大的一个"就是一句空话. 这种情况下的对策通常是：放松根据数据所构造的约束条件，保证解的存在性. 而约束条件一放松，解往往就有许多个甚至无穷多个. 于是，就可以根据"熵最大"的原则进行选择了. (3) 这里说的是全局的"熵最大"而不是局部的"熵极大". 具有最大熵的解是唯一的而具有极

大熵的解可能不是唯一的. 但是, 极大点往往是唯一的, 因而常常对"熵最大"和"熵极大"不加区分.

以上的论述也许有点太抽象了. 让我们来考虑一维(功率)频谱分析的例子(参见图 1.1.1).

图 1.1.1 中列举了三种情况.

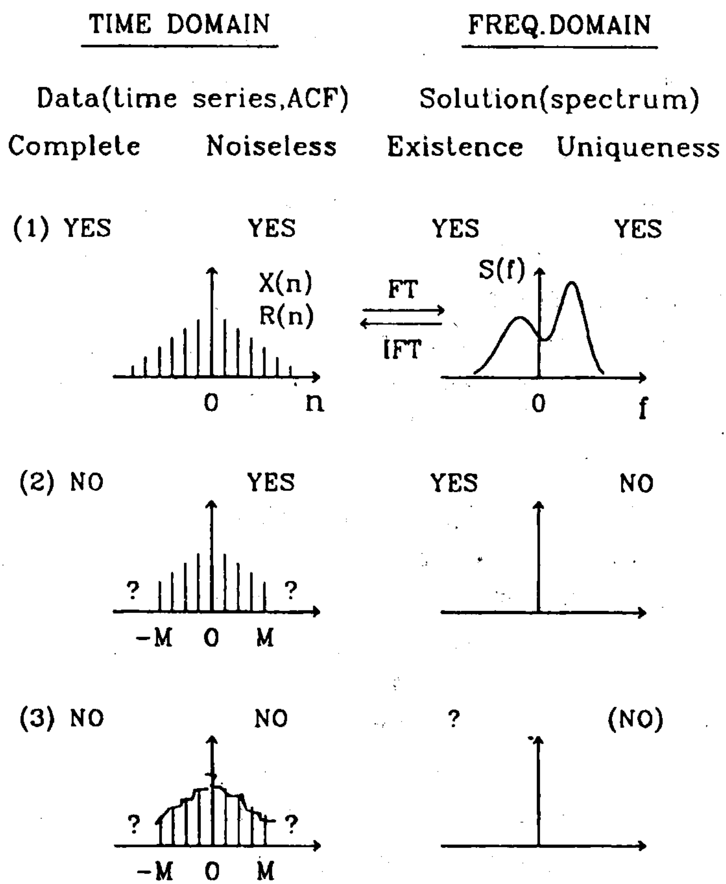

图 1.1.1 一维频谱分析中的不适定问题

(1) 时域的数据(时间序列或它的相关函数)完全而且无噪声. 这时频域的解(频谱)存在而且唯一. 数据和解通过付里叶变换(Fourier transform, 缩写 FT. IFT 即 inverse FT, 逆付里叶变换)相联系.

(2) 时域的数据不完全, 但是无噪声. 数据只在[−M,M]的区间上给定. 这时频域的解存在但不唯一. 在[−M,M]区间外不同的数据外推会导致不同的解. 一般地说来, 解有无穷多个.

(3) 时域的数据不完全而且有噪声. 即使在区间[−M,M]上给定的数据, 也包含有噪声而不精确. 这时频域的解可能不存在. 例如, 当噪声很强大时, 不管相关函数的外推如何, 用付里叶变换形式上求得的频谱不满足非负条件(即频谱值不能是负的); 唯一性更无从谈起. 即使噪声比较小, 非负解存在. 情况也不会比(2)好. 总而言之, 解的存在性不能保证, 而唯一性的要求肯定不能满足.

情况(1)是适定(well−posed)的数学问题, 不是最大熵方法的研究对象. 情况(2)和(3)是不适定问题, 正是最大熵方法的用武之地.

让我们首先考虑情况(2): 解存在但不唯一. 如前所述, 不同的数据外推会导致不同的解. 根据最大熵方法, 最直接了当的做法是: 对于每一个数据外推或每一个解, 计算它的熵. 将所有的熵的数值进行比较从而挑选出熵最大的数据外推或熵最大的解来. 这就是频谱分析中的最大熵方法. 但是, 这种简单的做法是不能付诸实用的, 因为要求进行无穷多次计算. 这就需要将这种方法表述为适当的数学问题. 简而言之, 就是把数据当作约束条件而求熵的最大值, 从而把解求出来. 还需要指出的是, 时域中熵最大的数据外推并不对应频域中熵最大的解. 这就导致最大熵方法中的两个学派 MEM1 和 MEM2. 这个问题将在第 4 节讨论.

