

非晶态物质径向分布函数 测定的最大熵方法

魏 文 锋

(中国科学院金属研究所)

1984年6月20日收到

提 要

本文导出一种计算非晶态物质径向分布函数的最大熵方法,给出非晶态 D_2O 径向分布函数的计算实例,并与传统的傅里叶变换法进行了比较. 结果表明,最大熵方法对结构信息的反映更为敏感,因而,在同样实验数据的条件下,最大熵方法可能比传统的傅里叶变换法提供更多的结构信息.

一、引 言

傅里叶变换是径向分布函数测定的数学基础. 若已知非晶态物质在整个倒空间的相干散射强度分布,则通过傅里叶变换,可获得唯一确定的径向分布函数. 然而,通常的衍射实验只能给出部分倒空间上的强度分布,因而,傅里叶变换的结果是真实结构与截断函数的傅里叶变换的卷积,通常会出现一些伪结构. 通过适当地选择截断函数虽然能减少或消除伪结构,但要以牺牲分辨率为代价. 从统计估测的观点来看,用传统的傅里叶变换法处理截断的衍射数据,意味着人为地假定在实验测量不到的倒空间上相干强度处处为零. 这种假设显然与事实不符. 这种使最终结果依赖于对未知信息的主观假设的做法,违背了数据变换的 Jaynes^[1] 原则. 这个原则指出,对实验数据施行任何变换的结果,都必须与全部已知信息相符合,同时最大限度地与未知信息无关. 本文是应用最大熵方法,在径向分布函数计算问题上实现 Jaynes 原则的一次尝试. 目的在于通过计算实例,探讨这种方法在非晶态物质径向分布函数分析中的潜力及可能的应用前景. 对于最大熵原理及其在功率谱估测、图像处理、光学变换等领域中的应用有兴趣的读者可参阅文献[1—6].

二、最大熵原理

假定所研究的系统中,可能出现 M 个不同的事件, $m_j (j = 1, 2, \dots, M)$, 每个事件出现的概率分别记为 p_1, p_2, \dots, p_M . 假设函数 $f_k(m_j)$ 的期望值 F_k 已知 ($k = 1, 2, \dots, L, L < M$). 根据 Jaynes 原则,我们求概率分布 $\{p_j\}$, 使其满足给定条件,即

$$\sum_{j=1}^M p_j f_k(m_j) = F_k \quad (1)$$

以及

$$\sum_{j=1}^M p_j = 1, \quad (2)$$

同时最大限度地与未知信息无关,即不确定性最大.

Shannon^[2] 指出,对满足已知信息的离散概率分布的不确定性,存在唯一的、明确的度量. 这种度量称为概率分布的熵,记为

$$H(p_1, p_2, \dots, p_M) \equiv -K \sum_{j=1}^M p_j \ln p_j, \quad (3)$$

其中, K 为大于零的常数.

显然,(3)式定义的熵是非负的,其值随概率分布的不确定性增加而增加,并且有加和性.

不难看出,对于完全确定的系统,即 $\{p_j\}$ 中仅有其中一个值为 1,其余皆为零时,熵值为零. 在其它情况下,熵值为正. 当 M 个事件以等概率出现时,熵值最大,此时系统的不确定性最大.

由上述讨论可知,满足 Jaynes 原则的 $\{p_j\}$ 就是满足全部已知信息,同时熵值最大的 $\{p_j\}$.

用拉格朗日乘法,可求出 $\{p_j\}$ 为

$$p_j = e^{\left\{ -1 - \mu_0 - \sum_{k=1}^L \mu_k f_k(m_j) \right\}}, \quad (4)$$

其中 $\mu_0, \mu_1, \dots, \mu_L$ 为拉格朗日乘子.

(4)式表明,除其它因素外, p_j 是 k 的函数. 在极端情况下,即当对所研究的系统一无所知时, $p_j = 1/M$. 也就是说,当已知信息量为零时,最大熵方法给出的解为均匀分布. 随着信息量的增加 (F_k 的数目增加), $\{p_j\}$ 一般也随之变化. 这就是说,最大熵方法可以给出与已知信息相适应的、同时又最大限度地与未知信息无关的解,从而避免了传统傅里叶变换法对未知信息的依赖性,因此,可以认为最大熵方法是一种客观的方法.

