

# 第一章 复杂系统的挑战

---

## 1.1 本书的结构

本书的目的是，建立一种用统一观点去处理复杂系统的概念和方法。全书由两部分组成：引论以定性的方式讨论复杂系统，而本书的其余部分采用定量方法。第一章，我们将介绍复杂系统的一些例子，以及处理复杂系统的某些典型方法，例如热力学和协同学。我们将讨论自组织概念，尤其是信息的各个方面。在本章的最后一节我们将提出新理论的纲要，可把它看作协同学的宏观方法。第二章，介绍协同学微观方法的概要，第三章介绍最大信息熵原理，它在本书中起着重要作用。第四章通过最大信息熵原理在热力学中的应用详细阐明这个原理。

本书的其余部分将论述我们的定量方法及其应用，我们将在其中给出一些物理学和生物学的详细例子。最后，我们要证明模式识别领域中一种重要方法，它作为一个特例，包含于我的一般理论之中。这样，我们便真正在科学上取得了惊人的统一。对定

性讨论不十分感兴趣的读者，可以不看引论性的第一章，而根据自己的知识水平直接阅读第二章，第三、四章和第五章。

现在让我们先看——

## 1.2 什么是复杂系统

首先，我们必须讨论什么是复杂系统。从朴素意义上看，可以把它们看作是由许多同类或不同类的部分（或基元、组分）组成的系统。这些部分或组分的相互联系可以是很复杂的，也可以是不太复杂的。科学的许多分支给我们提供了大量的例子，有些系统其实十分简单，而另一些系统却真正称得起是复杂的。

让我们从物理学中的一些例子开始讨论吧。气体由大量的分子组成，比如说每立方厘米中就有 $10^{22}$ 个分子。气体分子以完全无规则的形式飞来飞去，因此它们之间进行着频繁的碰撞（图1.1）。相比之下，晶体中的原子或分子排列得井然有序，它们仅轻微地振动（图1.2）。我们可能对一些特定的性质感兴趣，诸如气体压强和温度或晶体的压缩率等。或者，我们可从它们服务于某一目的的角度来考虑，这些系统，例如象水蒸气的气体可用于蒸汽机、晶体，可用做导电体等。有些物理系统主要是为某些目的服务而设计的，例如激光器（图1.3）就是一种用以产生特殊形式的光的新光源。

在化学中，我们又一次涉及到复杂系统。在化学反应中，大量分子参与反应，从而生成新的分子。生物学更充满了复杂系统。细胞由复杂的细胞膜、细胞核和细胞质组成，它们之中的每一个都含有许多低一层的组分（图1.4）。在细胞中，在同一时间有条不紊的进行着几十到几千个新陈代谢过程。器官由以非常规则

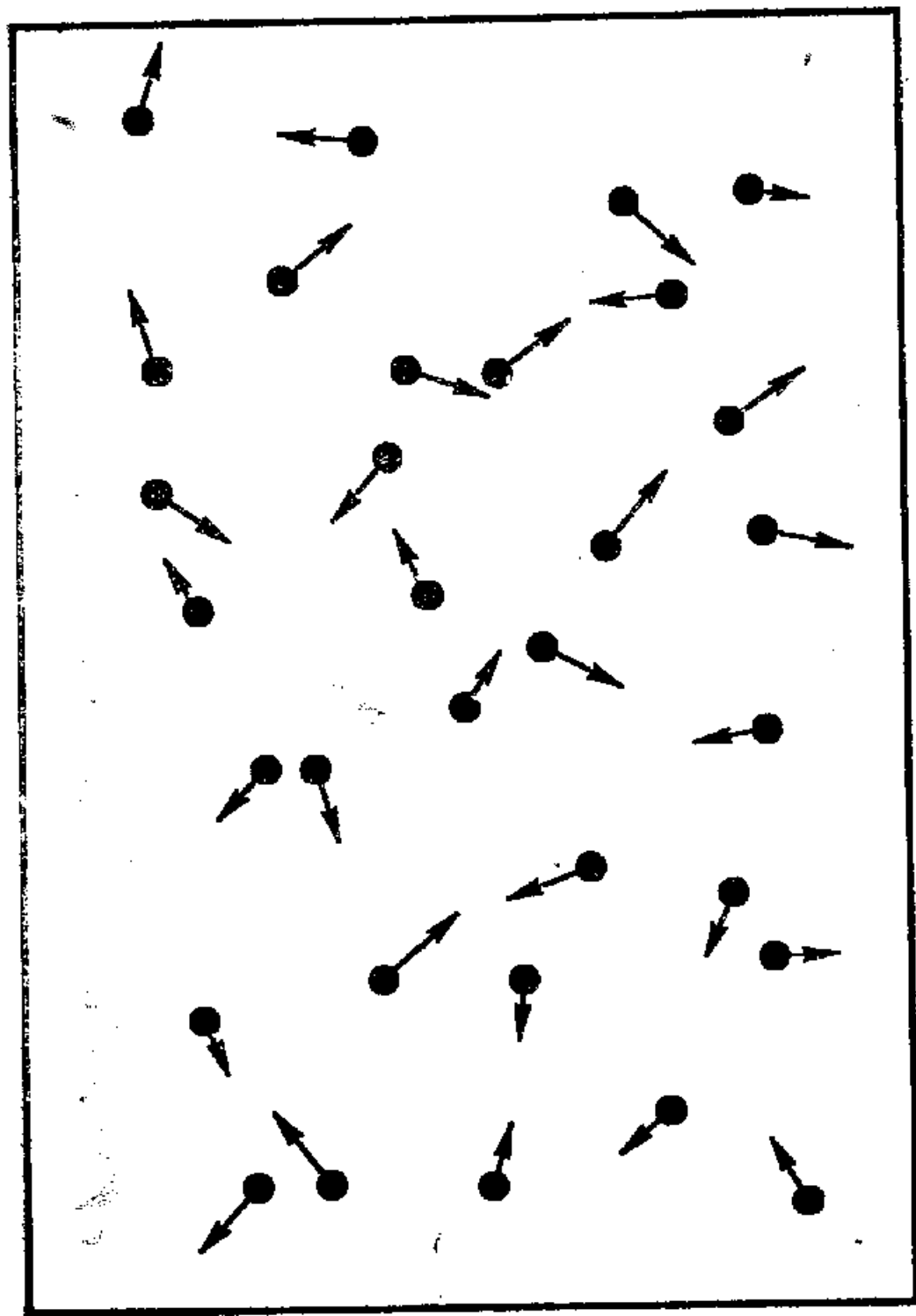


图1.1 在箱中动运着的气体原子

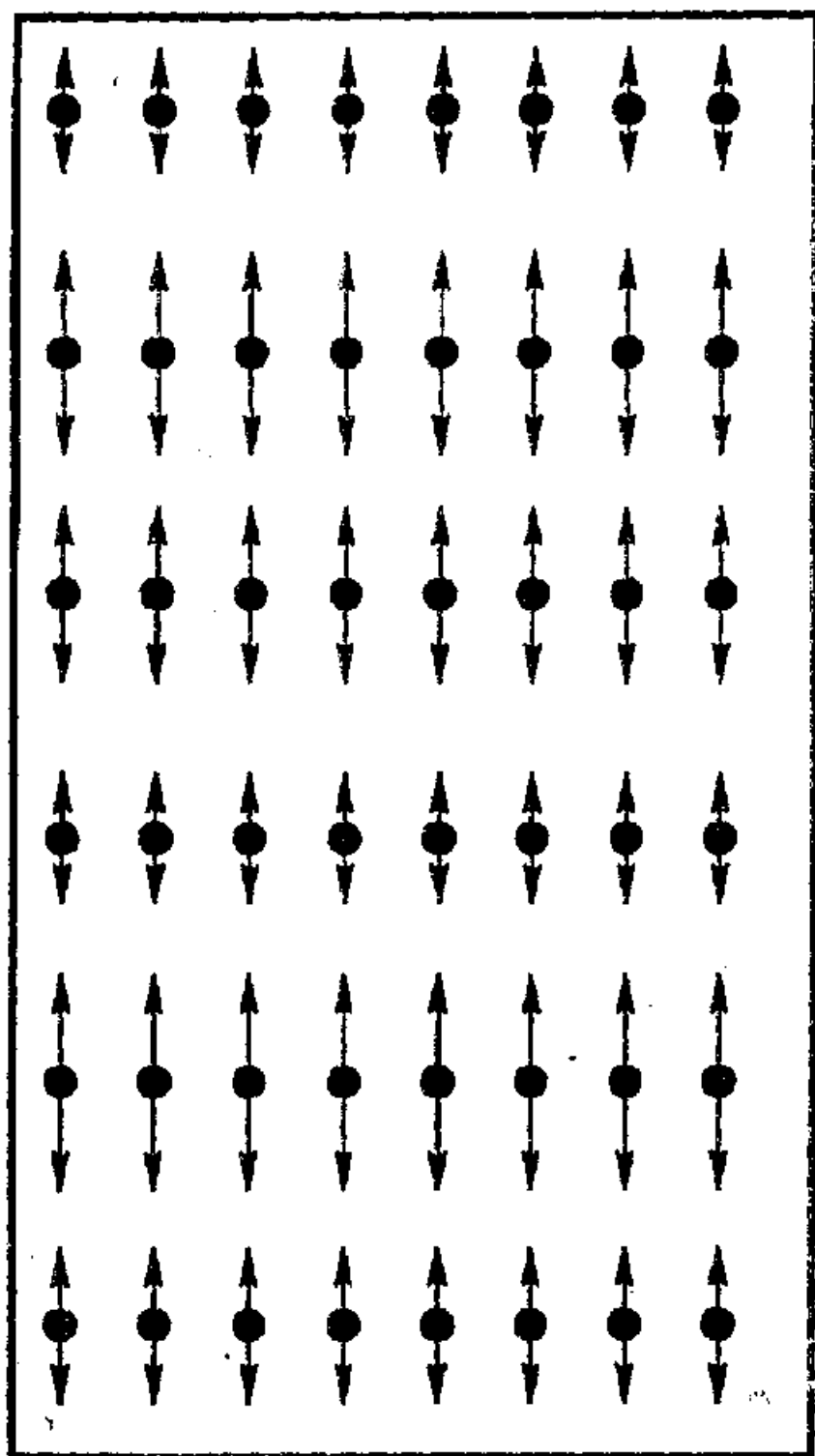


图1.2 晶体中的原子

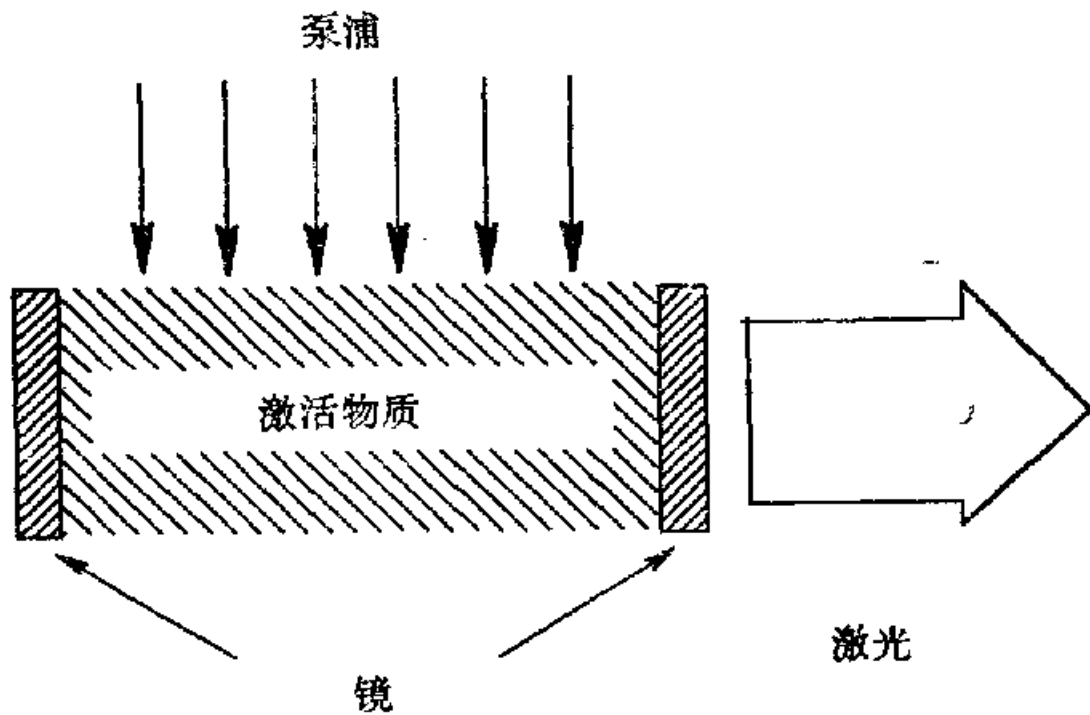


图1.3 激光器示意图

的方式合作着的大量细胞组成，各个器官又进一步服务于各种特定的目的，并在一个动物体内合作着，这些动物本身又组成了动物社会（图1.5）。世界上最复杂的系统大概要数人的脑子了，它由 $10^{10}$ 个甚至更多的神经细胞所组成（图1.6、图1.7）。它们的合作才使我们能够识别图样、说话，或者进行其它的思维活动。

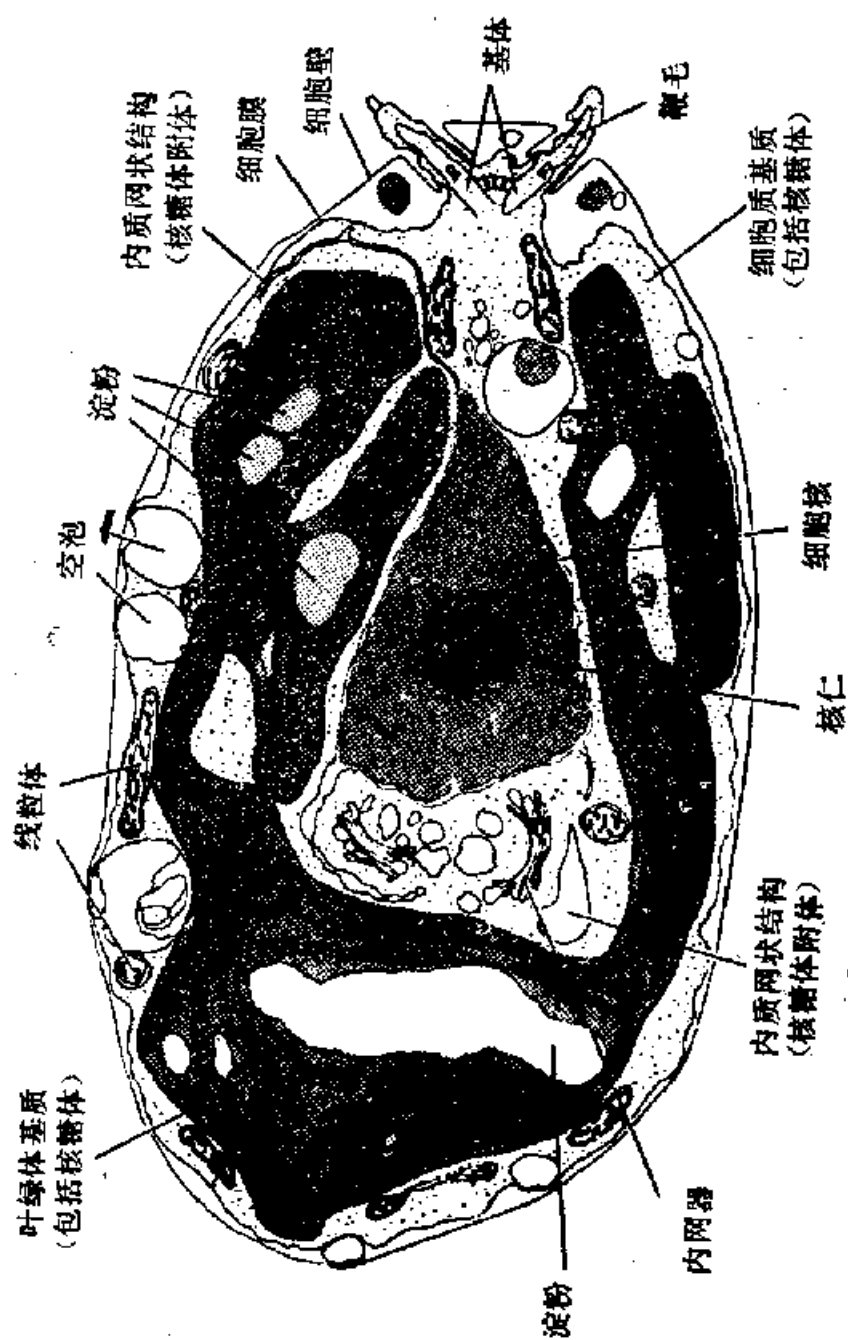
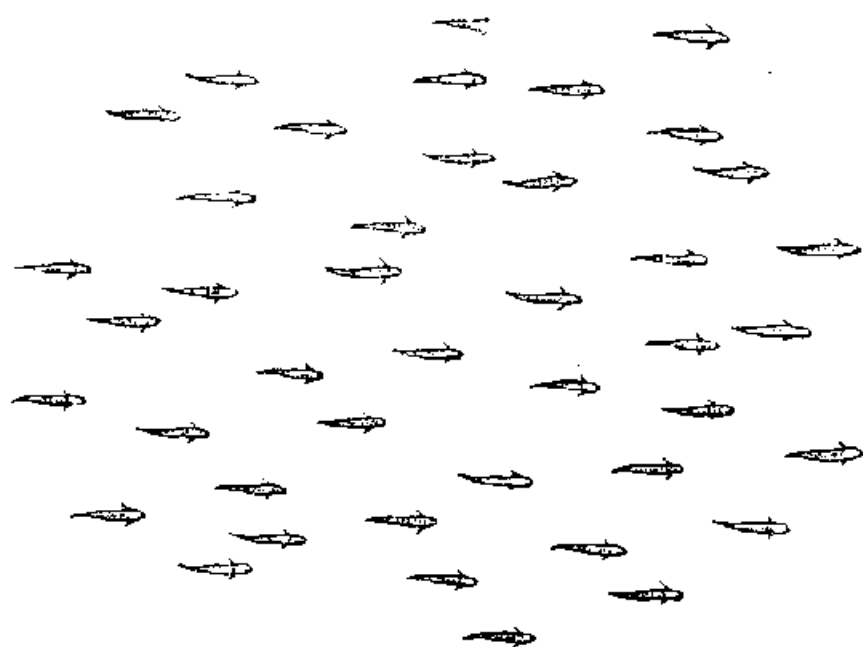
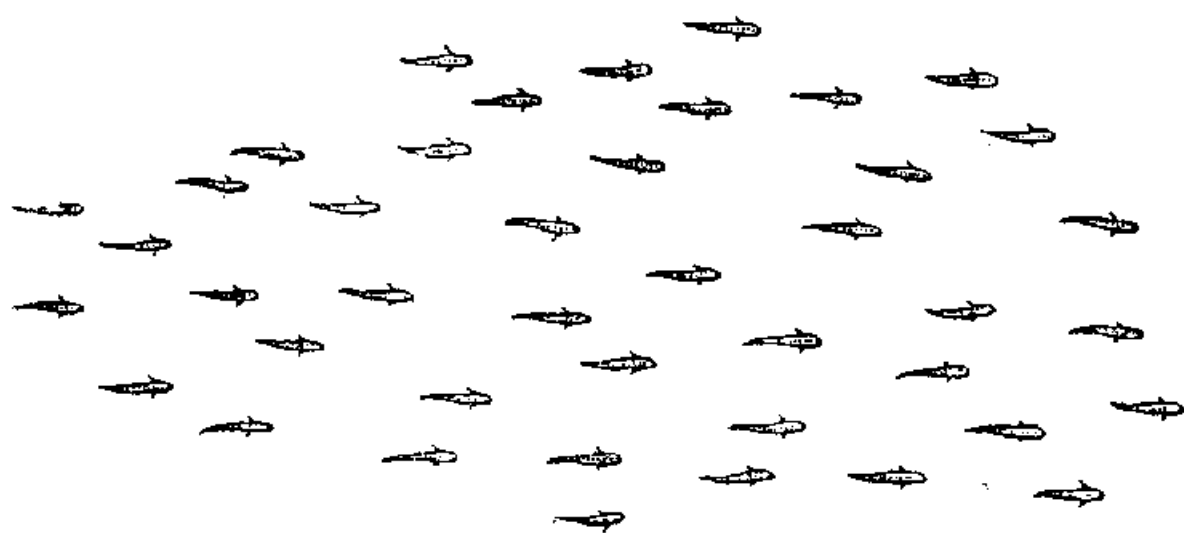


图1.4 一个典型的生物细胞



(a)



(b)

图1.5 鱼群



图1.6 神经细胞网之一



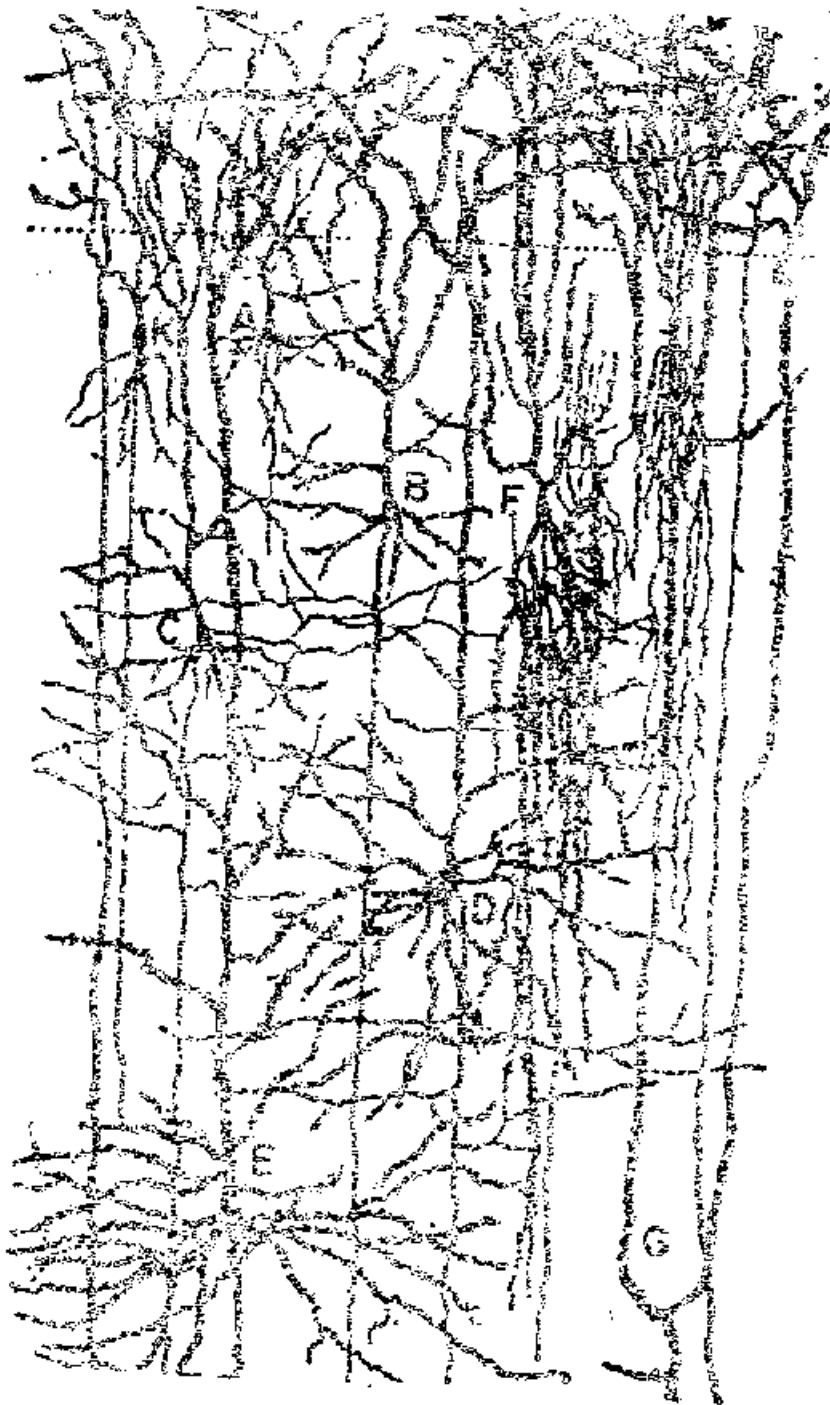


图1.7 神经细胞网之二

在工程学中，我们仍然要研究复杂系统。这类系统可能是机器，比如说一台汽车发动机，或者整个工厂，或者组成网络的一些发电厂。经济活动有无数的参与者，如商品和货币的流通、贸易、生产、商品的消耗和储备，它为我们提供了复杂系统的另一个范例。同样，包含有政治、宗教、职业或文化习惯的社会各种活动，是这种系统的又一例证。人们越来越清楚地认识到计算机是复杂系统，特别是所谓的第五代计算机，它将以知识处理来代替现代计算机的数字处理。

系统的复杂性不仅表现在它们是由大量数目的部分所构成的，而且还表现在它们具有复杂的行为。正如心理学中所研究的那样，人的行为的各种表现可以是很复杂的。但另一方面，我们也羡慕运动、呼吸等过程中肌肉的高度协调（图1.8）。最后我们要指出，现代科学本身就是一个复杂系统，这一点可以从它的各个分支的庞大数目明显地看出。

我们可能会问：为什么为数众多的系统是这样复杂？它们是怎样产生的呢？例如，无论在生物学中还是在工程技术中，我们都容易看到对复杂性的需要。这些系统服务于一些特定的目的，仔细考究，我们便会发现，这些目的仅能由复杂系统来满足，而这些系统由许多部分所组成，这些部分又以很有条理的形式相互作用。我们根据系统的产生方式，可以将其分为两大类。一类是人们设计和建造的人造系统，这些机器或结构因此而服务于特定的目的。但是在自然界中还存在着大量的由自然界本身所产生的系统，换句话说，另一大类系统是自组织的。在这里，特别是生物学中，进化论的观点，也就是达尔文主义，显然起着重要的作用，人们试图通过达尔文进化论理解系统为什么会向复杂的方向

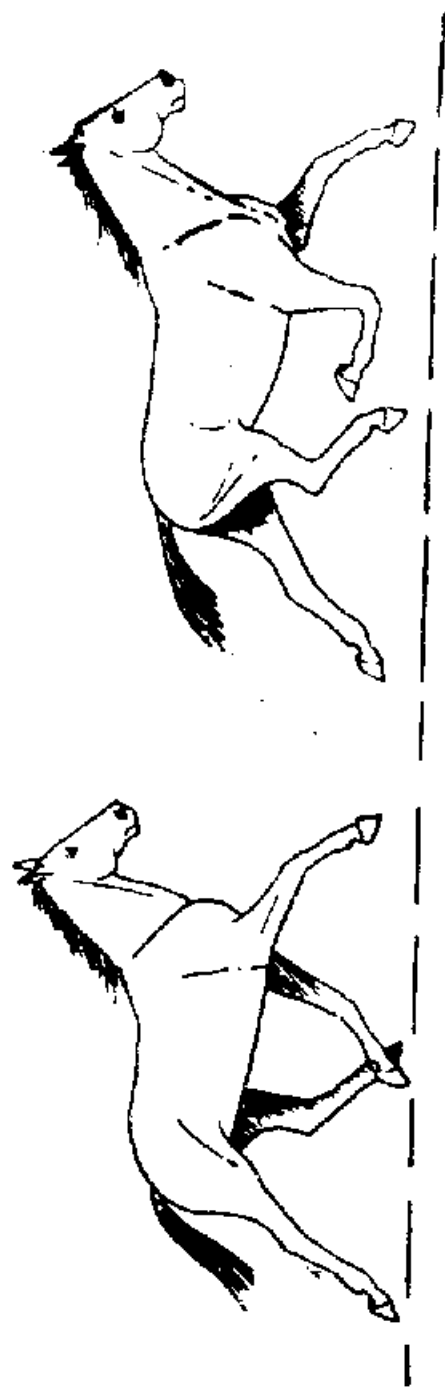


图1.8 疾走的马

演化，这些越来越复杂的系统又是怎样演化出来的。

对复杂系统作了这样十分肤浅和粗略的介绍之后，我们试图给出一种比较严格的定义。现代的定义是以代数复杂性的概念为基础的。至少在某种程度上，一个系统可以用一系列数据来描述，例如从恒星射来的光的涨落强度（图1.9），或者一个病人的发烧曲线，它们的数据可用数字表示。我们因此来研究一个数列，并试图定义这种数列的复杂性。考虑一个具体例子，比如说象1、4、9、16、25、36等数字，我们发现这个数列可以由简单的 $n^2$ 规律得到（ $n$ 是整数）。因而，每当给出一个数列时，我们总可以询问，是否存在一个计算机程序和一组初始数据，使得我们可以借以计算出整个数据组。当然，依照计算机的不同结构，一个计算机的程序可以长于另一个计算机的程序。

为了能够比较程序的长度，我们必须引入一种通用计算机。无需考虑细节，我们可以说，能够造出这种通用计算机，至少在图林(Turing)表明的思维实验中（图1.10）是如此。因而，我们把这种通用计算机称为图林机。于是我们力图把程序和初始数据组压

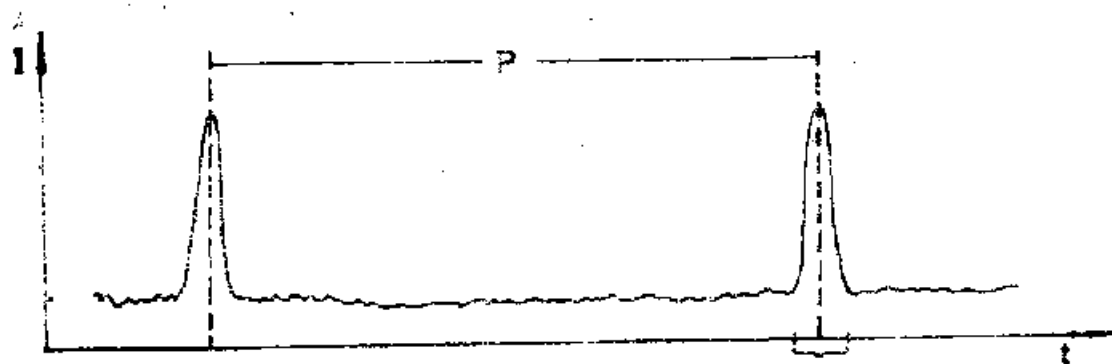


图1.9 脉冲星的光脉冲

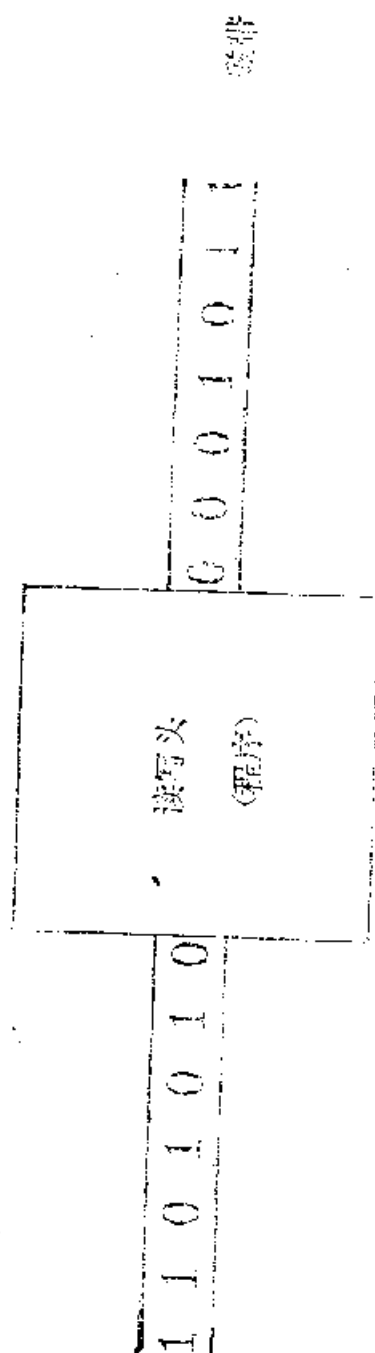


图1.10 图林机示意图

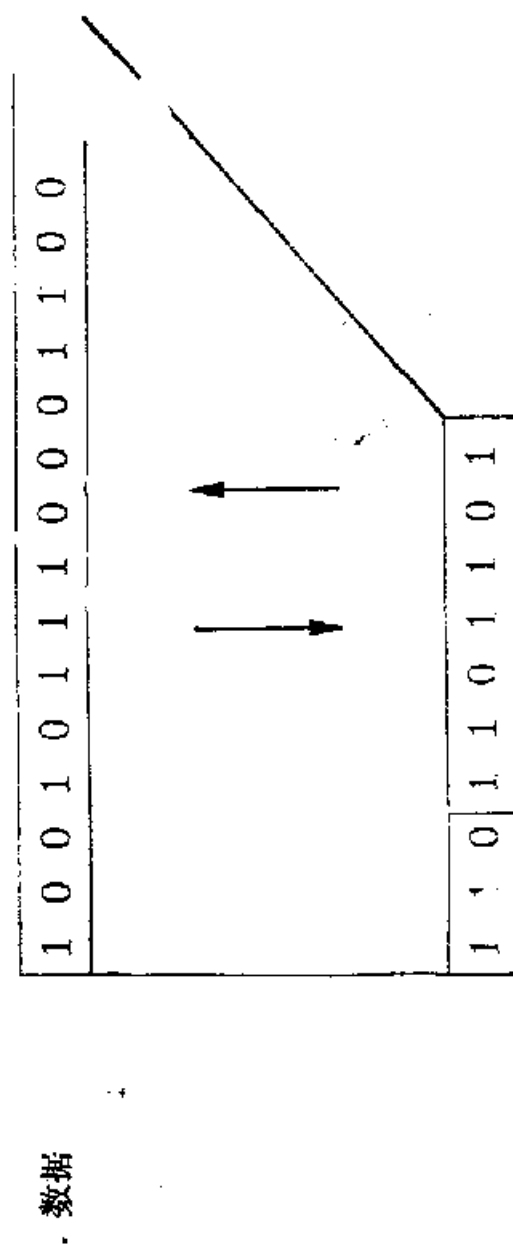


图1.11 一列数据压缩成一组最短程序及数据

缩到最低限度。程序和初始数据的最短长度便是代数复杂程度的量度（图1.11）。但是这种定义是有缺陷的。正如利用著名的哥德尔（Goedel）定理证明的那样，寻求最短程序和数目最少的初始数据，这个问题不能以普适的方式求解。换句话说，要求解这个问题，没有合适的一般算法，我们只能在特殊情况下发展这种算法。的确，我们有时能够找到捷径。我们来考虑气体。在这里，人们也许企图设法跟踪各个粒子的径迹和它们的碰撞，然后推导出单个粒子的速度分布函数。人们从微观描述出发，对这个问题的研究，至今尚未揭晓。尽管如此，在统计力学中我们已能够导出这种分布函数了，这就是玻尔兹曼分布，它具有十分简单而又优雅的形式，它的导出没有采用微观途径，但使用了熵的概念（请见下文）。类似的例子还可以举出一些，它们表明捷径是存在的，通过这些捷径，原来很复杂的问题可以用十分直接的方式求解。于是，我们认识到复杂性的概念是非常微妙的。确实，本书的主要目的是以统一的观点提供这种捷径，而这些捷径使我们能够处理复杂系统。

一个复杂系统可以从不同的角度来研究。例如，我们可以在宏观层次上通过生物宏观行为的研究来处理生物系统，或在中间层次上研究各个器官的功能，或者最后研究DNA的化学。需要搜集的数据似乎常常是无穷无尽的。此外，我们往往不能决定哪个方面的资料需要优先选择。看来要了解怎样处理复杂系统，我们必须经历一个学习和摸索的过程。

### 1.3 如何处理复杂系统

科学划分为越来越多的学科，寻求统一的原理就变得越发重

要。复杂系统的无处不在,使我们面临着挑战,我们必须寻求解决复杂系统的统一原理。为了从微观层次上描述复杂系统,需要极其庞大的数据,最终谁(甚至一个社会)也处理不了这些数据。因此,我们必须引用某种能搜集数据或思维的经济学。当我们找到可应用于完全不同的复杂系统的定律时,我们可期望对其深入地理解。

要寻求普适定律,聪明的办法是先问一下,我们希望在哪个层次上来表述定律,是微观层次还是宏观层次。相应地,我们可以用完全不同的方式描述系统。比如,气体在微观层次上完全无序,而在宏观层次上在我们看来事实上是均匀的,即无结构的。与此相反,晶体在微观层次上井井有条,而在宏观层次上仍然是均匀的。在生物学中,我们涉及到各级层次,从分子层次,细胞和器官层次到整个植物或动物。这样划分层次可能过于粗糙,而且对层次恰当的选择决不是无关紧要的。此外,“微观的”和“宏观的”变成了相对的概念。例如,一个生物分子,同组成它的原子相比,可以看作是“宏观的”,而同细胞相比,它就是“微观的”了,等等。顺便说说,在每一个层次上,我们都面临着一类特定的组织或结构。

现代西方科学的方法被称作是分析的方法。人们试图把一个系统分解成它的各个部分,从而了解整个系统的各种性质。在一些领域中,我们可以从第一原理(它属于基本定律)出发,进行研究。当然,在这方面最突出的领域是物理学,更具体地说是粒子物理学。通常,人们认为系统的各部分及其性质是“客观”给定的,“只”需要从部分的性质去推断整个系统的性质。这里,有必要做如下两点评论:首先,严格地说,我们事实上是从宏观数据推断微观事件,而且,去检验不同的微观模型是否可得到相同的宏观数据



组,这是一个有趣的问题;其次,分析方法是以还原性为基础的,或者说在极端情况下,是以还原论为基础的。但是,我们越发研究复杂系统,就越发认识到还原论有其局限性。例如,通晓化学并不意味着我们了解生命。相反,当我们从微观过渡到宏观层次时,系统一次又一次地出现新的特性——它们并不出现在微观层次上。

例如,虽然用波长和振幅可以描述波,但这些概念与个别粒子(诸如一个原子)是格格不入的。我们需要了解的不仅是甚或不是各个粒子的特性,而是它们一起演奏的管弦乐。要理解这曲管弦乐,在许多情况下,我们可以应用模型系统,通过它,我们能够详细研究复杂系统的特征。本章1.4节将介绍几种模型系统。宏观描述提供了解决复杂系统的另一种方法。例如,我们不是通过给出气体原子在每一时刻的所有坐标来描述气体,而是用宏观量,诸如压强和温度来描述。自然界本身已经给我们提供了测量或感知这些量的方法,这是一个惊人的特点。

为了研究复杂系统,我们往往还要寻找合适的变量或相关量去描述这些系统的性质。无论哪一种宏观描述都会使信息受到巨大的压缩,因此我们关心的不再是单个微观数据,而是整体的性质。处理复杂系统的一个重要步骤是建立各个宏观量之间的关系式。这些关系式是微观事件的结果,而关于这些微观事件,我们一无所知,或所知甚少。热力学给出了这种关系式的一些例子,例如,有关气体压强和温度的定律,它们可由统计力学的微观定律推导出来。一般来说,我们必须对最终导致宏观数据的微观事件作出猜想。