再来看看情况(3): 解不一定存在, 唯一性要求不能满足. 这种情况下的对策已经讨论过了. 具体说来, 就是在求熵最大值的时候, 不要求约束条件在每个数据点上都得到满足, 而只要求在区间[−M,M]上总的来说, 数据拟合的偏差不超出一定范围就行了,

从而保证解的存在性.

如上所述，最大熵方法要在所有可行解中挑选熵最大的一个. 那么熵是怎样定义的呢? 为什么要挑选熵最大的解呢? 下面两节就要回答这些问题.

§1.2 熵的定义

熵(entropy)是一个历史颇长的概念. 十九世纪中叶，克劳修斯(Clausius)首先把熵引进热力学. 接着，玻耳兹曼(Boltzmann)推导出熵在统计力学中的表达式即著名的玻耳兹曼关系：$S=k\log W$. 热力学和统计力学中的熵，都是描述热力学系统的无序程度(参阅文献[1.1]).

熵又是一个近代的概念. 1948 年仙农(Shannan)在创立信息论时，找到一个唯一的量来量度信源的不确定性. 这个量与热力学和统计力学中的熵数学形式和物理意义都相近，所以也称为熵. 不过为了明确起见，信息论中的熵有时也称为"信息熵"或"仙农熵". 最大熵方法中的熵正是信息熵. 因此，有必要作比较详细的介绍(参阅文献[1.2]).

一个离散信源可以表示为

$$X:\begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ p_1 & p_2 & \cdots & p_n \end{bmatrix},$$

即随机变量X取值(符号) x_i的先验概率为p_i, $i=1,\cdots,n$. 这里 $P(X=x_i\bigcap X=x_j)=0$, $i\neq j$; $\sum\limits_{i=1}^{n}p_i=1$. 我们对信源只有部分知识即信源具有不确定性，就用先验概率分布$(p_1,\cdots,p_n)$来描述. 如果要求不确定性的测度H满足以下三个条件，则 H 可以确定到只相差一个常数. 这三个条件是:

(1) $H(p_1,\cdots,p_n)$是 p_i, $i=1,\cdots,n$的连续函数.

(2) 如果所有p_i相等，则$H(1/n,\cdots,1/n)$是n的单调递增函数.

(3) 合成法则(composition law). 如果把符号(也就是值)任意分成 m 组($x_{11},\cdots,x_{1k_1}$), ($x_{21},\cdots,x_{2k_2}$), $\cdots$, ($x_{m1},\cdots,x_{mk_m}$), 对应的概率为 $w_1 = p_{11} + \cdots + p_{1k_1}$, $\cdots$, $w_m = p_{m1} + \cdots + p_{mk_m}$, 那么必须有

$$\begin{aligned} H(p_1,\cdots,p_n) = H(w_1,\cdots,w_m) \\ + w_1 H\left(p_{11} \middle| w_1,\cdots, p_{1k_1} \middle| w_1\right) + \cdots \\ + w_m H\left(p_{m1} \middle| w_m,\cdots, p_{mk_m} \middle| w_m\right), \end{aligned}$$

(式中竖线表示条件概率). 合成法则又称为一致性(consistency)或可加性(additivity). 意思是说, 不确定性的测度 H 应该与试验步骤无关.

满足以上三个条件的测度 H 具有唯一的形式

$$H(p_1,\cdots,p_n) = -k\sum_{i=1}^{n} p_i \log p_i. \tag{1.2.1}$$

常数 k 决定于所选用的单位, 通常取$k = 1$. 选取不同的对数的底, 熵就有不同的单位. 在最大熵方法中, 无须考虑单位. 为了数学运算的方便, 对数的底总是取为 e. 所以, 除非特别标明, 有 $\log \equiv \log_e \equiv \ln$.

(1.2.1)式中的量 H 称为(信息)熵. 它描述信源的不确定性. 对它的一种解释是: 收到从信源传来的一条消息(一串符号)以后, 不确定性消除了. 熵就是平均起来, 收到一个符号所消除的不确定性, 或所获得的信息量.

