

数学物理逆问题中测量信息的熵描述

李静, 黄卡玛*, 赵翔

四川大学电子信息学院, 成都 610064

* 联系人, E-mail: kmhuang@scu.edu.cn

收稿日期: 2007-03-06; 接受日期: 2007-07-08

国家自然科学基金重点项目资助(批准号: 60531010)

摘要 逆问题通常分为两大类: 优化设计和参数识辨. 在讨论数学物理逆问题的参数识辨前, 测量点的数目和位置的选取是对大量数据的分析和评估中两大至关重要的问题. 针对参数识辨中的这两大问题, 应用贝叶斯理论, 首次将信息熵引入到测量信息的描述中, 提出了描述测量点的数目和位置对重构精度影响的数学方法和手段. 最后通过具体的计算实例说明了该方法的有效性.

关键词

逆问题
信息熵
贝叶斯方法
评价因子
信息增量

逆问题是由果求因的问题. 从应用的角度, 我们可以将逆问题分为两大类: 优化设计问题和参数识辨问题^[1]. 优化设计问题是工程设计中的重要课题; 而参数识辨问题在地质勘探、无损检测和医学成像等工程技术领域有着更为重要的应用价值. 本文讨论的就是参数识辨中测量信息的描述问题.

在逆问题的解决过程中, 当得到测量信息之后, 通常需要先对大量数据筛选, 即进行参数识辨前大量数据的分析和评估, 以减小重构计算量的压力. 在参数识辨问题的研究中, 已有的大量文献主要就重构方法进行了讨论^[2,3], 而忽略了对参数识辨前大量数据的分析和评估. 这里我们以微波生物医学成像为例说明参数识辨前分析和评估的重要意义. 如图 1 所示, Ω_1 是包含被测生物体的计算区域, Ω_2 是均匀的背景媒质区域. 其原理是用微波照射被测生物体, 然后通过生物体外部区域 Ω_2 中的散射场的测量值来重构 Ω_1 内部的复介电常数 ϵ^* 的分布.

一般人们会认为, 为了精确重构 Ω_1 内部的复介电常数 ϵ^* 的分布, 所需区域 Ω_2 内散射场的测量点应该越多越好, 因此采集的 Ω_2 内散射场的测量数据将会非常多. 实际上, 测量点的增加会急剧加重重构计算的资源消耗和成本, 同时, 由于在

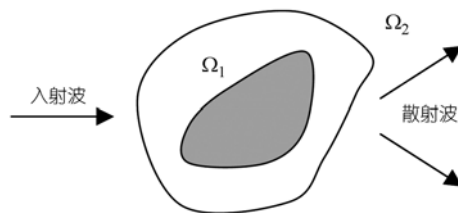


图 1 微波成像示意图

实际测量中总会有一定的误差, 测量点太多也会导致重构计算中误差的积累. 因此, 在数学物理逆问题参数识辨前的分析和评估中有两个问题必须得以解决: (i) 用多少测量点上的散射场值能够精确重构 Ω_1 内部的复介电常数分布; (ii) 由于测量点是有限的, 那么, 在什么区域上的测量值更有价值. 到目前为止, 我们还缺乏数学手段和方法来描述这些问题.

与以往文献中熵的用法^[4,5]相比, 本文首次提出将熵用在逆问题参数识辨前的分析和评估中, 以解决参数识辨中测量点数和测量位置的选取问题. 本文首先结合贝叶斯方法^[6~8], 提出采用待识辨参数分布的后验概率来定义参数样本出现的概率, 再计算出参数对应的最大熵, 求出包含不同测量点数的测量组对应的评价因子, 通过评价因子即可确定我们所需的测量点数, 解决分析和评估中的第一大问题; 其次, 利用前面得到的最大熵, 用信息熵的方法确定包含在不同测量区域中的信息增量, 即表示出不同位置测量区域的“贡献”大小, 解决分析评估中的第二大问题. 不失一般性, 本文在最后引入一实例对参数识辨遇到的上述两大问题进行说明. 计算结果表明: 我们可以根据一定的评价因子确定重构所需的测量点数, 根据信息增量确定实际测量区域对参数识辨的贡献大小.