本书中,我们打算说明对完全不同学科的系统,如何作出这种猜想。读者将同时看到,在足够抽象的水平上,复杂系统的行

为存在着深刻的类似性，换句话说，复杂行为可在极不相同的基  
质上实现。我们体会到越是复杂的系统，越具有与人类相似的行  
为特征。因此，我们被引向或者可能被错误地引到歧路：用人类  
的术语去描述复杂系统的行为。在自然科学中，极力避免拟人论，  
避免用机械论的观点提供所有的解释和概念，这已成为传统。本  
章后几节将讨论这种机械论与拟人论的困境，当涉及到信息，以  
及讨论意义和目的作用时，我们特别要论及这些。

让我们用一般性评论来结束本节。不久以前，人们或多或少  
地普遍感到，物理学（或自然科学）与人文科学差异巨大，而后  
者研究着真正的复杂行为和复杂系统。长久以来，鉴于物理学有  
预言未来事件的本领，它受到了赞美。但正如我们将要看到的那  
样，物理学越是涉及复杂系统，就越需要新的概念，而且它的某  
些主要特点（比如准确预言的本领）就会逐渐丧失。

#### 1.4 模型系统

物理学的巨大成功，依靠它的方法论。在物理学中，一个复杂  
的系统被分解为一些特定部分，仅仅改变一个或很少几个参数，  
就能对这些部分的行为进行可重复的研究。这种方法的著名例子  
是，伽利略的自由落体实验，或牛顿通过仅由太阳和一个行星组  
成的系统的研究，对整个行星运动所作的分析。换句话说用这种方  
法，牛顿处理了单体问题，至多是二体问题，然而却产生了牛顿力  
学。从牛顿力学的表述出发，拉普拉斯推断，牛顿力学意味着，一  
旦知道了初始时刻系统中各粒子的速度和位置，那么未来便可完  
全被预言。在近代物理学中，可预言性的概念曾受到两次冲击。  
量子力学告诉我们，不可能在同一时刻以无限的精确度测量一个

粒子的速度和位置，因此我们不能预言一个粒子的未来径迹。

最近，所谓的决定论性混沌理论表明，甚至在经典力学中，也不能以绝对的精度确保可预言性。考虑钢球落到竖直刀刃上，这种非常简单的例子（图1.12）。钢球由于相对于刀刃的精确位置的不同，其轨道可以向左边或右边倾斜。这意味着粒子（也就是钢球）的运动，以非常敏感的方式依赖于初始条件。初始条件的微小变化就可以导致完全不同的径迹。过去的几年里在物理学、化学和生物学中，已经发现的大量的这种例子，显示出对初始条件的敏感性（图1.13）。尽管这样，复杂系统的适当的模型系统具有普遍性，这种想法仍是可行的。

在此，我想列举几个大家熟知的例子。激光光源已成为相干过程的自组织的一个范例，因为其中各个原子以非常规则的方式相互作用着，由此产生了相干的激光波（图1.14）。宏观结构自发形成的另一个例子是流体。例如，从下面加热一层流体时，流体可以呈现出一些特定的空间图样，诸如卷筒或蜂窝状结构（图1.15）。继续加热，流体就会呈现出空-时图样，比如卷筒状的振荡。有些化学反应也会产生宏观图样，比如化学振荡，它的颜色出现周期变化，例如从红到蓝再到红，等等。此外，还有螺旋形图样或同心波的现象（图1.16）。在生物学中，清水中的动物水螅已成为形态形成的一种模型系统。把水螅切为两部分，留下脚的那部分，很快长出一个新头，反之亦然，留下头部的部分，上边长出了新的脚（图1.17）。

详尽的实验可以使我们在化学物质的产生和扩散形成的化学场的概念基础上，得出一些关于这种再生机理的结论。在形态形成中，自组织的另一个例子由粘性霉菌提供（图1.18）。通常，

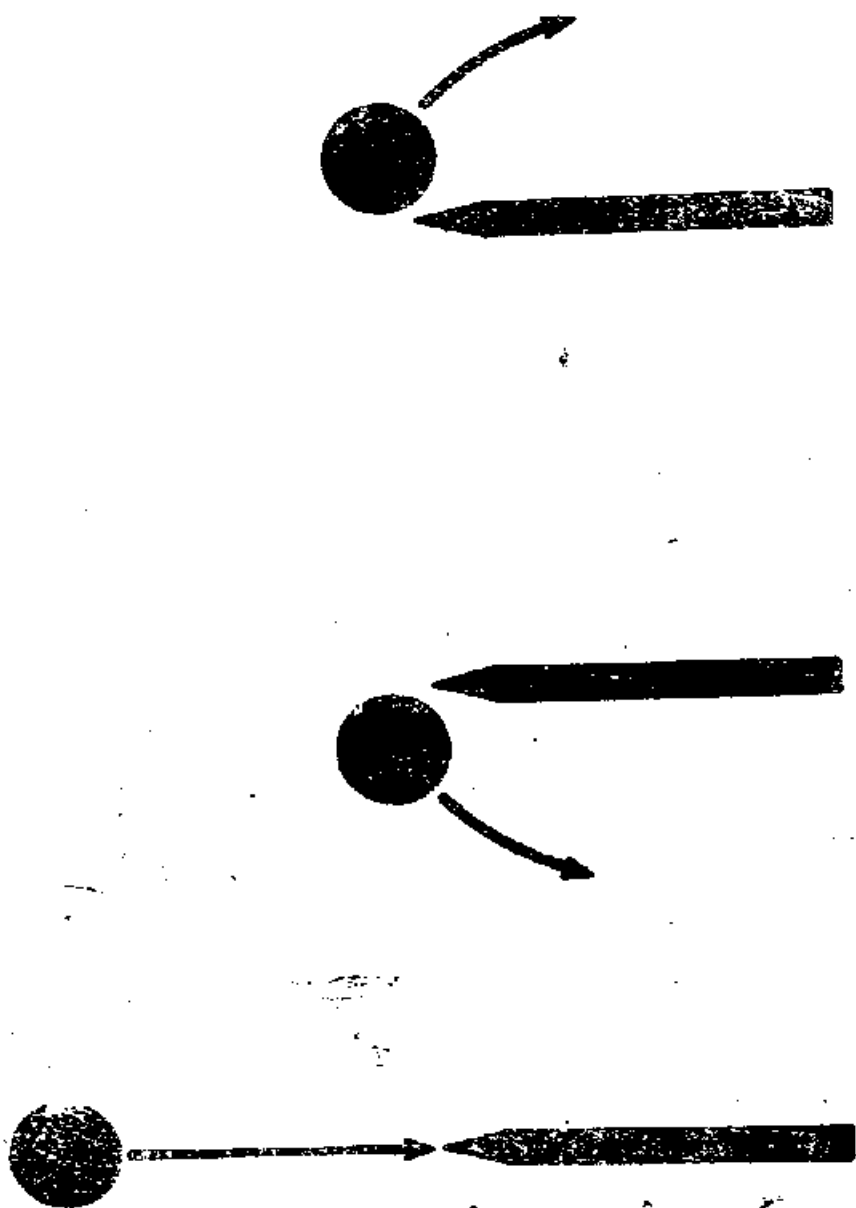


图1.12 钢球下落到刀刃上，依赖初始位置的不同，球会偏向左边或右边。

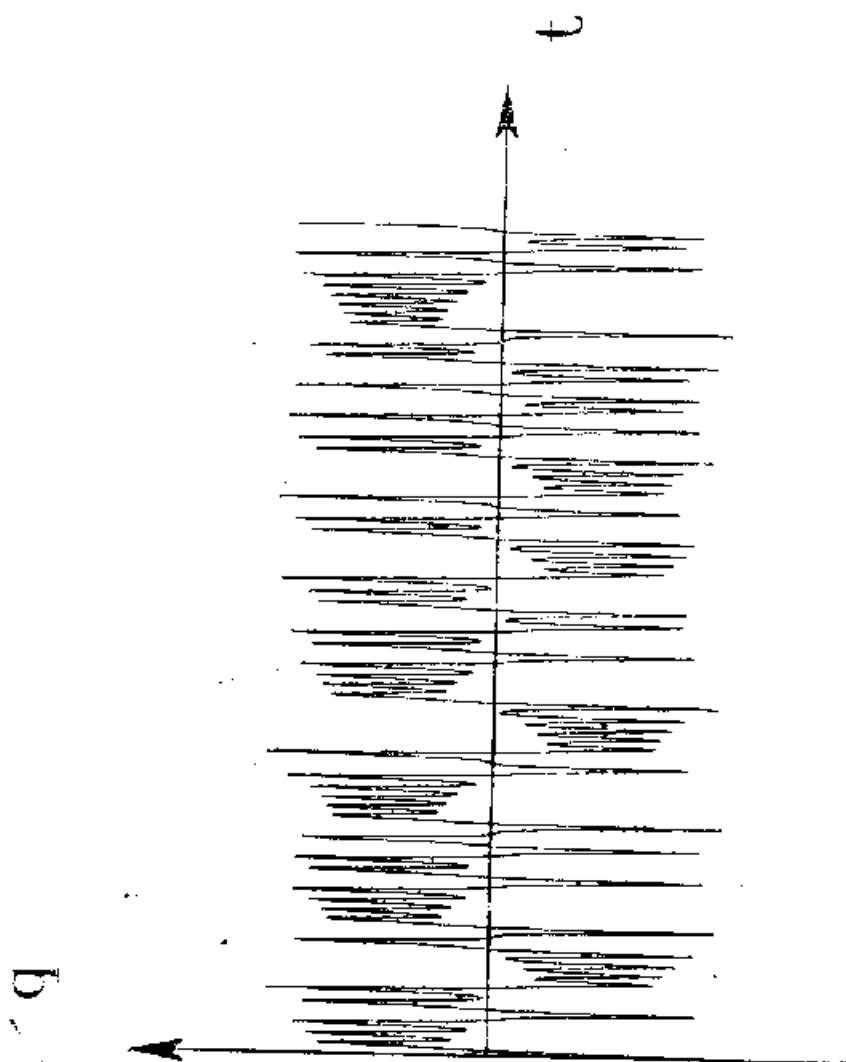
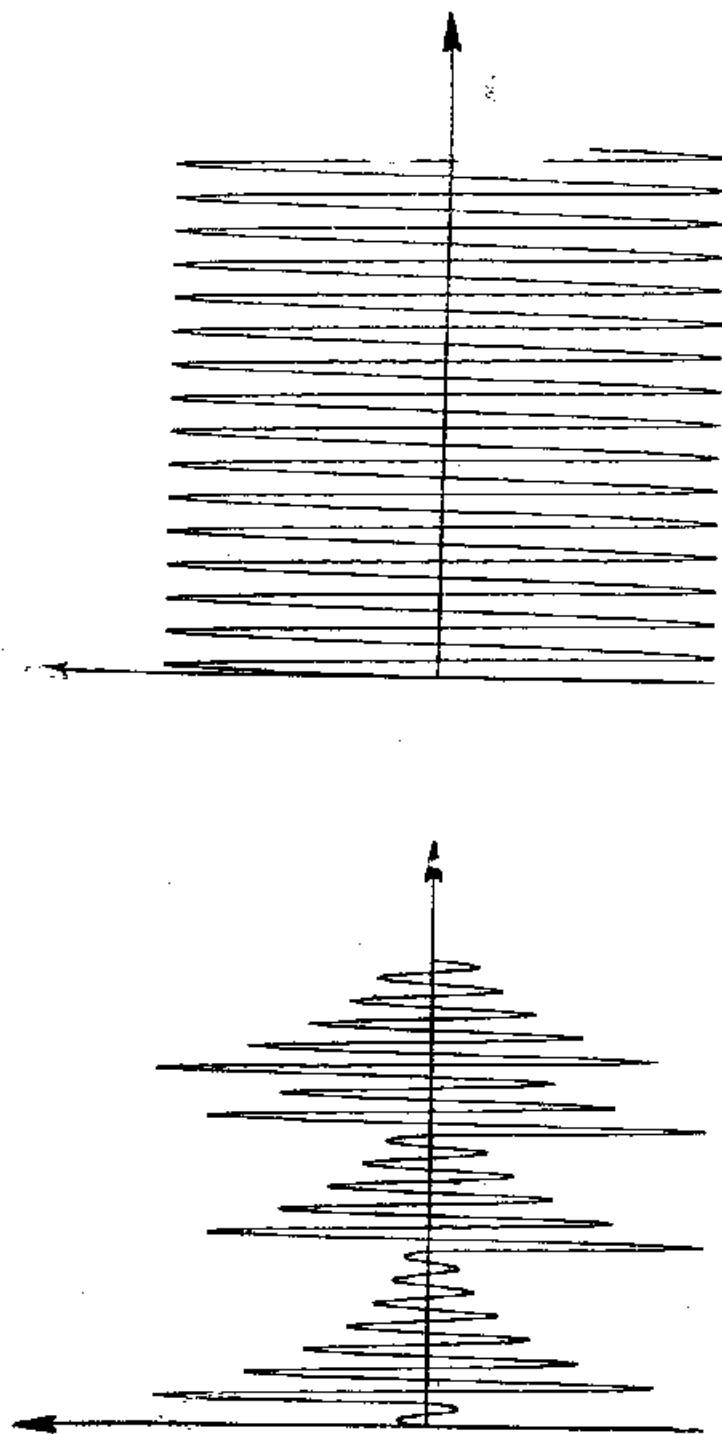


图1.13 混掩系统中量的变化

普通灯

激光



(a)

(b)

图1.14 灯光和激光的基本差别。图示了两种情况下场振幅与时间的关系曲线。(a)：各个无关波列构成普通灯光；(b)：激光中的单一无穷长正弦波。

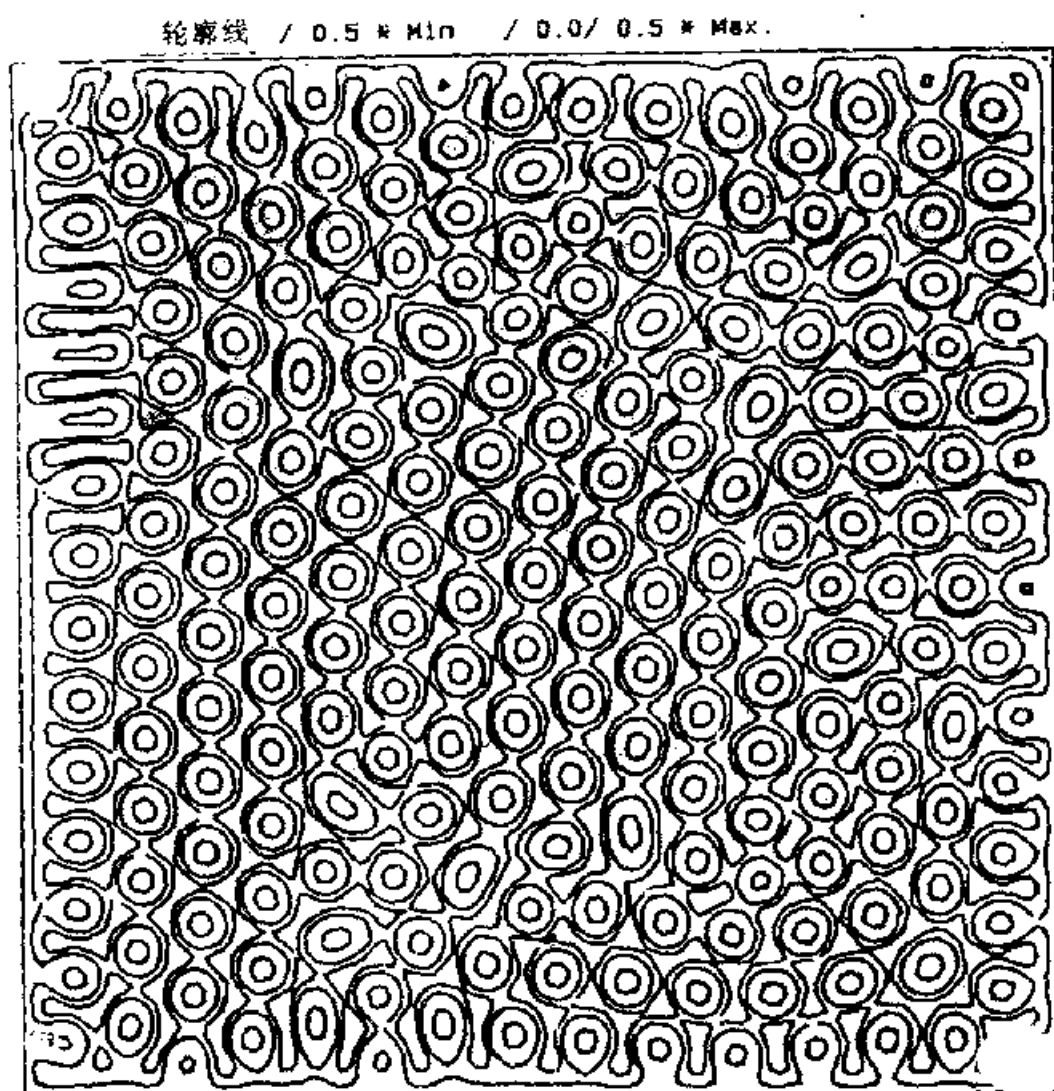


图1.15 流体力学中的一个图样

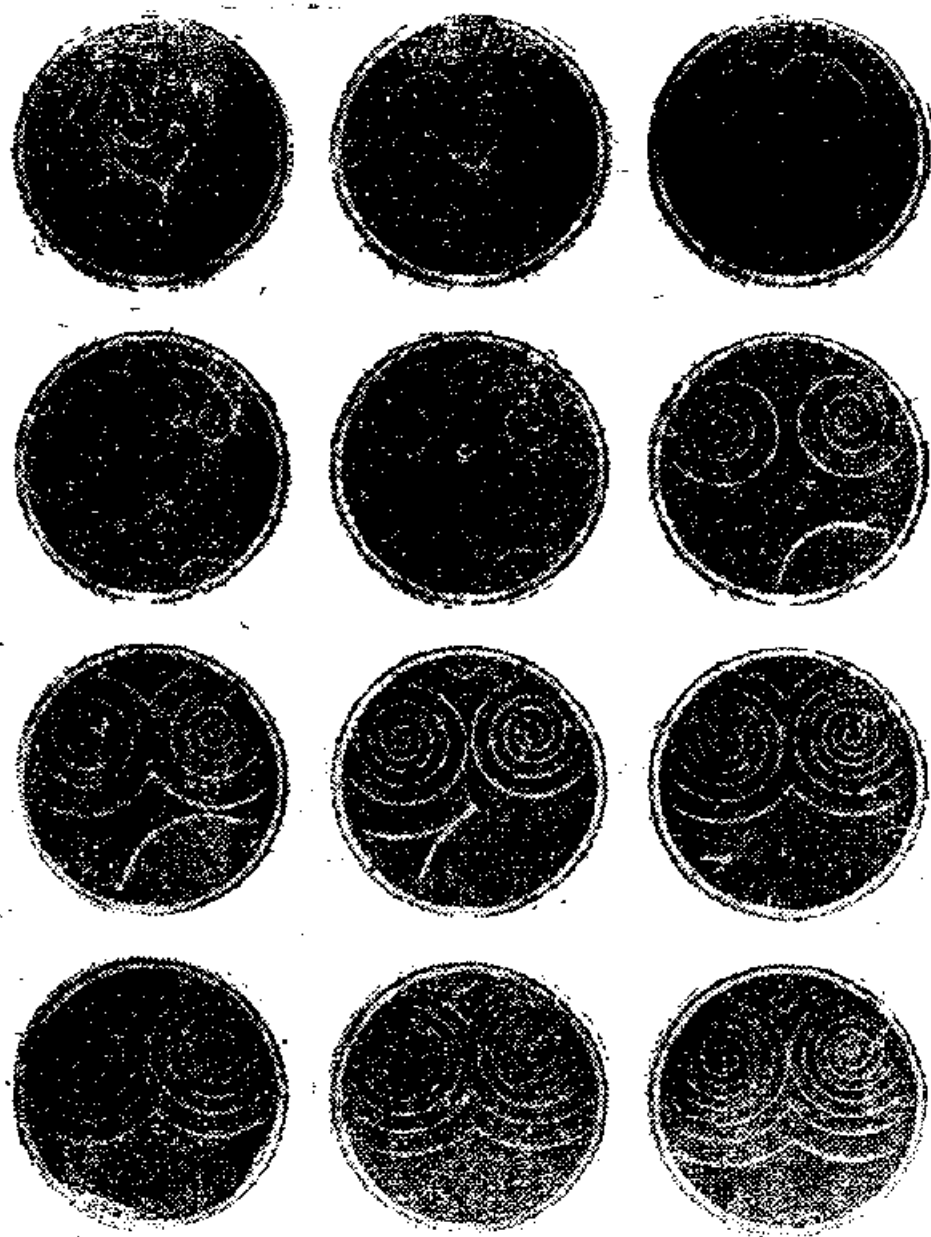


图1.16 B-Z反应的螺旋波



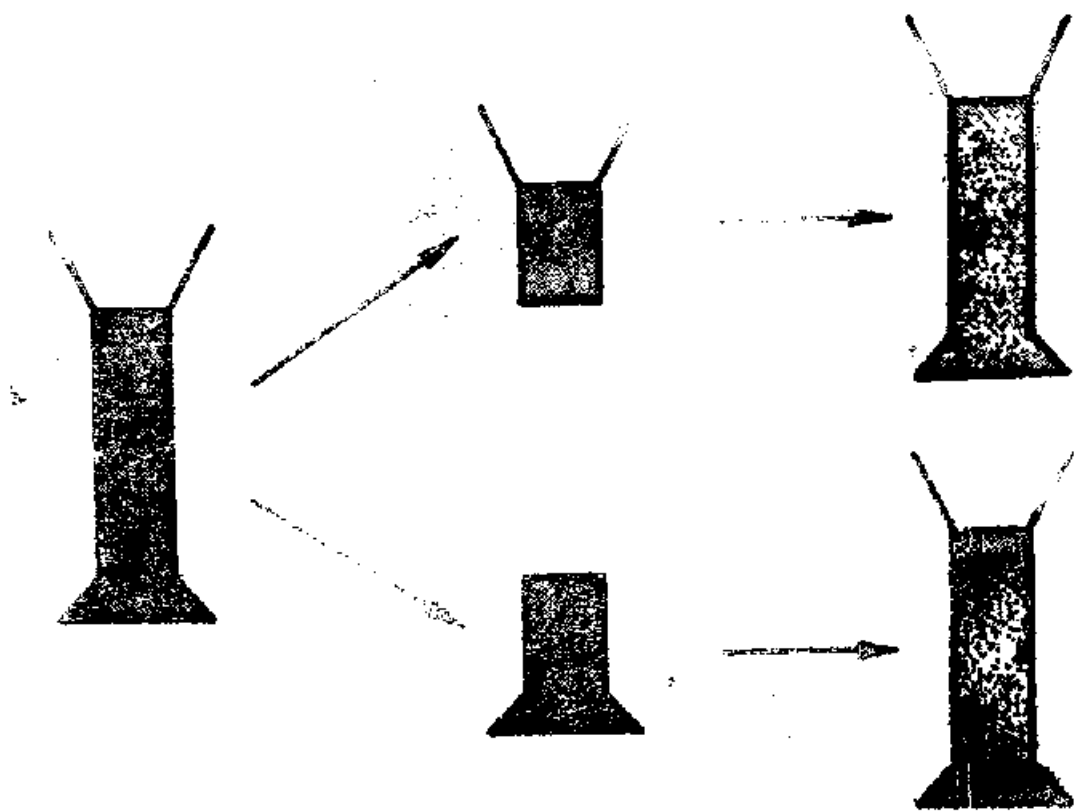


图1.17 水莖实验揭示出，其中变异的信息不可能存储于遗传因子中。从左到右，完整的水莖被从中间切断，过一会儿，上部重新长出了脚，下部重新长出了头。

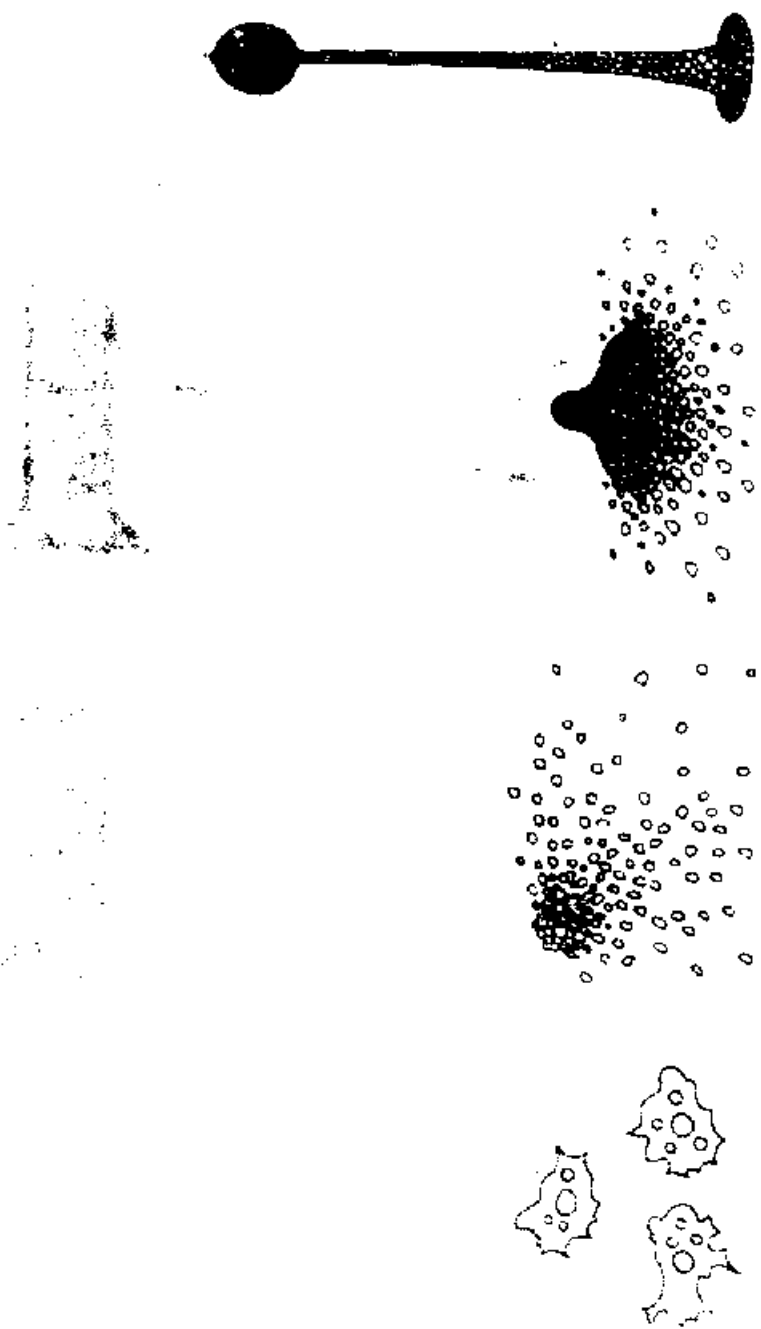


图1.18 粘性霉素的发育阶段。从左到右：细胞聚集起来，且越聚越紧密，最后分化成为蘑菇。

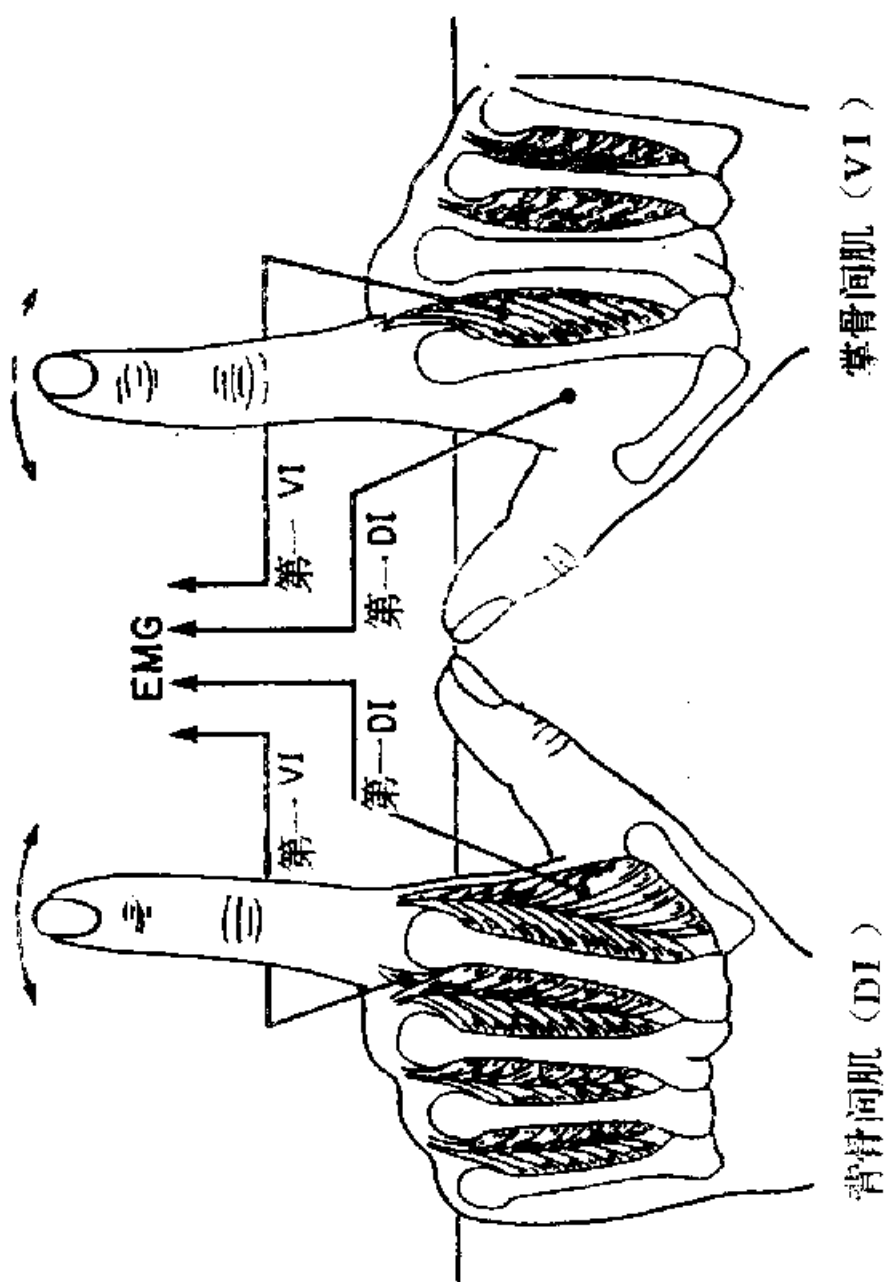


图1.19 研究手运动的实验

这种小蘑菇以单个细胞的形式生存，而这些细胞又生长在基质上。但在粘性霉菌的生命循环期内，各个细胞向一点聚集，分化并生成蘑菇，最后，这些蘑菇又散布它们的芽胞，接着又重新开始了整个生命循环。另一个模型系统是用来研究神经传导的鱿鱼轴突，或者大家都很熟悉的遗传发生学中果蝇的例子，在这里，巨型染色体、快速倍增率和引起变异的可能性，使得这个小动物成为该领域中的一个理想的研究对象。

最近，为了研究肌肉和神经细胞的配合，尤其是研究各种运动间的转变（图1.19），人手的运动已成为模型系统。我们将在本书的后面研究这个例子和几个其它的例子。手指运动的这种不自觉的改变使人强烈地联想到马、猫和其它四足动物步伐的变化。非常一般地说来，这些模型系统能使我们建立一些新的概念。我们首先可用各种各样的较简单系统检验这些概念，然后再把它们应用到真正的复杂系统。因此，后几章将专门用来建立这些新概念，并用一些具体的例子来说明它们的适用性。

## 1.5 自组织

如前所述，我们可将人造系统和自组织系统相区别。在本书中，我们讨论各种自组织系统。然而，要指出的是这两类系统之间的差异并不是十分明显的。例如，人可以这样来构造系统：通过适当的约束，系统便能以自组织形式产生特定的功能。前边提到的一个典型例子是激光器，激光器的特定装置利用它的反射镜使原子产生一种特殊的光。显然，从长远来看，有希望制造出以自组织方式执行程序计算机。

鉴于下文的需要，对自组织给予适当的定义将是很有用的。

如果系统在获得空间的、时间的或功能的结构过程中,没有外界的特  
定干预,我们便说系统是自组织的。这里的“特定”一词是指,那  
种结构和功能并非外界强加给系统的,而且外界是以非特定的方  
式作用于系统的。例如,产生六角形结构的流体是以一种完全均匀  
的方式从下面加热的,它便是通过自组织才获得其特定的结构的。  
在本书中,我们主要讨论一种特殊的自组织,即非平衡相变。

当我们改变参数,如改变温度时,熟平衡系统能够呈现出状  
态之间的特定转变。例如,加热时,冰将熔化而形成新的状态——  
液体,即水。加热水,使它变热,在一个特定的温度下,它将沸腾,  
并变成蒸气。因此,同样的微观基元(即各个水分子)可以产生完全  
不同的宏观状态,且这些状态之间会发生突然的转变,与此同时,  
出现了一些新的性质,例如,冰与水具有极不相同的机械性质。

以下,我们关心的主要是远离平衡系统所发生的类似转变。  
1.4节已提供了这方面的例子,例如,形成特定的空间图样的流  
体、发射相干光波的激光器,和经受分化转变、导致特定器官形  
成的生物组织等。

## 1.6 普适性的追求

### A. 热力学

热力学是一门以普适的观点处理任意复杂系统的学科。例  
如,我们可以谈一块石头或一辆汽车的温度,也可以谈一幅绘画  
或一头动物的温度。进一步,我们知道,当温度发生变化时,系  
统的重要性质将发生变化,如在熔解温度下冰的熔解,以及测  
量病人体温的意义。然而,这个例子同时说明温度肯定不足以  
在其它许多方面表征一辆车或一幅绘画。石头、车辆、服装和绘

画，它们具有处于热平衡状态的客体的性质，这样的平衡态可以通过将系统孤立来达到，也可将系统与一具有给定温度且处于热平衡的另一系统相接触而达到。

热力学的另一中心的重要概念是熵。熵是热平衡系统的一个概念，且这种热平衡能用温度 $T$ 来表征。熵的变化由著名的关系式 $dS = dQ_{rev}/T$ 给出。这里， $T$ 是绝对温度，而 $dQ_{rev}$ 是可逆地加于系统或从系统取出的热量。热力学的一般规律是：(i) 第一定律。第一定律指出孤立系统的能量是守恒的，这个能量可以取各种形式，如内能、所做的功或热量。联系 $dU$ ， $dA$ 和 $dQ$ 的典型关系式是

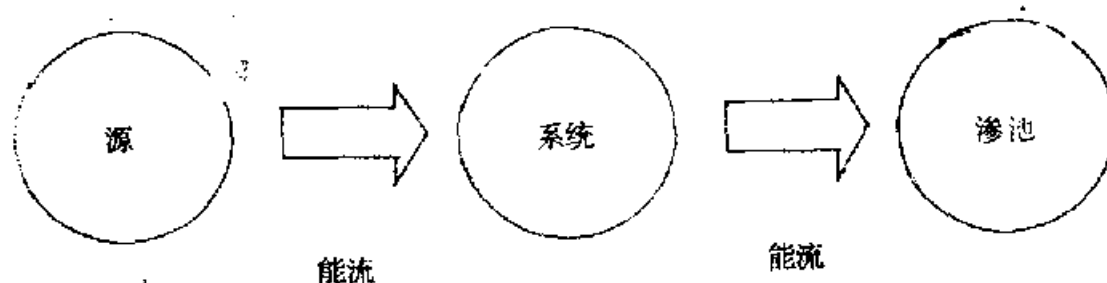
$$dU = dQ - dA \quad (1.1)$$

(ii) 第二定律。它告诉我们，封闭系统的熵决不会减少，而只能增加直至达到它的极大值。正如下文所见，守恒定律，比如能量守恒定律，与所谓的最大熵原理一起，使我们可以从宏观数据推导出系统的某些微观性质。例如，我们可以直接导出气体的速度分布函数。在本书中，我们实际上只关心开放系统（图1.20）。这些系统是靠连续的能量流和（或）物质流而保持它的特定状态的。正如我们将要看到的那样，传统的热力学不能用来处理这类系统，相反，我们必须建立某种新的热力学，这些将要在以后几章中详细阐明。

热力学可视为一种宏观的唯象理论。它以统计物理为基础，下一节将对此作出几点评述。

## B. 统计物理学

在这个领域，人们试图通过微观理论去推导唯象的热力学宏观定律。对于气体的单个粒子，这种微观理论可由牛顿力学，或



1.20 开放系统示意图。系统从源那里获得能量，并将其余能量通过渗池耗散掉。

量子力学所提供。通过适当的统计平均，便可从微观定律推导出宏观量。中心的概念仍然是熵  $S$ 。玻尔兹曼指出，熵与系统的相同宏观状态对应的不同微观状态数  $W$  有关，此定律可表述如下：

$$S = k \ln W \quad (1.2)$$

其中， $k$  是玻耳兹曼常数。一个关键的，至今尚未完全解决的问题在于解释，为什么宏观现象可以是不可逆的，而所有的基本定律又是可逆的。例如，牛顿力学定律对时间反演不变，也就是，当我们倒放电影时，牛顿力学允许所有的过程以相反的次序出现。另一方面，十分明显，宏观物理学中的过程是不可逆的。例如，当我们打开一个充满气体分子的气体容器的隔板时，气体将跑到第二个容器中去，并均匀地充满两个容器。相反的过程，即第二个容器自发地变空，所有的气体分子自动返回到原先的容器，这种现象从未在自然界出现过。

此外，我们并不能证明各种求平均的步骤总是合理的。另一个方面，我们用统计物理能够阐明不可逆热力学的很多现象，诸

如弛豫过程、热传导、分子扩散，等等。

### C. 协同学

协同学对复杂系统所遵从的普遍规律的定量表述进行了第三次探索。在这个领域中，我们研究能以自组织方式形成空间、时间或功能结构的各种系统。在物理学中，协同学研究远离热平衡的系统，典型的例子是从下面加热的液体，或激光。化学和生物系统也可想象为物理系统，因而仍然可用协同学处理，然而协同学也处理其它系统，诸如经济学或社会学中的系统。在协同学中，我们把注意力集中于宏观变化和质变，由此产生新的结构或功能。宏观变化和质变这些限制便是我们为了找到一般原理所付出的代价。

我们将在第二章让读者回顾协同学的主要方法。我们将看到，协同学以微观表述（如物理学中的微观运动方程）为出发点。在其它领域，诸如生物学或化学，中观方法可能是适当的，这时，我们从相应的子系统出发，例如在生物学中，从整个细胞的某些适当性质出发，然后，假定所研究的系统受到一些外界约束，例如输入系统的一定的能量。于是，当改变这个控制参数时，便可出现不稳定性，从而将系统驱于一种新的状态。