三、径向分布函数的最大熵算法

上节仅限于一般化的讨论. 下面我们考虑径向分布函数计算这个具体问题. 物质结构可以用散射元在正空间的密度分布来描述. 这里我们仅着眼于几何关系,因而散射元的物理本质对讨论问题并不重要,本文中散射元是指原子. 设总共有 N 个散射元以概率 $p_j (j = 1, 2, \dots, W)$ 分布在 W 个等体积的体积元内. 由于 N 是个大数,可以认为各体积元内所包含的散射元数目为 $n_j = Np_j$. 此时,(3)式可改写为如下等价形式:

$$H \equiv -K \sum_{j=1}^W Np_j \ln p_j. \quad (5)$$

仅仅具有短程序的非晶态物质,其结构可以用径向分布函数来描述. 以任一散射元为中心,将正空间分为 M 个体积为 $4\pi r_j^2 \Delta r$ 的球壳,径向分布函数给出 N 个散射元在 M 个球壳内的分布. 在这种情形下,(5)式可化为

$$H \equiv -K \sum_{j=1}^M 4\pi r_j^2 \rho(r_j) \ln \rho(r_j), \quad (6)$$

其中 $\rho(r_j)$ 是第 j 层球壳中的散射元密度。

设样品的平均密度为 ρ_0 , 则 $(\rho(r_j) - \rho_0)$ 表示第 j 层球壳中实际密度对平均密度的偏离。当 r_j 足够大时, 对非晶态物质, 可以认为 $\rho(r_j) = \rho_0$, 即

$$\sum_{j=1}^M 4\pi r_j^2 (\rho(r_j) - \rho_0) = 0. \quad (7)$$

设实验观测的干涉函数为 $F(s_k)$, ($k = 1, 2, \dots, L$), $\rho(r_j)$ 与 $F(s_k)$ 的关系可写为

$$\sum_{j=1}^M 4\pi r_j^2 (\rho(r_j) - \rho_0) \sin(s_k \cdot r_j) / (s_k \cdot r_j) = F(s_k). \quad (8)$$

以(7)和(8)式为约束, 求 $\rho(r_j)$, 使(6)式达最大。应用拉格朗日乘数法有

$$E = - \sum_{j=1}^M 4\pi r_j^2 \rho(r_j) \ln \rho(r_j) - \mu_0 \sum_{j=1}^M 4\pi r_j^2 (\rho(r_j) - \rho_0) - \sum_{k=1}^L \mu_k \left[\left(\sum_{j=1}^M 4\pi r_j^2 (\rho(r_j) - \rho_0) \sin(s_k \cdot r_j) / (s_k \cdot r_j) \right) - F(s_k) \right]. \quad (9)$$

其中为方便起见, 将(6)式中常数 k 取为 1。

令 $\frac{\partial E}{\partial \rho(r_j)} = 0$, 可得

$$\rho(r_j) = \exp \left\{ -1 - \mu_0 - \sum_{k=1}^L \mu_k \sin(s_k r_j) / (s_k r_j) \right\}. \quad (10)$$

(10)式中共有 $(L + 1)$ 个待定系数 $\mu_0, \mu_1, \dots, \mu_L$, 可通过求解 $(L + 1)$ 个约束方程得到。只需将(10)式代入(7)和(8)两式, 便构成 $(L + 1)$ 个约束方程:

$$\sum_{j=1}^M 4\pi r_j^2 \left(e^{\left\{ -1 - \mu_0 - \sum_{k=1}^L \mu_k \sin(s_k r_j) / (s_k r_j) \right\}} - \rho_0 \right) = 0 \quad (11)$$

以及

$$\sum_{j=1}^M 4\pi r_j^2 \left(e^{\left\{ -1 - \mu_0 - \sum_{k=1}^L \mu_k \sin(s_k r_j) / (s_k r_j) \right\}} - \rho_0 \right) \sin(s_k r_j) / (s_k r_j) = F(s_k) \quad k = 1, 2, \dots, L. \quad (12)$$

令 $s_0 = 0$, 则 $F(s_0) = 0$, 将(11)和(12)两式合并, 有

$$\sum_{j=1}^M 4\pi r_j^2 \left(e^{\left\{ -1 - \sum_{k=0}^L \mu_k \sin(s_k r_j) / (s_k r_j) \right\}} - \rho_0 \right) \sin(s_k r_j) / (s_k r_j) = F(s_k) \quad k = 0, 1, \dots, L. \quad (13)$$

有多种算法可用来解方程组(13)。我们发现, 改进的牛顿法^[7]是一种稳定而且有效的解法。若选密度为 ρ_0 的均匀分布为初始态, 即取迭代初值 $\mu_0 = -1 - \ln \rho_0$ 和 $\mu_1, \mu_2, \dots, \mu_L = 0$, 通常经 5 次迭代可获满意的解。

作为最大熵方法的计算实例, 我们采用 Chowdhury 等人^[8]的中子散射数据, 计算了非晶态 D_2O 的简约分布函数 $G(r)$ ($G(r) \equiv 4\pi r(\rho(r) - \rho_0)$)。为了在数据严重截断的