1 测量区域和测量点数的选取及贝叶斯方法

1.1 信息熵的定义

在信息论中, 信息熵就是随机事件不确定性大小的度量^[9]. 设逆问题的模型用算子方程表示为 $Af = y$, f 是待识辨参数, y 是测量项, 它关于 f 的有限次测量值是 $y_1, y_2, \dots, y_m$, A 是将 f 映射到测量值 $y_1, y_2, \dots, y_m$ 的正向算子. 如果设随机事件 f 具有多种可能的结果 $f_1, f_2, \dots, f_n$, 每种结果出现的概率为 $P_1, P_2, \dots, P_n$, 则该随机事件对应的信息熵定义为

$$H(f) = -\sum_{i=1}^n P_i \ln P_i = \sum_{i=1}^n P_i \ln \frac{1}{P_i}. \quad (1)$$

如果设参数 f 在获得某一个测量信息 Y 之前的熵为 $H(f)$, 得到测量信息 Y 之后熵应减小, 设为 $H_Y(f)$, 那么, 包含在测量信息 Y 中关于参数 f 的信息增量为

$$I(Y, f) = H(f) - H_Y(f), \quad (2)$$

但是, 如何量化测量信息 Y 对 $H_Y(f)$ 中 P_i 的影响呢? 根据贝叶斯方法, 本文采用随机事件的后验概率来定义 $H_Y(f)$ 中的 P_i ^[10]. 这样, 就可以得到参数 f 在已知每个测量区域的测量值后的信息熵, $I(Y, f)$ 越大的测量区域对参数识辨的贡献也就越大.

1.2 贝叶斯方法和最大熵

贝叶斯方法所考虑的先验概率常常伴有主观成分, 先验信息包括主观性先验和客观性先验, 其中, 主观性先验有专家的经验 and 意见等, 客观性先验有相关的历史数据. 根据贝叶斯假设^[6], 当无任何先验信息时, 参数 f 可能的结果 $f_1, f_2, \dots, f_N$ 对应的先验概率 $\pi_i(f)$ ($i=1, 2, \dots, N$)应在 f 的取值范围 D 内是“均匀”分布的, 所以 $\pi_i(f)$ 可写成

$$\pi_i(f) = \begin{cases} \frac{1}{N}, & \text{当 } f \in D, \\ 0, & \text{当 } f \notin D, \end{cases} \quad i = 1, 2, \dots, N, \quad (3)$$

其中 N 为参数 f 所有的样本数. 我们令 $H_{\max}(f)$ 为待识辨参数对应的最大熵(无任何先验信息时的熵), 则最大熵 $H_{\max}(f)$ 应为

$$H_{\max}(f) = \sum_{i=1}^N \pi_i(f) \ln \frac{1}{\pi_i(f)} = \sum_{i=1}^N \frac{1}{N} \ln \frac{1}{\frac{1}{N}} = \ln N. \quad (4)$$

贝叶斯方法的基本思想是将包含先验信息的先验概率和包含测量信息的似然概率进行融合, 得到问题的后验概率, 利用测量信息对先验信息进行更新和修正以获得对问题的最终认识.

根据贝叶斯方法的原理, 定义 f 为待识辨参数, $y_1, \dots, y_m$ 为已知测量值, 有贝叶斯公式:

$$\pi_i(f|y_1, \dots, y_m) = \frac{\pi_i(f)\pi_i(y_1, \dots, y_m|f)}{\sum_{i=1}^N \pi_i(f)\pi_i(y_1, \dots, y_m|f)}, \quad (5)$$

此时, 上式右端的分母是一个常数, 于是 $\pi_i(f|y_1, \dots, y_m)$ 正比于分子, 即

$$\pi_i(f|y_1, \dots, y_m) \propto \pi_i(f)\pi_i(y_1, \dots, y_m|f), \quad (6)$$

其中 $\pi_i(f)$ 是 f 的先验概率, 反映了 f 的先验信息; $y_1, \dots, y_m$ 是试验发生的事件, 是测量信息; $\pi_i(y_1, \dots, y_m|f)$ 为包含了测量信息的似然概率; $\pi_i(f|y_1, \dots, y_m)$ 是 f 的后验概率, 即(1)式中的 P_i , 它综合了过去提供的经验——参数 f 的先验信息和 $y_1, \dots, y_m$ 提供的新信息——测量信息, 反映了参数 f 的综合信息, 它定义在整个解空间, 表示了问题的“完全”解, 即由所有可能出现的解组成的解集中任意一个解出现的概率.