正如协同学中证明的那样，在这种不稳定点处，一般会有几个集体模变为不稳定的，它们将成为描述宏观模式的“序参量”。同时，这些宏观变量，即序参量通过“伺服原理”支配微观成分的行为。以这种方式，序参量出现，并以其支配能力使系统获得它自己的结构。当控制参数一再改变时，系统会经历一个不稳定性序列，同时伴随着一系列的结构更替。

协同学大概是一个无止境的领域，而我们仅仅迈出了最初的



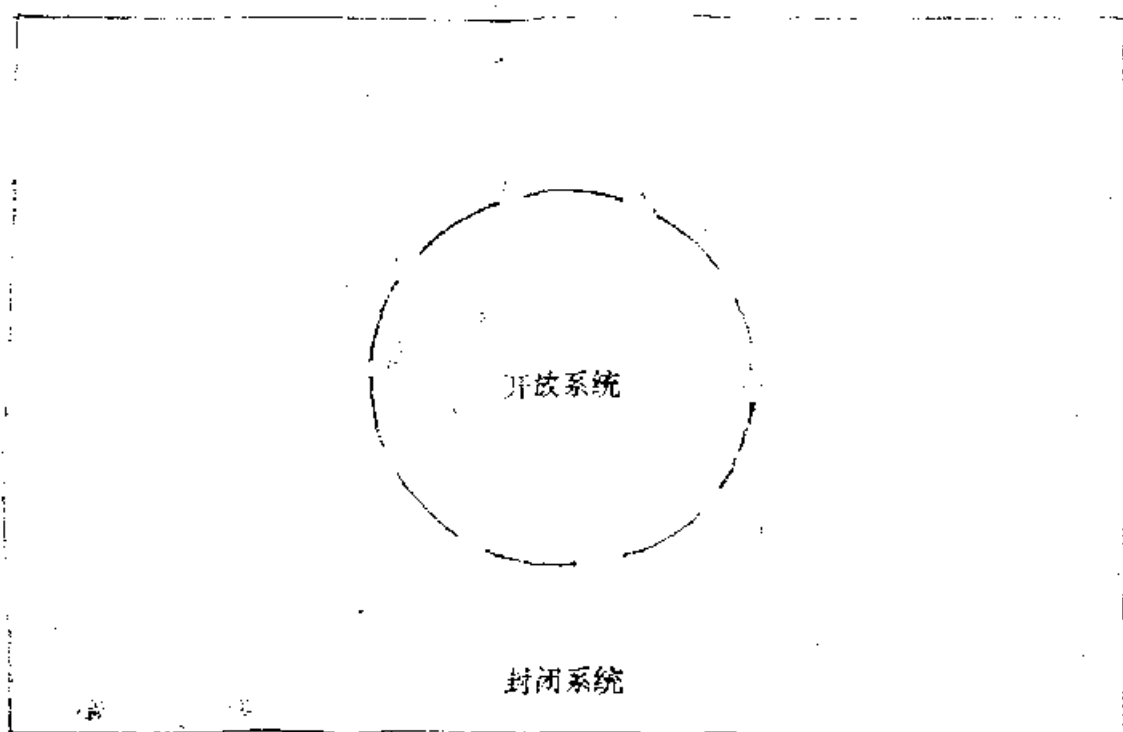


图1.21 嵌入一个封闭系统中的开放系统

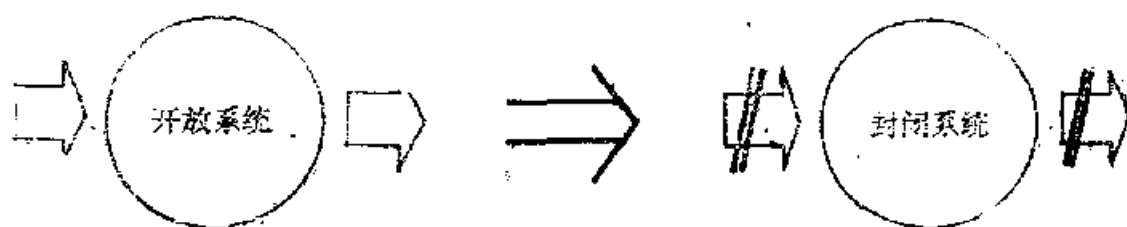


图1.22 封闭系统可视为将流入流出开放系统的能流切断的极限情况

几步。在过去一二十年中，我们证明了大量系统的行为服从协同学的一般规律，我相信，在将来会找到更多这样的例子。另一方面，我们必须认识到，仍有可能发现更多、更普遍的规律。

如前所述，热力学研究热平衡系统，而协同学研究远离热平衡的系统。这里出现相当奇特的情形。一方面，我们总可以将开放系统嵌入一个更大的封闭系统（图1.21）。例如，地球是一个开放系统，它从太阳那里吸收能量，并在夜间把能量辐射到宇宙空间。但如果我们将太阳和宇宙的一部分视为一个系统，则我们可以把整个系统看作一个封闭系统，热力学定律对它适用。到此为止我们看到，协同学的规律必须与热力学定律不矛盾。但在另一方面，在能量或物质流趋于零的极限情况下，任何开放系统都将成为封闭系统，因此，最终我们处理的是一个封闭系统（1.22）。因而，热力学的一般定律必须是协同学规律的极限情况。

读者会注意到，这个方案至今尚未完成，还有待于将来去探索。在这本书中，我们的目的是探索协同学的第二个基点，我们在1.5到1.7节中详细讨论这一点，更详细的讨论放在以后几章。这条途径的出发点是信息概念，我们在下节将涉及到它的一些最重要的方面。

## 1.7 信息

信息一词的使用是很混乱的，这是由于人们在极不相同的意义上使用信息一词所造成的。在日常语言中，信息含有消息或通知的意思。信、电视、电话传送等均携带着信息。在下面，我将讨论科学上使用的信息一词，我们从申农信息的概念出发，这里使用的信息不含任何意义。接着，我们将简短地研究关于消息

的信息，在最后讨论意义的自创生问题。

显然，当涉及物理系统时，我们希望摆脱各种拟人论，因为我们想以尽可能客观的方式描述物理系统。然而生物学中这种类似的倾向相当明显，因此最后我们对生物系统得出一种物理的、甚至力学的图景。但是，非常奇怪，在现代计算机比如第五代计算机的发展上，似乎我们又希望重新引入意义、关联等概念。因而本节打算讨论怎样从消除了意义的信息概念回到意义的自创生。

### A. 申农信息：意义的消除

第三章，我们将详细讨论申农信息的概念。为了使现在的讨论有坚实的基础，我们用几个例子阐明申农信息的概念。掷一个硬币，有两种结局；扔一个骰子，有六种结局。掷硬币的情况下，我们有两种信息，正面或背面；掷骰子的情况下，我们有从 1 到 6 的某个数目将会出现的信息。我们同样可以回答“是”或“非”等等。申农信息的概念正是来源于可能性的数目  $Z$ ，掷硬币时，它是 2，掷骰子时，它是 6。正如我们将在下文中看到的那样，对信息的正确量度不是数  $Z$  本身，而是它的对数，通常取以 2 为底的对数，也就是把信息定义为

$$I = \log_2 Z \quad (1.3)$$

这个定义可以化为另一种形式，本书中它将一再出现。举一个语言的例子，比如，英语。我们可以用数字  $j=1, 2, \dots$  标记字母  $a, b, c, \dots$ ，即  $a \rightarrow 1, b \rightarrow 2$  等。然后可以计算这些字母在书中或图书室里或其它场合出现的次数  $N_j$ 。我们定义指标为  $j$  的字母出现的相对次数为

$$P_j = \frac{N_j}{N} \quad (1.4)$$

其中  $N$  为所有字母出现的数目， $N = \sum N_j$ 。因此，在该书中（或图

书室里，等等）含有每个字母的平均信息为

$$i = - \sum_j P_j \ln P_j \quad (1.5)$$

这公式的推导请参看第三章。申农曾使用他的概念，去研究甚至在噪声影响下传输信息通道的容量。申农信息的下述特点在下文中相当重要。（i）申农信息同任何意义无关。因此，不存在诸如有意义或无意义、有目的等概念。（ii）申农信息是对封闭系统而言。只有一个不变的消息库，它的数目是 $Z$ 。

### B. 信息的效应

在这一小节中，我们打算引入一种新的方法，它朝着包括语义学的信息概念前进了一步。注意到，只有当我们将消息接收者的反应也考虑进去时，才能赋予消息一意义，这样，我们便得出一基本概念，这就是下面将要说明的消息的“相对重要性”概念。

考查一组消息，给每个消息都标定一列数字。中心的问题在于模拟接收机。为此我们借用现代动力系统理论的概念，或一般地说，借用协同学的概念，我们用一动力系统来模拟接收机。以后（即第二章）我们将用数学来描述这种系统。为了满足现在的需要，只需作几点一般性的评述，就足以够用了。我们来讨论一个系统，比如气体、生物细胞或者经济学，其状态可以在微观、中观或宏观层次上用一组量 $q_i$ 来表征，随着时间的推移，诸 $q_i$ 可能改变。我们可以把诸 $q_i$ 放在一起，写成态矢量 $\underline{q}(t) = (q_1(t), \dots, q_N(t))$ 。 $\underline{q}$ 的时间演化（即系统的动力学）便由微分方程

$$\frac{d\underline{q}}{dt} = \underline{N}(\underline{q}, \alpha) + \underline{F}(t) \quad (1.6)$$

来确定，这里的 $\underline{N}$ 是决定论部分，而 $\underline{F}$ 代表涨落力。我们目前

需要完全了解的是下述问题：如果没有涨落力，只要给定初始时刻的  $q$  值，而且所谓的控制参数  $\alpha$  保持不变，则  $q$  的未来进程是唯一确定的。随着时间的推移， $q$  将趋于一种吸引子态。为把这种吸引子的一个简单例子形象化，请看用纸模拟的带有一些丘陵和山谷的小型地貌（图1.23），固定  $\alpha$  就意味着选择了一种特殊的地貌，在重力（以及在摩擦力）作用下，小球可以在地貌中滑动。把  $q$  调到初始时刻的意思是指最初把球置于一个特定的位置。例如放在山丘的斜坡上（图1.24）。球将从这里下滑到谷的底部，这个底部就是一个吸引子。正如行家们所知的那样，各种动力系统也可能具有其它类型的吸引子，比如极限环——系统持续进行的一种振荡，或者更为复杂的吸引子，例如“混沌吸引子”。在下面，通过研究象谷底的那种吸引子（所谓的不动点）可以阐明我们的概念。当出现涨落  $F$  时，球可能从一个吸引子态跳到另一个吸引子态（图1.25）。



图1.23 用纸折成的微型地貌图形象地表示具有不动点吸引子的动力系统

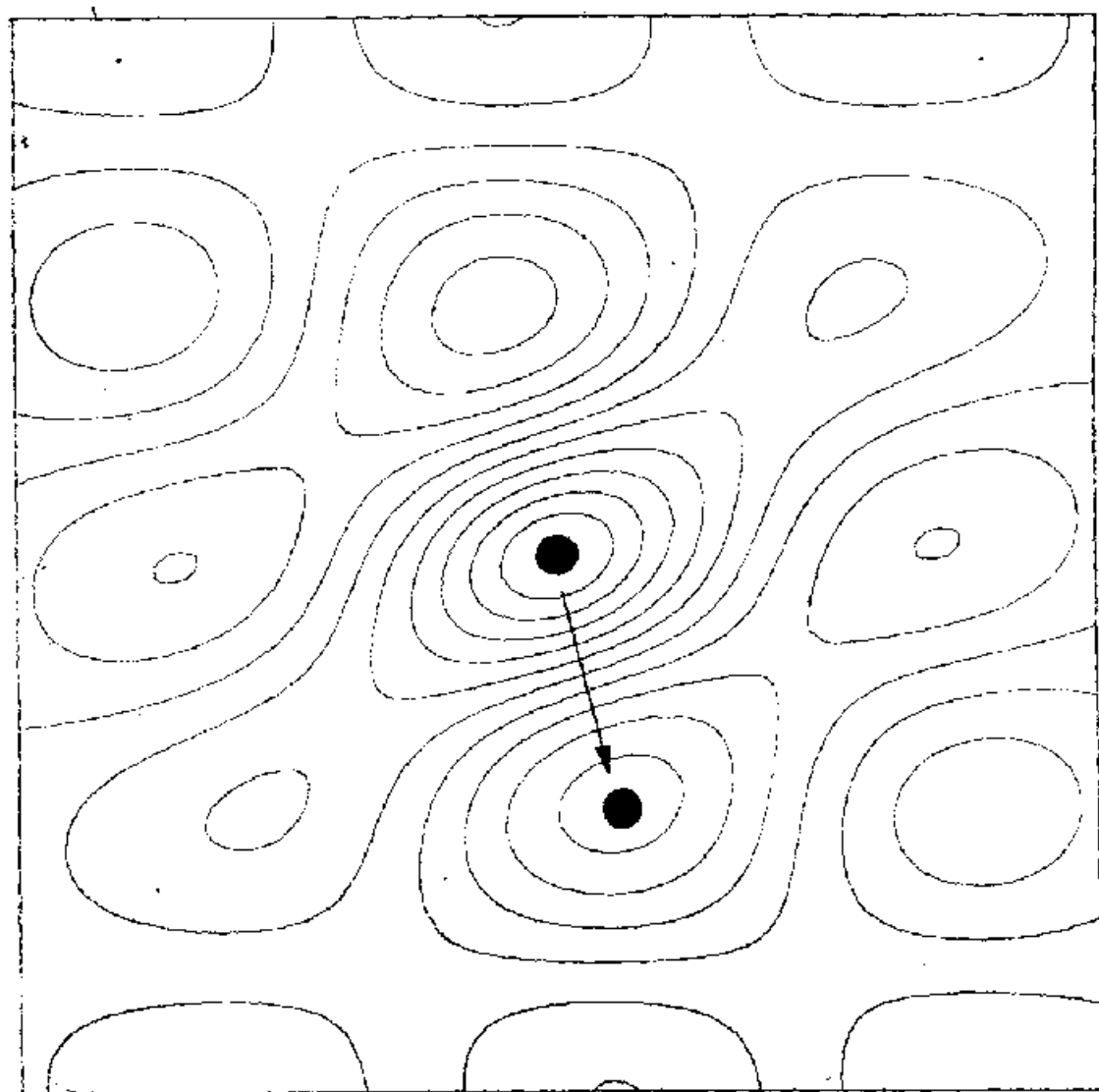
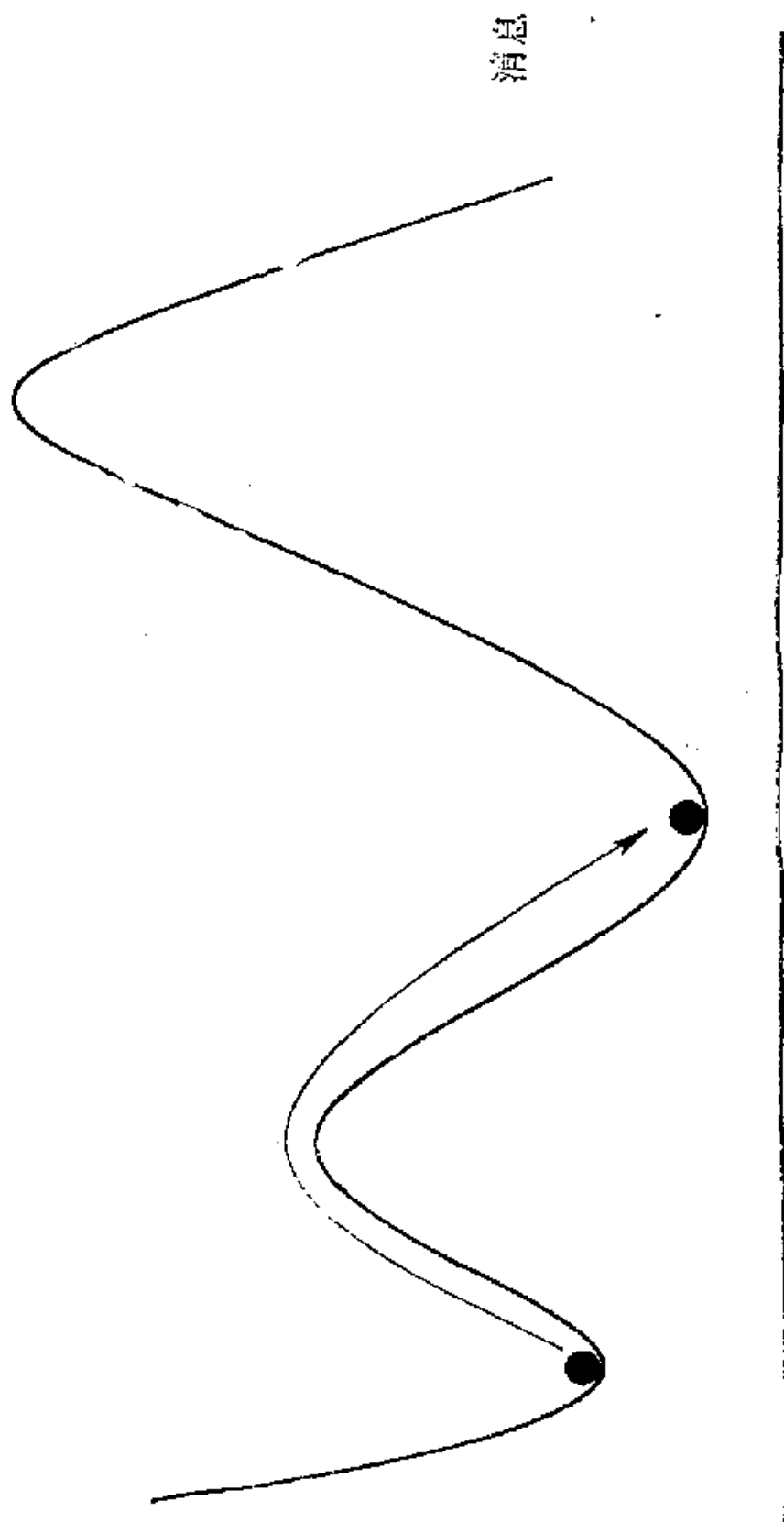


图1.24 图1.23的等高图



### 在吸引子之间的跃迁

图1.25 系统在两个吸引子之间的跃迁

做了上述准备之后，我们讨论原来的问题——给消息赋予意义的问题。我们假定，系统收到消息就意味着参数 $\alpha$ 和 $\underline{q}$ 的初始值被该消息所确定。我们暂时假定，这些参数是唯一确定的。把这种理论推广到消息不完整的情况是很直截了当的事（见下文）。我们先忽略涨落，并假定，在消息来到之前，系统已处于一个吸引子态，这称为中性状态。吸引子可能是一种静止的状态，也就是一个不动点，同样也可能是一个极限环、环面或奇怪吸引子，或者尚待动力系统理论发现的某种形式的吸引子。我们称其为吸引子 $\underline{q}_0$ 。收到消息后，参数 $\alpha$ 和初始值 $\underline{q}_0$ 可能会重新取值，原则上，这里会发生两件事。假定允许等待某一度量时间，以致我们可以假定动力系统处于一个可能的吸引子态，这时，情况之一是，消息使系统仍留在 $\underline{q}_0$ 状态。在这种情况下，消息显然是无用的或没有意义的。

另一种情况是，系统进入一种新的吸引子。我们首先假定，这个吸引子是由事件的消息唯一确定的。显然，不同的消息可能导致相同的吸引子。在这种情况下，我们将称为消息过剩。

最后，尤其在生物学领域至今仍有一个谜：信息是如何产生的？然而，如果假定事件的消息如图1.26所描绘的那样（它显然是含糊的）绘出，则容易把这个问题形象化。依据系统内本身的涨落，两个新的稳定点（或吸引子）可以被实现。这里，事件的消息包含通常词义的信息，但它是含糊的，这种含糊性是由系统的涨落来澄清的。不严格地说，原来的信息加倍了，因为这时有两个可能实现的吸引子。

在生物学的情形下，变异产生了这种涨落。另一方面，在物理学领域中，我们应该确切地说，这涉及到对称破缺效应。

把所有这些不同的过程放在一起，我们得到如图1.27的基本



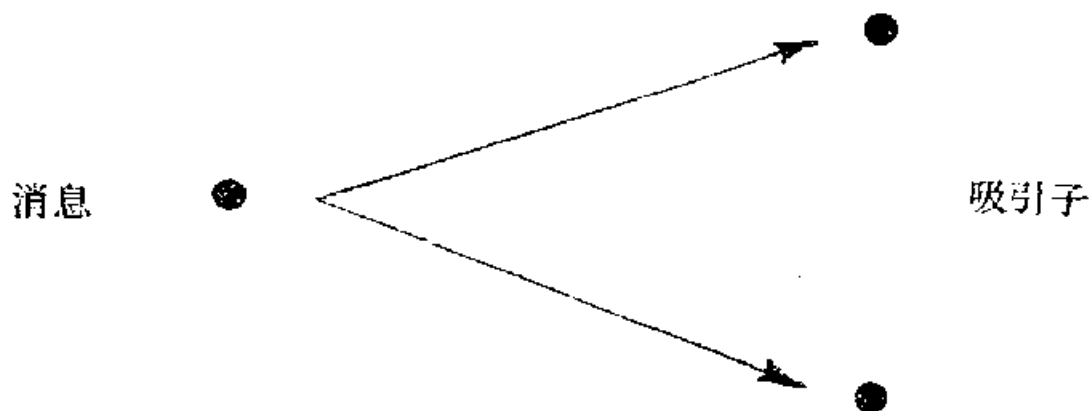


图1.26 通过图1.25所示的过程，一个消息可由于系统内的涨落到达两个不同的吸引子上，由此可见两个吸引子容易受影响。

流程，当然，当我们研究不同消息的效应时，可能形成更复杂的流程，诸如图1.28所示的情形。

现在我们将用下述方式处理问题，使得我们可以赋值予事件消息，或更确切地说，我们打算定义一个“消息的相对重要性”。为此，我们首先必须引入各个吸引子的“相对重要性”，实际上，各个吸引子将是新消息的起源，于是可把它置于新的动力系统，这一过程可无穷往复。但对实际应用来说，我们必须在某一级将这种链式序列截断，并在这一级对每一吸引子赋以相对重要性之值。考虑单步过程，就己能清楚地阐明我们的步骤了。下面，我们详细地分析这一过程。

我们把相对重要性赋予各个吸引子，第0个吸引子  $q_0$  的数值为0，而其它一些吸引子的可能数值为  $0 \leq P_i' \leq 1$ ，我们把它归一化，即

$$\sum_i P_i' = 1 \quad (1.7)$$

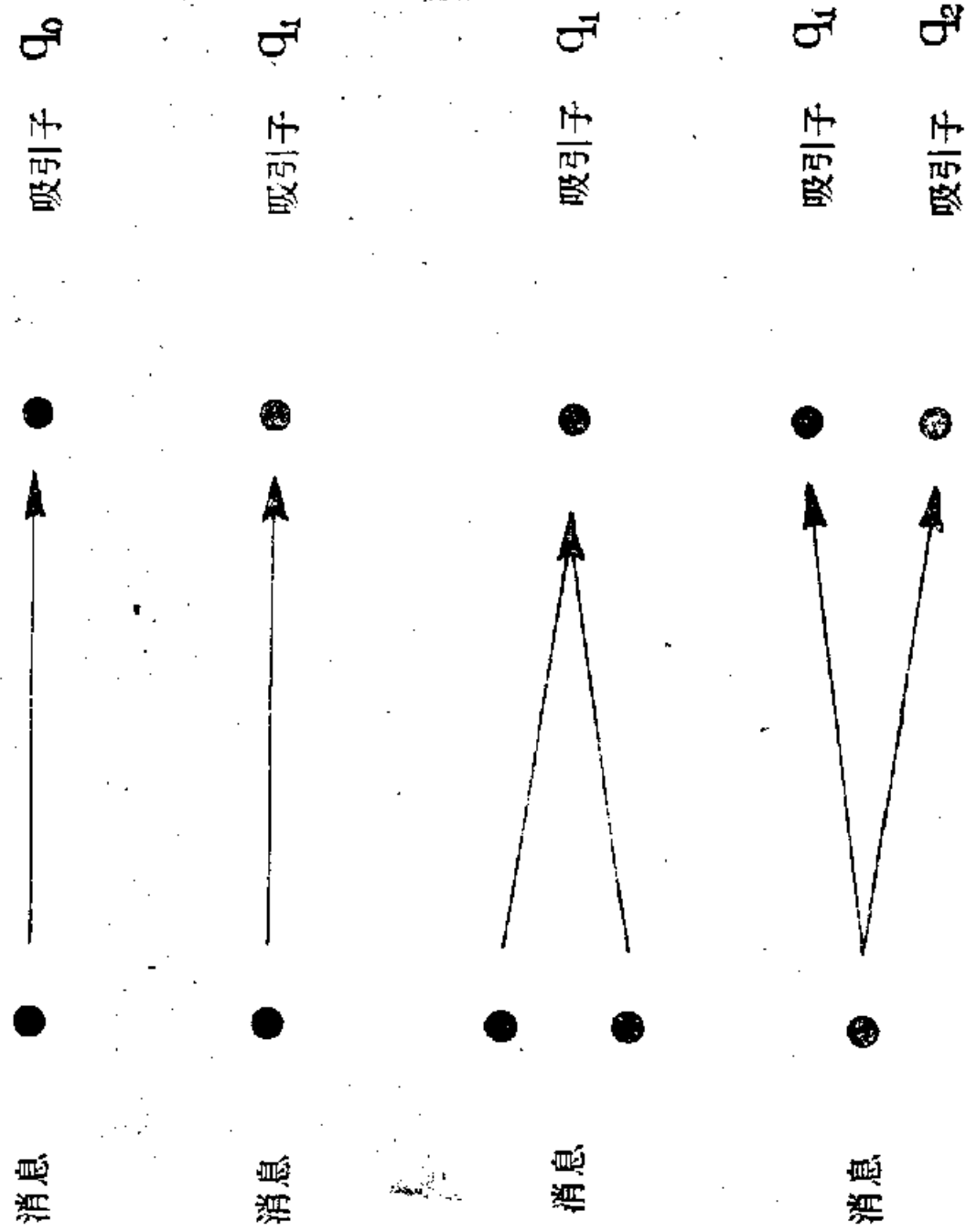


图1.27 消息到达吸引子的各种可能性

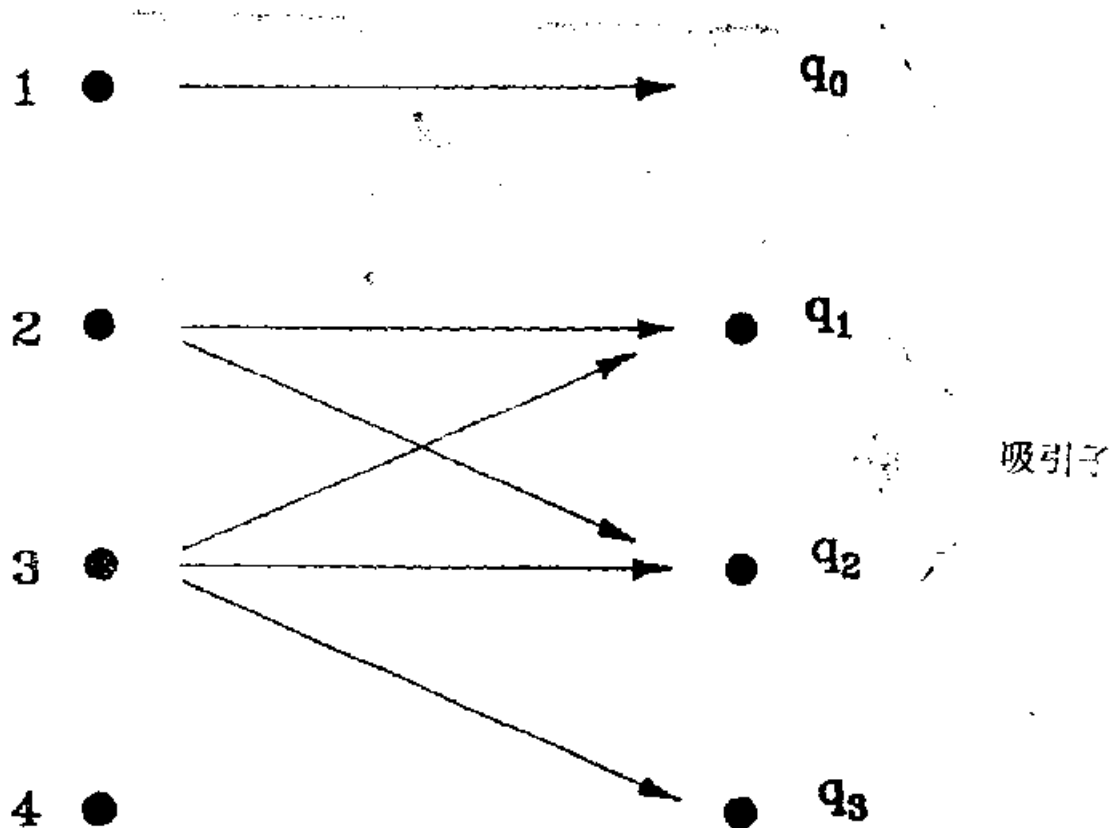


图1.28 消息达到吸引子的另一个例子

$P_i'$  的确定依赖于动力系统必须完成的任务。我们可以设想这是一个特定吸引子恰巧能完成的一项特定任务，或者是一个任务系综。任务或任务系综的实现，具有给定的相对重要性。显然，消息的相对重要性  $P_i$  不仅依赖于动力系统，而且依赖于它必须完成的任务。现在的问题是：各消息的数值  $P_i$  是什么？为此，我们研究消息和吸引子（动力系统在收到消息后，被驱动于该吸引子）的联系。如果接收到消息  $i$  后，系统达到吸引子  $k$ （包括第 0 吸引子）上，我们对该过程赋以矩阵元  $M_{ik} = 1$ （或  $= 0$ ）。如果考虑到系统

的内部涨落，单个信息便可借助于涨落将系统驱动到几个不同的吸引子上，其中，每个吸引子具有分岔率 $M_{ik}$ ，且 $\sum_k M_{ik} = 1$ 。我们定义相对重要性 $P_i$ ，

$$P_i = \sum_k L_{ik} P_k' = \sum_k \frac{M_{ik}}{\sum_{j'} M_{j'k} + \varepsilon} P_k' \quad (1.8)$$

这里令 $\varepsilon \rightarrow 0$ （这个技巧是为了达到这样的目的：当分子、分母同时等于零时比值保持确定）。首先假定，对任一个 $P_k' \neq 0$ 至少有一个 $M_{ik} \neq 0$ 。我们很容易验证 $P_i$ 是归一化的，这可按下述步骤证明：

$$\sum_i P_i = \sum_k \frac{\sum_i M_{ik}}{\sum_{j'} M_{j'k} + \varepsilon} P_k' \quad (1.9)$$

$$= \sum_k \left( \sum_{i'} \frac{M_{i'k}}{\sum_{j'} M_{j'k} + \varepsilon} \right) P_k' \quad (1.10)$$

$$= \sum_k P_k' \quad (1.11)$$

其中，式（1.10）中的括号等于1。

现在来研究在 $P_k' \neq 0$ 时，对于一切 $j$ ， $M_{jk} = 0$ 的情况。在这种情况下，当式（1.9）和式（1.10）对 $k$ 取和时，有些 $P_k' \neq 0$ 的系数等于零，但由于 $\sum P_k' = 1$ ，我们得到 $\sum P_i < 1$ 。如果此不等式成立，我们就叫做信息欠缺。

以较抽象的方式，我们可以把式（1.8）左边等式作为一种基本定义，其中假定

$$\sum_k L_{ik} \leq 1 \quad (1.12)$$

其中等号在不存在信息欠缺的情况下成立。

注意，规定  $M_{jk} = 1$ （在事件消息达到单个最终吸引子的情况下），可以推广为

$$0 < M_{jk} \leq 1 \quad (1.13)$$

当几个系统依次耦合时，式（1.8）的左边等式使我们可以立刻写出公式。例如在两步过程中，我们立刻得到

$$P_j = \sum_k L_{jk}^{(1)} P_k' = \sum_{kk'} L_{jk}^{(1)} L_{kk'}^{(2)} P_k'', \quad (1.14)$$

我们很容易验证，只要  $\sum P_k' = 1$  和  $\sum L_{jk} = 1$ ，则  $\sum P_j = 1$ 。则单个步骤写成

$$\begin{aligned} \sum_j P_j &= \sum_j \sum_{kk'} L_{jk} L_{kk'} P_k'' \\ &= \sum_{kk'} \underbrace{\left( \sum_j L_{jk} \right)}_{=1} L_{kk'} P_k'' \end{aligned} \quad (1.15)$$

$$= \sum_{k'} \underbrace{\sum_k L_{kk'}}_{=1} P_k'' = 1 \quad (1.16)$$

我们可以定义

$$L'_{jk} = \sum_k L_{jk}^{(1)} L_{kk'}^{(2)} \quad (1.17)$$

因为诸  $L$  是正定的，我们得到

$$L'_{jk} \geq 0 \quad (1.18)$$

根据归一化性质（在没有信息欠缺的情况下），

$$\begin{aligned}\sum_i L'_{ik} &= \sum_k \sum_i L_{ik}^{(1)} L_{ik}^{(2)} \\ &= 1 \\ &= \sum_k L_{kk}^{(2)} = 1\end{aligned}\quad (1.19)$$

我们直接得到

$$L'_{ik} \leq 1 \quad (1.20)$$

因此  $L'_{ik}$  遵守不等式

$$0 \leq L'_{ik} \leq 1 \quad (1.21)$$

应当注意从  $P''$  (或更高阶  $P^{(n)}$ ) 到  $P$  的递推关系依赖于路径。

上述方法不仅引入了消息相对重要性的概念, 而且给我们提供了确定  $P_i$  的算法, 而  $P_i$  具有一定的概念意义和实际意义。

知道了某种任务或任务系统时, 这种算法使我们可以挑选出要发送的消息, 即具有最大  $P_i$  的消息。如果有同样大小的几个  $P_i$ , 那么, 发送哪个消息是无所谓的。从概念的角度看, 我们可以确定, 动力系统是湮灭了信息, 还是使信息守恒, 或是产生了信息。为此, 我们使用传统信息理论意义上的信息概念。但代替由符号的相对频率引起的信息量, 我们使用一组消息内的相对重要性的概念, 也就是我们将引入量

$$S^{(0)} = - \sum P_i \ln P_i \quad (1.22)$$

$$\text{和} \quad S^{(1)} = - \sum P_{i'} \ln P_{i'} \quad (1.23)$$

其中,  $P_i$ ,  $P_{i'}$  已在式 (1.8) 中引入了。若按我们的假定,  $\sum P_{i'} = 1$  且当  $\sum P_i < 1$  时, 则出现信息欠缺。在  $\sum P_i = 1$  的情况下, 如果

$$S^{(1)} \leq S^{(0)} \quad (1.24)$$

成立，我们就称作信息湮灭。如果

$$S^{(1)} = S^{(0)} \quad (1.25)$$

成立，我们就称作信息守恒。如果

$$S^{(1)} > S^{(0)} \quad (1.26)$$

成立，我们就称作信息产生。当我们讨论一些特殊情况时，这些定义的含义很快就会清楚。例如，如果两个消息导致同一个吸引子，这时系统过剩，信息量（就传统的意义上说）变小。它从

$$\begin{aligned} S^{(0)} &= -k \left( \frac{1}{2} \ln \left( \frac{1}{2} \right) + \frac{1}{2} \ln \left( \frac{1}{2} \right) \right) \\ &= k \ln 2 \end{aligned} \quad (1.27)$$

减少到

$$S^{(1)} = -k \cdot 1 \cdot \ln 1 = 0 \quad (1.28)$$

在  $P_i$  到  $P_i'$  的一对一映象中，我们发现将  $\{P_i\}$  变换为同一组的  $\{P_i'\}$ ，大概只不过是指标进行置换，即对状态重新编号而已。

在这种情况下，显然式(1.25)成立。最后在式(1.26)的情况下，例如只有一个  $P_1 = 1$ ，而所有其它的  $P_i$  等于零，变换后的情况为， $P' = P'' = \frac{1}{2}$ ，而所有其它的等于零。于是从  $S^{(0)} = -k \cdot 1 \cdot \ln 1 = 0$  增大到

$$\begin{aligned} S^{(1)} &= -k \left( \frac{1}{2} \ln \frac{1}{2} + \frac{1}{2} \ln \frac{1}{2} \right) \\ &= k \ln 2 \end{aligned} \quad (1.29)$$

当然，举这些例，并不意味着已经证明了式(1.24)一式(1.26)所示的定义，而只不过是说明一下它们的含意罢了。

基于协同学的方法具有一些鼓舞人心的进一步的特点。语义学已成为研究动力系统响应（吸引子）的一个问题，系统可以是校正错误的（或者是补充所缺少的信息的）。如果事件消息没有将初始状态  $\underline{q}$  置于吸引子上（这是不正确的），它可能将  $\underline{q}$  置

入吸引域内，亦即置于特定谷底（即吸引子或不动点）周围的山坡上，这种情况下，系统会把态矢量拽回到该吸引域内的吸引子上，即进入正确状态。一个有趣的问题是确定实现一个给定吸引子（或一个给定的“相对重要性”之值）所需的最小彼特数。

我们现在的方案也可模拟系统的认识过程。对于消息 $j$ ，一个系统能够被“敏化”或“钝化”，例如通过较多或较少的参数对特定消息起反应而“敏化”或“钝化”。

在上述处理中，我们假设了消息值是对接收者的同一初始态来测量的。在下一步考虑时，我们假定消息施在处于另一初始态的接收者上，该初始状态是由以前的消息所设置的。

以这种方式，我们得消息的干涉与消息的相对重要性依赖于接收者以前接收到的消息。特别是在一般情况下，消息的相对重要性将以不可对易的方式依赖于后继消息。这样，接收者一再被消息所变革。显然，消息的相对重要程度将成为时间的函数。

下述评论对协同过程尤为重要。一个协同系统必须不仅是一个表现出极限环或混沌行为的动力系统，而且也可能是一种发生不可逆过程的系统，该不可逆过程导致系统从无组织状态进入有结构状态，例如从液体变成固体。

作为本部分的结束，让我们评论一下模式识别吧。对于模式识别，我们将在本书中以多种观点阐明它，它可视为是接收者（例如大脑或机器）对入射消息的一种加工处理。因此，用以上所概述的想法讨论模式识别将是一件有趣的工作。我提醒大家，至少一般地说来，模式识别是一个多步过程，其中接收者扮演着很活跃的角色”。第一步，接收者在一个全局的层次上收到图样，这里，一般有几个吸引子可以达到。然后，要求感觉系统将其注意