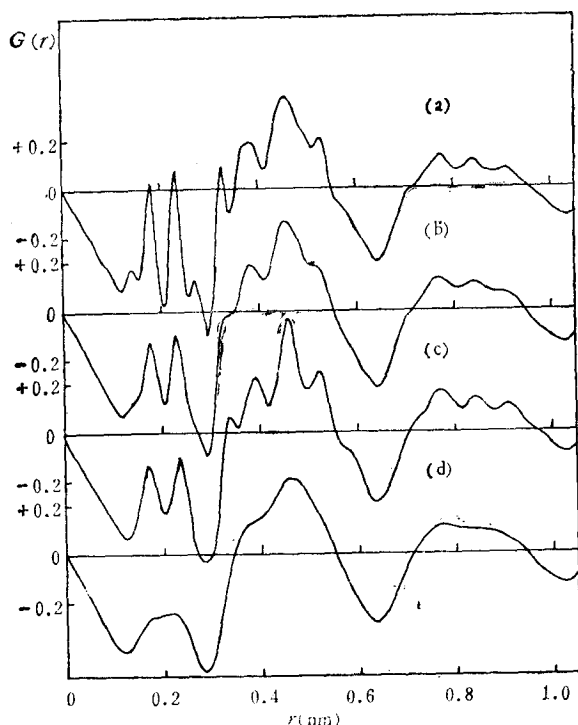


图1 非晶态 D_2O 原子简约分布函数

(a) 最大熵方法, $s_{\max} = 190 \text{ nm}^{-1}$; (b) 傅里叶变换法, $s_{\max} = 190 \text{ nm}^{-1}$,
(c) 最大熵方法, $s_{\max} = 95 \text{ nm}^{-1}$, (d) 傅里叶变换法, $s_{\max} = 95 \text{ nm}^{-1}$

条件下比较最大熵方法和传统的傅里叶变换法,除了采用全部衍射数据 ($s_{\max} = 190 \text{ nm}^{-1}$) 进行计算和比较外,也计算和比较了仅用一半衍射数据 ($s_{\max} = 95 \text{ nm}^{-1}$) 的结果. 图 1(a) 和图 1(b) 分别是当 $s_{\max} = 190 \text{ nm}^{-1}$ 时,用最大熵方法和傅里叶变换法计算的简约分布函数. 图 1(c) 和图 1(d) 分别是当 $s_{\max} = 95 \text{ nm}^{-1}$ 时,用最大熵方法和傅里叶变换法计算的简约分布函数. 图 1(b) 和图 1(d) 的计算公式为^[8]

$$G(r_j) = \frac{2}{\pi} \sum_{k=1}^L s_k F(s_k) W(s_k) \sin(s_k r_j),$$

其中

$$W(s_k) = \sin(\pi s_k / s_{\max}) / (\pi s_k / s_{\max}).$$

图 1 指出,在实验数据相同的条件下,最大熵方法可以比傅里叶变换法给出更多的结构信息. 数据截断得越多,这种趋势越显著. 比较图 1(c) 和 (d) 可以发现,当实验数据被截断至 $s_{\max} = 95 \text{ nm}^{-1}$ 时,最大熵方法仍至少可以提供六个相关距离,而傅里叶变换法仅给出模糊的包络. 由此,我们可以设想,当由于实验条件的限制,只能获得极为有限的可用信息时,借助最大熵方法可以弥补实验上的不足. 从正空间看,最大熵方法提高了结构信息的分辨率,这等价于把倒空间上的强度分布扩展到实验测量不到的区域. 从这种意义上讲,最大熵方法可以用来解决实验数据的外推问题^[3].

图 1(b) 和 (c) 是用截断点不同的两组实验数据计算的结果. 前者是全部实验数据

的傅里叶变换,后者是仅采用一半实验数据的最大熵结果。可以看出,虽然最大熵方法仅用一半实验数据,但它提供的结构信息已经和用全部实验数据的傅里叶变换法相当。这表明,最大熵方法的分辨率高于傅里叶变换法。这个结果为简化衍射实验提供了可能。由于非晶态物质相干散射强度一般较弱,而且随衍射角增大而迅速衰减,因而强度测量通常十分困难,特别是高角范围内的测量更加困难。借助于最大熵方法,可以将实验测量中止在某一中等角度上。我们对若干种非晶态物质的计算结果表明,用最大熵方法,用截断在 $s_{\max} = 90 \text{ nm}^{-1}$ 的数据就足以获得傅里叶变换法用 $s_{\max} = 160 \text{ nm}^{-1}$ 的数据得到的结果。