当无任何先验信息时, 如果略去(3)式中概率取 0 的部分, (3)式可写成

$$\pi_i(f) = \frac{1}{N}, \quad \text{当 } f \in D, \quad (7)$$

将(6)与(7)式结合, (6)式可简化为

$$\pi_i(f|y_1, \dots, y_m) \propto \pi_i(y_1, \dots, y_m|f), \quad (8)$$

此时, 已知 m 个测量点的测量值后参数 f 对应的信息熵为

$$H_m(f) = \sum_{i=1}^N \pi_i(y_1, y_2, \dots, y_m|f) \ln \left(\frac{1}{\pi_i(y_1, y_2, \dots, y_m|f)} \right). \quad (9)$$

为了描述似然概率的本质, 我们令

$$\pi_i(y_1, \dots, y_m|f) \propto \exp(-\tau_i), \quad (10)$$

而误差 τ_i 可由算子方程左右两端差值的范数的平方给出^[10]:

$$\tau_i = \|Af_i - y\|^2, \quad (11)$$

将(10)与(11)式结合, 代入(9)式, 有

$$H_m(f) = \sum_{i=1}^N \exp(-\tau_i) \ln \left(\frac{1}{\exp(-\tau_i)} \right), \quad (12)$$

该式即为逆问题中已知 m 个测量点的测量值后参数 f 的信息熵.

通常对于实际问题的分析评估, 必须要对测量点的数目和位置给出统一的描述方法. 我们令 m 个相邻的测量点为一测量区域, 当 m 相同而测量位置不同时, $H_m(f)$ 也会有所不同, 将 $H_{\max}(f)$ 和 $H_m(f)$ 代入(2)式, 我们就可以得到包含在不同测量区域中关于参数 f 的信息增量:

$$I = H_{\max}(f) - H_m(f), \quad (13)$$

通过它便能判别不同测量区域对参数识辨的贡献大小. $H_m(f)$ 越小, I 越大, 则该测量区域对参数 f 识辨的贡献也就越大. 接着, 我们定义

$$\eta = -\ln \left| \frac{H_{\max}(f) - H_m(f)}{H_{\max}(f)} \right| \quad (14)$$

为已知 m 个测量点的测量值后的评价因子, 根据评价因子可以判断重构实际所需的测量点数.

通常情况下, 由于待识辨参数 f 的样本空间 N 的值都很大, 我们不能对所有样本进行逐一计算, 所以我们采用马尔可夫链蒙特卡罗法——Gibbs 抽样器对总体样本进行抽样, 产生一组服从后验概率 $\pi_i(f|y_1, \dots, y_m)$ 的样本. 下面用一个具体实例进行说明.

2 实例的讨论

这里我们采用数值实验, 利用一个一维的参数识辨模型来说明参数识辨前测量点的数目和测量位置的选取方法. 本例的模型为

$$\begin{cases} \frac{d^2 y(x)}{dx^2} + f(x) \cdot y(x) = 0, & x \in [-1, 1], \\ y(x) = 1.71653, & x = \pm 1, \end{cases} \quad (15)$$

$f(x)$ 为待识辨参数, $y(x)$ 为测量项.

当已知 $f(x)$ 的解析式为

$$f(x) = \cos(x) - \sin^2(x), \quad x \in [-1, 1] \quad (16)$$

时, 结合(15)式我们可以得到 $y(x)$ 的准确测量值分布, 如表 1 所示.

为了验证前面关于逆问题参数识辨前测量信息的熵描述, 下面我们假设 $f(x)$ 的解析式未知. 已知表 1 中 $y(x)$ 的 7 组测量值, 将用有限差分法求得的 $f(x)$ 的解 $f^c(x)$ 与其解析解比较, 得到参数 $f(x)$ 的重构误差:

$$\delta = \left| \frac{f(x) - f^c(x)}{f(x)} \right| \times 100\%, \quad (17)$$

再将重构误差和用熵描述的评价因子进行对比, 从而说明用评价因子描述实际所需测量点数

的有效性.

表 1 包含不同测量点数的测量组

x_i	$y(x)$						
	3 个测量点	4 个测量点	5 个测量点	6 个测量点	7 个测量点	8 个测量点	9 个测量点
$x_1 = -1$	-	1.71653	1.71653	1.71653	1.71653	1.71653	1.71653
$x_2 = -3/4$	2.40508	-	-	2.07859	2.07859	2.07859	2.07859
$x_3 = -1/2$	-	2.40508	2.40508	2.40508	2.40508	2.40508	2.40508
$x_4 = -1/4$	-	-	-	-	-	2.63508	2.63508
$x_5 = 0$	2.71828	2.71828	2.71828	2.71828	2.71828	2.71828	2.71828
$x_6 = 1/4$	-	-	-	-	2.63508	-	2.63508
$x_7 = 1/2$	-	2.40508	2.40508	2.40508	2.40508	2.40508	2.40508
$x_8 = 3/4$	2.40508	-	-	-	-	2.07859	2.07859
$x_9 = 1$	-	-	1.71653	1.71653	1.71653	1.71653	1.71653