力集中于探测其它特征元，以便选择出一组更精细的吸引子。为更明确起见，现举例说明。第一步，我们确定了物体的总体轮廓线，例如，近似于一个圆形、一个矩形，等等。然后，如果是圆的，便有几个吸引子：苹果、脸、车轮、树。这时，接收者要求给出进一步的细节，如颜色、垂直线（会不会是鼻子？）等，这样，此过程可继续下去。

注意到，我们对模式识别的解释与“传统的”方法有所不同，我们将在十二章回到后者。在传统的方法中，首先将模式分解为“本原素”或“特征元”。而这里，我们从总体模式（轮廓线）开始，然后越来越细。

我们的方法还对人类识别图样提供了一个解释（或至少是一个启示），这就是为什么中断的轮廓线得到补足，使得我们所“看”到的是连续的线。

### C. 意义的自创生

如前已及，协同学可视为关于在宏观层次上新质出现的理论。如果合理解释协同学的结果，我们可因此将意义的出现作为系统新质的出现来研究，或换句话说，称为意义的自创生。为了探其究竟，我们拟将物理系统（如激光）与几个生物学模型系统作一比较。作为开始，让我们对信息在生物系统的作用做些一般评述。

生物系统最惊人的特点之一在于各部分之间的高度协调。在一个细胞中，数以千计的代谢过程在同一时刻有条不紊地进行着。在动物体内，成千上万个神经和肌肉细胞密切协作，产生了井然有序的运动，心跳，呼吸或血流。识别是一种高度合作的过程，人类的思维和言谈也无不如此。显然，所有这些高度协调、密切相关的过程只有通过交换信息才可能实现。例如，我们将会

看到，这些信息被产生、传输、接收、处理，还要转换成信息的新质，并同时在系统的不同部分之间和不同的层次之间交流。我们因此得出：信息是生命赖以存在的至为关键的元素。

信息的概念异常微妙。进一步阐明其某些方面将是本节的目的。我们会看到，信息不只与通道容量相联系，也不只与系统的中枢神经对各部分的发号施令有关，它还具有“媒介”的作用，系统的各部分对此媒介的存在作出贡献，又从它那里得到怎样以相干的、合作的方式来行动的信息。在这个层次上，出现了语义学。

让我们先看看物理学。对于封闭系统，第二定律是适用的，它告诉我们，结构将会解体，至少在宏观层次上系统趋于均匀。在微观层次上，出现了一派混沌。由于这个原因，信息不可能产生于封闭的热平衡系统。热平衡系统甚至也不能储存信息。让我们看一个典型的例子，如一本书，乍一看，它似乎是处于热平衡态，确实，我们可以测量其温度。尽管如此，它还没有达到最终的热平衡态。随着日走星移，每一字母的墨渍将会扩散消失，直至达到一个均匀的状态。

这一简单例子告诉我们，任何由封闭系统构成的存储器都不是处于平衡态的。无论如何，我们都必须问一问，在每一给定情况下，信息能够储存多久。因此，让我们考虑一个开放系统，它靠能量流和（或）物质流维持在远离热平衡的状态。如前所及，在开放系统中，甚至在无生命界，特定的空间或时间结构会以自组织的方式产生。产生相干光的激光器、形成特定空间或时间图样的流体以及呈现出连续振荡、空间涡旋或同心波的化学反应都是这方面的例子。在这个层次上，我们已经可以在某种程度上谈信息的产生或储存了。但另一方面，我们很难为这些过程赋予如关

系、目的或意义等字眼。

让我们更详细地讨论一下激光，因为它能使我们引入一种对生物或其他系统也很有用的术语。在激光器的晶体（例如红宝石）中嵌入大量的原子（见图1.29），当从外界激发这些原子

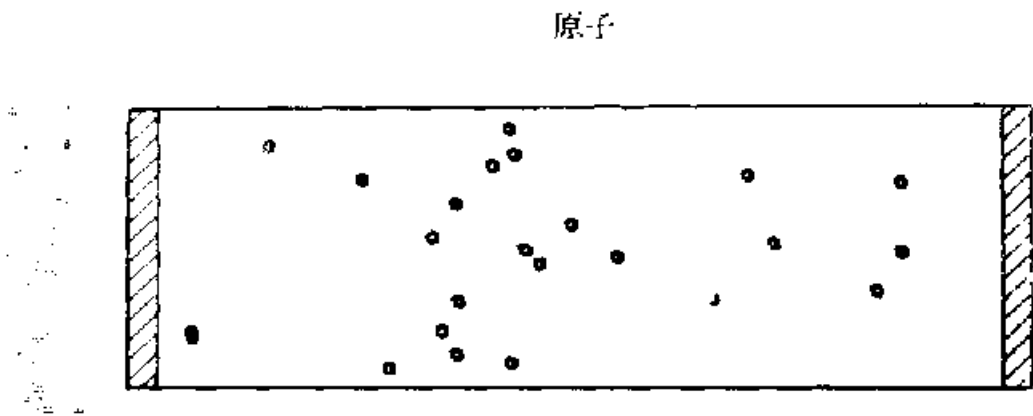


图1.29 嵌入一激光器的晶体中的原子

时，原子会发出一些独立的波列（见图1.30）。因此，每一原子便发射一信号，即它们创生了由光场所携带的信息。这些所发出

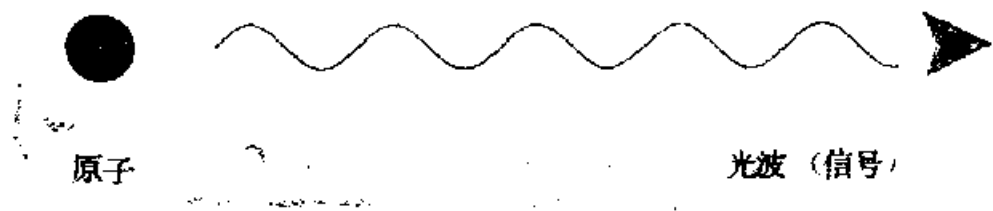


图1.30 一个激发原子发射的光波（信号）

的波列会碰上其他被激发的原子，并迫使这些原子放大原来的波

列（如图1.31所示）。这样，信号便得以加强，因而信息起着加强信号的作用（见图1.32）。由于每一原子彼此独立地发出光波，这些光波分别被不同的激发原子所放大，因此便产生了虽然是放大的，但却互不相关的波的叠加，这时观察到的是极不规则的图样，如图1.33所示。



图1.31 当光波碰上激发原子后，它可能使原子将原光波放大。



图1.32 一个链式放大过程

但是，当信号的幅度足够大以后，一个崭新的过程出现了。这时，原子开始相干振荡，场本身也变得相干，它不再由互不相关的波列构成，而变成实际上是无穷长的正弦波，如图1.34所示。

这里，我们看到了自组织的一个典型例子，没有外界的干涉，却出现了相干波的时间结构，建立了序。详细的数学分析表明，出现的相干光波起着序参量的作用，它强迫原子相干振荡，

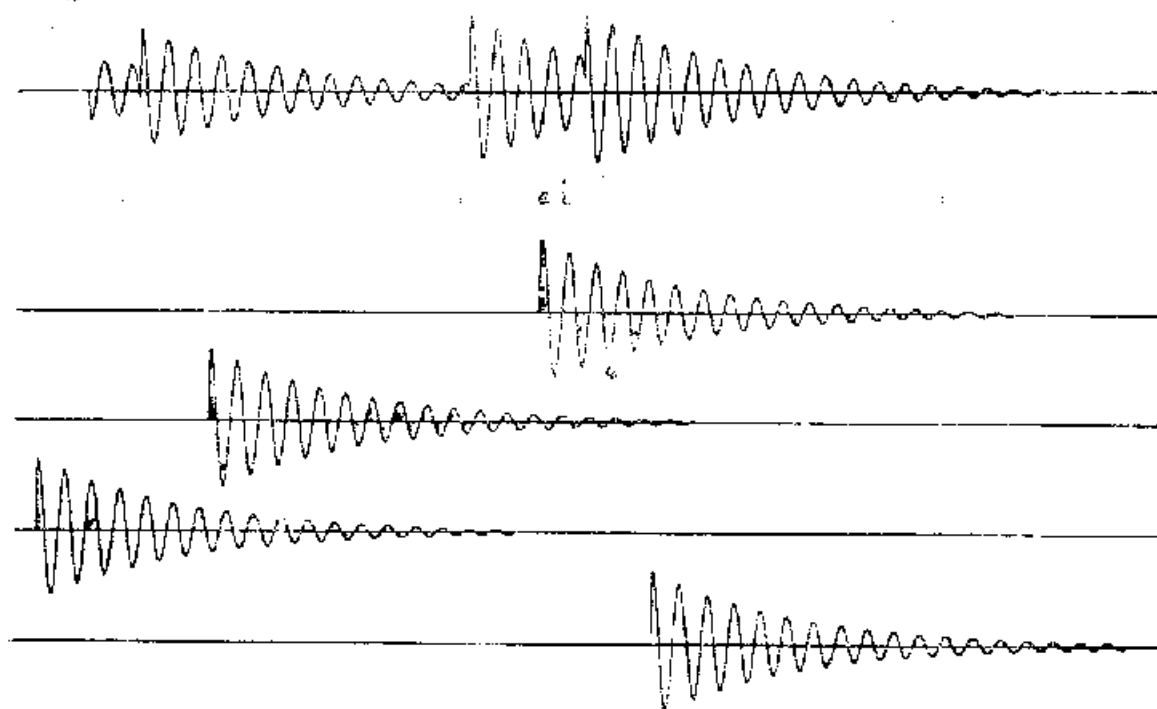


图1.33 通过光波的非相干叠加，发射出更无规则的光。

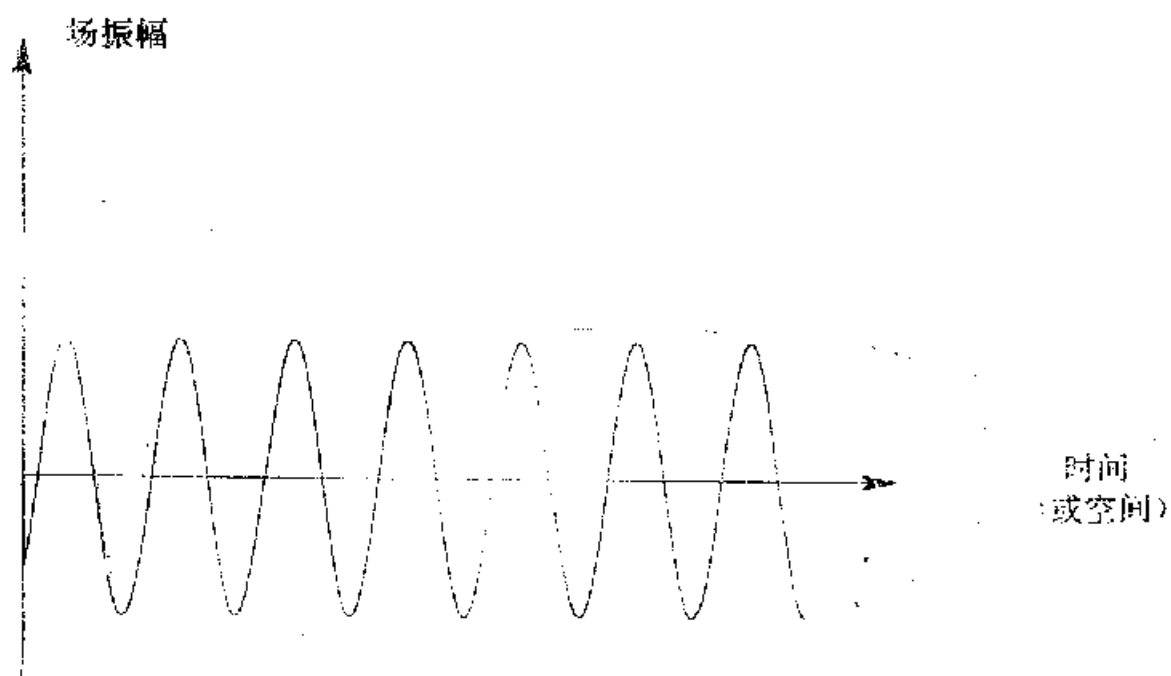


图1.34 在激光中，场振幅是实际上无限长的正弦波，只有很小的位相涨落。

或换句话说，它使役着原子（图1.35），注意，这里我们遇到了一

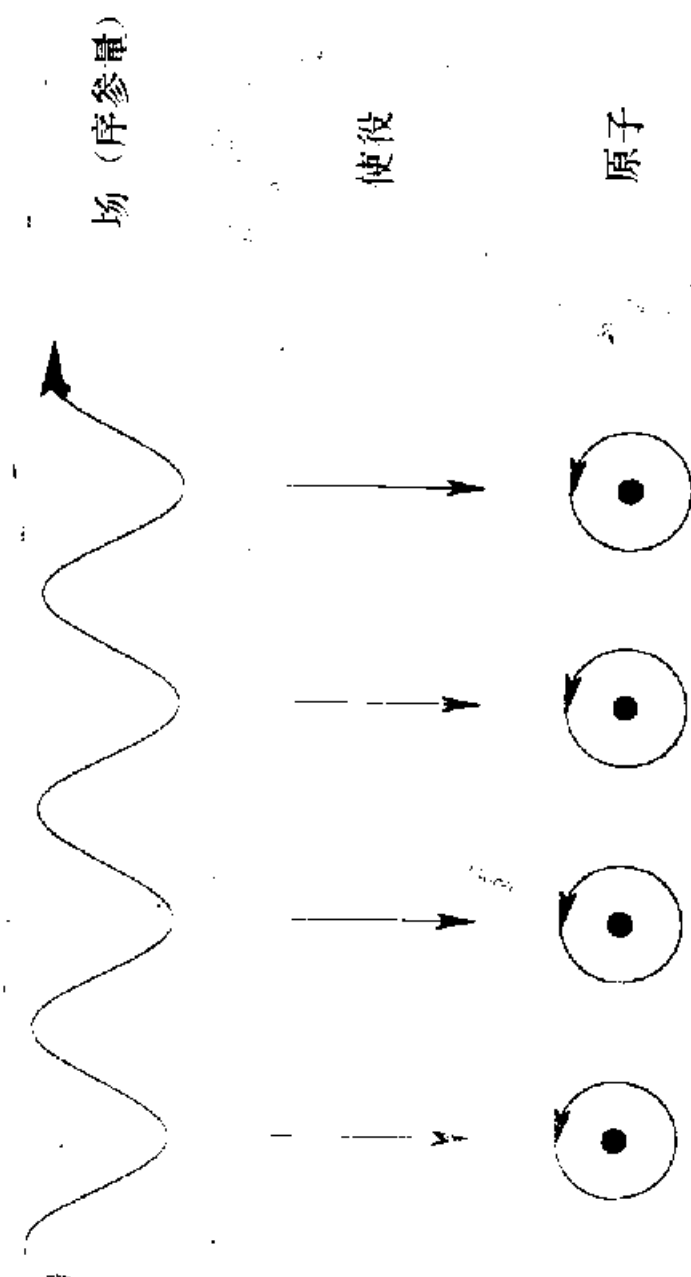


图1.35 示意。场起着序参量的作用，它决定了电子的运动，换句话说，场使役着原子。

种循环因果性，一方面序参量使役着原子；但另一方面，它又是由原子的联合行动产生的（图1.36）。

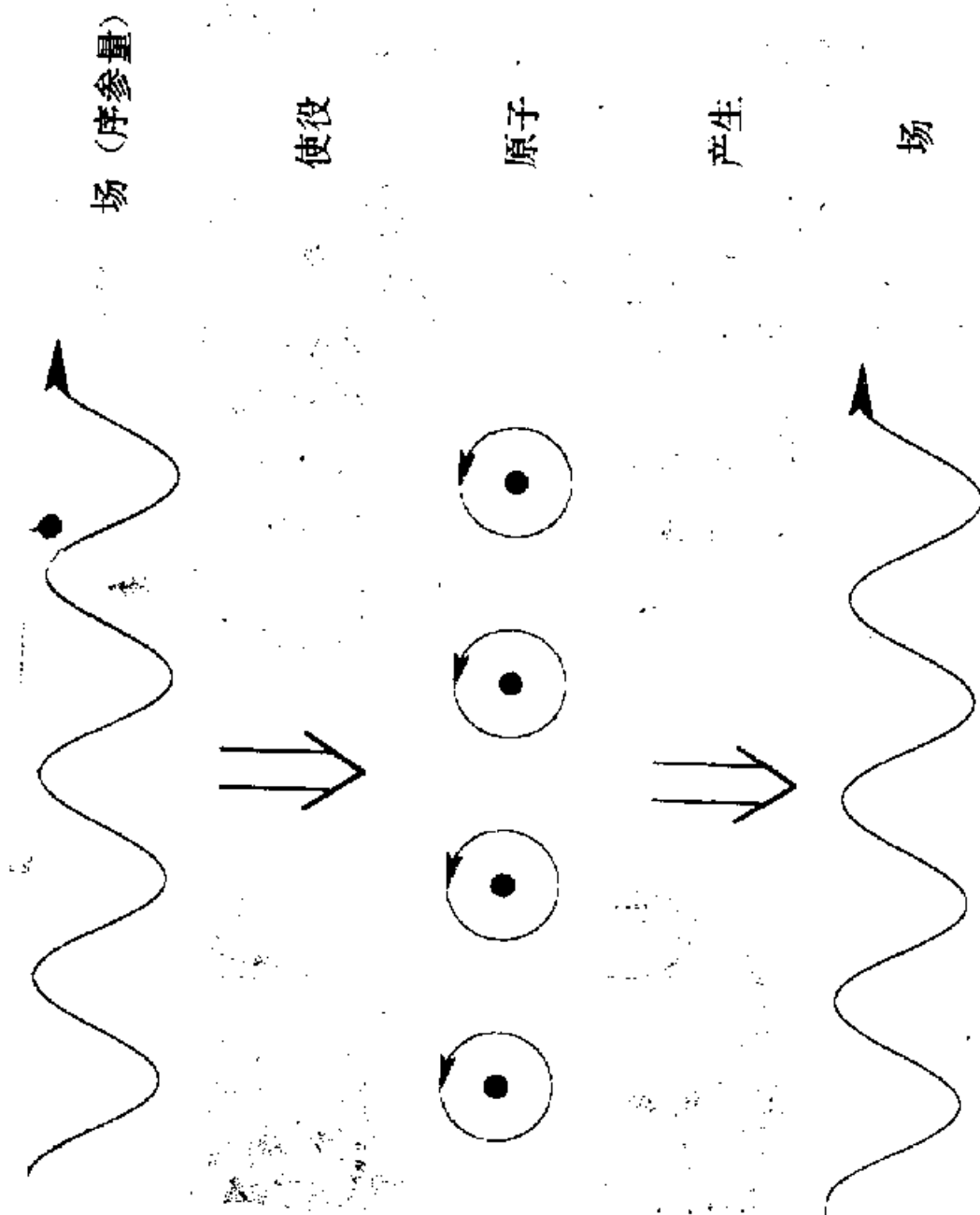


图1.36 循环因果性示意。一方面场作为序参量使役着原子，另一方面原子通过受激辐射产生了场。

从信息的观点看来，序参量起着双重作用：它通知各原子如何行动，此外，它又告诉了观察者系统的宏观有序态情况。虽然要描述所有原子状态，我们需要大量的信息，但一旦有序态已建立，便只有一个量是必须的，这就是总的光场的位相，即信息被大大压缩了。我们可以称序参量为“信息子”。在过去，我们已证明序参量的概念适用于一大类极不相同的物理、化学和生物系统。

为了说明在我们所考虑的层次上，信息交换的作用，我们考虑一个粘性霉菌的例子，通常粘性霉菌的细胞是独立地生活在基质上的，但如食物变少，它们会在某一特定点聚集起来。这种自聚集的机理如下：

起初，单个细胞发射一种物质，称为cAMP（环状单磷酸腺苷），即它们发出一种信号或消息，这就是信息。一旦cAMP分子碰上其它细胞，便会增殖，其方式与激光原子放大入射信号非常类似（见图1.37）。显然，元素本身意识不到信息的意义，但

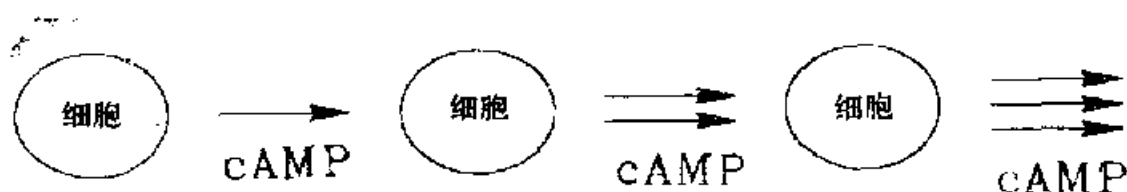


图1.37 粘性霉菌细胞中cAMP分子数目的放大示意

通过cAMP分子的发射、放大或扩散，便形成了cAMP分子浓度的螺旋图样，即产生了较高层次上的信息，如图1.38所示。由于此信息是由系统的合作性所产生的，故我们称它为协同学信息。螺旋波形成某种梯度场（信息子），这种梯度场可被单个细胞所测



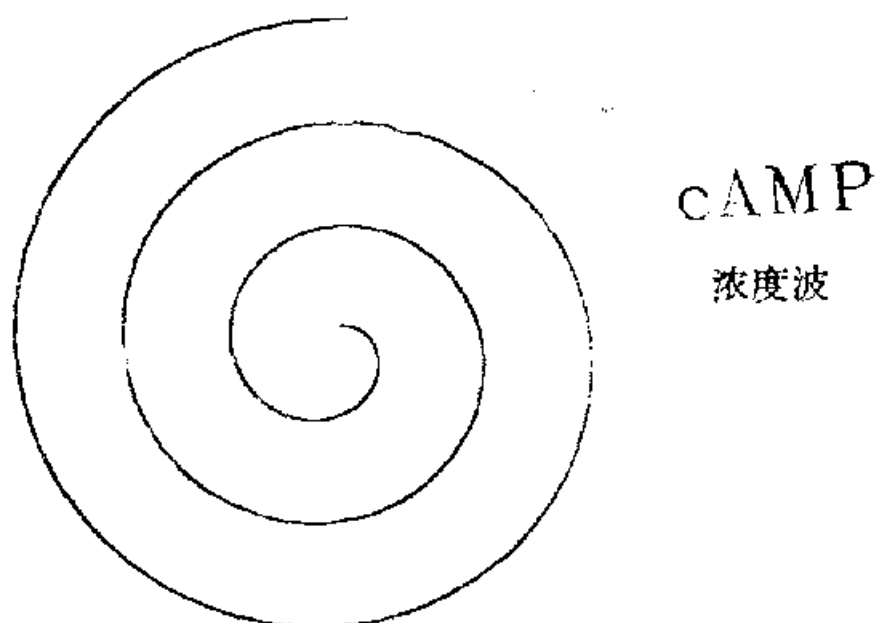


图1.38 粘性霉菌形成过程中cAMP的浓度波示意图

知，因此这些细胞便移到梯度场的最高处。这里，我们可以清楚地区分出信息的产生者、信息携带者和接收者，它们分别是细胞、cAMP和细胞。然而，在下一层次，我们会看到新的意义出现了，即所产生的分子浓度图样起着将细胞引导于聚集中心的作用。

同样的想法对位置信息的概念也基本适用。这里，我们假定，化学物质的产生和扩散已形成了预图样的化学场，生物组织中的各个细胞便从它那里接收到信息。一般地，假设有两种物质，即活性剂分子和抑制剂分子。在活性剂分子浓度高的地方，假定引起细胞变异的基因能被接通。以这种方式，这里的预图样起着信息子的作用。水螅是一个特殊的模型系统。

回忆一下迄今我们已建立了哪些概念，这是大有好处的。十分明显，存在着一个信息层次链。在最低层次，各个部分能够发射触发系统其它部分的信息。这样的信息能在特定的基元对间发生转移，或由一般的携带者转移。第一种情况的例子是每次连结

两个神经细胞的神经纤维，第二种情况的例子是向血液释放的内分泌，或者是释放到空气中的外激素。

虽然，在所有这些情况下，最初的信息交换可能是偶然的。但各个信号之间展开了竞争或合作，因而最终达到一种新的合作状态，它在本质上不同于先前的无序的或无关联的状态。这种新的状况由一个或一组序参量来描述，或者等价地用一种或几种信息子来描述。各个部分的状态可借助于伺服原理来决定。但是，我们可用另一种方式来表述这件事，即，系统的各个部分达成了特定的一致，或者说发生了自组织。同时，发生了信息压缩。在宏观层次上，信息出现了，而且在许多情况下，它增强了系统的可靠性或效率，或使两者都增强，或用于如上所述的其它特定目的。

这个新的集体层次从外部世界能够观察到，而且通过这种联系的建立达到了新的语义层次。顺便提一下，这种联系可与外部世界建立，同样也可以在同一系统内建立。因此，在这里可以使用象有用、无用或关联之类的词。以激光的例子来看这是十分明显的，合作状态达到了很高的效率。在生物系统的类似情况中，这种行为对整个系统都是有用的。在不稳定点之后，系统可能取得各种可能的状态，要选择这些状态，系统需要额外的信息。一种可能性是，这种信息是遗传来的，或者由系统其它各部分形成的约束来提供。但在退化的情况下，环境起重要作用，或换句话说，判断将要建立的态的价值，环境是一种重要的依据，按照我的见解，正是在这里，出现了生物学意义上的信息。通过不稳定性，形成了集体状态，但只有相对于环境，它才获得了意义，在某种程度上，只有相对于整个系统存在的价值，它才获得了意义。

这些评论也适用于遗传密码，虽然它的真正起源至今还不太

清楚。我们可以推测，第一次涨落的出现创造了具有一些特定性质的生物大分子。其中最重要的性质是，它们能以自催化的形式当然倍增。这个分子为其表型输送的信息的价值由环境来判断，同时还有其它一些分子及其表型也有可能属于同一环境。通过变异和选择的相互作用，产生了新型分子及其表型，这样，我们便观察到信息的创生。但是，这些信息是否有用呢？这只有通过这种物种与其环境的相互作用来检验。

上述考虑中，我们描述了形成有序或有结构的合作状态的头一步。但与我们上边曾经介绍过的物理系统（诸如激光、流体动力学或化学反应）相比，生物学中出现了一个新的特点，即固化作用。例如，活化剂分子引入细胞的基因时，该细胞分化为特殊的细胞——它不能再变更或不能再变为原先那种细胞。在某种程度上，动力学过程可能导致固态结构，如骨骼、器官等。与此类似，信息在DNA（即遗传密码）中以刚性形式存储。仿佛按照遗传密码给出的规则可构造出神经系统为刚性“排布”的低级动物。

另一方面，在高级动物中，除神经系统的刚性排布外，还出现了相当多的自组织。系统与其环境的相互作用与储存于系统的遗传信息一起导致了新信息的形成。通过环境对产生并存储于大脑的新信息的不断检验，从而建立了新的联系，因而产生了一种新的语义学。然而我们可以预料，“固化作用”出现在语义信息的各个层次，起了使系统更可靠和存储信息（记忆）的作用。而且海比（Hebb）的联系概念（通过它的使用而使其强化）可能是一个正确的概念。语义学的建立要求系统内部高度合作，要求系统与外界反复进行相互作用。在这方面，语义信息不是一种静态性质，而是一种诸关系和关联一再经受检验，增强和消失的

过程。我由此认为，意识不是一种静止的状态，而是一个过程，在其中信息不断地在大脑的各部分之间转移并一再被加工。

现在，讲一下模式识别可能是合适的。各种低等动物会立即对输入的信息（诸如闪光）作出反应，为了对信号作出反应，仅需少数几个特征，比如强度的阈值等等。但在高级动物中，进入的信息肯定要与已存贮的信息进行比较。但是，如何进行比较，我们要稍改变一下图象。

往往假定，进入的模式可与样板比较。但是，样板的存贮会需要相当多的信息量。因此，人们可能会考虑，在协同学的意义，是否只有一些特定的特征元被以序参量的形式储存下来，但当将它们调出时，却会产生详细的图象。按照这种意义，模式识别成为一种活的过程，其中大脑以自组织的方式产生了新的模式，并利用某种假设，不断地将新模式与接收的模式比较。例如，大家知道，当人们观察脸时，注意力集中于特殊部位——眼睛、鼻子或嘴等，并一再观察这些部位。

最后，让我们讨论一下人类。与动物不同，他们不仅能够通过遗传密码传递信息，而且能通过教育来传递信息，这一方式在动物世界中的应用是非常有限的。我们如此丰富的文化是世代以这种新方式传递信息的结果。但在此出现了一个极大的困难，这就是人类积累了惊人的知识量。因此，与协同学的精神一致，寻求统一的概念和原理来对付这样的世界将是很重要的。

此外，上述方法为我们提供的一种图象，与通常从生物系统中得到的极不相同。传统上往往假定存在着一个单控制中心（比方说在大脑中），由它来安排所有的行为，而我们所强烈支持的模式则要求自组织过程，最近我们通过手的运动的关联及其变化

的特定实验获得的定量理论证明了这一假设。这些由 S. 凯尔索 (S. Kelso) 进行的实验, 要求试验者平行地摆动手指, 当频率增加到一定的值时, 手的运动不由自主地变为一种反位相振荡。根据神经和肌肉行为的自组织假定, 可以解释这种转变的所有细节。

这当然是一种极端的情况, 但大概必须认为, 生物系统中信息的产生和转移有两种方式: 一种是传统的方式, 其中特定的运动, 神经程序产生特定的作用, 而其他现象则完全以自组织的形式出现。我们可以假设, 生物系统的信息处理, 在很大程度上表现为自组织的作用。例如, 生物系统极大的灵活性以及它的适应性、可塑性就是明证。

按照我的看法, 对生物系统的信息研究, 对于现代社会也很有意义, 社会正常职能依赖于信息的产生、转移和加工过程。也许表现出来的最重要的特点是循环因果性, 它导致集体状态, 在社会学中, 这种集体状态可能代表社会风气、公共舆论、民主或专政。

#### D. 维持一有序态需多少信息

让我们考虑一个标准的例子, 即激光。我们假定激光器中的原子是二能级的。低能态的所有原子数为  $N_1$ , 高能态的原子数为  $N_2$ 。我们有

$$N_1 + N_2 = N \quad (1.30)$$

就量子力学的意义上说, 可把占据数目  $N_1$ ,  $N_2$  与单个原子的占据概率

$$P_j = \frac{N_j}{N}, \quad j=1,2 \quad (1.31)$$

联系起来。因此每个原子的信息为

$$i = -P_1 \ln P_1 - P_2 \ln P_2 \quad (1.32)$$

而所有原子的信息为

$$I = -N(P_1 \ln P_1 + P_2 \ln P_2) \quad (1.33)$$

众所周知，受激原子或者通过自发辐射发射一个光子，或在已有其它光子的情况下，通过受激辐射发射光子。我们可把单个光子看作携带一个信息元的符号。正如我们所知，光子可以通过反射镜跑掉。因此，我们要问：为了维持相干状态，光子的产生率必须是多大？

由激光理论可知，我们不仅要引入光子数 $n$ ，并将它作为一个变量，而且还要引入反转数。反转数的定义是上、下两态的占据数目之差

$$D = N_2 - N_1 \quad (1.34)$$

根据激光理论，光子的增加率遵从方程

$$\frac{dn}{dt} = W D n - 2\kappa n \quad (1.35)$$

右边第一项表示光子的产生率，其中 $W$ 是光子产生的速率常数，而第二项表示光子通过反射镜逃逸所造成的损失，于是有效增加率便减少了。正如激光理论中所证明的那样，方程(1.35)描述了相干光子的增加，而忽略了不相干光子的增加。由于激光过程中所发生的作用，反转也随时间变化。其改变率由方程

$$\frac{dD}{dt} = (D_0 - D)/T - 2W D n \quad (1.36)$$

确定， $D_0$ 是由不导致激光发射的泵浦过程和弛豫过程决定的反转数。 $T$ 是任意反转向反转 $D_0$ 漂移的弛豫时间。方程(1.36)中的后一项来源于产生光子的激光过程。一般来说，衰减常数 $\kappa$ 远远小于率常数 $1/T$ 。这样一来，我们便可使用绝热近似，令

$$\frac{dD}{dt} \approx 0 \quad (1.37)$$

将式 (1.37) 代入式 (1.36)，即可对  $D$  求解方程 (1.36)，于是得到

$$D = \frac{D_0}{1 + 2TWn} \quad (1.38)$$

当激光器处于激光阈值以上不远时，我们可将分母展开为光子数  $n$  的幂级数，由一级近似得到

$$D \approx D_0 - 2D_0TWn \quad (1.39)$$

将这个结果代入光子的增加率方程 (1.35)，可直接得到

$$\frac{dn}{dt} = (WD_0 - 2\kappa)n - 2TW^2D_0n^2 \quad (1.40)$$

式 (1.40) 中的第二项总是引起光子产生率的减少，只要不等式

$$WD_0 - 2\kappa > 0 \quad (1.41)$$

成立，第一项将总是导致正的产生率。式 (1.41) 与激光条件相同，它确保了完全正的光子产生率，或确保了，完全正的信号产生率。这对于维持非零的光子流从而维持激光的有序状态是必须的。按照式 (1.41)，仅当  $D_0$  足够大时这个条件才能成立，而  $D_0$  很高是通过泵浦很大来实现的。由

$$n = \frac{WD_0 - 2\kappa}{2TW^2D_0} \quad (1.42)$$

我们可以推导出，条件 (1.41) 确保激光器在一切时刻必定存在非零的光子数。

我们来研究当增大泵浦率（换句话说，增大反转  $D_0$ ）时，信息 (1.32) 或 (1.33) 的行为。为此，将式 (1.42) 代入式 (1.39)，当  $n \geq 0$  和  $n = 0$  时分别得到

$$D \approx \frac{2\kappa}{W} = \text{常数!} \quad (1.43)$$

$$\text{和} \quad D = D_0 \quad (1.44)$$

换句话说, 在低泵浦率开始时,  $D_0$  是小量, 且不出现光子。然后  $D$  与  $D_0$  以同一比率增大。然而, 一旦出现激光作用, 由式 (1.43) 知, 反转  $D$  变为一个常量, 不再增大。所有输入到激光器的多余能量都被转换为相干光子。由式 (1.34) 和式 (1.30), 有关系式

$$N_1 = \frac{1}{2} (N - D) \quad (1.45)$$

$$N_2 = \frac{1}{2} (N + D) \quad (1.46)$$

成立。于是, 根据式 (1.31), 可变换为占据概率

$$P_1 = \frac{1}{2} \left(1 - \frac{D}{N}\right) \quad (1.47)$$

$$P_2 = \frac{1}{2} \left(1 + \frac{D}{N}\right) \quad (1.48)$$

将式 (1.47) 和式 (1.48) 代入式 (1.32), 得到

$$i = -\frac{1}{2} \left(1 - \frac{D}{N}\right) \ln \frac{1}{2} \left(1 - \frac{D}{N}\right) - \frac{1}{2} \left(1 + \frac{D}{N}\right) \ln \frac{1}{2} \left(1 + \frac{D}{N}\right) \quad (1.49)$$

以后, 我们将使用由

$$\gamma = \frac{D}{N} \quad (1.50)$$

定义的参数。因为  $D$  的变化范围是

$$D: -N, \dots, +N \quad (1.51)$$

则参数  $\gamma$  的变化范围为

$$\gamma: -1 \cdots +1 \quad (1.52)$$

作为  $\gamma$  的函数,  $i$  的行为描绘于图 1.39 中。式 (1.49) 与式 (1.43) 和式 (1.44) 联立, 我们便可得到当泵浦率  $D_0$  增加时, 按照



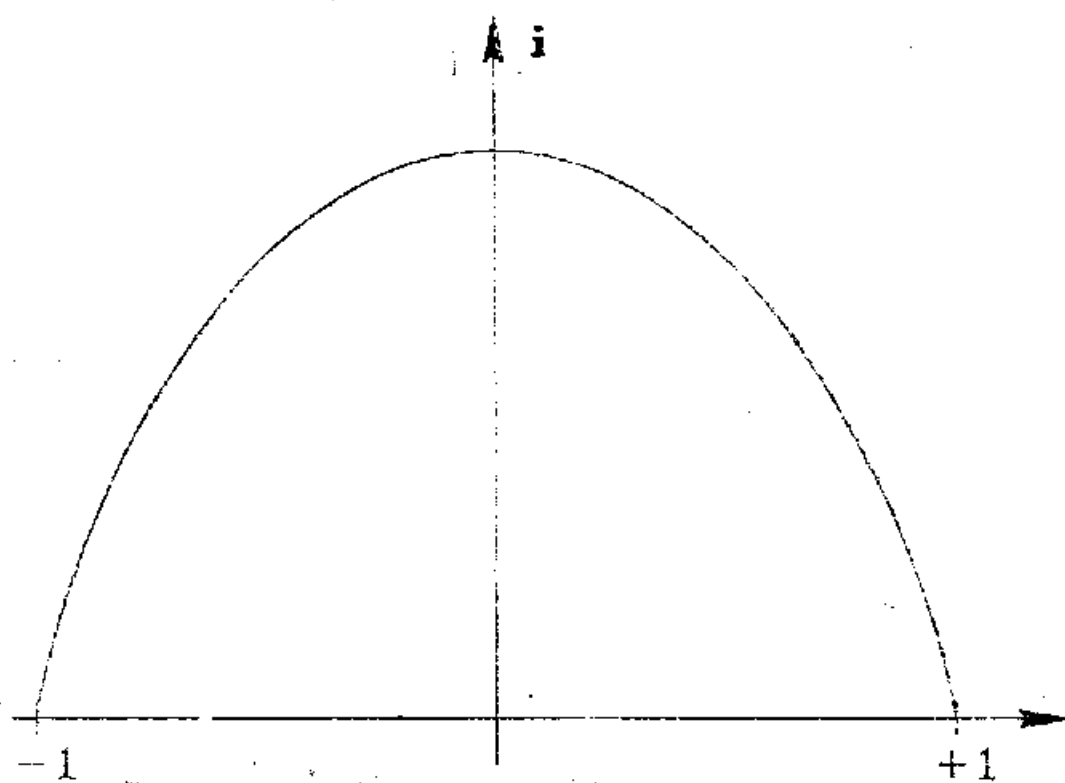


图1.39 一个原子的信息随参数 $\gamma$ 的变化

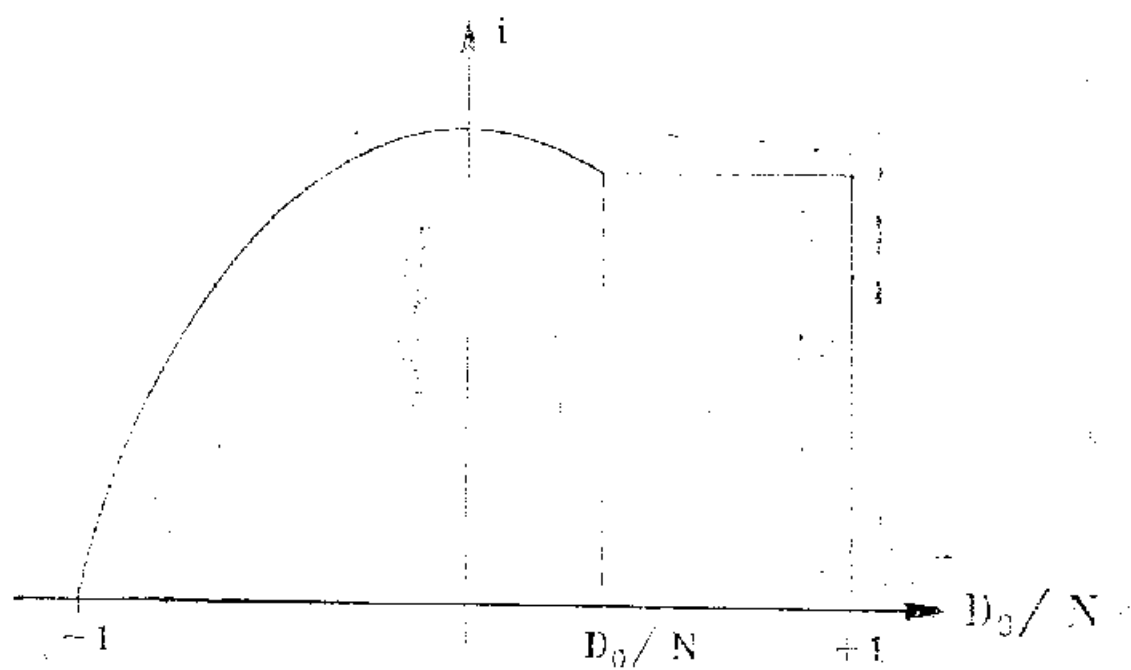


图1.40 激光中单原子信息的行为

图1.40, 信息先增长, 然后越过极大值而达到饱和。我们当前的方法不能处理包含于光场的信息, 因为直到现在我们不曾考虑任何涨落, 即光子数目  $n$  的任何概率分布。本书主要任务之一是把信息作为泵浦强度的函数来研究, 不仅对原子是这样, 而且对光子也是这样。