图 1(a) 是用全部实验数据计算的最大熵分布。显然,它比图 1(b) 反映更多的结构细节。图 1(a) 给出一个新的相关距离 ($r = 0.27 \text{ nm}$), 这个峰可解释为两种原子相关的总和—O-O 第一近邻和 D-D 第二近邻的叠加。这与 Chowdhury 等人用 CRN 模型计算的结果一致^[8]。傅里叶变换法没有分辨出这个峰,其原因显然是由于卷积效应,使位于 0.23 nm 的 D-D 第一近邻相关峰宽化,从而掩盖了位于 0.27 nm 处的这个弱峰。图 1(a) 在约 0.14 nm 处还出现一个小峰。这个峰可能是由于强度数据的误差造成的伪峰^[8]。由于最大熵方法客观地反映全部已知信息,因而已知信息中的误差信息也不可避免地要出现在结果中。从计算方法上考虑,用最大熵原理可以减少或消除误差给结果带来的影响,这一点从理论上讲是可行的。这项工作目前正在进行中。

四、结 语

我们引入了最大熵原理,讨论了如何将这一原理应用于非晶态物质径向分布函数测定,用有限的实验数据获得较多的结构信息。简言之,首先定义概率熵,用来度量待测系统的不确定性。每个实验值都使这种不确定性下降。如果全部实验数据不足以消除系统的不确定性,那么,最大熵原理可用来确保所得的结果仅仅反映由已知信息决定的统计性质,而与未知信息无关。

本文对非晶态 D_2O 径向分布函数的计算结果表明,同传统的傅里叶变换法相比,最大熵方法对结构信息的反映更为敏感,因而在实验数据相同的条件下,最大熵方法可能提供较多的结构信息,特别是结构的细节。当可用的实验数据较少时,最大熵方法的这一特点更为突出。因而,最大熵方法更加适用于衍射数据被严重截断的情形。

值得指出,由于最大熵方法适于处理严重截断的数据,因而使简化衍射实验有了可能。这将使实验易于操作并节约大量时间。

虽然本文集中讨论最大熵方法在非晶态结构分析中的应用,但其可能应用的领域并不局限于此。事实上,凡涉及有限数据的变换问题都可以用最大熵方法进行处理。例如,涉及面很广的分解卷积问题,可望用最大熵方法求得良好的解。我们对计算机模拟数据进行的卷积—分解卷积的计算结果表明,最大熵方法优于目前普遍采用的傅里叶变换法和迭代法。用最大熵方法进行光电子谱和 X 射线谱的分解卷积的研究正在进行中。

本工作是在郭可信先生的指导和支持下进行的。黑祖昆同志提出不少建议;王之明,李静环和

胡魁义同志为使用计算机提供良好的条件。作者在此一并致谢。

参 考 文 献

- [1] E. T. Jaynes, *Phys. Rev.*, **106** (1957), 620.
- [2] C. E. Shannon, *Bell System Tech. J.*, **27**(1948), 379; 623.
- [3] J. P. Burg, Maximum Entropy Spectral Analysis, in Proceedings of the 37th Meeting of the Society of Exploration Geophysicists (1967).
- [4] T. J. Ulrych and T. N. Bishop, *Rev. Geophysics and Space Phys.*, **13** (1975), 183.
- [5] B. R. Frieden, *J. Opt. Soc. Am.*, **62** (1972), 511.
- [6] S. W. Wilkins, J. N. Varghese and M. S. Lehman, *Acta Cryst.*, **A39** (1983), 47.
- [7] H. Spath, *Comm. ACM*, **10** (1967), 726.
- [8] M. R. Chowdhury, J. C. Dore and J. T. Wenzel, *J. Non-Cryst. Solids*, **53** (1982), 247.

MAXIMUM ENTROPY METHOD IN STRUCTURAL ANALYSIS

WEI WEN-DO

(*Institute of Metal Research, Academia Sinica, Shenyang, China*)

Abstract

Information theory provides a constructive criterion for setting up probability distributions on the basis of partial knowledge, and leads to a type of statistical inference which is called the maximum entropy estimate. It is the least biased estimate possible on the given information, i.e., it is maximally noncommittal with regard to missing information.

The principle of maximum entropy, which has proven useful in other contexts, is adopted here to design a procedure for obtaining structural information from an incomplete set of diffraction data. A comparison is made between the proposed procedure and the conventional Fourier inversion method used in RDF analysis of non-crystalline materials. The maximum-entropy method is found to have a higher resolution and also the advantage of no adjustable parameters with a high degree of reliability.

This method is iterative and uses more computer time than direct techniques; however, a number of comparative examples indicate that a significant improvement on the resultant structure in quality and resolution is possible with only a few iterations.