为了方便计算, 我们先将 $f(x)$ 沿 x 轴在 $[-1, 1]$ 区间均匀离散, 离散间隔 $\Delta x = 1/4$, 则离散后的分量个数 $k = 9$. 根据先验信息, $f(x)$ 的取值范围 $D = [-1/5, 1]$, 再将 $f(x)$ 在 $[-1/5, 1]$ 范围内等间隔离散, 离散间隔 $\Delta f = 2/15$, 则离散后的分量个数 $l = 10$. 所以, 离散后 $f(x)$ 的总样本数 $N = l^k = 10^9$. 此外, 根据离散间隔 Δx , 将 3 个相邻的点划为一个测量区域, 则在 $[-1, 1]$ 内共有 7 个测量区域.

本例中, 按 0.2% 的概率选取的随机抽取样本数 $M = 2 \times 10^6$, 根据(4)式, 待识辨参数对应的最大熵 $H_{\max}(f) = \ln M = \ln 2 \times 10^6 \approx 14.50866$. 然后运用有限差分对 $\frac{d^2 y(x)}{dx^2}$ 展开, 再将(15)式代入(11)式, 此时 τ_i 即为

$$\tau_i = \left\| \frac{y(x_j + 2\Delta x) - 2y(x_j) + y(x_j - 2\Delta x)}{(2\Delta x)^2} + f_i(x_j) \cdot y(x_j) \right\|^2, i = 1, \dots, M; j = 1, \dots, k, \quad (18)$$

其中 k 是 $f(x)$ 沿 x 轴在 $[-1, 1]$ 区间均匀离散后的分量个数.

对于表 1 中包含不同测量点数的测量组, 分别将 $f(x)$ 的 M 个样本代入(18)式求得 M 个 τ 值, 结合信息熵的计算式(12)和评价因子的计算公式(14), 将 τ_i 代入, 便得到包含不同测量点数的测量组对应的评价因子, 结果如图 2 所示. 由于已知测量信息是离散的, 图 2 中及后面出现的重构误差均是在 $y(x)$ 相应测量点上 $f(x)$ 重构误差的平均.

通过图 2 我们可以看出, 评价因子能够较好地反映重构误差的变化规律. 结合表 1 和图 2, 不难得到这样的结论: 在不考虑测量误差的情况下, 已知的测量点越多, 其对应的重构误差越低, 评价因子越低. 由图 2 还可以直观地看到, 要实现 5% 以下的重构误差, 即达到小于 4.5 的评价因子, 我们只需要 6 个测量数据即可, 相对 9 个测量数据可以节省 33.3% 的资源消耗和计算成本. 这就说明了用评价因子可以对逆问题参数识辨前分析和评估中的测量点数进行描述.

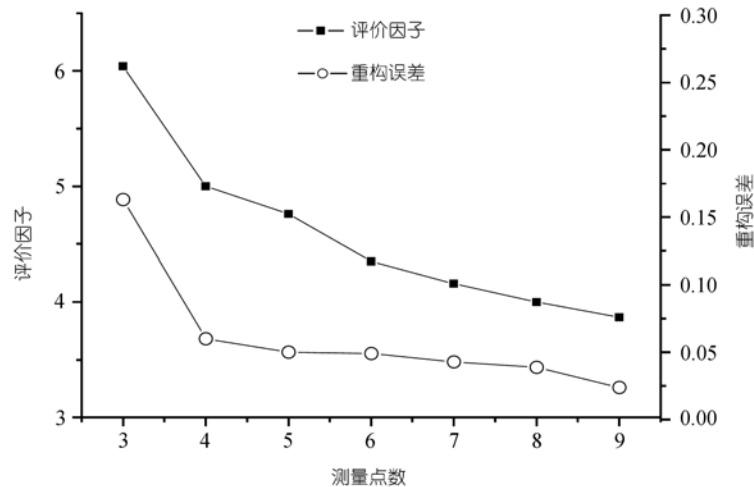


图2 不同测量点数对应的评价因子和重构误差

对于之前划分好的7个测量区域,使用与7个测量组相同的信息熵的计算方法,分别得到已知7个不同测量区域的测量值后参数 f 的信息熵,再将得到的信息熵代入信息增量 I 的计算公式(13),即可得到包含在每个测量区域中关于参数 f 的信息增量,具体结果见表2.