的确, 我们将得到惊人的结果, 即, 当趋于激光作用的阈值时, 令人感兴趣的信息包含在光子中, 而不是在原子中。

## 1.8 协同学的第二个基点

前几节讨论了信息和自组织的软的方面, 现在想介绍我们方法的硬核, 本书的其余部分将继续探讨这个问题。让我们简短地回顾一下, 在协同学中, 迄今做了哪些工作。以往, 我们从微观或中观层次出发, 并在该层次上列出方程。然后, 由不稳定性、序参量和伺服概念(它可以表示成严格的数学形式), 我们已能证明, 宏观层次上结构的出现, 以及相伴随的新质的出现。

在某种意义上, 这种方法可与统计力学方法相平行。现在, 我们打算建立一种与热力学类似的方法。即, 用宏观观察量来处理复杂系统。然后推测产生宏观结构或行为的过程的微观结构。为达到此目的, 我们欲使用的工具是最大熵原理或最大信息熵原理, 它已由吉尼斯相当普遍地发展起来。

第三章将详细介绍这个原理, 因此, 这里只需概述一下基本思想就足够了。我们的出发点是宏观可观测量——它可能出现涨落, 但它们的平均值已知。用脚标  $k$  标定各宏观量, 用  $f_k$  表示它们的平均值。然后, 我们想推测系统各状态(以脚标  $i$  标记)的概率分布  $P_i$ 。通过使信息

$$i = - \sum_j P_j \ln P_j \quad (1.53)$$

在约束条件

$$\sum_j P_j f_j^{(k)} = f_k \quad (1.54)$$

下达到极大的方法，我们达到了上述目的。显然， $f_j^{(k)}$  就是状态  $j$  对以  $k$  标记的宏观变量的贡献。再者，我们要求概率分布归一化，即

$$\sum_j P_j = 1 \quad (1.55)$$

正如吉尼斯所证明的那样（将在第四章介绍），这个原理使我们能以简洁而优雅的方式导出热力学基本公式。在这个推导过程中，约束指的是封闭系统的各个守恒量，也就是能量、粒子数，等等。要把最大熵原理推广到远离热平衡系统，甚至于非物理系统，关键在于适当选择约束条件。

正如我们将要看到的那样，迄今使用的能量甚至流入系统的能量流守恒的约束不能用来处理开放系统，尤其不能处理从无结构转变到有结构状态这样的非平衡相变。最大熵原理不时地遭到非难，因为约束的选择似乎带有一种特殊的主观性，其中的约束条件被认为是随观察者意愿选择的，并未根据客观标准来进行。对这种非难，吉尼斯作了详细的论辩，这里我想补充另一种观点。这就是科学界内部的一致性。在科学界，结论经过普遍同意才具有客观性。人们或许会把这种观点称为“相对客观主义”。但实际上，对于任何自然理论，我们顶多只能说这些，因为在自然科学中，一种理论绝不能被证真，而只能被证伪，这一点正是由波普

尔 (Popper) 正确加以指出的。因此, 我们必须采取一种正确选择合适约束的学习过程。这实际上就是热力学中已发生的情形。迄今, 我们都知道, 热力学中合适的约束就是守恒定律。

在非平衡相变领域中, 或一般地说在开放系统领域中, 我们第一步要做的, 就是说明这些约束是什么。的确, 当我们仅限于分析非平衡相变时, 就会发现, 只要微观分布函数能够得到, 熵原理的宏观方法与微观理论推导, 二者的结果总是完全一致。因而, 我坚信在这里能够找到一致点。

从数学观点来看, 还有另外一点值得一提。即, 当我们用按实验给出的平均值来规定特定约束时, 由最大(信息)熵原理总会再得到这些平均值的分布函数。到此为止, 我们进行着一种同义反复。但是, 下一步, 借助于概率分布便可导出新的平均值, 而且可以作出能够被实验证明的预言。一旦这些预言未被证明, 就可以将这些新的实验数据作为补充约束, 然后导出修正的概率分布函数。用这种方式, 这个过程可以无穷反复。但是, 尽管如此, 一旦将讨论限于一类指定的现象, 我们总能在选择合适的一组有限约束上求得一致。

如上所及, 这类现象之一就是热力学中具有适当约束的封闭系统。另一类是将要处理的非平衡相变。正如我们将看到的那样, 这后一类包括了各种领域中的许多现象, 比如空间图样、新型信息以及振荡现象的出现。对于导致决定论性混沌的过程, 选择适当的约束是(至少部分是)今后的一个任务。

正如我们将看到的那样, 按照我们的方法, 约束被重新理解, 尤其是在以下的方面: 我们能初步地推断, 合适的约束必须包括宏观变量, 即序参量。但我们知道, 在非平衡相变中,

出现了序参量的临界涨落，换句话说，它的涨落变为宏观变量。的确，我们将会证明，计入序参量的涨落是寻求这类现象的合适的约束的关键。

在微观理论成果指导下，还可导出许多结果，这时，我们甚至在一开头就不需要序参量。相反，我们的方法将从相关函数，即观测量的矩出发，便可重新导出序参量和伺服模。而且，我们可顺便构造出宏观模式，换句话说，我们可自动识别出在非平衡相变中产生的演化模式。

在结束时，我们讨论分析（或微观）方法和整体（或宏观）方法的关系，并进一步指出宏观方法的优点。子系统本身往往非常复杂，于是很难或甚至不可能具体列出微观或中观方程。当我们研究极端状态如人脑时，子系统，比如神经细胞（神经原）本身就是复杂系统。一个神经细胞包含细胞体、一个轴突和数目多到 8 万的树枝状晶体——它把一个神经细胞和其它神经细胞连接了起来。在人脑中有数以亿万计的神经细胞。协同学认为尽管如此复杂，仍有许多行为模式可用序参量概念来处理，在那里现在已用唯象方式建立了序参量方程。

最近，我们找到一个范例，即，手指运动的协调，尤其是在手指运动之间的不自觉转变。虽然子系统数目相当巨大（它由神经细胞、肌肉细胞和其它组织组成），但其行为却能用单个序参量表示出来。我们将在本书中介绍这些实验，并通过这个例子阐明一般方法。进一步的例子将取自激光物理学和流体动力学。在协同学从微观的观点出发研究的一大批例子中，普遍地应用我们的宏观方法已并不困难。另外还能找到为数众多的只有宏观方法才适用的例子（尤其在生物学中）。

在结束本章时，我提请大家注意一点，我们将交替使用“最大信息原理”与“最大熵原理”，然而我们将会明白，对非平衡系统来说，用信息一词代替熵更为恰当。

**表1.1 系统与变量**

化学反应	不同相的分子密度
凝 固	
流 体	速 度 场
火 焰	温 度 场
等离子体	电场和磁场
激光、参量振子	原子偶极矩、反转数
固体物理学 耿氏振子、灯丝的形成	电子和空穴的密度
形态发生学	细胞组织中的细胞密度
生物（出现）前进化	生物分子数目
群体动力学	动物数目
神经网络	神经细胞的激发率
肌肉的运动	伸长、收缩
经 济 学	货 币 流
社 会 学	具有特定态度的人数
协同计算	元件的活化

## 第二章 从微观世界到宏观世界

---

### 2.1 描述的层次

在本章，我要提出一些基本的概念和方法，这些概念和方法是协同学以微观的方法来研究自组织时所采用的。熟悉这些内容的读者可以跳过本章，直接阅读第三章。另一方面，对于那些不很熟悉这些概念和方法，但又想钻研得更深一些的读者，建议他们去读我的《协同学导论》和《高等协同学》两本书，那里有这些概念的详细介绍。

当我们描述一个系统时，首先必须确认描述系统的变量。这种描述，可以在相互联系着的不同层次上来进行。让我们首先看一下图2.1所示的流体的例子。在微观层次上，流体由单个分子所组成。因此，为了完整地描述流体，我们必须处理单个分子的位置和速度。然而，对于我们想要达到的许多目的，在一个中观的层次上处理流体就足够了。这时，我们要应用体积元的概念。这种体积元与整个流体的尺度相比足够小，但同时，这种体积元

又要足够大，以便我们能可靠地使用密度、速度场或区域温度等概念。最后，我们的兴趣在于宏观的层次上，我们想研究宏观层次上结构或图样的形成，例如底部被加热的流体的一块六角形图样。对于生物系统，我们也可类似地区分描述的层次。但是，读者很快便会认识到，这里，我们在选择层次上有相当大的任意性。

让我们考虑一个生物组织的例子。在微观层次上，我们可以论及生物分子或细胞器等。在中观的层次上，我们可以论及细胞，最后，在宏观层次上，我们关心的是这些细胞所构成的整个组织或器官。很显然，在从中观向宏观过渡时，我们必须忽略细胞及其组成成份的许多细节特征，而选取那些与器官形成有关的特征。因此，在中观层次上，我们已经在进行着信息的巨大压缩。

虽然在一些情况下，从微观到宏观层次的直接过渡也是可能的，但在本章，我们将着重于从中观到宏观层次的过渡。在物理学中，中观层次可通过统计力学来研究，在那里，我们引入并处理某些相关变量。然而，在许多情况下，例如化学和生物中，我们将采用相应变量的唯象方程。

现在列举一些中观变量的例子（如图2.1），以便能给读者一些我们将要处理的各种问题的印象。

## 2.2 朗之万方程

粒子的布朗运动，例如浸在液体中的尘埃的运动，为我们提供了一个中观描述的非常典型的例子。它的运动由朗之万方程来描述，其中变量 $q$ 就是粒子的速度。所有液体分子的运动有两种效应：它们一方面引起布朗粒子的速度衰减；另一方面对布朗粒子施一随机力。则朗之万方程写为，



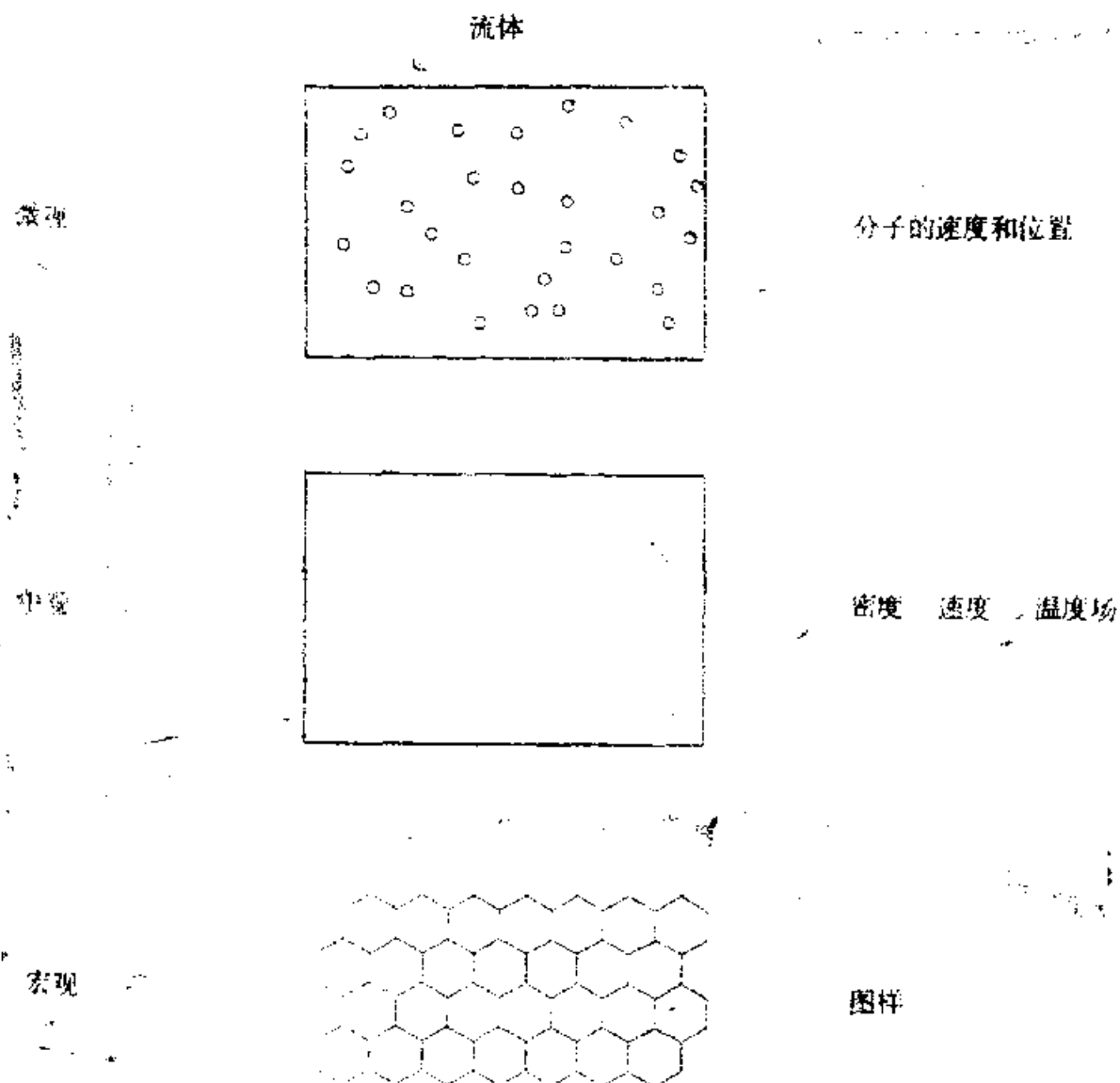


图2.1 通过流体的例子来说明微观、中观和宏观的方法

$$\dot{q} = K(q) + F(t) \quad (2.1)$$

对于布朗粒子，其中的 $K$ 由下式给出：

$$K(q) = -\gamma q \quad (2.2)$$

在协同学中，下述非线性表达式是我们经常非常关心的一种形式，

$$K(q) = \alpha q - \beta q^3 \quad (2.3)$$

式 (2.1) 中的涨落力 $F$ 由下述性质来表征：

$$\langle F(t) \rangle = 0 \quad (2.4)$$

$$\langle F(t) F(t') \rangle = Q \delta(t-t') \quad (2.5)$$

其中平均是对整个随机过程进行的。如果要考虑的变量有几个，我们将  $q_1, \dots, q_N$ ，写成一矢量  $\underline{q}$ ，这时朗之万方程写为

$$\dot{\underline{q}} = \underline{K}(\underline{q}) + \underline{F}(t) \quad (2.6)$$

相应的涨落力具有性质：

$$\langle F_i(t) \rangle = 0 \quad (2.7)$$

$$\langle F_i(t) F_{j'}(t') \rangle = Q_{ij} \delta_{ij} \delta(t-t') \quad (2.8)$$

注意到  $\underline{q}$  可能是一个高维空间的矢量，因而代表一个非常复杂的系统。在有些情况下，涨落力本身是依赖于态变量  $\underline{q}$  的，在这种情况下，出现了一些特定的问题，它们通过伊藤-斯特拉特洛维奇 (Stratonovich) 积分来解决。在伊藤积分中，式 (2.1)，即朗之万方程必须由下述方程来代替，

$$d\underline{q}(t) = \underline{K}(\underline{q}(t)) dt + \underline{g}(\underline{q}(t)) d\underline{w}(t) \quad (2.9)$$

其中  $\underline{K}$  和  $\underline{g}$  一般是  $\underline{q}$  的非线性函数， $d\underline{w}$  描述一个随机过程，我们假设

$$\langle d\underline{w} \rangle = 0 \quad (2.10)$$

$$\langle d\underline{w}(t) d\underline{w}(t') \rangle = dt \delta(t-t') \quad (2.11)$$

在伊藤方程中，我们假定出现在式 (2.9) 的最后一项中的  $\underline{q}(t)$  和  $d\underline{w}(t)$  是统计上无关联的。我们处理一个多维的态矢量，其分量记为  $q_1, q_2, \dots, q_m$ ，伊藤方程写为

$$dq_i(t) = K_i(\underline{q}(t)) dt + \sum_m g_{im}(\underline{q}(t)) dw_m(t) \quad (2.12)$$

其中描述随机过程的  $d\underline{w}_m$  具有性质

$$\langle d\underline{w}_m \rangle = 0 \quad (2.13)$$

$$\langle dw_i(t)dw_m(t) \rangle = \delta_{im} dt \quad (2.14)$$

为了显示伊藤手续的特点，让我们考虑一个任意可微的函数

$$\tilde{u} = u(\tilde{q}) \quad (2.15)$$

及其微分。由于式 (2.14) 所示的性质，我们必须按下式计算到二阶导数

$$du_i = \sum_k \frac{\partial u_i}{\partial q_k} dq_k + \frac{1}{2} \sum_{kl} \frac{\partial^2 u_i}{\partial q_k \partial q_l} dq_k dq_l \quad (2.16)$$

将式 (2.12) 代入式 (2.16) 中，保留含  $dt$  的项，忽略所有高阶项，我们得到

$$\begin{aligned} du_i = & \sum_k \frac{\partial u_i}{\partial q_k} [K_k(q)dt + \sum_m g_{km} dw_m(t)] \\ & + \frac{1}{2} \sum_{kl} \frac{\partial^2 u_i}{\partial q_k \partial q_l} [\sum_m g_{km} g_{lm} dw_m dw_m] \end{aligned} \quad (2.17)$$

让我们简单提示一下斯特拉特洛维奇方法，单变量的随机方程是

$$dq = K(q)dt + g(q)dw(t) \quad (2.18)$$

对于一个矢量，分量

$$dq_i = K_i(\tilde{q})dt + \sum_m g_{im}(\tilde{q})dw_m(t) \quad (2.19)$$

与伊藤积分相反，对式 (2.18) 或式 (2.19) 中的最后一项，现在要给以不同的解释，即它们必须按照中点规则来计值。

$$g\left(q\left(\frac{t_i + t_{i-1}}{2}\right)\right)dw(t_i) \quad (2.20)$$

即  $q$  和  $dw$  不再是统计独立的。以后将会看到，我们可以通过宏观讨论来恢复伊藤方程。

## 2.3 福克-普朗克方程

在许多应用中，特别是当我们的问题是非线性时，即  $K$  是  $q$  的非线性函数时，较好的办法是利用分布函数  $f(q, t)$  满足的福克-普朗克方程。 $f(q, t)$  描述  $t$  时刻发现变量  $q$  在  $q \rightarrow q + dq$  区间的概率。

与式 (2.1) 的朗之万方程相应的福克-普朗克方程是

$$\frac{\partial f}{\partial t} = - \frac{\partial}{\partial q} (K(q)f) + \frac{Q}{2} \frac{\partial^2 f}{\partial q^2} \quad (2.21)$$

其中右边第一项称为漂移项，最后一项称为扩散项。

定态解满足

$$\frac{\partial f}{\partial t} = 0 \quad (2.22)$$

如果边界条件

$$f(q) \rightarrow 0$$

$$\text{当 } q \rightarrow \pm \infty \quad (2.23)$$

时，得满足，定态解很易得到，它是

$$f = N \exp \left[ - \int_{-\infty}^q 2K(q')/Q \, dq' \right] \quad (2.24)$$

其中归一化因子  $N$ ，用以保证  $f$  的积分是 1。在多维态矢量情形，我们必须找到

$$f(\underline{q}, t) \quad (2.25)$$

所满足的福克-普朗克方程，它是

$$\frac{\partial f}{\partial t} = - \sum_i \frac{\partial}{\partial q_i} (K_i f) + \frac{1}{2} \sum_{ij} Q_{ij} \frac{\partial^2 f}{\partial q_i \partial q_j} \quad (2.26)$$

式 (2.26) 的明显解只是一些特殊情况下可以得到, 例如, 当  $K$  是变量  $\underline{q}$  的线性函数,  $Q_k$  与  $\underline{q}$  无关时。在这种情况下, 含时解和定态解可以明显地构造出来 (见我的《高等协同学》一书)。如果满足所谓细致平衡律, 式 (2.26) 的定态解也可以明显地构造出来 (见 2.4 节)。

当我们从伊藤微分方程出发, 相应的福克-普朗克方程具有形式

$$\frac{\partial f}{\partial t} = - \sum_k \frac{\partial}{\partial q_k} [K_k(\underline{q})f] + \frac{1}{2} \sum_{kl} \frac{\partial^2}{\partial q_k \partial q_l} [\sum_m g_{km} g_{lm} f] \quad (2.27)$$

在斯特拉特洛维奇积分的情形, 可以证明, 福克-普朗克方程是

$$\begin{aligned} \frac{\partial f}{\partial t} = & - \sum_i \frac{\partial}{\partial q_i} \{ [K_i(\underline{q}) + \frac{1}{2} \sum_{kj} \frac{\partial g_{ij}}{\partial q_k} g_{kj}] f \} \\ & + \frac{1}{2} \sum_{lm} \frac{\partial^2}{\partial q_l \partial q_m} ( \sum_i g_{li} g_{mi} f ) \end{aligned} \quad (2.28)$$

## 2.4 具有细致平衡的系统的福克-普朗克方程的严格定态解

在本节, 我们证明, 在细致平衡条件下, 通过积分可以明显得到福克-普朗克方程的定态解。

虽然细致平衡原理几乎在所有实际上的热平衡系统中都成立, 但在远离热平衡的系统中却不尽然。因此, 对每一情形必须分别仔细考察, 看这一原理是否成立 (例如, 通过对称性考虑)。另外, 对福克-普朗克方程的结构进行考察也能决定细致平衡是否能满足。

## A. 细致平衡

我们用  $\underline{q}$  表示一组变量  $q_1, \dots, q_N$ , 用  $\tilde{\underline{q}}$  表示时间反演下相应的变量,

$$\tilde{\underline{q}} = \{\varepsilon_1 q_1, \dots, \varepsilon_N q_N\} \quad (2.29)$$

其中  $\varepsilon_i = -1$  ( $+1$ ), 对应于时间反演下坐标  $q_i$  改变符号 (不改变符号)。进一步, 我们用  $\tilde{\underline{\lambda}}$  表示一组由外界决定的参量, 时间反演后记为

$$\tilde{\underline{\lambda}} = \{\lambda_1 \nu_1, \dots, \lambda_M \nu_M\} \quad (2.30)$$

其中,  $\nu_i = -1$  ( $+1$ ) 取决于外参量在时间反演下的对称性。我们将发现, 系统在  $t_1$  时处于  $\underline{q}$ , 在  $t_2$  时处于  $\underline{q}'$  的联合概率为

$$P(\underline{q}', \underline{q}; t_2, t_1) \quad (2.31)$$

以下, 我们考虑一个定常系统, 于是联合概率只依赖于时间差  $t_2 - t_1 = \tau$ , 因此, 式 (2.31) 可以写为

$$P(\underline{q}', \underline{q}; t_2, t_1) = W(\underline{q}', \underline{q}; \tau) \quad (2.32)$$

我们现在表述细致平衡原理, 我们采用下面两种定义:

(i) 细致平衡原理的第一种表述:

$$W(\underline{q}', \underline{q}; \tau, \underline{\lambda}) = W(\tilde{\underline{q}}, \tilde{\underline{q}}'; \tau, \tilde{\underline{\lambda}}) \quad (2.33)$$

联合概率可以表示为定态分布  $f(\underline{q}, \underline{\lambda})$  乘以条件概率  $P$ , 其中定常性是通过下式显示的,

$$P = P(\underline{q}' | \underline{q}; \tau, \underline{\lambda}) \quad (2.34)$$

因此, 式 (2.31) 可以重新表述为

$$P(\underline{q}' | \underline{q}; \tau, \underline{\lambda}) f(\underline{q}, \underline{\lambda}) = P(\tilde{\underline{q}} | \tilde{\underline{q}}'; \tau, \tilde{\underline{\lambda}}) f(\tilde{\underline{q}}', \tilde{\underline{\lambda}}) \quad (2.35)$$

这里及以后，我们假定福克-普朗克方程具有唯一的定态解，这时，可直接证明下式成立

$$f(\underline{q}, \lambda) = f(\underline{\tilde{q}}, \underline{\tilde{\lambda}}) \quad (2.36)$$

定义每秒的跃迁概率

$$w(\underline{q}', \underline{q}; \lambda) = \left[ \frac{d}{d\tau} P(\underline{q}' | \underline{q}; \tau, \lambda) \right]_{\tau=0} \quad (2.37)$$

在式 (2.35) 的两边对  $\tau$  求导，并令  $\tau=0$  (但  $\underline{q}' \neq \underline{q}$ )，我们得到

(ii) 细致平衡原理的第二种表述：

$$w(\underline{q}', \underline{q}; \lambda) f(\underline{q}, \lambda) = w(\underline{\tilde{q}}, \underline{\tilde{q}}'; \underline{\tilde{\lambda}}) f(\underline{\tilde{q}}', \underline{\tilde{\lambda}}) \quad (2.38)$$

显然，其意义非常简单，左边描述从态  $\underline{q}$  到新态  $\underline{q}'$  的总跃迁速率。细致平衡原理要求，这个速率应该等于当运动相反（动量相反）时从  $\underline{q}'$  到  $\underline{q}$  的速率。可以证明，以上两种表述是等价的。

### B. 所要求的福克-普朗克方程的结构及其定态解

利用条件概率  $P$ （它就是格林函数），我们将福克-普朗克方程写为

$$\frac{d}{d\tau} P(\underline{q}' | \underline{q}; \tau, \lambda) = L(\underline{q}', \lambda) P(\underline{q}' | \underline{q}; \tau, \lambda) \quad (2.39)$$

其中，我们假定算子  $L$  具有形式

$$L(\underline{q}) = - \sum_i \frac{\partial}{\partial q_i} K_i(\underline{q}, \lambda) + \frac{1}{2} \sum_{ik} \frac{\partial^2}{\partial q_i \partial q_k} Q_{ik}(\underline{q}, \lambda) \quad (2.40)$$

我们总可以假定扩散系数是对称的

$$Q_{ik} = Q_{ki} \quad (2.41)$$

我们定义以下新系数：

(i) 不可逆漂移系数

$$D_i(\underline{q}, \underline{\lambda}) = \frac{1}{2} [K_i(\underline{q}, \underline{\lambda}) + \varepsilon_i K_i(\underline{q}, \widetilde{\underline{\lambda}})] \equiv D_i^+ \quad (2.42)$$

(ii) 可逆漂移系数

$$J_i(\underline{q}, \underline{\lambda}) = \frac{1}{2} [K_i(\underline{q}, \underline{\lambda}) - \varepsilon_i K_i(\underline{q}, \widetilde{\underline{\lambda}})] \equiv D_i^- \quad (2.43)$$

$J_i$  在时间反演下象  $q_i$  一样变换。

我们将福克-普朗克方程的定态解写为

$$f(\underline{q}, \underline{\lambda}) = N e^{-\Phi(\underline{q}, \underline{\lambda})} \quad (2.44)$$

其中,  $N$  是归一化常数,  $\Phi$  可以解释为广义热力学势。细致平衡原理成立的必要充分条件是

$$Q_{ik}(\underline{q}, \underline{\lambda}) = \varepsilon_i \varepsilon_k Q_{ik}(\underline{q}, \widetilde{\underline{\lambda}}) \quad (2.45)$$

$$D_i - \frac{1}{2} \sum_k \frac{\partial Q_{ik}}{\partial q_k} = \frac{1}{2} \sum_k K_{ik} \frac{\partial \Phi}{\partial q_k} \quad (2.46)$$

$$\sum_i \left( \frac{\partial J_i}{\partial q_i} - J_i \frac{\partial \Phi}{\partial q_i} \right) = 0 \quad (2.47)$$

如果扩散矩阵  $Q_{ik}$  有逆存在, 由式 (2.46) 可以将  $\Phi$  的梯度解出

$$\frac{\partial \Phi}{\partial q_i} = \sum_k (Q^{-1})_{ik} \left( \sum_l \frac{\partial Q_{kl}}{\partial q_i} - 2D_k \right) \equiv A_i \quad (2.48)$$

这表明, 式 (2.48) 意味着可积性条件

$$\frac{\partial A_j}{\partial q_i} = \frac{\partial A_i}{\partial q_j} \quad (2.49)$$

它是一个由式 (2.48) 所定义的漂移和扩散系数成立的条件。将

(2.48) 代入式 (2.47), 则式 (2.47) 的形式为

$$\sum_i \left[ -\frac{\partial J_i}{\partial q_i} - J_i \sum_k (Q^{-1})_{ik} \left( \sum_l \frac{\partial Q_{kl}}{\partial q_i} - 2D_k \right) \right] = 0 \quad (2.50)$$



因此, 细致平衡成立的条件最终由式 (2.45)、式 (2.49)、式 (2.50) 给出。式 (2.46) 或式 (2.48) 使我们可以由纯粹积分来确定  $\phi$ , 因此, 福克-普朗克方程的定态解可以明显地确定。

## 2.5 路径积分

福克-普朗克方程的含时解可以用路径积分的形式表示出。为了简单起见, 以下我们将考虑扩散常数  $Q$  与  $q$  无关的情形。让我们首先考虑一维情形, 即  $q$  是单一变量。我们将时间间隔  $t$  分成等距步长:

$$t_0, t_1 = t_0 + \tau, \dots, t_N = t_0 + \tau N \quad (2.51)$$

时刻的分布函数  $f$  可被构造为对所有中间位置的一个多重积分

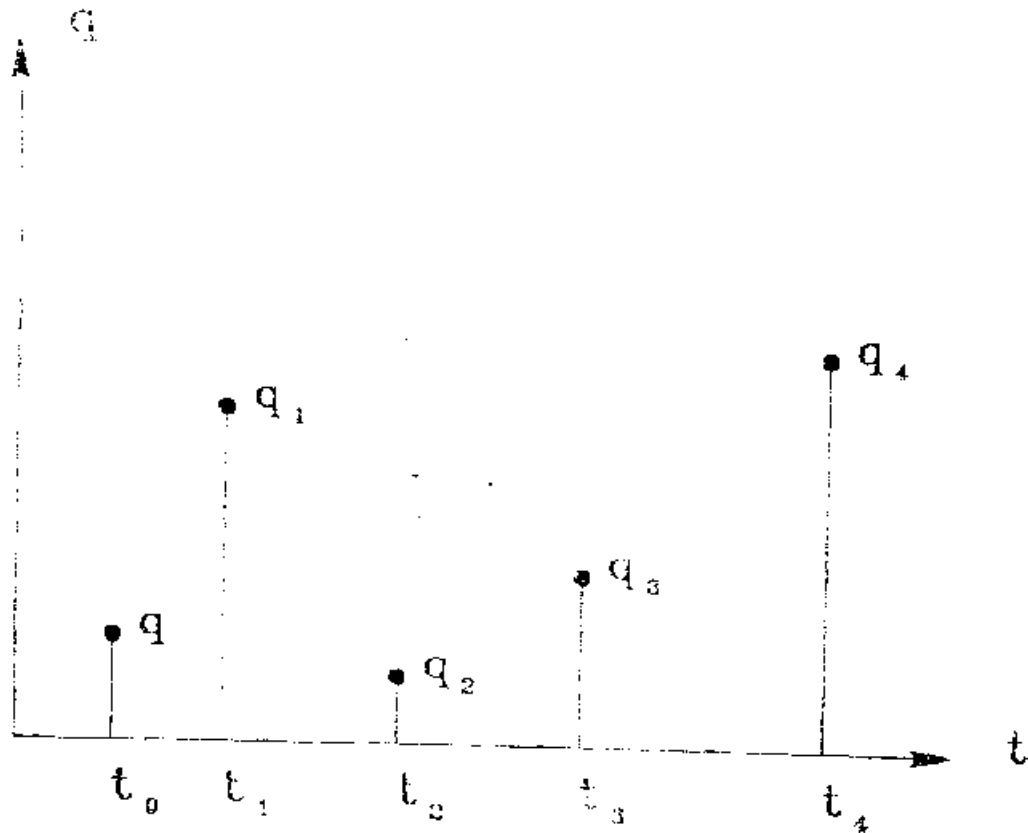


图2.2 粒子的位置或系统的态随时间的变化

(见图2.2)，路径积分的明显形式是

$$f(q, t) = \lim_{\substack{N \rightarrow \infty \\ N\tau = t}} \int \cdots \int Dq \exp(-\frac{1}{2} 0) f(q', t_0) \quad (2.52)$$

其中我们使用了简写，

$$Dq = (2Q\tau\pi)^{-N/2} dq_0, \dots, dq_{N-1} \quad (2.53)$$

和

$$0 = \sum_v \tau \{ (q_v - q_{v-1})/\tau - (Kq_{v-1}) \}^2 Q^{-1} \quad (2.54)$$

公式(2.52)的具体推导在《协同学导论》中给出。现在，将式

(2.52)推广到具有几个分量的多维矢量 $\underline{q}$ 的情形。仍然假定 $Q_{mn}$ 是与态矢量 $\underline{q}$ 无关的，我们这时得到下面结果：

$$f(\underline{q}, t) = \lim_{\substack{N \rightarrow \infty \\ N\tau = t}} \int \cdots \int Dq \exp(-\frac{1}{2} 0) f(\underline{q}', t_0) \quad (2.55)$$

其中用到了下列缩写：

$$Dq = \prod_{\mu=0}^{N-1} \left\{ (2\pi\tau)^{-\frac{n}{2}} (\det Q)^{-\frac{1}{2}} \right\} (dq_1, \dots, dq_N)_\mu \quad (2.56)$$

$$\underline{q}_N = \underline{q}, \quad \underline{q}_0 = \underline{q}' \quad (2.57)$$

$$0 = \tau \sum (\dot{\underline{q}}_v^T - K_{v-1}^T) Q^{-1} (\dot{\underline{q}}_v - K_{v-1}) \quad (2.58)$$

$$\dot{\underline{q}}_v = \tau^{-1}(\underline{q}_v - \underline{q}_{v-1}), K_{v-1} = K(\underline{q}_{v-1}) \quad (2.59)$$

$T$ 表示转置矢量。

最后，让我们向读者提示一下主方程。我们考虑一个分立的态空间矢量 $\underline{m}$ 。我们感兴趣的是概率分布 $P(\underline{m}, t)$ 。如果我们处理一个马尔科夫过程， $P$ 满足主方程

$$\frac{dP(\underline{m}, t)}{dt} = \sum_{\underline{n}} w(\underline{m}, \underline{n}) P(\underline{n}) - \sum_{\underline{m}} w(\underline{n}, \underline{m}) P(\underline{m}) \quad (2.60)$$

同样，找出式(2.60)的明显解是很困难的。在细致平衡的特殊情况下，定态概率分布可以明显地构造出。当有细致平衡时， $P$ 满足下面关系：

$$w(\underline{n}, \underline{m}) P(\underline{m}) = w(\underline{m}, \underline{n}) P(\underline{n}) \quad (2.61)$$

那么，式(2.60)的定态解可以明显地写为

$$P(\underline{m}) = N \exp [\Phi(\underline{m})] \quad (2.62)$$

其中  $\Phi(\underline{m})$  定义为

$$\Phi(\underline{m}) = \Phi(\underline{n}_0) + \sum \ln \{w(\underline{n}_{j+1}, \underline{n}_j) / w(\underline{n}_j, \underline{n}_{j+1})\} \quad (2.63)$$

$$\underline{m} = \underline{n}_N \quad (2.64)$$

## 2.6 复杂性的约简·序参量·伺服原理

在本节，我们处理包含许多部分的系统。我们想研究系统行为的定性变化。为此，我们将变量及其方程做几次变换。然后，在2.7节和2.8节，我们将给出一些重要的应用。我们从一个态矢量开始，它在微观或中观的层次上描述总系统，

$$\underline{q} = \underline{q}(\underline{x}, t) \quad (2.65)$$

在以下，我们假定态矢量是空间坐标  $\underline{x}$  的函数，这样，我们就可以一般地假定  $\underline{q}$  服从下面形式的演化方程

$$\dot{\underline{q}} = N(\underline{q}, \alpha) + F(t) \quad (2.66)$$

其中， $N$  是  $\underline{q}$  的非线性函数，它也可能含有微分算子，例如拉普拉斯算子，它是  $\underline{q}$  对空间坐标的微商， $\alpha$  是控制参量，例如，激

光中的功率输入，或对流体加热的热量，或对生物系统刺激的信号。 $\tilde{F}(t)$  是涨落力。我们现在分几步进行。

### A. 线性稳定性分析

在下面，我们假定，对于控制参量的一固定值 $\alpha_0$ ，决定论方程的解是已知的，即， $\tilde{q}_0$ 满足方程

$$\dot{\tilde{q}} = N(\tilde{q}, \alpha_0) \quad (2.67)$$

我们然后研究当控制参数变化时解的行为。为此我们假设，

$$\alpha: \quad \tilde{q} = \tilde{q}_0 + \tilde{w} \quad (2.68)$$

我们假定 $\tilde{q}_0$ 随 $\alpha$ 平滑地变化，为

$$\tilde{q}_0(\alpha) \quad (2.69)$$

为了研究解 $\tilde{q}_0$ 的稳定性，我们假设将式(2.68)代入式(2.66)，并去掉其中的涨落力，我们得到

$$\dot{\tilde{q}}_0 + \dot{\tilde{w}} = N(\tilde{q}_0 + \tilde{w}, \alpha) \quad (2.70)$$

在 $\tilde{w}$ 是一个小量的假设下，我们可以将式(2.70)的右边展成关于 $\tilde{w}$ 的幂级数，只保留最前面两项，

$$\dot{\tilde{q}}_0 + \dot{\tilde{w}} = N(\tilde{q}_0, \alpha) + L(\tilde{q}_0) \tilde{w} \quad (2.71)$$