以上讨论的是离散信源的情况. 在连续信源的情况下, X 的分布用概率密度$p(x)$(连续函数)来描述. 这时, 如果用离散信源来逼近, 即在(1.2.1)式中令 $p_i = p(x_i)\Delta x_i$, $n \to \infty$, $\Delta x_i \to 0$, $\sum \to \int$, 那么积分发散. 对付这个困难的通常做法是: 仿照(1.2.1)式而写出 X 具有连续分布情况下熵的表达式

$$H(p(x)) = -\int_{-\infty}^{+\infty} p(x)\log p(x)dx. \tag{1.2.2}$$

要注意的是, 这时的 H 并不表示信源的不确定性或所获得的信息

量的绝对数值，而只能用来计算它们在变化过程中的增减值. 另外一个困难是：在连续分布情况下的(1.2.2)式，并不象在离散分布情况下的(1.2.1)式对变量之间的一一对应变换不变. 克服这个困难的办法是引入一个测度$m(x)$:

$$H(p(x)) = -\int_{-\infty}^{+\infty} p(x)\log\left(\frac{p(x)}{m(x)}\right)dx. \tag{1.2.3}$$

(式中$m(x)$不一定要归一化.) 这样，当$p(x)$变化时，$m(x)$依同一方式变化而保持 H 的值不变，以后将会看到，在最大熵方法中，$m(x)$是所谓的对解的“先验知识”. (1.2.1)式亦可改造成

$$H(p_1,\cdots,p_n) = -\sum_{i=1}^{n} p_i\log\left(\frac{p_i}{m_i}\right). \tag{1.2.4}$$

(1.2.3)式的不变性证明如下.

设$x=f(y)$是 x 和 y 的一一对应变换. y 的概率密度函数$q(y) = p(x)\left|f'_y\right|$，对测度亦有$n(y)=m(x)\left|f'_y\right|$，因此有

$$\begin{aligned} H(p(x)) &= -\int_{-\infty}^{+\infty} q(y)\log\left(\frac{p(x)\left|f'_y\right|}{m(x)\left|f'_y\right|}\right)\mathrm{d}y \\ &= -\int_{-\infty}^{+\infty} q(y)\log\left(\frac{q(y)}{n(y)}\right)\mathrm{d}y \\ &= H(q(y)). \end{aligned}$$

顺便指出，用随机向量 $\mathbf{x}$ 代替(1.2.3)中的 x，H 的不变性仍然成立. 这时 $\mathbf{y}$ 也是随机向量，$\mathbf{x}=\mathbf{f}(\mathbf{y})$. 在上面的证明中只须将变换 $\mathbf{f}$ 的雅可比行列式的模$|J|$代替$\left|f'_y\right|$就行了.

以上讨论的是随机变量 X 的熵，所以也称为概率熵. 在最大熵方法中，X 可以是概率场. p_i 也可以是频率，或比例，或所谓的“相信程度”(degree of belief). 在求最大值时，不一定用熵的原始定义(1.2.1)–(1.2.4)而用熵的导出表达式.

§ 1.3 为什么要用最大熵方法

为什么要在所有可行解中，选择熵最大的一个呢？如在第一

节中所述，对付不适定问题，有各种各样的方法. 用“熵最大”作为选择解的判据，有什么好处？对于这个问题，有各种各样的回答. 归纳起来，有如下五点.

(1) 熵增加原理

熵增加原理指出，一个孤立系统的熵永远不会减少，即趋于最大. 所以，最大熵方法对解所作的选择是“合乎自然”的. 当然，如果已知某一特定的系统(不可能是孤立系统)在演变过程中熵在减少，就应该选择熵最小的解.

(2) 第一原理

“第一原理”的意思是：在数据不充分的情况下求解，解必须和已知的数据相吻合，而又必须对未知的部分作最少的假定，即对数据的外推或内插采取最超然的(maximally noncommittal)态度. 求解可以认为是从数据中提取信息的过程. 信息来自两个部分：一是已知数据，二是由于数据不完全而不得不对未知部分所作的假定. 这种假定相当于人为地“添加”了一些信息(不管是真是假). 熵最大就意味着获得的总信息量最少，也就是“添加”的信息最少. 所以，最大熵的解是最超然的.

(3) 最大多重性原理

在统计力学和图像处理等领域中，状态的熵 H，状态的多重性 W 和系统的粒子或像素的数目 N 有如下关系：

$$H \sim N^{-1}\log W,$$

($a \sim b$ 表示当 $N \to \infty$ 时，$\frac{a}{b} \to 1$). 状态的多重性(multiplicity)是指系统演变成这种状态的可能的途径的数目. 状态的多重性越大，则系统最终处于这种状态的可能性也就越大. 显然，最可能的状态具有最大多重性. 这就是最大多重性原理. 由上式有

$$W \sim e^{NH}.$$

显然，H 最大对应 W 最大. 因此，最大熵方法是最大多重性原理的必然结果. 当 N 很大时，H 对 $H_{\max}$ 的微小偏差，将引起 W 值极