表2 7个不同测量区域对应的信息增量和重构误差

测量区域	第一区域 (x_1, x_2, x_3)	第二区域 (x_2, x_3, x_4)	第三区域 (x_3, x_4, x_5)	第四区域 (x_4, x_5, x_6)	第五区域 (x_5, x_6, x_7)	第六区域 (x_6, x_7, x_8)	第七区域 (x_7, x_8, x_9)
I	0.04124	0.17631	0.19164	0.27315	0.19164	0.17631	0.04124
Δ	2.33%	8.40%	84.65%	92.58%	84.65%	8.40%	2.33%

Δ 是在表1中的9个测量点的基础上去掉表2中相应测量区域的测量点后的重构误差,该误差越小,说明该测量区域对重构影响越小,即对参数识辨贡献越小.通过表2我们可以看出,对于测量点数相同但测量位置不同的测量区域,包含在该区域的信息增量有一定的差异,根据重构误差 Δ 比较各测量区域的信息增量,我们就可以判别不同测量区域的贡献大小.一般情况下,包含在测量区域中的信息增量越大,则该区域对参数识辨的贡献就越大.就本例来说,由 x_4, x_5, x_6 组成的第四个测量区域对参数识辨贡献最大,由 x_1, x_2, x_3 组成的第一个测量区域和由 x_7, x_8, x_9 组成的第七个测量区域对参数识辨的贡献最小.此时,说明了用信息增量可以对不同测量区域的贡献大小进行有效地描述.

3 结论

在数学物理逆问题的工程应用中,测量点的数目和位置的选取是在参数识辨前对大量数据的分析和评估中两大至关重要的问题.本文将信息熵引入到了数学物理逆问题中,结合贝叶斯理论,将待识辨参数的先验信息描述为先验概率,测量信息表示为似然概率,二者共同得到包含待识辨参数综合信息的后验概率.结合后验概率和熵的计算公式,得到已知不同测量

信息时参数的信息熵,最后利用最大熵和参数的信息熵可以推出不同测量组对应的评价因子和包含在不同测量区域的信息增量,根据这两种方法即可选取合适的测量点数和测量位置.本文最后通过具体的计算实例表明:用信息熵表示的评价因子能够较好地反映重构误差的变化规律,可以用来描述参数识辨中实际所需的测量点数;另外,将信息熵引入逆问题的参数识辨中,还能够有效地描述不同测量区域对参数识辨的贡献大小.该项研究对逆问题的工程研究具有重要的指导意义.

参考文献

- 1 黄卡玛,赵翔. 电磁场中的逆问题及应用. 北京: 科学出版社, 2005. 13—17
- 2 Dudley D G, Habashy T M, Wolf E. Linear inverse problems in wave motion: Nonsymmetric first-kind integral equations. *IEEE Trans Ant Prop*, 2000, 48(10): 1607—1617 [\[DOI\]](#)
- 3 Serinagaoglu Y, Brooks D H, MacLeod R S. Bayesian solutions and performance analysis in bioelectric inverse problems. *IEEE Trans Biomed Eng*, 2005, 52(6): 1136—1139 [\[DOI\]](#)
- 4 Rebollo N L, Fernandez R J. The continuous wavelet transform as a maximum entropy solution of the corresponding inverse problem. *IEEE Trans Signal Proc*, 1999, 47(7): 2046—2050 [\[DOI\]](#)
- 5 Chen X A, Zhang W X. Solution of inverse problems by the general maximum entropy method. *Ant Prop Soc Int Symp*, 1991, 2: 1308—1311
- 6 蔡洪,张士峰,张金槐. Bayes 试验分析与评估. 长沙: 国防科技大学出版社, 2004. 1—15
- 7 张尧庭,陈汉峰. 贝叶斯统计推断. 北京: 科学出版社, 1991. 17—27
- 8 Fitzpatrick B G. Bayesian analysis in inverse problems. *Inverse probl*, 1991, 7(5): 675 [\[DOI\]](#)
- 9 王彬. 熵与信息. 西安: 西北工业大学出版社, 1994. 121—150
- 10 Mohammad D A. Joint estimation of parameters and hyperparameters in a Bayesian approach of solving inverse problems. In: *Proceedings of IEEE International Conference on Image Processing*. Genoa: IEEE, 1996, 2: 473—476