由于式(2.67)成立，上式左右两边第一项抵消，我们得到

$$\dot{\tilde{w}} = L(\tilde{q}_0) \tilde{w} \quad (2.72)$$

注意，依赖于 $\tilde{q}_0$ 的 $L$ 可能仍然含有作用于 $\tilde{w}$ 中的空间坐标的微分算子。然而，式(2.72)的一般解可以写为

$$\tilde{w}(t) = e^{\lambda t} \tilde{V} \quad (2.73)$$

注意到，我们同时也处理了更一般的情形，即 $\tilde{q}_0$ 是一个周期

或准周期的函数。关于详细的结果，读者可参考我的《高等协同学》一书。为了以下的目的，重要的是区分出所谓的稳定模和不稳定模，不稳定模是那些满足

$$\lambda > 0 \quad (2.74)$$

的模。它们被表示为

$$\lambda_u, \underline{V}_u \quad (2.75)$$

稳定模用

$$\lambda < 0 \quad (2.76)$$

来表征，并记为

$$\lambda_s, \underline{V}_s \quad (2.77)$$

注意，这里“稳定的”，和“不稳定的”只是对线性分析而言的，事实上，我们将证明，一般情况下，所谓的不稳定模将通过与稳定模的相互作用而变得稳定。进一步，我们的方法是完全非线性的，线性稳定性分析只提供了一个表示式 (2.66) 的所希望的解的合适参考系。

## B. 演化方程的变换

为了在非线性和随机的情况下求解式 (2.66)，我们假设：

$$\underline{q} = \underline{q}_0 + \sum_i \xi_u(t) \underline{V}_u + \sum_s \xi_s(t) \underline{V}_s \quad (2.78)$$

在  $L$  含有作用于空间变量的微分算子时， $V$  是空间变量的函数。

$$\underline{V} = \underline{V}(\underline{x}) \quad (2.79)$$

当我们将式 (2.78) 代入式 (2.26)，投影并展开函数  $\underline{V}_u$  和  $\underline{V}_s$ ，我们得到模幅度  $\xi_u$  和  $\xi_s$  的方程

$$\dot{\xi}_u = \lambda_u \xi_u + N_u(\xi_u, \xi_s) + F_u(t) \quad (2.80)$$

$$\text{和} \quad \dot{\xi}_s = \lambda_s \xi_s + N_s(\xi_u, \xi_s) + F_s(t) \quad (2.81)$$

指标 $u$ 和 $s$ 有两重目的,一方面它说明我们处理的是不稳模的幅度还是稳定模的幅度;另一方面,它们是 $\xi_u$ 和 $\xi_s$ 的单个分量的序号,例如,我们可以让 $u$ 取 $1, \dots, M$ ,  $s$ 取 $M+1, \dots$ 。我们将表明,指标 $u$ 或 $s$ 被赋予何种意义。幅度 $\xi_u$ 被称为序参量。

### C. 伺服原理

方程(2.66)的变换(2.78)不能使复杂性得到约简,方程(2.80)和(2.81)完全等价于式(2.66)。然而,协同学的伺服原理却可以使我们从式(2.80)和式(2.81)中通过一个明显的公式

$$\dot{\xi}_s(t) = f_s(\xi_u(t), t) \quad (2.82)$$

来消取伺服模。我的《高等协同学》一书给出了 $f_s$ 的明显构造,一些特殊情形下的构造,也可以在《协同学导论》一书中找到。这里,我们只通过一个例子来说明式(2.82)的内容,我们只给出伺服原理的首项。为此,让我们考虑下面形式的 $\xi_u$ 和 $\xi_s$ 的方程

$$\dot{\xi}_u = \lambda_u \xi_u + h_u(\xi_u, \xi_s) + F_u(t) \quad (2.83a)$$

$$\dot{\xi}_s = \lambda_s \xi_s + g_s(\xi_u) + q_s k_s(\xi_u) + F_s(t) \quad (2.83b)$$

这里,我们假定 $h_u$ 是一个非线性函数,其最低次非线性项是二次的。 $g_s$ 也是类似的非线性函数。这时,可以证明, $\xi_s$ 的最低次项是 $\xi_u$ 的二次项。伺服原理的最简单形式归结为令式(2.83b)中的 $\dot{\xi}_s$ 为零。保留打头项,我们容易得到

$$\xi_s = -\frac{1}{\lambda_s} g_s(\xi_u) - \frac{1}{\lambda_s} F_s(t) \quad (2.84)$$

可以严格证明,此结果是 $F_s$ 和 $\xi_u$ 的最低阶。

我们现在想研究,对于福克-普朗克方程的解,伺服原理意

意味着什么。为此,我们将福克-普朗克方程从旧变量 $\underline{q}$ 变换成新变量 $\underline{\xi}_u, \underline{\xi}_s$ ,

$$\underline{q} \rightarrow \underline{\xi}_u, \underline{\xi}_s \quad (2.85)$$

这时,福克-普朗克方程的一般形式是

$$\dot{P}(\underline{\xi}_u, \underline{\xi}_s; t) = L(\underline{\xi}_u, \underline{\xi}_s) P(\underline{\xi}_u, \underline{\xi}_s; t) \quad (2.86)$$

其中 $L$ 是一个线性算子。让我们考虑式(2.86)的定态解,它总可以写成如下形式:

$$P(\underline{\xi}_u, \underline{\xi}_s) = P(\underline{\xi}_s | \underline{\xi}_u) f(\underline{\xi}_u) \quad (2.87)$$

其中左边是一个联合概率分布,而右边的 $P$ 是一个条件概率。 $f$ 只是序参量的分布函数。伺服原理意味着式(2.87)右边的条件概率 $P$ 可以更进一步写为

$$P(\underline{\xi}_s | \underline{\xi}_u) = \prod_i P_i(\xi_i | \xi_u) \quad (2.88)$$

我们的结果式(2.84)现在可以给出一个具体的例子,以便我们了解伺服原理的最低阶式(2.88)究竟是什么样子。象通常一样,涨落力是高斯分布的,即是说,在 $F$ 到 $F+dF$ 区间发现 $F_s$ 的概率是

$$P(F \leq F_s \leq F + dF) = N' \exp(-F_s^2/Q') dF \quad (2.89)$$

现在,我们可以在式(2.84)中将 $F_s$ 解出,

$$F_s = -\lambda_s(\xi_s - \frac{1}{\lambda_s} g_s(\xi_u)) \quad (2.90)$$

这使得我们能够利用式(2.89)来确定条件概率。这样,我们得到

$$P(\xi_s | \xi_u) = N \exp[-(\xi_s - \frac{1}{\lambda_s} g_s(\xi_u))^2/Q] d\xi_s \quad (2.91)$$

其中我们已利用了简写

$$Q^{-1} = Q'^{-1} \lambda_s^2 \quad (2.92)$$

$$N = N' \lambda, \quad (2.93)$$

## 2.7 非平衡相变

在许多感兴趣的实际问题中，序参量的数目很少甚或为 1，而伺服模的数目仍然很大。让我们考虑一个序参量的情况，为了简单起见，我们将丢掉指标  $u$ ，

$$\xi_u \rightarrow \xi, F_u \rightarrow F, \lambda_u \rightarrow \lambda \quad (2.94)$$

一个典型的序参量方程是

$$\dot{\xi} = \lambda \xi - \beta \xi^3 + F(t) \quad (2.95)$$

如协同学中表明的那样。涨落力  $F(t)$  服从关系

$$\langle F(t)F(t') \rangle = Q \delta(t-t') \quad (2.96)$$

此外，如果我们假设  $F$  是高斯分布的，我们可以建立与式(2.95)相应的福克-普朗克方程

$$\dot{f}(\xi, t) = -\frac{\partial}{\partial \xi} [(\lambda \xi - \beta \xi^3) f] + \frac{Q}{2} \frac{\partial^2}{\partial \xi^2} f \quad (2.97)$$

让我们考虑定态解

$$\dot{f} = 0 \quad (2.98)$$

这时式 (2.97) 易于积分，我们得到

$$f = N \exp [Q^{-1}(\lambda \xi^2 - \beta \xi^4/2)] \quad (2.99)$$

式 (2.99) 为我们提供了出现在式 (2.87) 中的分布函数的一个既明显又典型的例子。如果细致平衡原理成立，我们还可以求出几个序参量的分布函数。我们假定  $\lambda_u$  是实的，细致平衡原理的一个特殊情形可以表示如下，此时，朗之万方程是

$$\dot{\xi}_u = \lambda_u \xi_u + \frac{\partial V}{\partial \xi_u} + F_u(t) \quad (2.100)$$



其中  $V$  是  $\xi_u$  的非线性函数, 我们假定

$$\langle F_u(t) F_{u'}(t') \rangle = \delta_{uu'} Q \delta(t - t') \quad (2.101)$$

这时, 与式 (2.100) 的朗之万方程相应的福克-普朗克方程的解可以一般地写为

$$f(\xi_u) = N \exp [-2V(\xi_u)/Q] \quad (2.102)$$

甚至当细致平衡原理不成立时, 只要分布函数是单连的, 这时解总可以写为

$$f(\xi_u) = N \exp [-\Phi(\xi_u)] \quad (2.103)$$

其中  $\Phi$  起着广义热力学势的作用。现在让我们考虑, 当我们改变控制参数时, 一个软的相变发生, 系统将变得不稳定, 进入一个新的区域。在这种情况下, 我们可以假定序参量仍然很小, 例如, 条件

$$\xi_u \propto \lambda_u^{\kappa} \quad (\kappa > 0) \quad (2.104)$$

成立。这时, 我们可以将  $\Phi$  展成关于  $\xi_u$  的幂级数, 即

$$\begin{aligned} \Phi(\xi_u) = & \sum_u \lambda_u \xi_u^2 + \sum_{uu'} c_{uu'} \xi_u \xi_{u'} \\ & + \sum_{uu'v} c_{uu'v} \xi_u \xi_{u'} \xi_v + \sum_{uu'v'w} c_{uu'v'w} \xi_u \xi_{u'} \xi_v \xi_{w'} \\ & \xi_u \xi_{u'} \xi_{u''} \xi_{u'''} \dots \end{aligned} \quad (2.105)$$

在非平衡相变的许多情形中, 保留到四阶就足够了, 但在一些例外的情况下, 更高阶也是需要的。一般情况下, 系统具有内部对称性, 这将在每一阶的系数  $c$  之间增加一些特定关系。

## 2.8 图样形成

现在, 我们要阐明, 以上的表述怎样用来描述图样的形成。

如果我们处理一个连续的广延系统，这时，我们需要用空间坐标描述，线性化方程 (2.72) 中的算子  $L$  一般地将含有空间坐标的导数。这种情况下，出现在式 (2.73) 中的  $\tilde{V}$  和  $\tilde{w}$  变成空间坐标的函数 (2.79)。

非线性方程 (2.66) 的解可以非常一般地写为

$$\tilde{q} = \tilde{q}_0 + \sum_u \xi_u(t) \tilde{V}_u(\tilde{x}) + \sum \xi_v(t) \tilde{V}_v(\tilde{x}) \quad (2.106)$$

它只是以前的公式 (2.78) 的重复。

结果表明，一般情况下， $\xi_u$  比  $\xi_v$  高一阶，即演化着的图样主要由式 (2.106) 中第一个对  $u$  的求和来决定，我们因此称它为模的骨架。如果只有一个序参量， $\tilde{V}_u$  的形式为

$$\tilde{V}_u = L^{-\frac{1}{2}} \sin kx \quad (2.107)$$

式 (2.106) 实际上是

$$\tilde{q} = \tilde{q}_0 + \xi_u(t) L^{-\frac{1}{2}} \sin kx \quad (2.108)$$

在许多情况下， $\xi_u$  服从式 (2.95) 的方程，它描述  $\xi_u$  从一个初始的涨落长成最后的样子 (如图 2.3 和图 2.4 所示)。非常明显，如果序参量的个数更多，或者函数  $\tilde{V}_u$  更复杂，我们能够得到的图样比式 (2.107) 和式 (2.108) 所描述的复杂得多。因此，这个理论能够导出复杂系统出现的空间结构。我们这里的例子只是简单提醒读者，我的《协同学导论》和《高等协同学》为读者提供了什么内容。我只想让读者知道，那里已处理了非常一般的情况，例如演化着的含时图样，如极限环、准周期运动或混沌。但在以后的许多情况下，这里的简略提示已足够了。

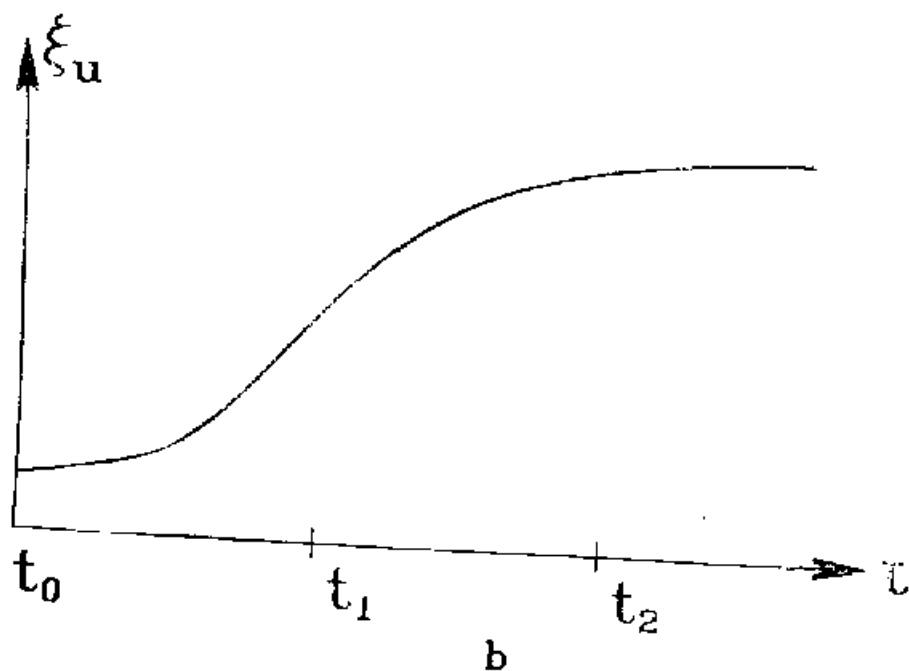
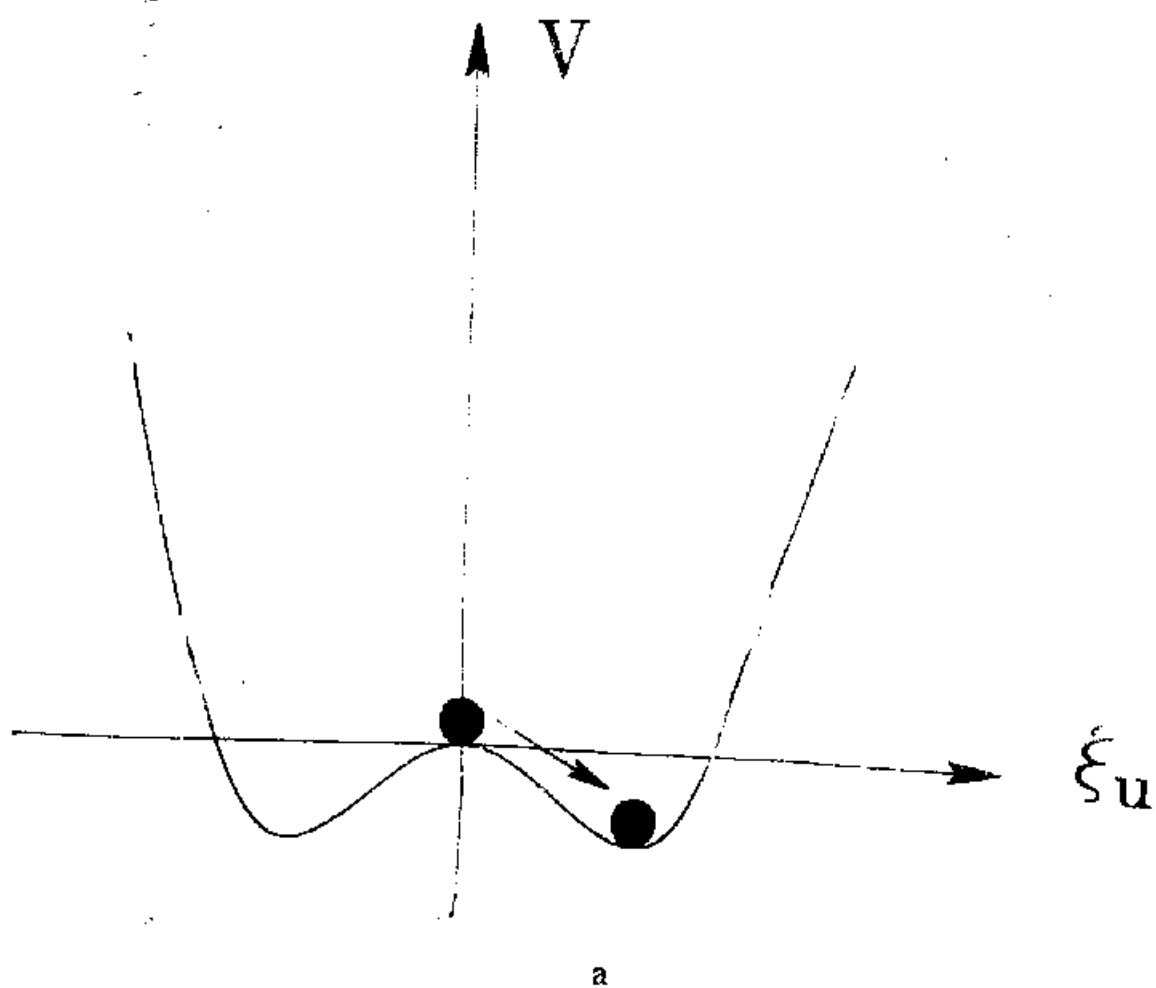


图2.3 序参量的行为随时间的变化。(a)势函数,其中坐标为 $\xi_u$ 的虚拟粒子在运动,(b) $\xi_u$ 的时间演化。

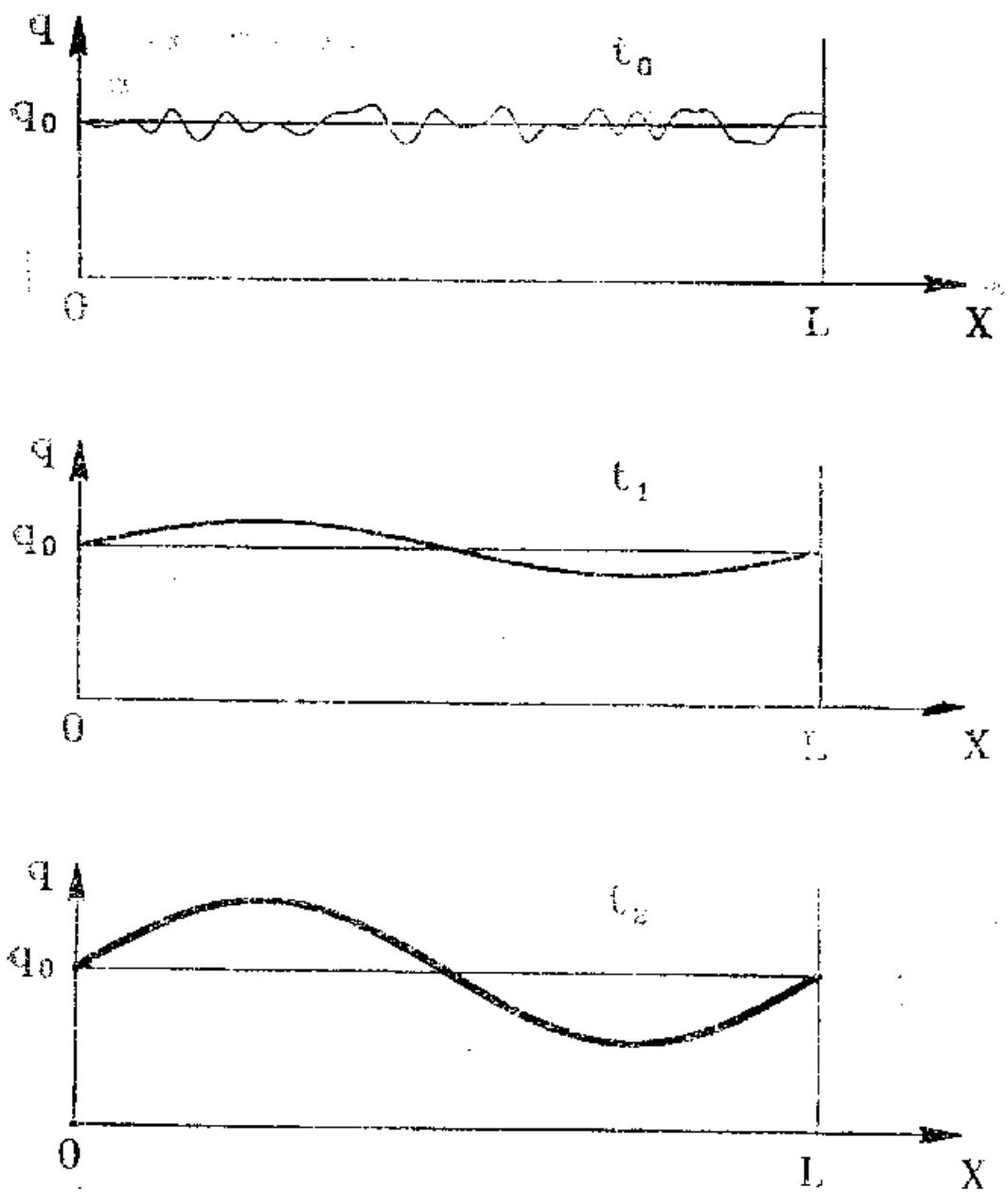


图2.4 由式(2.108)描述的空间图样随时间的演化, 它由图2.3所示的序参量所描述。

## 第三章 最大信息原理(MIP)

---

### 3.1 若干基本概念

本章讨论下述问题:

在一个系统的若干宏观量已经给定的条件下,我们希望设计一种方法,用以导出宏观变量乃至微观变量的概率分布。换句话说,希望从宏观世界出发,导出微观世界的一些结论。由于所研究系统的类型不同,所需的宏观量有很大差异。对于应用热力学的封闭物理系统,这些宏观量是能量、粒子数,等等。第四章将以此种系统为例讨论处理问题的一般方法。对于开放系统,比如物理学或生物学中的开放系统,将证明,所需的宏观量是强度量和强度量的涨落,等等。实际上,处理开放系统是以下几章的主要论题。信息概念是我们研究方法的出发点,本节将导出这一概念。

根据概率论的一些新的解释,可以深刻理解信息论这门看上去迥然不同的学科。考虑有 0 和 1 两种结果的抛掷硬币序列,并

把它们理解为莫尔斯 (Morse) 电码的线和点。众所周知, 莫尔斯电码可以用来传送讯号。因此, 只要赋予确定的符号序列以确定的含义, 就可以传送信息。也就是说, 确定的符号序列载有信息。在信息论中, 我们试图找出信息量的度量。

考虑一个具有  $R_0$  种可能事件 (实现) 的简单例子。这些事件是先验等概率的。例如, 抛掷硬币时, 可能事件为 0 和 1, 这时,  $R_0 = 2$ 。而抛掷骰子时, 有 6 种不同的可能结果, 于是  $R_0 = 6$ 。抛掷硬币或骰子的结果可以理解为接收到一个讯号, 而且在  $R_0$  种可能结果中, 只有一种是真正实现了的。很显然,  $R_0$  愈大, 在接收到讯号之前, 事件的不确定性就愈大, 从而接收到的讯号的信息量愈大。因此, 可以以如下方式解释整个过程: 初始时我们没有接收到任何信息, 即  $I_0 = 0$ , 此时可能出现的结果有  $R_0$  种。末了时, 我们收到信息量为  $I_1$ ,  $I_1 \neq 0$ , 而  $R_1 = 1$ , 亦即出现的结果只有一种。现在, 引入信息量  $I$  的度量, 显然,  $I$  与  $R_0$  必定有联系。那么,  $R_0$  与  $I$  是如何联系的呢? 我们认为对于独立事件, 要求  $I$  具有相加性。因此, 对于这样两组可能实现数分别为  $R_{01}$  和  $R_{02}$  的独立事件, 其总的可能实现数应是

$$R_0 = R_{01} R_{02} \quad (3.1)$$

按相加性要求

$$I(R_{01} R_{02}) = I(R_{01}) + I(R_{02}) \quad (3.2)$$

只要选择

$$I = K \ln R_0 \quad (3.3)$$

关系式 (3.2) 就可以得到满足, 这里  $K$  是常数。还可以证明式 (3.3) 是式 (3.2) 的唯一解, 常数  $K$  是任意的, 但可以通过某种定义给定。通常采用下面的定义。考虑一个只有二种符号 (或

字母)的所谓二元系统。例如,一个硬币的正面和背面,或者回答中的是和否,或者二进制中的数码“0”和“1”。当我们构造字长为 $n$ 的“词”(数列)时,所有可能实现数 $R = 2^n$ 。在此二元系中,我们欲令 $I$ 与 $n$ 相等,因此,要求

$$I \equiv K \ln R = Kn \ln 2 = n \quad (3.4)$$

当

$$K = 1/\ln 2 = \log_2 e \quad (3.5)$$

时,上式成立。这时,式(3.4)可写为形式

$$I = \log_2 R$$

对于二元系,符号序列中的一个位置称为“彼特”(“bit”),因此,信息量 $I$ 便直接用“彼特”表示。这样, $R = 8 = 2^3$ ,则 $I = 3$ 彼特。一般地说,若 $R = 2^n$ ,则 $I = n$ 彼特。式(3.3)中的信息量很容易推广到初态有 $R_0$ 种等几率事件,末态有 $R_1$ 种等几率事件的情形。这时信息量

$$I = K \ln R_0 - K \ln R_1 \quad (3.6)$$

当 $R_1 = 1$ 时,此式简化成式(3.3)的形式。掷骰子就是这种情况的一个简单例子。设计一个游戏,偶数表示赢,奇数表示输,则 $R_0 = 6$ , $R_1 = 3$ 。这种情况的信息量与掷硬币只有二种可能结果情况相同。

现在推导信息量的一种更为方便的表达式。

为此,首先考虑一个简化的莫尔斯电码的例子,其中只包括“线”和“点”(实际的莫尔斯电码还包括第三个符号——“间歇”)。对一个字长为 $N$ 的词,它由 $N_1$ 条线, $N_2$ 个点组成,即

$$N = N_1 + N_2 \quad (3.7)$$

那么,接收到这个词的信息量是多少呢?根据信息论的精神,必

须计算 $N_1, N_2$ 给定时, 可能构造的词의总数。处理方法与2.1节类似。将“线”和“点”分配到 $N$ 个位置上的可能方式有

$$R = \frac{N!}{N_1! N_2!} \quad (3.8)$$

种。换句话说,  $R$ 是 $N_1$ 条线和 $N_2$ 个点所能传送的讯号的总数。下面推导单位符号的信息量 $i, i = I/N$ 。将式(3.8)代入公式(3.3)得到

$$I = K \ln R = K [\ln N! - \ln N_1! - \ln N_2!] \quad (3.9)$$

利用斯特林 (Stirling) 近似公式

$$\ln N! = N(\ln N - 1) \quad (3.10)$$

(对于 $N > 100$ , 这是很好的近似), 易得

$$I \approx K [N(\ln N - 1) - N_1(\ln N_1 - 1) - N_2(\ln N_2 - 1)] \quad (3.11)$$

利用式(3.7), 得到

$$i \equiv I/N \approx -K \left[ \frac{N_1}{N} \ln \frac{N_1}{N} + \frac{N_2}{N} \ln \frac{N_2}{N} \right] \quad (3.12)$$

现在引入一个量, 可以理解为发现信号“线”或“点”的概率。此概率等于“线”或“点”出现的频率

$$P_j = N_j/N, \quad j = 1, 2 \quad (3.13)$$

借此, 最后一个公式取下列形式

$$i = I/N = -K(P_1 \ln P_1 + P_2 \ln P_2) \quad (3.14)$$

此表达式很容易推广到多于两种符号的情形, 例如字母表中的字母。非常类似地, 可得到单位符号的信息量的表达式, 即

$$i = -K \sum_j P_j \ln P_j \quad (3.15)$$

其中 $P_j$ 是符号出现的相对频率。根据这种解释, 显然, 单位符号信息量 $i$ 可用于信息传输等问题。



继续讨论之前，先说明一下这里所讲的信息这个词的含义。必须注意这个理论并不包含“有用”或“无用”、“有意义”或“无意义”等概念。例如，上面定义的莫尔斯电码中，很多词就可能没有意义。因此可以说，这里用到的信息是指事件的不确定性。这样看来，我们的理论似乎受到很大的限制，但事实上，终将证明它是非常有用的。

可以从两个完全不同的角度来理解信息表达式。其一，赋予  $P_i$  一个数值，利用式 (3.3)，相应地可以得出信息量  $I$  的一个值。第二种理解更为重要。将  $I$  看作  $P_i$  的函数，亦即，若改变  $P_i$  的值，则  $I$  的值随着作相应的改变。为了解释得更清楚，我们介绍一例应用，这例应用以后还要详细讨论。考虑在容器中作自由运动的气体原子（我们关心这些原子的空间分布），把容器划分成  $M$  个等体积元，第  $k$  个体积元中的粒子数记为  $N_k$ ，粒子总数为  $N$ ，在第  $k$  个体积元中发现一个粒子的相对频率是

$$N_k/N = P_k, \quad k = 1, 2, \dots, M \quad (3.16)$$

可以认为， $P_k$  是粒子遍布在各体积元  $k$  的分布函数。由于各体积元大小相同，物理性质亦无差异，因此，可以断定在各体积中发现粒子的概率相等，即

$$P_k = 1/M \quad (3.17)$$

现在欲由信息的性质导出公式 (3.17)。实际上，信息可能是这样的。在未接收到信息时或未作测量之前，存在  $R$  种可能结果，或者说， $K \ln R$  是我们无知（无法确知实现结果）的度量。另一种理解方法是： $R$  是原则上可能实现的数目。

现在考察一个共有  $M$  个容器的系统，每个容器有  $N$  个气体原子，假定每个容器中粒子数按不同的分布函数  $P_k$  分布，即

$$P_k(1), P_k(2), P_k(3), \dots$$

与此相应,得到不同的实现数目,即有不同的信息量。例如,如果  $N_1 = N, N_2 = N_3 = \dots = 0$ , 则  $P_1(1) = 1, P_2(1) = P_3(1) = \dots = 0$ , 那么  $I(1) = 0$ , 反之, 如果  $N_1 = N_2 = N_3 = \dots = N/M$ , 则  $P_1(2) = 1/M, P_2(2) = 1/M, \dots$ , 那么  $I(2) = -M \log_2 1/M = M \log_2 M$ , 当容器的数目  $M$  很大时, 信息量  $I$  是一个很大的数。

因而, 对于任何一个气体容器, 气体分子按以上第二种情形分布的可能性比按第一种情况分布的可能性大得多, 这意味着, 具有最大可能实现数  $R$  的概率分布  $P_i$  出现的概率占绝对优势。因而这种分布有最大信息, 于是, 要求

$$-\sum P_i \ln P_i = \text{Extr!} \quad (3.18)$$

成立, 并满足概率  $P_i$  的总和等于 1 的约束条件

$$\sum_{i=1}^M P_i = 1 \quad (3.19)$$

上述原理将被证明为是实际的物理、化学、生物系统的基本原理, 我们将在后面讨论它。

利用拉格朗日乘子法来求解式 (3.18) 和式 (3.19), 方法是将待定的  $\lambda$  乘以式 (3.19), 并与式 (3.18) 左边相加, 要求总的表达式仍为极大值, 这时允许所有  $P_i$  是相互独立的, 不再考虑式 (3.19) 的约束条件。对等式

$$-\sum P_i \ln P_i + \lambda \sum P_i = \text{Extr!} \quad (3.20)$$

左边求变分, 相当于对  $P_i$  求偏导, 得出

$$-\ln P_i - 1 + \lambda = 0 \quad (3.21)$$

由此得出解

$$P_i = \exp(\lambda - 1) \quad (3.22)$$

解  $P_i$  与下标 “ $i$ ” 无关, 即  $P_i$  是常数, 把它代回式 (3.19), 便可

定出 $\lambda$ 值,

$$M \exp(\lambda - 1) = 1 \quad (3.23)$$

于是, 求出

$$P_i = 1/M \quad (3.24)$$

正如所料, 它与式 (3.17) 一致。

### 3.2 信息增益

信息表达式 (3.15) 可以解释为 $f_i$ 的平均值

$$\bar{i} = \sum P_i f_i \quad (3.25)$$

这里  $f_i = -K \ln P_i, P_i \neq 0$  (3.26)

$f_i$ 的权重因子是 $P_i$ 。这意味着可把式 (3.26) 解释为指标 $i$ 的符号出现的概率为 $P_i$ , 其信息量 $f_i$ 。

假定一组测量中得到指标为 $i$ 的符号的相对频率为 $P_i$ , 而在不同的条件下, 可能测得另一新的相对频率值 $P_i'$ 。那么, 与此相应的信息变化量 $\Delta_i$ 是多少?

采用式 (3.26) 的解释, 它可直接定义为

$$\Delta_i = K \ln P_i' - K \ln P_i \quad (3.27)$$

为了得到信息变化量的平均值, 把式(3.27)按新的分布函数 (相对频率)  $P_i'$  求平均, 就得所谓信息增益(或“库拜克(Kullback) 信息”)

$$K(P', P) = \sum_i P_i' \Delta_i = K \sum_i P_i' \ln P_i' / P_i \quad (3.28)$$

当然, 有

$$\sum_i P_i = 1 \quad (3.29)$$

$$\sum_i P_i' = 1 \quad (3.30)$$

信息增益 $K(P', P)$  具有如下重要性质:

$$K(P', P) \geq 0 \quad (3.31)$$

当且仅当

$$P' = P$$

时, 即对所有 $k$ 的值都有 $P_k' = P_k$ 时, 式(3.31)中等式成立。

### 3.3 信息熵和约束条件

这一节和下一章, 我们举一个信息概念在物理学中应用的例子。习惯上, 将信息记为 $S$ , 并令式(3.3)中的常数 $K$ 等于玻尔兹曼常数 $k_B$ 。后面将会看到, 由于诸多原因,  $S$ 被称为信息熵。由于化学和生物系统可以视为物理系统, 因此, 我们的研究同样适合于这两种系统。本章的一般公式同样适用于其它学科, 如信息过程等。现在, 我们从基本表达式

$$S = -k_B \sum_i P_i \ln P_i \quad (3.32)$$

出发。其中下标“ $i$ ”可以看作是用来描述粒子或子系统个体特征的。对此, 我们作一些解释。例如下标 $i$ 可以描述气体粒子的位置或速率, 或者同时描述位置和速率。在更一般的描述中, 下标 $i$ 代表可能获得的随机变量的值。为简单起见, 这里假设 $i$ 是离散的。

本书要解决的中心问题之一是找出决定 $P_i$ 的方法(比如求容器中气体分子的位置)。在很多学科中我们所面临的问题是如何作出正确的估计, 使导出的 $P_i$ 与关于系统的所有已知知识相一致。考虑一种一维理想气体。例如, 可测量的是重心, 这时, 约束条件的表达式如下

$$\sum_i P_i q_i = M \quad (3.33)$$

其中 $q_i$ 是体积元 $i$ 的位置,  $M$ 是定值, 等于 $Q/N$ ,  $Q$ 为重心坐标,

$N$ 为粒子数。当然，满足式(3.33)的 $P_i$ 有很多组，可以任意选定一组 $\{P_i\}$ ，究竟选定哪一组，这纯粹是出于偏爱，就象在日常生活中一样。如何才能避免这种偏爱呢？让我们再回到气体原子的例子上来。引用3.1节所提出的原理，得出：式(3.32)取极大值时所对应的分布占出现几率的绝对优势。但由于有约束条件式(3.33)的限制，并非所有的分布都需要考虑。因此，必须找出满足式(3.33)，同时使式(3.32)取极大值的分布函数。此原理可以推广到有一组约束条件的情形。例如，用变量 $i$ 区分不同的速度，粒子动能 $E_i$ 取值一定，就是一个约束条件。粒子的动能用粒子质量 $m$ 和粒子速度 $v_i$ 表为 $f_i$  ( $f_i = mv_i^2/2$ )。每个粒子平均动能为

$$\sum_i P_i f_i = E_{kin} \quad (3.33a)$$

一般地，对单个系统，量 $i$ 可由量 $f_i^{(k)}$ ， $k=1, 2, \dots, M$ （位置、动能或其它典型特征量）来表征。如果这些特征量满足相加性，其总和 $f_k$ 取值一定，则有如下形式约束的条件：

$$\sum_i P_i f_i^{(k)} = f_k \quad (3.34)$$

通常还要加上概率分布归一化条件

$$\sum_i P_i = 1 \quad (3.35)$$

在条件(3.34)和(3.35)的约束之下，求解式(3.32)的极值问题可用拉格朗日乘子法，其参数为 $\lambda_k$ ， $k=1, 2, \dots, M$ （参见3.1节）。将式(3.34)左边乘以 $\lambda_k$ ，式(3.35)左边乘以 $(\lambda - 1)$ ，将所得二表示式相加，然后从 $(1/k_B)S$ 中减去此和数，因子 $1/k_B$ 相当于 $\lambda$ ， $\lambda_k$ 的某种归一化，然后，对 $P_i$ 求总和并作变分：

$$\delta \left[ -\frac{1}{k_B} S - (\lambda - 1) \sum_i P_i - \sum_k \lambda_k \sum_i P_i f_i^{(k)} \right] = 0 \quad (3.36)$$

上式对  $P_i$  微分, 并令所得到的表达式为零, 有

$$-\ln P_i - 1 - (\lambda - 1) - \sum_k \lambda_k f_i^{(k)} = 0 \quad (3.37)$$

由此立即解出  $P_i$ , 为

$$P_i = \exp \left\{ -\lambda - \sum_k \lambda_k f_i^{(k)} \right\} \quad (3.38)$$

把式 (3.38) 代入式 (3.35), 得到

$$e^{-\lambda} \sum_i \exp \left\{ -\sum_k \lambda_k f_i^{(k)} \right\} = 1 \quad (3.39)$$

为了方便起见, 把式 (3.39) 中对  $P_i$  的求和  $\sum_i$  缩记为

$$\sum_i \exp \left\{ -\sum_k \lambda_k f_i^{(k)} \right\} = Z(\lambda_1, \dots, \lambda_M) \quad (3.40)$$

$Z$  称为配分函数。把式 (3.40) 代入式 (3.39), 则

$$e^{-\lambda} = Z \quad (3.41)$$

或者

$$\lambda = \ln Z \quad (3.42)$$

因此, 只要确定了所有  $\lambda_k$  的值,  $\lambda$  的值便可确定。为了求  $\lambda_k$  所满足的方程, 把式 (3.38) 代到约束方程 (3.34), 即知

$$\langle f_i^{(k)} \rangle = \sum_i P_i f_i^{(k)} = e^{-\lambda} \sum_i \exp \left\{ -\sum_l \lambda_l f_i^{(l)} \right\} f_i^{(k)} \quad (3.43)$$

式 (3.43) 与式 (3.40) 有类似的结构。而两者的差异来源于式

(3.43) 中要以  $f_i^{(k)}$  乘各指数函数。然而, 把式 (3.40) 对  $\lambda_k$  求微分, 则很容易导出式 (3.43) 中的求和。根据式 (3.41), 可用配分函数来表示式 (3.43) 中等式右边的第一个因子, 从而有

$$\langle f_i^{(k)} \rangle = \frac{1}{Z} \left( -\frac{\partial}{\partial \lambda_k} \right) \underbrace{\sum_i \exp \left\{ -\sum_l \lambda_l f_i^{(l)} \right\}}_Z \quad (3.44)$$

或写为更简洁的形式

$$f_k \equiv \langle f_i^{(k)} \rangle = - \frac{\partial \ln Z}{\partial \lambda_k} \quad (3.45)$$

由于上式左边的量已知（参见式（3.34）），且式（3.40）又给出以 $\lambda_k$ （ $k = 1, \dots, M$ ）为自变量的函数 $Z$ 的特殊形式，因此，式（3.45）是关于 $\lambda_k$ （ $k = 1, \dots, M$ ）的一组方程，它们具有简洁的形式。

我们进一步引用一个以后非常有用的公式。把式（3.38）代入式（3.32），得到

$$\frac{1}{k_B} S_{\max} = \lambda \sum_i P_i + \sum_k \lambda_k \sum_i P_i f_i^{(k)} \quad (3.46)$$

利用式（3.34）和式（3.35），上式可重写为

$$\frac{1}{k_B} S_{\max} = \lambda + \sum_k \lambda_k f_k \quad (3.47)$$

于是，最大信息熵可用 $f_k$ 的平均值和拉格朗日参数 $\lambda_k$ 表出。熟悉力学中第一类拉格朗日方程的读者一定记得，拉格朗日参数有其物理意义，在力学中，它们就代表力。同样，在后面的讨论中，我们将会看到，这里的拉格朗日参数 $\lambda_k$ 也有其物理（或化学的、生物的）意义。通过对上面公式[即式（3.38）、式（3.42）和式（3.32），式（3.45）和式（3.47）]求导，我们最初提出要推求 $P_i$ 和 $S_{\max}$ 的任务就完成了。

现在进一步推导一些有用的公式。

首先看式（3.34），当函数 $f_i^{(k)}$ 和 $f_k$ 发生变化时，最大信息熵 $S_{\max}$ 是如何随之改变的。根据式（3.47）， $S$ 不仅依赖于 $f_k$ （ $k = 1, \dots, M$ ），而且还依赖于 $\lambda$ 和 $\lambda_k$ ，而它们又是 $f_k$ 的函数，因此，对 $f_k$ 求偏导时，必须特别小心。首先由式（3.42）计算 $\lambda$ 的

变化

$$\delta\lambda = \delta \ln Z = \frac{1}{Z} \delta Z$$

将Z用式 (3.49) 表示, 得到

$$\delta\lambda = e^{-\lambda} \sum_i \sum_k \{ -\delta\lambda_k f_i^{(k)} - \lambda_k \delta f_i^{(k)} \} \exp \{ - \sum_i \lambda_i f_i^{(1)} \}$$

又利用式 (3.38)  $P_i$  的定义, 则上式变为

$$\delta\lambda = - \sum_k [ \delta\lambda_k \sum_i P_i f_i^{(k)} + \lambda_k \sum_i P_i \delta f_i^{(k)} ]$$

利用式 (3.43) 和对  $\langle \delta f_i^{(k)} \rangle$  的类似定义, 上式可写成:

$$\delta\lambda = - \sum_k [ \delta\lambda_k \langle f_i^{(k)} \rangle + \lambda_k \langle \delta f_i^{(k)} \rangle ] \quad (3.48)$$

将此式代入  $\delta S_{\text{max}}$  [式 (3.47)], 这时有关  $\lambda_k$  的变分  $\delta\lambda_k$  已消去, 得到

$$\delta S_{\text{max}} = k_B \sum_k \lambda_k [ \delta \langle f_i^{(k)} \rangle - \langle \delta f_i^{(k)} \rangle ] \quad (3.49)$$

将它写成下面形式

$$\delta S_{\text{max}} = k_B \sum_k \lambda_k \delta Q_k \quad (3.50)$$

其中由下式定义了“广义热”  $\delta Q_k$ ,

$$\delta Q_k = \delta \langle f_i^{(k)} \rangle - \langle \delta f_i^{(k)} \rangle \quad (3.51)$$

在后面, 与热力学联系起来讨论时, “广义热”的含义会更加明确。与式 (3.45) 类似, 可以导出  $f_i^{(k)}$  的均方差的简单表示式

$$\langle f_i^{(k)2} \rangle - \langle f_i^{(k)} \rangle^2 = \frac{\partial^2 \ln Z}{\partial \lambda_k^2} \quad (3.52)$$

在许多实际应用中,  $f_i^{(k)}$  依赖于另外的量  $\alpha$  (或者一组这样的量  $\alpha_1, \alpha_2, \dots$ )。我们要求得  $\alpha$  变化时, 平均值 (3.34) 的变化。把

$f_{i,k}^{(k)}$  对  $\alpha$  求偏导, 并取平均值, 得到



$$\left\langle \frac{\partial f_{i,a}^{(k)}}{\partial \alpha} \right\rangle = \sum_i P_i \frac{\partial f_{i,a}^{(k)}}{\partial \alpha} \quad (3.53)$$

利用式 (3.38) 关于  $P_i$  的表示和式 (3.41), 式 (3.53) 的右边可写成如下形式,

$$\frac{1}{Z} \sum_i \frac{\partial f_{i,a}^{(k)}}{\partial \alpha} \exp \left( - \sum_k \lambda_k f_{i,a}^{(k)} \right) \quad (3.54)$$

很容易将它表示成  $Z$  对  $\alpha$  的偏导形式, 即

$$(3.54) = - \frac{1}{Z} \frac{1}{\lambda_k} \frac{\partial Z}{\partial \alpha} \quad (3.55)$$

于是, 得到最后公式

$$\left\langle \frac{\partial f_{i,a}^{(k)}}{\partial \alpha} \right\rangle = - \frac{1}{Z} \frac{1}{\lambda_k} \frac{\partial Z}{\partial \alpha} \quad (3.56)$$

如果有几个参量  $\alpha_i$  存在, 将上式左右两边的  $\alpha$  换成  $\alpha_i$ , 则可以推广这个公式。

我们已经多次看到, 量  $Z$  [见式 (3.40)] 或它的对数是很有用的 [见式 (3.45)、式 (3.52) 和式 (3.56)]。我们希望证明,  $\ln Z \equiv \ln$  [参见式 (3.42)] 可以直接用变分原理确定。很容易看出, 对式 (3.36), 可用如下方法加以解释: 求下式的极大值

$$\frac{1}{k_B} S - \sum_k \lambda_k \sum_i P_i f_i^{(k)} \quad (3.57)$$

满足仅有的约束条件

$$\sum_i P_i = 1 \quad (3.58)$$

利用式 (3.34)、式 (3.47) 和式 (3.42), 可以得出式 (3.57) 的极大值就等于  $\ln Z$ 。注意, 对于  $\ln Z$  的变分与对  $S$  的变分不同。在前一种情形, 我们必须求出在式 (3.34) 和式 (3.35) 的约束下,  $S$  的极大值, 这时  $f_i$  固定而  $\lambda_k$  未知。而这里, 只有式 (3.58) 一个约束条件, 同时假定  $\lambda_k$  是给定的, 那么, 这种从一组固定量到另

一组固定量的变换如何进行呢？通过下一章的一个物理例子，我们就会弄清这一点，而且，前面提到的更多方面将通过这个例子予以阐明。

### 3.4 连续变量

在许多已知的应用中，变量 $\xi$ 是连续的，而不是离散的。因而很容易证明，在这种情况下，一般说来，信息量是发散的，这是由于存在很多连续变化的状态。有必要简要地讨论对于连续变量的信息量的定义。我们从概率密度的定义

$$\text{Prob}(X \leq \xi \leq X + \Delta\xi) = P(\xi) \Delta\xi \quad (3.59)$$

着手，来研究这个问题。现在引入精度区间(*Interval of accuracy*)的概念，并重新定义概率分布：

$$P_\varepsilon(j) = \int_{\xi_j - \varepsilon/2}^{\xi_j + \varepsilon/2} P(\xi) d\xi \quad (3.60)$$

假定 $P(\xi)$ 是连续的，式(3.60)可近似表示为

$$P_\varepsilon(j) \approx P(\xi_j) \cdot \varepsilon \quad (3.61)$$

现在对于精度区间 $\varepsilon$ ，利用下式定义信息量

$$I_\varepsilon = - \sum_j P_\varepsilon(j) \ln P_\varepsilon(j) \quad (3.62)$$

把式(3.61)代入式(3.62)，得到

$$I_\varepsilon = - \sum_j \varepsilon P(\xi_j) \ln P(\xi_j) - \sum_j \varepsilon P(\xi_j) \ln \varepsilon \quad (3.63)$$

从而得到最终结果

$$I_\varepsilon = - \int d\xi P(\xi) \ln P(\xi) - \ln \varepsilon \quad (3.64)$$

以后，我们将略去常数和不感兴趣的项 $-\ln \varepsilon$ 。显然，上式可以推广到多个变量的情形。

## 第四章 物理学中的一个 例子: 热力学

---

为了使  $i$  的含义具体化, 假定它代表一个粒子的速度。在更加近代的理论中,  $P_i$  是一个多粒子体系中量子态  $i$  的占有概率。进一步假定  $f_{i,\alpha}^{(k)}$  代表能量  $E$ , 参量  $\alpha$  代表体积, 于是令

$$f_{i,\alpha}^{(k)} = E_i(V), \quad k=1, \quad (4.1)$$

另外, 有对应关系

$$f_1 \longleftrightarrow U \equiv \langle E_1 \rangle; \quad \alpha \longleftrightarrow V; \quad \lambda_1 = \beta \quad (4.2)$$

其中, 已特别令  $\lambda_1 = \beta$ 。借此, 可将前面的一些公式用我们在热力学、统计物理学中熟知的关系式直接表示出来。代入式(3.38), 我们有

$$P_i = e^{-\lambda - \beta E_i} \quad (4.3)$$

这正是著名的玻尔兹曼分布函数。方程 (3.47) 成为

$$\frac{1}{k_B} S_{\max} = \ln Z + \beta U \quad (4.4)$$

或经适当移项后,

$$U - \frac{1}{k_B \beta} S_{\max} = - \frac{1}{\beta} \ln Z \quad (4.5)$$

这是热力学和统计物理学中熟知的公式。其中第一项可理解为内能 $U$ ,  $1/\beta$ 为绝对温度 $T$ 乘以玻尔兹曼常数 $k_B$ ,  $S_{\max}$ 为熵。方程右边代表自由能 $F$ , 这样采用热力学符号, 式(4.5)写为

$$U - TS = F \quad (4.6)$$

与式(4.5)比较, 知

$$F = -k_B T \ln Z \quad (4.7)$$

及 $S = S_{\max}$ , 可将下标“max”略去。式(3.40)变为

$$Z = \sum_i \exp(-\beta E_i) \quad (4.8)$$

这正好是通常的配分函数, 应用以上这些公式, 很容易检验其它一些热力学等式。

唯一费思考的问题在于鉴别独立变量与非独立变量。先考虑信息熵 $S_{\max}$ 。在式(3.47)中它是 $\lambda$ ,  $\lambda_k$ 和 $f_k$ 的函数, 但是,  $\lambda$ 和 $\lambda_k$ 本身是由包含 $f_k$ 和 $f_i^{(k)}$ 的方程决定的, 方程中,  $f_k$ 和 $f_i^{(k)}$ 作为给定的量[参见式(3.34)、式(3.42)、式(3.43)], 因此独立变量是 $f_k$ ,  $f_i^{(k)}$ , 非独立变量是 $\lambda$ ,  $\lambda_k$ , 并且由式(3.47)知,  $S_{\max}$ 是独立变量 $f_k$ ,  $f_i^{(k)}$ 的函数。事实上,  $f_i^{(k)}$ 是 $i$ 的确定函数(例如状态“ $i$ ”的能量), 但仍然依赖于 $\alpha$ [例如体积, 参见式(4.1)], 这样, 在我们的讨论中, 真正的独立变量是 $f_k$ 和 $\alpha$ 。于是, 最后得到 $S = S(f_k, \alpha)$ 。本例中,  $f_1 = E = U$ ,  $\alpha = V$ , 因此

$$S = S(U, V) \quad (4.9)$$

将普遍关系式 (3.49) 应用于我们的特定模型, 假设只改变内能  $U$ , 而  $V$  保持不变, 则

$$\delta \langle f_i^{(1)} \rangle \equiv \delta f_i \equiv \delta U \neq 0 \quad (4.10)$$

和

$$\delta f_{i,2}^{(1)} \equiv \delta E_i(V) = (\delta E_i(V)/\delta V) \delta V = 0 \quad (4.11)$$

因此

$$\delta S = k_B \lambda_1 \delta U$$

或

$$\frac{\delta S}{\delta U} = k_B \lambda_1 (\equiv k_B \beta) \quad (4.12)$$

根据热力学, 式 (4.12) 左边定义为绝对温度的倒数

$$\frac{\delta S}{\delta U} = 1/T \quad (4.13)$$

如前所料, 得到  $\beta = 1 / (k_B T)$ , 另一方面, 改变  $V$  而保持  $U$  不变, 即

$$\delta \langle f_i^{(1)} \rangle = 0 \quad (4.14)$$

而

$$\langle \delta f_i^{(1)} \rangle = \langle \delta E_i(V)/\delta V \rangle \cdot \delta V \neq 0 \quad (4.15)$$

使式 (3.49) 可写为

$$\delta S = k_B (-\lambda_1) \langle \delta E_i(V)/\delta V \rangle \cdot \delta V$$

或

$$\frac{\delta S}{\delta V} = -\frac{1}{T} \langle \delta E_i(V)/\delta V \rangle \quad (4.16)$$

由热力学知

$$\frac{\delta S}{\delta V} = \frac{p}{T} \quad (4.17)$$

其中 $p$ 是压强, 与式(4.16)比较得

$$\langle \delta E_i(V)/\delta V \rangle = -p \quad (4.18)$$

将式(4.13)和式(4.17)代入式(3.49), 得到

$$\delta S = \frac{1}{T} \delta U + \frac{1}{T} p \delta V \quad (4.19)$$

在热力学中, 等式右边也等于 $dQ/T$ , 其中 $dQ$ 是热量, 以上, 我们按照式(3.51)定义“广义热”正是出于这个原因。以上讨论可以推广到具有多种粒子的系统, 其粒子数 $N_k$  ( $k = 1, \dots, m$ )是给定的。令 $f_i$ 为能量 $E$ ,  $f_{k'+1}$ 为平均粒子数 $N_{k'}$ ,  $k' = 1, \dots, m$  (注意下标的移动)。因为不同种类的粒子 $j$ 可由不同的粒子数 $N_i$ 表示, 所以可将下标 $i$ 推广到 $i$ ,  $N_1, \dots, N_m$ , 即令

$$f_i^{(k+1)} \rightarrow f_{i, N_1, \dots, N_m}^{(k+1)} = N_k$$

为了与热力学对应, 我们令

$$\lambda_{k+1} = -\frac{1}{k_B T} \mu_k \quad (4.20)$$

$\mu_k$ 称作化学势。

由等式(3.47)和(3.42)得出(等式两边同时乘 $k_B T$ )下式

$$TS = \underbrace{k_B T \ln Z + U}_F - \mu_1 \bar{N}_1 - \mu_2 \bar{N}_2 - \dots - \mu_m \bar{N}_m \quad (4.21)$$

由式(3.49), 可以得出等式

$$\frac{\delta S}{\delta \bar{N}_k} = -k_B \lambda_{k+1} = \frac{1}{T} \mu_k \quad (4.22)$$

配分函数写为

$$Z = \sum_{N_1 N_2 \dots N_m} \sum_i \exp \left\{ -\frac{1}{k_B T} [E_i(V) - \mu_1 N_1 - \dots - \mu_m N_m] \right\} \quad (4.23)$$

虽然以上讨论对于不可逆热力学是非常有用的，但是在热力学中，独立变量与非独立变量所起作用发生某些调换，不同的调换将给出不同的热力学势（Gibbs, Helmholtz 势等）。但作这类调换并非我们的任务。在此，只提一下其中一种重要情形：我们不取  $U, V$ （和  $N_1, \dots, N_m$ ）作为独立变量，而是引入  $V$  和  $T = (\partial S / \partial U)^{-1}$ （及  $N_1, \dots, N_m$ ）作为新的独立变量。作为一个例子，考虑  $U-V$  情形（形式上令  $\mu_1, \mu_2, \dots = 0$ ）。式（4.7）中的自由能  $F$  是直接作为  $T$  的函数给出的，对  $T$  求偏导，得出

$$-\frac{\partial F}{\partial T} = k_B \ln Z + \frac{1}{T} \frac{1}{Z} \sum_i E_i \exp(-\beta E_i)$$

右边第二项正是  $U$ ，于是

$$-\frac{\partial F}{\partial T} = k_B \ln Z + \frac{1}{T} U \quad (4.24)$$

与式（4.5）比较，其中  $1/\beta = k_B T$ ，得到重要关系

$$-\frac{\partial F}{\partial T} = S \quad (4.25)$$

其中省略了下标“max”。

对以上理论在不可逆热力学中，即弛豫现象中的应用感兴趣的读者可参阅作者的另一本书《协同学导论》。该书涉及另一个非常不同的领域，即远离平衡的物理体系和生物体系的非平衡相变，以及非物理系统（例如经济学）的非平衡相变。





# 第五章 最大信息原理在自组织 系统中的应用

---

## 5.1 引言

根据第一章，自组织系统是那些不需外界特定的干预，能够通过内部过程产生宏观的空间、时间或空时结构的系统。迄今为止，控制宏观结构的序参量的分布函数只是通过微观理论来计算（参见第二章）。在本节，我们将从宏观量导出它们，我们将通过阈值附近的单模和多模激光明显地演示这一程序。

最大信息熵原理能使人们对系统的微观态的概率分布做出公正的估计，而对这些系统，如果没有这一原理，我们只知道一些与宏观观测量相应的平均值。如我们在前一节所看到的，这一原理为人们提供了与热力学的概念和基本关系的非常漂亮的联系，即它能够非常好的应用于热平衡系统。另一方面，关于这一原理对远离热平衡系统的一般应用方面，还没有什么成功的尝试。

在本节，我要向读者显示，上述这一原理是怎样成功地应用于那些通过自组织产生宏观结构的非平衡系统的。用这种方法，我们将重新得到这些系统的分布函数，大家熟知的这些分布函数，以前曾从微观理论导出（见2.4和2.7节）。同时，我们还可以将这些函数推广。为了说明我们的程序，让我们先看激光。

## 5.2 自组织系统中的应用：单模激光

自组织系统的特征是能够用合适的序参量描述宏观结构的出现。我们将通过具体例子来说明我们的程序，而不去进行抽象的考虑。重要的一点是严格依附最大信息原理的基本概念，即考虑宏观上观测到的量。

实验上观测到的单模激光的量是定态的强度和强度的二阶矩。众所周知，只有当我们除了测量强度的一阶矩外，还测量强度的二阶矩时，普通灯泡的光和激光的基本差别才可显示出来。进一步测量的量是强度关联，但由于我们考虑的是与时间无关的理论，我们这里将忽略这些信息。

单模激光的含时间和空间的电场强度的形式是

$$E(x, t) = E(t) \sin kx \quad (5.1)$$

其中幅度  $E(t)$  可按式分解为正频和负频部分

$$E(t) = \underbrace{B e^{-i\omega t}}_b + \underbrace{B^* e^{i\omega t}}_{b^*} \quad (5.2)$$

如果我们对光场强度的测量是在较振荡周期大，但较  $B(t)$  的涨落时间小的时间间隔进行的话，则输出光强正比于  $B^* B$  和激光的损耗速率  $2\kappa$ 。为简单起见，我们丢掉所有其他常数，令

$$I = 2\kappa B^* B \quad (5.3)$$

类似地，如果对强度的平方在同一时间区间平均，结果是

$$I^2 = 4\kappa^2 B^{*2} B^2 \quad (5.4)$$

由于激光的涨落， $B^*$ 和 $B$ 是一个定态过程的随机变量，这导致我们在式(3.32)中将 $P_i$ 的 $i$ 认做 $B^*$ 和 $B$ ，那里我们令 $k_B = 1$ ，并将右边解释为信息 $i$ 。由于随机变量 $B$ 不再是分立的而是连续的，我们必须将对 $i$ 的求和用下面的积分代替：

$$i = - \int P(B, B^*) \ln P(B, B^*) d^2 B \quad (5.5)$$

方程(3.34)可以被解释为对 $d^2 B$ 的积分，将几率 $P(B, B^*)$ 看做权重函数。用括号表示这个平均，我们考虑下面两个约束，

$$f_1 = \langle 2\kappa B^* B \rangle \quad (5.6)$$

$$f_2 = \langle 4\kappa^2 B^{*2} B^2 \rangle \quad (5.7)$$

进一步，通过同样的类比，导致下面 $f_i^{(k)}$ 的定义

$$f_{B, B^*}^{(1)} = 2\kappa B^* B \quad (5.8)$$

$$f_{B, B^*}^{(2)} = 4\kappa^2 B^{*2} B^2 \quad (5.9)$$

现在我们可以立即应用公式(3.38)，我们得到

$$P(B, B^*) = \exp(-\lambda - \lambda_1 2\kappa B^* B - \lambda_2 4\kappa^2 (B^* B)^2) \quad (5.10)$$

用稍微不同的记号，上式可写为

$$P(B, B^*) = N \exp(-\alpha |B|^2 - \beta |B|^4) \quad (5.11)$$

这一函数在激光物理中是大家熟悉的，它是 H. 里斯肯 (H. Risken) 通过求解与本作者得到的激光的朗之万方程相应的福克-普朗克方程得到的。

我们注意到在激光的情形下,  $\alpha$  必须为负, 但在阈值附近,  $\alpha$  可取从负到正的值。

### 5.3 没有位相关系的多模激光

在这种情形下, 场强按模式分解为

$$E(x, t) = \sum_{\mathbf{l}} E_{\mathbf{l}}(t) \sin k_{\mathbf{l}} x \quad (5.12)$$

为简单起见, 上式只考虑轴向模, 模的幅度仍可分为正频和负频部分, 即

$$E_{\mathbf{l}}(t) = B_{\mathbf{l}}(t) e^{-i\omega_{\mathbf{l}} t} + B_{\mathbf{l}}^*(t) e^{i\omega_{\mathbf{l}} t} \quad (5.13)$$

在较振动周期长但较涨落周期短的时间区间上, 强度的平均值是

$$I_{\mathbf{l}} = 2\kappa_{\mathbf{l}} B_{\mathbf{l}}^* B_{\mathbf{l}} \sim n_{\mathbf{l}} \quad (5.14)$$

如果我们将  $n_{\mathbf{l}}$  或者  $B_{\mathbf{l}}^* B_{\mathbf{l}}$  认做是随机变量, 5.2节的结果可直接推广, 我们得到

$$f_{\mathbf{l}} = \langle n_{\mathbf{l}} \rangle \quad (5.15)$$

$$f_{\mathbf{l}, \mathbf{l}'} = \langle n_{\mathbf{l}}, n_{\mathbf{l}'} \rangle$$

$k$  和  $i$  分别是

$$k \longleftrightarrow \mathbf{l}, \mathbf{l}' \quad (5.16)$$

$$i \longleftrightarrow (n_1, n_2, \dots, n_M) = n \quad (5.17)$$

我们有

$$\underset{\sim}{f}_{\mathbf{n}}^{(\mathbf{l})} = n_{\mathbf{l}} \quad (5.18)$$

$$\underset{\sim}{f}_{\mathbf{n}}^{(\mathbf{l}, \mathbf{l}')} = n_{\mathbf{l}}, n_{\mathbf{l}'} \quad (5.19)$$

$$P_i \rightarrow P(\underline{n}) \quad (5.20)$$

式 (3.38) 的直接应用得到最后结果

$$\begin{aligned} P(\underline{n}) &\equiv P(n_1, \dots, n_M) \\ &= \exp(-\lambda - \sum_i \lambda_i n_i - \sum_{i,j} \lambda_{i,j} n_i n_j) \end{aligned} \quad (5.21)$$

式 (5.21) 可以等价地写为

$$N \exp(-\sum_i \alpha_i n_i - \sum_{i,j} \beta_{i,j} n_i n_j) \quad (5.22)$$

这一形式也可通过特殊形式的多模福克-普朗克方程得到, 即通过细致平衡原理可明显地构造出此解。

#### 5.4 周期性序参量的过程

我们要证明, 如果所考虑的过程的序参量是周期性的, 如何来猜测分布函数。

让我们考虑一个角坐标  $\varphi$  的例子和矩的周期是  $2\pi$  的情形。这问题本身提示我们考虑周期函数的矩而不是  $\varphi$  的矩。即

$$\langle \sin \varphi \rangle, \langle \cos \varphi \rangle, \dots \quad (5.23)$$

或者更一般地表为

$$\langle \sin n\varphi \rangle, \langle \cos n\varphi \rangle \quad (5.24)$$

其中  $n$  是正整数。

为了说明我们的程序, 让我们考虑特殊情形, 由于对称的原因, 对于所有  $n$ ,

$$\langle \sin n\varphi \rangle = 0 \quad (5.25)$$

成立。在以上处理中, 我们只保留最初几个矩。如果只保留式 (5.24) 中最低的两项, 我们得到

$$P(\varphi) = \exp\{\lambda + \lambda_1 \cos\varphi + \lambda_2 \cos 2\varphi\} = \exp\{V(\varphi)\} \quad (5.26)$$

很明显, 最大信息(熵)原理允许我们以非常直接的方式导出象激光那样的非平衡系统的分布函数的一般形式。在某些限制条件下, 结果与我们从微观理论得到的分布函数一致。现在, 如何对其他情形讨论已相当清楚。系统总的态 $q(\underline{x}, t)$ 必须投影到描述所观测到的宏观空间或时间模式的函数上去。这样, 我们可以得到一些振幅, 其矩可被测量, 因而可用来定义 $f_k$ 以及函数 $f_k^{(t)}$ 。

尽管我们应用这一原理已取得了成功, 但是, 必须切记, 关于拉格朗日乘子我们能说的还很少, 当然, 它们现在可由实验来测量。另一方面, 微观理论的优点一直是, 它可以根据第一原理来确定这些常数, 因此, 它能预言 $\alpha_i$ 在临界点时改变符号。然而, 从宏观理论导出这些性质也是很诱人的。

我们的方法可应用于诸如流体的对流, 化学反应的图样形成以及形态发生学的生长等一系列问题。

尽管我们的结果与一些热力学的结果形式上相似, 但仍然存在一些基本的差别。首先, 我们意识到, 这里的常数与热平衡系统中的极不相同。例如, 象激光这样的非平衡系统, 我们必须处理输出光强, 而在平衡系统中, 我们处理的是能量。这一点也可从微观层次的处理得到反映。在微观处理中, 能量起着决定性的作用, 而在非平衡系统中, 速率常数和生长速率决定着演化着的图样。

我们可以得出一意义很深远的结论。迄今, 最大信息(熵)原理已被应用于热力学和不可逆热力学, 但没有应用于非平衡相变。在前两个领域中, 能量是所用的约束之一。现在的情形是, 我们处理输出强度 $I$ 。但更重要的是, 我们现在必须采用二阶矩的形式,  $\langle I^2 \rangle$ 包含了强度关联。而平衡热力学中的最大熵原理

绝没有这样做。但现在，我们看到，对非平衡相变来说， $\langle I^2 \rangle$  的引入是相当明显的，也是很必要的。在阈值处和阈值以上， $\lambda_1 \geq 0$ ，对  $\exp(\lambda_1 |b|^2)$  的积分将发散，这反映了临界涨落效应。由于这一点，为了考虑由于饱和引起的限制， $\langle I^2 \rangle$  必须得考虑。换句话说，接近非平衡相变点，涨落变得“可观测”，因此，必须通过合适的约束予以考虑。在我看来，对平衡系统的相变也是同样的，因此，那里也需要将此原理推广。





# 第六章 非平衡相变的最大信息原理：序参量、伺服模式和模式的确定

---

## 6.1 引言

前一章表述了在非平衡相变点附近，远离平衡系统的分布函数的约束。只要可能与以前微观理论的结果进行比较，我们发现二者总是完美的一致。另一方面，我们的表述仅限于序参量，即只能推导出序参量的合适的分布函数。假定序参量可由实验来确认，本章将表明，我们的处理可以被推广，而不要求预先知道序参量的知识。相反，求助于合适的关联函数，我们将能够确定序参量、伺服模以及出现的模式。

## 6.2 一般方法

我们假定要研究的系统由态矢量

$$\underline{q} = (q_1, \dots, q_N) \quad (6.1)$$

来描述。其分量是可观测的。 $q_i$  的指标  $i$  可代表细胞，或不同类型的物理量和其他量。我们进一步假定  $q_i$  的统计平均以及它们的四阶以下的各阶矩都是已知的。我们引入下面  $f$  量作为约束。

$$f_i = \langle q_i \rangle ; \quad f_i^{(1)} = q_i \quad (6.2)$$

$$f_{ij} = \langle q_i q_j \rangle ; \quad f_{ij}^{(2)} = q_i q_j \quad (6.3)$$

$$f_{ijk} = \langle q_i q_j q_k \rangle ; \quad f_{ijk}^{(3)} = q_i q_j q_k \quad (6.4)$$

$$f_{ijkl} = \langle q_i q_j q_k q_l \rangle ; \quad f_{ijkl}^{(4)} = q_i q_j q_k q_l \quad (6.5)$$

约束 (6.2) — (6.5) 的作用将通过拉格朗日乘子考虑，利用拉氏乘子使信息熵最大，我们得到信息

$$i = \exp \{ V(\underline{\lambda}, \underline{q}) \} \quad (6.6)$$

其中  $V$  由下式定义，

$$\begin{aligned} V(\underline{\lambda}, \underline{q}) = & \lambda + \sum_i \lambda_i q_i + \sum_{ij} \lambda_{ij} q_i q_j + \sum_{ijk} \lambda_{ijk} q_i q_j q_k \\ & + \sum_{ijkl} \lambda_{ijkl} q_i q_j q_k q_l \end{aligned} \quad (6.7)$$

为了与非平衡相变发生联系，我们求  $V$  的极值

$$\frac{\partial V}{\partial q_i} = 0, \quad i = 1, \dots, N \quad (6.8)$$

一般地，我们期望有几个极值，其位置记为  $\underline{q}^0$ 。由于我们要考虑非平衡相变，我们选择  $\underline{q}^0$  使得  $V(\underline{q}^0 + \underline{w})$  关于  $\underline{w}$  有最高对称性。与最大信息原理一致，这一选择意味着它是一个无偏向的。只有较低的对 称性有利于特定的模式，从而引入偏向。另一种将

$\tilde{q}^0$ 适当地定义为合适的极值位置的方法是从一个无结构的态开始,改变控制参数来跟踪 $\tilde{q}^0$ 。令

$$\tilde{q} = \tilde{q}^0 + \tilde{w} \quad (6.9)$$

我们可以将 $V$ 重新写为

$$V(\tilde{\lambda}, \tilde{q}) = \tilde{V}(\tilde{\lambda}, \tilde{w}) \quad (6.10)$$

其中

$$\begin{aligned} \tilde{V}(\tilde{\lambda}, \tilde{w}) = & \tilde{\lambda} + o + \sum_{ij} \tilde{\lambda}_{ij} w_i w_j \\ & + \sum_{ijk} \tilde{\lambda}_{ijk} w_i w_j w_k \\ & + \sum_{ijkl} \tilde{\lambda}_{ijkl} w_i w_j w_k w_l \end{aligned} \quad (6.11)$$

其中,  $\tilde{\lambda}_{ij}$ 由下式给出:

$$\tilde{\lambda}_{ij} = \frac{1}{2} \frac{\partial^2 V}{\partial q_i \partial q_j} \bigg|_{\tilde{q}^0} \quad (6.12)$$

同时, 旧的约束 (6.2) — (6.5) 又可写为

$$f_i = \left\langle \frac{\partial V}{\partial \lambda_i} \right\rangle \quad (6.13)$$

$$f_{ij} = \left\langle \frac{\partial}{\partial \lambda_{ij}} V \right\rangle \quad (6.14)$$

等等, 它们可以变换为新的约束

$$\tilde{f}_{ij} = \left\langle \frac{\partial \tilde{V}(\tilde{\lambda}, \tilde{w})}{\partial \tilde{\lambda}_{ij}} \right\rangle \quad (6.15)$$

等等。

### 6.3 序参量、伺服模和出现的模式的确定

因为约束 (6.14) 和 (6.15) 对指标  $i, j$  是对称的, 因此, 拉氏乘子也是如此,

$$\tilde{\lambda}_{ji} = \tilde{\lambda}_{ij} \quad (6.16)$$

因而矩阵

$$\Delta = (\tilde{\lambda}_{ij}) \quad (6.17)$$

可用实本征值  $\hat{\lambda}_\kappa$  来对角化。对角化可通过变换

$$w_i = \sum_\kappa a_{i\kappa} \xi_\kappa \quad (6.18)$$

来实现, 其中

$$\tilde{V}(\tilde{\lambda}, \tilde{w}) = \hat{V}(\hat{\lambda}, \xi) \quad (6.19)$$

且  $a_{i\kappa}$  是正交的。

借助式 (6.18), 表示式 (6.11) 可按照式 (6.19) 来变换, 其中右边明显地写为

$$\begin{aligned} \hat{V}(\hat{\lambda}, \xi) = & \hat{\lambda} + \sum_\kappa \hat{\lambda}_\kappa \xi_\kappa^2 + \sum_{\kappa, \lambda, \mu} \hat{\lambda}_{\kappa, \lambda, \mu} \xi_\kappa \xi_\lambda \xi_\mu \\ & + \sum_{\kappa, \lambda, \mu, \nu} \hat{\lambda}_{\kappa, \lambda, \mu, \nu} \xi_\kappa \xi_\lambda \xi_\mu \xi_\nu \end{aligned} \quad (6.20)$$

一般地,  $\hat{V}$  代表  $\xi = 0$  附近的鞍点。相应地, 我们区分出正的和负的  $\lambda$ ,

$$\begin{aligned} \hat{\lambda}_\kappa &\geq 0; \quad \kappa \rightarrow u, \quad \text{总数为 } N_u \\ \hat{\lambda}_\kappa &< 0; \quad \kappa \rightarrow s, \quad \text{总数为 } N_s \end{aligned} \quad (6.21)$$

通过与微观理论的结果比较, 我们可以采用非平衡相变的术语, 将那些  $\hat{\lambda} \geq 0$  的  $\kappa$  用  $u$  表示 ( $u$  表示不稳定), 并称  $\xi_u$  相应地为序参量。进一步, 将那些  $\hat{\lambda} < 0$  的  $\kappa$  用  $s$  表示, 并称  $\xi_s$  为伺服模的幅度。与此分解相应, 我们将  $\hat{V}$  写为

$$\hat{V}(\hat{\lambda}, \xi) = \tilde{\lambda} + \hat{V}_u(\hat{\lambda}_u, \xi_u) + \hat{V}_s(\hat{\lambda}_s, \hat{\lambda}_u; \xi_s, \xi_u) \quad (6.22)$$

其中第一部分仅与序参量有关, 而

$$\begin{aligned} \hat{V}_u = & \sum_{\kappa} \hat{\lambda}_{\kappa} \xi_{\kappa}^2 + \sum_{u,v,w} \hat{\lambda}_{u,v,w} \xi_u \xi_v \xi_w \\ & + \sum_{u,v,w,x,y} \hat{\lambda}_{u,v,w,x,y} \xi_u \xi_v \xi_w \xi_x \xi_y \end{aligned} \quad (6.23)$$

$\hat{V}_s$  更明显地写为

$$\begin{aligned} \hat{V}_s = & \sum_{\kappa} (-|\lambda_{\kappa}| \xi_{\kappa}^2) \\ & + \sum_{u,v,w} 3\hat{\lambda}_{u,v,w} \xi_u \xi_v \xi_w \\ & + \sum_{u,v,w,x} 4\hat{\lambda}_{u,v,w,x} \xi_u \xi_v \xi_w \xi_x \\ & + \text{乘积 } \xi_u \xi_v \xi_w, \xi_u \xi_v \xi_w \xi_x, \xi_u \xi_v \xi_w \xi_x \xi_y, \\ & \xi_u \xi_v \xi_w \xi_x \xi_y, \xi_u \xi_v \xi_w \xi_x \xi_y \xi_z \text{ 的求和} \end{aligned} \quad (6.24)$$

积分

$$\int \exp(\hat{V}_s) d^N \xi_s = g(\xi_u) > 0 \quad (6.25)$$

定义了一个仅含序参量  $\xi_u$  的函数。

通过

$$g(\underline{\xi}_u) = \exp(-h(\underline{\xi}_u)) \quad (6.26)$$

我们引入函数 $h$ 。通过

$$h(\underline{\xi}_u) + \hat{V}_u = W_u(\underline{\xi}_u | \underline{\xi}_u) \quad (6.27)$$

引入一个新函数 $W_u$ 。这个定义保证了

$$P(\underline{\xi}_u | \underline{\xi}_u) = \exp\{W_u(\underline{\xi}_u | \underline{\xi}_u)\} \quad (6.28)$$

对任何 $\underline{\xi}_u$ 在伺服模的空间都是归一化的。为了使式(6.22)在引入 $h$ 后仍保持不变, 我们通过

$$\hat{\lambda} + \hat{V}_u(\hat{\lambda}_u, \underline{\xi}_u) - h(\underline{\xi}_u) = W_u(\underline{\xi}_u) \quad (6.29)$$

引入新函数 $W_u$ 。最后, 我们可将式(6.22)重写为

$$\hat{V}(\hat{\lambda}, \underline{\xi}) = W_u(\underline{\xi}_u) + W_u(\underline{\xi}_u | \underline{\xi}_u) \quad (6.30)$$

这使得我们得到

$$\exp \hat{V} = P(\underline{\xi}_u) P(\underline{\xi}_u | \underline{\xi}_u) \quad (6.31)$$

其中

$$P(\underline{\xi}_u) = \exp W_u \quad (6.32)$$

而 $P(\underline{\xi}_u | \underline{\xi}_u)$ 由式(6.28)来定义。

显然,  $P(\underline{\xi}_u | \underline{\xi}_u)$ 是条件概率, 而 $P(\underline{\xi}_u)$ 只是序参量的分布函数。迄今, 我们的方法是相当一般的。它允许我们决定序参量的分布函数及伺服模的条件分布。特别是, 当把式(6.31)写得更明显一些时, 它代表着伺服原理的特殊情形(参见2.6节)。

## 6.4 近似

为了与微观理论的结果及伺服原理的最低阶近似相联系, 我

们引入下面近似。我们假定 $\xi_u$ 和 $\xi_v$ 是小量，以致于大小的量级是 $\xi_u \sim \xi_u^2$ 。这样，我们将忽略 $\hat{V}_u$ 中的 $\xi_u^0, \xi_u^0$ 以及更高次项。由于归一化条件，我们容易得到

$$\hat{V}_u \approx V_{u, \text{eff}} = - \sum_i |\lambda_i| (\xi_u - f_i(\xi_u)) \quad (6.33)$$

其中

$$\begin{aligned} f_i(\xi_u) = \frac{1}{2|\lambda_i|} & \left( \sum_{u,v} 3 \hat{\lambda}_{u,u} \xi_u \xi_u \right. \\ & \left. + \sum_{u,v} 4 \hat{\lambda}_{u,u,v} \xi_u \xi_u \xi_u \right) \end{aligned} \quad (6.34)$$

注意，出现在式(6.26)中的 $h$ 由下式给出

$$h(\xi_u) = \sum_i |\lambda_i| f_i^2(\xi_u) \quad (6.35)$$

我们引用另一种方法近似表示 $\hat{V}_u$ ，它更符合最大信息原理的精神，为此，我们用

$$\bar{P} = \exp \left\{ \bar{\lambda}(\xi_u) + \sum_i \bar{\lambda}_i(\xi_u) \xi_u + \sum_i \bar{\lambda}_{ii}(\xi_u) \xi_u^2 \right\} \quad (6.36)$$

近似表示 $P(\xi_u | \xi_u)$ ，以使约束

$$\int \bar{P} d^N \xi_u = 1 \quad (6.37)$$

$$\int \bar{P} \xi_u d^N \xi_u = \int P(\xi_u | \xi_u) \xi_u d^N \xi_u \quad (6.38)$$

$$\int \bar{P} \xi_u^2 d^N \xi_u = \int P(\xi_u | \xi_u) \xi_u^2 d^N \xi_u \quad (6.39)$$

得到满足，就是说，我们现在满足于使四个矩中的最初两个矩的约束得到满足。这样，我们就得到式(6.36)中的指数平方项，它出现在伺服原理的最低阶近似中。

## 6.5 空间模式

以上方法使我们能够决定序参量、伺服模以及它们的分布函数。当我们用指标 $i$ 代表格点，式(6.18)中的系数 $\alpha_{ik}$ 确定了与指标为 $k$ 的序参量相应的空间模模式。因此，当 $k$ 跑遍序参量的指标， $\alpha_{ik}$ 便确定了模的骨架（见2.8节）。式(6.18)的全部求和确定了对一组给定序参量及相应的伺服模所产生的模式，其中我们假定了 $\tilde{V}$ 的对称性被一组特定的与 $\tilde{V}$ 的一个极小值对应的序参量所破缺。从我们的表述中看出，我们的方法不仅在确定物理系统中序参量支配的演化着的模式时有用，而且是识别一般模式的主要结构的一个工具，我们希望这一方法能成功地应用于机器的图样辨认和人及动物的图样识别理论（又见第12章）。

## 6.6 与朗道相变理论的关系及福克-普朗克方程的猜测

我们可以从一个新的角度来看待热平衡相变的朗道理论，在那里，出发点是自由能作为指数的热平衡分布。这个自由能被展成序参量的幂级数。非常明显，求助于序参量的四阶以下的矩，我们得到了完全相同的表达式。因此，最大信息原理，以一种全新的解释指出了一条通向朗道相变理论的新的捷径。

一个有趣的问题是，我们是否可以决定一个福克-普朗克方程，它的定态解恰好是最大信息原理所决定的分布函数。为了说明这一问题，让我们考虑单变量 $\xi_i \equiv \xi$ 的情形，扩散“常数”与变量有关的福克-普朗克方程的一般形式是



$$\frac{dP}{dt} = - \frac{d}{d\xi} (K(\xi)P) + \frac{d}{d\xi} D(\xi) \frac{d}{d\xi} P \quad (6.40)$$

将其定态解与我们上面决定的分布函数(6.32)相比较, 容易得到

$$W_0(\xi_0) = - \int^{\xi_0} \frac{K(\xi)}{D(\xi)} d\xi \quad (6.41)$$

换句话说, 仍然未知的漂移系数 $K$ 和扩散系数 $D$ 必须服从关系

$$-K/D = W' \quad (6.42)$$

此式可用许多方式来满足。

最大信息原理对合理猜测恰当的 $K$ 和 $D$ 无济于事, 满足式(6.42)的所有测猜实际上是同样可能的。因此, 我们必须寻求其它判据, 这就是简单性。为此, 我们令

$$D = Q = \text{常数} \quad (6.43)$$

这时, 福克-普朗克方程约化成为

$$\frac{dP}{dt} = - \frac{d}{d\xi} (K(\xi)P) + Q \frac{d^2 P}{d\xi^2} \quad (6.44)$$

其中 $Q$ 仍然是一个固定而未知的参量, 与式(6.44)相应的朗之万方程是

$$\dot{\xi} = K(\xi) + F(t) \quad (6.45)$$

其中漂移系数由下式给出 (由于式(6.42)):

$$K \propto -\partial W_0 / \partial \xi \quad (6.46)$$

$F$ 是高斯白噪音。更仔细地考察式(6.45)和式(6.46)以及 $W_0$ 的形式(6.41) (其中 $D = Q = \text{常数}$ ), 我们发现, 我们正好得到了与朗道二级相变的朗道泛函相应的朗道方程。这使得我们可重新解释朗道理论。以上是对接近相变点的朗之万方程形式上的

一个特殊的猜测。非常明显，以上列举的单变量的方法不是很令人满意的，特别是当我们处理多变量的福克-普朗克方程时，因为这时选择合适的漂移和扩散系数的可能范围非常大。

显然，为了确定漂移和扩散系数，需要新的判据或约束。我们将在第九章证明，如果我们考虑含时关联函数，这一点确实可以做到。

## 第七章 自组织系统在接近不稳定点时的信息、信息增益和效率

---

### 7.1 引言

本节我们打算详细说明上一章的结果。那里已经证明,我们可以把联合概率分布函数 $P(\xi_u, \xi_s)$ 分解为一种乘积的形式:

$$P(\xi_u, \xi_s) = \prod_i P(\xi_s | \xi_u) f(\xi_u) \quad (7.1)$$

其中,  $\xi_u$ 表示序参量,  $\xi_s$ 表示伺服模的振幅。这种分解是与微观理论的伺服原理一致的。我们现在想证明,按照公式(7.1)可以把信息和信息增益分解为两部分:一部分是仅涉及序参量的,另一部分是在序参量分布上求过平均的伺服模信息的和。正如我们将要看到的那样,在接近不稳定性点时,序参量的信息显著地改变,而伺服模的信息却没有变化。因此,在接近这些不稳定性

点时只要研究序参量的信息和信息增益的行为就可以对一大批经历非平衡相变的系统，具体地计算出它们的信息和信息增益。我们将要证明，如何直接测量信息和信息增益以及（在本节所定义的意义上的）效率。

信息的定义是

$$i = - \sum_j P_j \ln P_j \quad (7.2)$$

而信息增益的定义是

$$K = \sum_j P_j \ln(P_j/P_j') \quad (7.3)$$

我们把  $P_j$  解释为脚标  $j$  所表征状态的概率分布。在式 (7.3) 中， $P_j$  和  $P_j'$  是两种不同的概率分布。为了紧扣上述结果，我们令脚标  $j$  与随机变量  $\xi_j$ ， $\xi_j$  的许可值一致。象通常一样，我们也用  $\xi_j$ ， $\xi_j$  表示那些许可值。注意，一旦知道了  $\xi_j$  和  $\xi_j$ ，用式 (2.78) 即可确定状态矢量  $q$ 。

## 7.2 伺服原理及其对信息的应用

我们想证明，借助于式 (7.1) 能够把信息 (7.2) 变成一种特定的形式。为此将式 (7.1) 代入式 (7.2)，并把乘积的对数写成对数之和，我们得到

$$i = - \sum_{\xi_j, \xi_j} \prod_j P_j(\xi_j | \xi_j) f(\xi_j) \left\{ \ln f(\xi_j) + \sum_j \ln P_j(\xi_j | \xi_j) \right\} \quad (7.4)$$

利用归一化条件

$$\sum_{\xi_s} P_s(\xi_s | \xi_{\sim s}) = 1 \quad (7.5)$$

将式 (7.4) 变为

$$\begin{aligned} i = & - \sum_{\xi_{\sim s}} f(\xi_{\sim s}) \cdot \ln f(\xi_{\sim s}) \\ & - \sum_{\xi_{\sim s}} f(\xi_{\sim s}) P_s(\xi_s | \xi_{\sim s}) \ln P_s(\xi_s | \xi_{\sim s}) \end{aligned} \quad (7.6)$$

上式可写为

$$i = I_r + \sum_{\xi_{\sim s}} f(\xi_{\sim s}) I_s(\xi_{\sim s}) \quad (7.7)$$

其中

$$I_r = - \sum_{\xi_{\sim s}} f(\xi_{\sim s}) \ln f(\xi_{\sim s}) \quad (7.8)$$

是序参量的信息，而

$$I_s = - \sum_{\xi_s} P_s(\xi_s | \xi_{\sim s}) \ln P_s(\xi_s | \xi_{\sim s}) \quad (7.9)$$

是脚标为  $s$  的伺服子系统或伺服模的信息。正如下面所证明的那样，当接近不稳定点时， $I_r$  显著变化，而  $I_s$  只有微弱的改变。信息  $I_s$  显然是在  $\xi_{\sim s}$  具有特定值的假定下的一种信息。按照我们的解释，序参量以一种特定方式使役着伺服模，同时又通过自组织确保了宏观结构。

### 7.3 信息增益

我们用脚标  $n$  和  $a$  来区分分布  $P$  和  $P'$ ，按照定义 (7.3)，信

息增益为

$$K = \sum_{\xi_u, \xi_s} P_n(\xi_u, \xi_s) \cdot \ln [P_n(\xi_u, \xi_s) / P_s(\xi_u, \xi_s)] \quad (7.10)$$

利用式 (7.1), 式 (7.10) 可写为:

$$K = \sum_{\xi_u, \xi_s} f_n \prod_i P_{i..} \{ \ln f_n + \sum_i \ln P_{i..} - \ln f_s - \sum_i \ln P_{i..} \} \quad (7.11)$$

在这里乘积的对数已写为对数之和。利用式 (7.5), 稍加重排, 便得

$$\begin{aligned} (7.11) = & \sum_{\xi_u} f_n \ln(f_n/f_s) \\ & + \sum_i \sum_{\xi_u} f_n \sum_{\xi_s} P_{i..} \ln(P_{i..}/P_{i..}) \end{aligned} \quad (7.12)$$

上式可化为最后形式

$$K = K_f + \sum_i \sum_{\xi_u} f_n K_i \quad (7.13)$$

其中,  $K_f$  定义为

$$K_f = \sum_{\xi_u} f_n \ln(f_n/f_s) \quad (7.14)$$

而  $K_i$  定义为

$$K_i = \sum_{\xi_s} P_{i..}(\xi_s | \xi_u) \ln(P_{i..}(\xi_s | \xi_u) / P_{i..}(\xi_s | \xi_u)) \quad (7.15)$$

#### 7.4 一个例子：非平衡相变

本节，我们打算阐明，前面通过非平衡相变导出的公式的用处（参看2.7节）。我们假定  $\lambda_0$  是实数。为了更具体起见，我们写出非平衡相变系统的典型演化方程。利用序参量  $\xi_u$  和伺服模振幅  $\xi_s$ ，可把这些方程写成

$$\dot{\xi}_u = \lambda_u \xi_u + h_u(\xi_u, \xi_s) + F_u(t) \quad (7.16)$$

$$\dot{\xi}_s = \lambda_s \xi_s + g_s(\xi_u) + \xi_s \cdot k_s(\xi_u) \cdots + F_s(t) \quad (7.17)$$

如果系统由外部控制参数控制，则式 (7.16) 和式 (7.17) 右边的所有量会以不同的形式依赖于控制参数。

把控制参数进行规范，使不稳定性发生在  $\alpha = 0$  处，则  $\lambda_u$  和  $\lambda_s$  以下面形式依赖于  $\alpha$ ，

$$\lambda_u \propto \alpha^\kappa, \quad \lambda_s = \lambda_s(0) + o(\alpha) \approx \lambda_s(0) \quad (7.18)$$

其中， $\kappa$  为某一大于零的指数。显然， $\lambda_u$  非常敏感的依赖于  $\alpha$ ，而  $\lambda_s$  只是弱依赖于  $\alpha$ ——因为第一项是非零常数。同样，函数  $h_u$ ， $g_s$ ， $k_s$ ， $F_u$  和  $F_s$  仅弱依赖于  $\alpha$ 。为了进一步明确起见，我们只应用伺服原理到头几项，其结果与令

$$\dot{\xi}_u \approx 0 \quad (7.19)$$

的绝热消去法相同。这样，我们就可求解式 (7.17) 至第一级，结果是

$$\xi_s \approx -\frac{1}{\lambda_s(0)} g_s(\xi_u) - \frac{1}{\lambda_s(0)} F_s(t) \quad (7.20)$$

在2.6节的方程式(2.91)中,我们导出了它的条件概率分布

$P_1(\xi_1 | \xi_0)$ 。它是

$$P_1(\xi_1 | \xi_0) = N \exp \left[ -(\xi_1 - g_1(\xi_0)/\lambda_1)^2 Q^{-1} \right] \quad (7.21)$$

现在可以把这个显结果代入式(7.9)的 $I_1$ 中,引入新的变量

$$\tilde{\xi}_1 = \xi_1 - \frac{g_1(\xi_0)}{\lambda_1} \quad (7.22)$$

我们便可消去伺服变量的概率分布对 $\xi_0$ 的依赖关系,于是 $I_1$ 便与 $\xi_0$ 无关。因而可在式(7.7)中对第二项中的 $\xi_0$ 积分。由此得到

$$I = I_1 + \sum_i I_i \quad (7.23)$$

至少在现在这种近似上,第二部分不依赖于 $\alpha$ 。因而在接近不稳定点时,信息的改变仅由序参量支配,

$$I(\alpha_1) - I(\alpha_2) \approx I_1(\alpha_1) - I_1(\alpha_2) \quad (7.24)$$

用同样的近似处理,我们可将信息增益写为

$$K = K_1 = \int d\tilde{\xi}_1 f(\tilde{\xi}_1, \alpha_2) \ln [f(\tilde{\xi}_1, \alpha_2)/f(\tilde{\xi}_1, \alpha_1)] \quad (7.25)$$

后面我们还要详细研究 $I_1$ 。在下面的7.11节,我们要研究 $I_1$ 对 $\alpha$ 的依赖关系。正如我们将要看到的那样, $I_1$ 随 $\alpha$ 单调的变化,而 $I_1$ 表现出某种“奇异”行为。

## 7.5 软模的不稳定性

正如我们在上一节看到的那样,在接近不稳定性点时,整个系统的信息的改变和信息增益分别与序参量的信息的改变和信息增益等同。在本节中,我们打算利用序参量分布函数的显式来计



算  $i$  和  $K$ , 其中以单序参量的二阶非平衡相变为例。如我们在2.7节中所见, 分布函数的形式是

$$f(\xi) = N \exp [\alpha \xi^2 - \beta \xi^4] \quad (7.26)$$

将式 (7.26) 代入信息的定义式 (7.8), 我们容易得到

$$I_i = -\ln N - \alpha \langle \xi^2 \rangle + \beta \langle \xi^4 \rangle \quad (7.27)$$

其中, 我们用了简写符号

$$\langle \xi^n \rangle = \int_{-\infty}^{+\infty} f(\xi) \xi^n d\xi \quad (7.28)$$

同样, 信息增益 (7.14) 具有显式

$$\begin{aligned} K(f_\alpha, f_{\alpha_0}) &= \int f_\alpha \ln(f_\alpha/f_{\alpha_0}) d\xi \\ &= \ln N(\alpha) - \ln N(\alpha_0) + (\alpha - \alpha_0) \langle \xi^2 \rangle_\alpha \end{aligned} \quad (7.29)$$

## 7.6 我们能够测量信息和信息增益吗

### A. 效率

对自组织系统行为 (比如激光的行为) 的研究表明, 序参量的平方是这种系统的宏观作用的量度。例如在激光中, 场模起着序参量的作用, 激光的作用可以用场模幅的平方来度量。这就使我们产生了用

$$\Omega(\alpha) = \langle \xi^2 \rangle \quad (7.30)$$

来度量宏观作用的想法, 更确切地说, 是用它来度量系统的功。平均定义为

$$\langle \xi^2 \rangle = \int \xi^2 f(\xi) d\xi \quad (7.31)$$

为明显起见, 我们使用分布函数

$$f(\xi) = N \exp [\alpha \xi^2 - \beta \xi^4] \quad (7.32)$$

其中，归一化因子定义为

$$N^{-1} = \int_{-\infty}^{+\infty} \exp [\alpha \xi^2 - \beta \xi^4] d\xi \quad (7.33)$$

因为在许多自组织系统中， $\alpha$  与输入功率有关。当改变控制参数  $\alpha$  时，我们把  $\Omega$  的变化率定义为效率。注意，我们暂时把  $\alpha$  直接看成输入功率。

$$W = \frac{d\Omega}{d\alpha} \quad (7.34)$$

利用关系式 (7.31) — (7.33)，我们容易得到

$$W = \frac{\int \xi^4 \exp[\alpha \xi^2 - \beta \xi^4] d\xi}{\int \exp[\alpha \xi^2 - \beta \xi^4] d\xi} - \frac{\left( \int \exp[\alpha \xi^2 - \beta \xi^4] \xi^2 d\xi \right)^2}{\left( \int \exp[\alpha \xi^2 - \beta \xi^4] d\xi \right)^2} \quad (7.35)$$

上式能够写成（我们用  $\xi_u$  代替  $\xi$ ）

$$W = \langle \xi_u^4 \rangle - \langle \xi_u^2 \rangle^2 \quad (7.36)$$

如果控制参数  $\alpha$  以隐含形式写进分布函数 (7.32) 中，如

$$\exp [\lambda(\alpha) \xi_u^2 - \beta \xi_u^4] \quad (7.37)$$

所示，就必须把式 (7.36) 用

$$W = \frac{d\lambda}{d\alpha} (\langle \xi_u^4 \rangle - \langle \xi_u^2 \rangle^2) \quad (7.38)$$

来代替。在远离不稳定点以下和以上来计算  $W$  的值并不困难，我们将在下面给出相应的结果。

## B. 信息和信息增益

我们现在来计算信息  $I_t$ 。利用显式 (7.32) 与 (7.33) 容易

得到

$$-I_f = \ln N + \alpha \langle \xi^2 \rangle - \beta \langle \xi^4 \rangle \quad (7.39)$$

虽然二阶矩和四阶矩可以直接测量, 但  $N$  的对数有待于与可观测  
量相联系。为此, 我们首先将  $\ln N$  对  $\alpha$  求导,

$$\frac{d \ln N}{d\alpha} = - \frac{\int \xi^2 \exp [\alpha \xi^2 - \beta \xi^4] d\xi}{\int \exp [\alpha \xi^2 - \beta \xi^4] d\xi} \quad (7.40)$$

显然上式可写成

$$\frac{d \ln N}{d\alpha} = - \langle \xi^2 \rangle \quad (7.41)$$

对式 (7.41) 积分得到

$$\ln N = - \int_{\alpha_0}^{\alpha} \langle \xi^2 \rangle d\alpha + \ln N(\alpha_0) \quad (7.42)$$

把这个结果代入式 (7.39), 得到

$$-I_f = - \int_{\alpha_0}^{\alpha} \langle \xi^2 \rangle d\alpha + \ln N(\alpha_0) + \alpha \langle \xi^2 \rangle - \beta \langle \xi^4 \rangle \quad (7.43)$$

为了去掉对数  $\ln N$ , 我们计算与  $\alpha_0$  相应的信息, 从而得到最后结果

$$\begin{aligned} -I_f(\alpha) + I_f(\alpha_0) &= - \int_{\alpha_0}^{\alpha} \langle \xi^2 \rangle d\alpha + \alpha \langle \xi^2 \rangle_{\alpha} - \beta \langle \xi^4 \rangle_{\alpha} \\ &\quad - \alpha_0 \langle \xi^2 \rangle_{\alpha_0} - \beta \langle \xi^4 \rangle_{\alpha_0} \end{aligned} \quad (7.44)$$

或写成微分形式

$$- \frac{dI_f}{d\alpha} = \alpha \langle \xi^2 \rangle^2 + \alpha \langle \xi^4 \rangle + \beta \langle \xi^2 \rangle \langle \xi^4 \rangle - \beta \langle \xi^6 \rangle \quad (7.45)$$

同样, 信息增益的形式是

$$K = - \int_{\alpha_0}^{\alpha} \langle \xi^2 \rangle_{\alpha'} d\alpha' + (\alpha - \alpha_0) \langle \xi^2 \rangle_{\alpha} \quad (7.46)$$

## 7.7 几个序参量的情形

现在打算把7.6节的结果推广到有几个序参量的情形。我们采用下面形式的分布函数  $f$ ,

$$f(\underline{\xi}_k) = N \exp \left[ \sum_k \Delta_k(\alpha) \xi_k^2 + h(\underline{\xi}_k) \right] \quad (7.47)$$

在许多情况下, 上式能够从微观模型具体导出 (参看2.4节, 或者按照第六章的最大信息原理)。

我们将给出信息增益的具体结果, 并能证明, 它们只要不依赖于控制参数, 则对非线性函数  $h$  的特定形式并不敏感。实际上, 它们会依赖于  $\alpha$ , 但通常只是微弱的相关。把(7.47)代入  $K_I$  的定义式, 我们容易得到

$$\begin{aligned} K_I &= \int f \ln(f/f_0) d^n \xi = \int N \exp \left[ \sum_k \Delta_k(\alpha) \xi_k^2 + h(\underline{\xi}_k) \right] \\ &\quad \times \left\{ \ln N - \ln N_0 + \sum_k (\Delta_k(\alpha) - \Delta_k(\alpha_0)) \xi_k^2 \right\} d^n \xi \end{aligned} \quad (7.48)$$

上式可以写成

$$K = \ln N - \ln N_0 + \sum_k (\Delta_k(\alpha) - \Delta_k(\alpha_0)) \langle \xi_k^2 \rangle_{\alpha} \quad (7.49)$$

其中  $\ln N$  仍有待计算。为此, 我们微商  $\ln N$  而得到

$$\frac{d \ln N}{d\alpha} = - \frac{\sum_k (\partial \Delta_k / \partial \alpha) \int \xi_k^2 d^n \xi \exp [\dots]}{\int d^n \xi \exp [\dots]} \quad (7.50)$$

它可写成

$$\frac{d \ln N}{d\alpha} = - \sum_k \langle \partial \Delta_k / \partial \alpha \rangle \langle \xi_k^2 \rangle \quad (7.51)$$

将式 (7.51) 对  $\alpha$  积分, 然后把它的结果代入式 (7.49), 我们得到最后结果

$$K = - \int_{\alpha_0}^{\alpha} \sum_k \langle \partial \Delta_k / \partial \alpha \rangle \langle \xi_k^2 \rangle_{\alpha} d\alpha \\ + \sum_k \langle \Delta_k(\alpha) - \Delta_k(\alpha_0) \rangle \langle \xi_k^2 \rangle_{\alpha} \quad (7.52)$$

最后, 我们打算计算  $\xi_1$  的效率。它的输出定义为

$$\Omega(\alpha) = \langle \xi_1^2 \rangle \quad (7.53)$$

简单计算表明, 效率为

$$W \equiv \frac{d\Omega}{d\alpha} = \sum_k \langle \partial \Delta_k / \partial \alpha \rangle (\langle \xi_1^2 \xi_k^2 \rangle - \langle \xi_1^2 \rangle \langle \xi_k^2 \rangle) \quad (7.54)$$

本节已经证明, 信息变化、信息增益和效率能够用序参量来表示。这一点是关于自组织系统行为的新发现, 相对于上述量说来, 自组织系统似乎仅由少数几个自由度所支配。特别是, 我们证明了信息、信息增益和效率是可观测量, 因此可以预期, 这些量可以象热平衡系统的熵一样进行测量。但是, 按照内部机制来说, 非平衡系统中产生的有序完全不同于热平衡系统所建立的有序。同样, 热力学熵和信息的物理意义是不同的。

## 7.8 单序参量信息的具体计算

现在具体计算接近非平衡相变时单序参量的信息, 也就是计算式 (7.27)。我们还将借助于这个结果来讨论, 为什么我们将

有关的量解释为信息，而不是熵。

回顾一下序参量连续的情况，在这种情况下，我们必须用定义

$$I_f = - \int d\xi f(\xi) \ln f(\xi) - \ln \varepsilon \quad (7.55)$$

其中， $\varepsilon$ 是一个间隔（无量纲单位），它是测量时的精度。

我们用式 (7.26) 来计算式 (7.27)

$$I_f = - \ln N - \alpha \langle \xi^2 \rangle + \beta \langle \xi^4 \rangle - \ln \varepsilon \quad (7.56)$$

在远离阈值以下或阈值以上的区域，可以通过简单的形式具体计算式(7.56)

#### A. 远离阈值以下的区域

在这个区域中， $\alpha$  是负的而且它的绝对值相当大，因此在式 (7.26) 中只有 $\xi$ 的平方项是重要的。这样一来，我们必须用高斯分布

$$f = N \exp(-|\alpha| \xi^2) \quad (7.57)$$

计算式 (7.56)。归一化因子易于确定，它是

$$N = \sqrt{|\alpha|/\pi} \quad (7.58)$$

式 (7.56) 中的积分也可以方便地计算，特别是，我们得到

$$|\alpha| \langle \xi^2 \rangle = \frac{1}{2} \quad (7.59)$$

和

$$\beta \langle \xi^4 \rangle = \frac{3}{4} \frac{\beta}{|\alpha|^2} \quad (7.60)$$

当  $|\alpha|$  足够大时，与式 (7.59) 相比，可忽略式 (7.60)。因此，留下的最后结果是

$$I_f = -\frac{1}{2} \ln |\alpha| + \frac{1}{2} + \frac{1}{2} \ln \pi - \ln \varepsilon \quad (7.61)$$

现在讨论  $\ln |\alpha|$  和  $\ln \varepsilon$  的行为。由于  $\varepsilon$  是精度的量度，我们可以假定  $\varepsilon$  是小量，于是

$$\varepsilon \longrightarrow 0$$

$$\text{而} \quad -\ln \varepsilon \longrightarrow +\infty \quad (7.62)$$

另一方面，因为系统远处于阈值以下，我们可以假定

$$\begin{aligned} |\alpha| &\rightarrow \infty \\ -\frac{1}{2} \ln |\alpha| &\rightarrow -\infty \end{aligned} \quad (7.63)$$

当我们保持  $\varepsilon$  不变而让  $|\alpha|$  增大， $I_t$  将变成 0，并最终会变为负的，这在数学上是不允许的，因为信息是正量。然而更重要的是，由于物理上的原因， $|\alpha|$  的增大不能超过某一值，即分布函数 (7.26) 仅在测量精度范围  $\varepsilon$  内，才实质上不等于 0。这意味着，在这个间隔内分辨系统的不同状态就失去意义了，换句话说，几乎可以肯定在该间隔内系统恰处于一个特定状态。此时，我们对怎样来固定  $\alpha$ ，也就是对怎样定义“几乎可肯定”一语的含义，仍然含糊不清。例如，我们可以要求

$$I_t = 0 \quad (7.64)$$

由此可推出  $\alpha$  的条件

$$-\frac{1}{2} \ln |\alpha| + \frac{1}{2} + \frac{1}{2} \ln \pi - \ln \varepsilon = 0 \quad (7.65)$$

由于我们只考虑信息的变化，因而可以使信息保持有一个常数的不确定性，以便同样要求

$$I_t = \frac{1}{2} + \frac{1}{2} \ln \pi \quad (7.66)$$

选择此式的方法是，使  $\alpha$  和  $\varepsilon$  之间的关系式变得特别简单。即

$$-\frac{1}{2} \ln |\alpha| - \ln \varepsilon = 0 \quad (7.67)$$

由此容易导出

$$|\alpha|^{\frac{1}{2}} = \frac{1}{\varepsilon} \quad (7.68)$$

## B. 远离阈值以上的区域

稍微研究一下分布函数 (7.26) 的行为就会看出, 只要  $\alpha$  足够大, 它就可用高斯函数来近似表示。为此, 应用最陡下降法而将式 (7.26) 写成形式

$$f(\xi) = N \exp(g(\xi)) \quad (7.69)$$

当然, 其中

$$g(\xi) = \alpha \xi^2 - \beta \xi^4 \quad (7.70)$$

式 (7.70) 的极值由

$$g'(\xi) = 0 = 2\alpha\xi - 4\beta\xi^3 \quad (7.71)$$

给出, 而式 (7.71) 具有解

$$\xi_0 = \pm \sqrt{\frac{\alpha}{2\beta}} \quad (7.72)$$

在这里我们暂取加号。我们通过下式引入一个新的变量  $\xi$ ,

$$\xi = \xi_0 + \eta \quad (7.73)$$

利用展开式

$$g(\xi) = g(\xi_0 + \eta) = g(\xi_0) + g'(\xi_0)\eta + \frac{1}{2}g''(\xi_0)\eta^2 \quad (7.74)$$

我们可把式 (7.69) 化为

$$f = N \exp \left[ \frac{1}{4} \frac{\alpha^2}{\beta} - 2\alpha\eta^2 \right] \quad (7.75)$$

现在, 要在区间  $-\infty < \xi < +\infty$  上进行归一化。归一化因子是

$$N = \frac{2\alpha}{\pi} \exp \left[ -\frac{1}{4} \frac{\alpha^2}{\beta} \right] \quad (7.76)$$

迄今, 我们仅处理了  $f$  具有一个极大值的情况。我们首先来计算只能出现一个极大值情况下的信息, 即我们人为地破坏对称性。

在这种情况下, 可以方便地计算  $I_t$ , 其结果为

$$I_t = \frac{1}{2} \ln \alpha - \frac{1}{2} \ln 2 + \frac{1}{2} \ln \pi + \frac{1}{2} - \ln e \quad (7.77)$$

我们现在来讨论最感兴趣的情况, 即整个分布函数呈现出两个极大值, 粗略地说, 系统可取两种状态 (参看图7.1)。显然, 我们



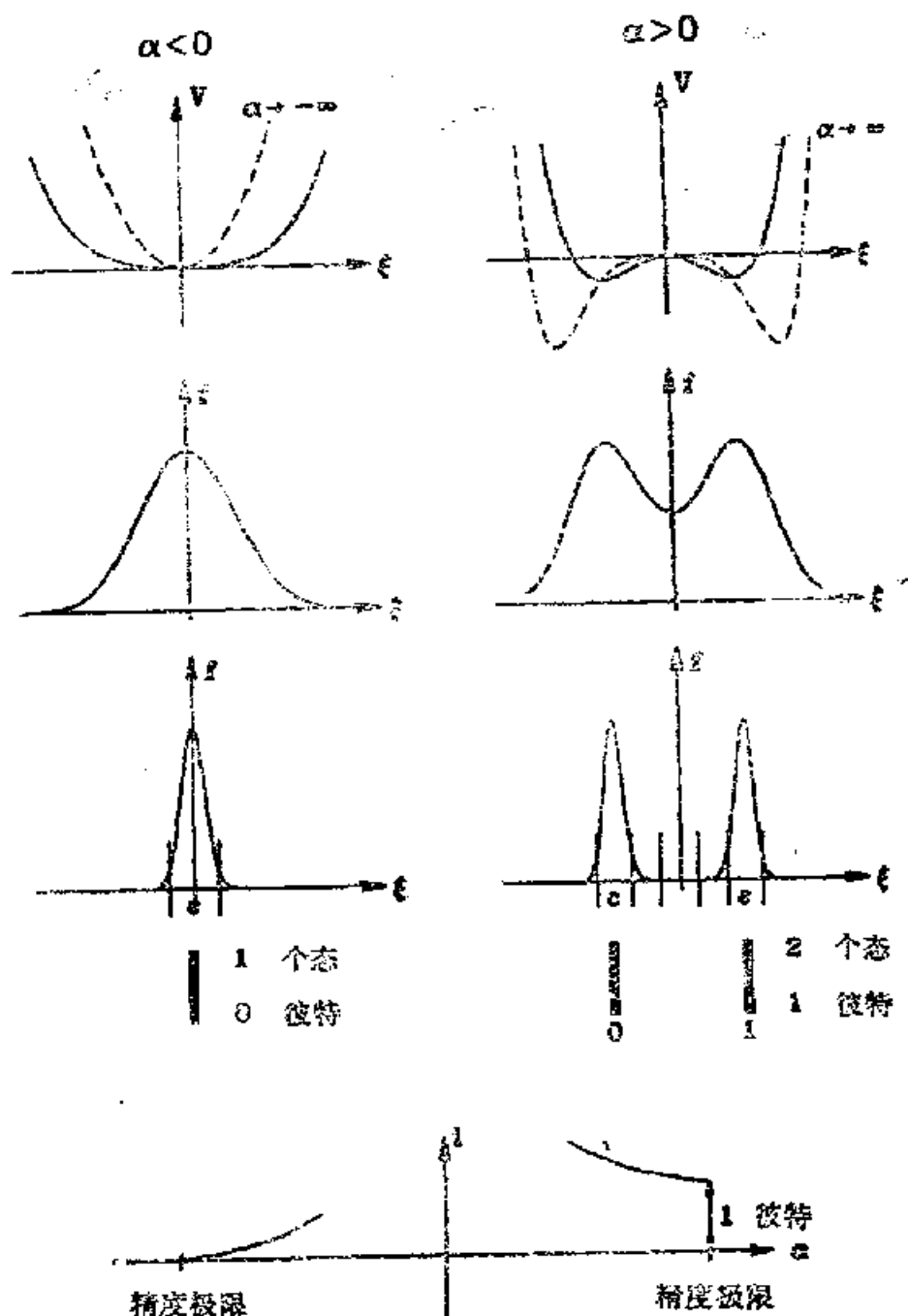


图7.1 第一行：如所周知，分布函数 (7.69) 和 (7.70) 一起可以解释为一个在势阱中运动同时受到一涨落力的粒子的坐标。左图是  $\alpha < 0$  的情形，实线是  $\alpha$  接近于零的情形，而虚线是  $|\alpha|$  较大的情形。显然，只有一个粒子可以趋于其中的极小值。右边是  $\alpha > 0$  的情形，同样，实线的  $\alpha$  较小，而虚线对应于较大的  $\alpha$ ，这时，势阱变得更深，且呈现两个极小值。第二行：左边，与只有一个极小值的势相应的分布函数只有一个极大值；右边，对于  $\alpha > 0$ ，出现了两个不同的极大值。第三行：左边，随  $|\alpha|$  增加，分布函数变窄，只有一个峰；右边，双峰分布函数变窄，每一个都落入精度范围  $\varepsilon$  内。注意  $f$  的标度与上一行不同。第四行：信息相对于  $\alpha$  的变化，图中表示了接近精度极限时信息的行为。

可在此系统中存储信息， $\eta$  为我们可以用 0 表示一种状态而用 1 表示另一种状态。

归一化因子  $N$  由下式决定，

$$\int_{-\infty}^{+\infty} N \exp(\alpha \xi^2 - \beta \xi^4) d\xi = 1 \quad (7.78)$$

由于极大值处对原点  $\xi = 0$  对称，我们可把式 (7.78) 的左边写为

$$2N \int_0^{\infty} \exp(\alpha \xi^2 - \beta \xi^4) d\xi \quad (7.79)$$

在只对一个极大值积分的这种规定下，我们将从 0 到  $+\infty$  的积分用从  $-\infty$  到  $+\infty$  的积分来代替。

$$\int_0^{\infty} \approx \int_{-\infty}^{+\infty} \exp(\alpha \xi^2 - \beta \xi^4) d\xi \quad (7.80)$$

按照上式计算，这个积分可近似表为

$$\exp \left[ -\frac{1}{4} \frac{\alpha^2}{\beta} \right] \int_{-\infty}^{+\infty} \exp\{-2\alpha \eta^2\} d\eta \quad (7.81)$$

经计算，式 (7.81) 中的积分为

$$\sqrt{\frac{\pi}{2\alpha}} \quad (7.82)$$

使用关系式 (7.79) — (7.82)，我们可由式 (7.78) 计算归一化因子  $N$ ，得到

$$N = \exp \left[ -\frac{1}{4} \frac{\alpha^2}{\beta} \right] \frac{1}{2} \sqrt{\frac{2\alpha}{\pi}} \quad (7.83)$$

为了计算  $I_1$ ，我们以同样的方式重复这些步骤，得到的最后结果是

$$I_1 = -\frac{1}{2} \ln 2\alpha + \frac{1}{2} \ln \pi + \ln 2 + \frac{1}{2} - \ln \epsilon \quad (7.84)$$

现在比较一下，远离阈值之上的式 (7.84) 与远离阈值之下的式 (7.61)。

我们首先讨论精度极限  $\varepsilon$  的作用。在小节A中，我们建立了  $\varepsilon$  和  $\alpha$  的关系式， $\alpha$  显然描述高斯函数下降的快慢程度。比如从式 (7.81) 中显然可以看出，远离阈值之上的衰减常数是  $2\alpha$ ，而远离阈值之下的是  $|\alpha|$ 。根据与以前相同的物理原因，我们必须要求

$$-\frac{1}{2} \ln 2\alpha - \ln \varepsilon = 0 \quad (7.85)$$

将式 (7.84) 和式 (7.61) 相减，并利用式 (7.85) 和式 (7.67)，我们得到

$$\Delta I_t = \ln 2 \quad (7.86)$$

如果回想一下信息的正规定义，我们应该用以 2 为底的对数，而不是自然对数，这就得到

$$\Delta I_t = \log_2 2 = 1 \quad (7.87)$$

它意味着，远离阈值之上的系统能够可靠地存储 1 彼特信息，而远离阈值之下的系统不能存储信息。下面的小节D将讨论对它的解释。

### C. 数值结果

在图7.2中，我们取  $\varepsilon = 1$  和  $\beta = 1$  计算了  $I_t$ 。对于小节A和B的结果，值得注意的是， $\alpha$  从负值变为正值的过程中信息经过了一个峰值。其物理原因是清楚的。当接近  $\alpha = 0$  时，出现了平衡相变中熟知的临界涨落，它引起  $f$  的弥散因而使  $I_t$  增加。当  $\alpha$  到达负值时， $f$  围绕着一个最大值 (参看图7.1)，于是  $I_t$  下降。对于增加的正  $\alpha$  值， $f$  有两个不同的极大值且信息再次下降——然而它比相应负  $\alpha$  的信息要大。为完整起见，我们也计算了信息增益，它

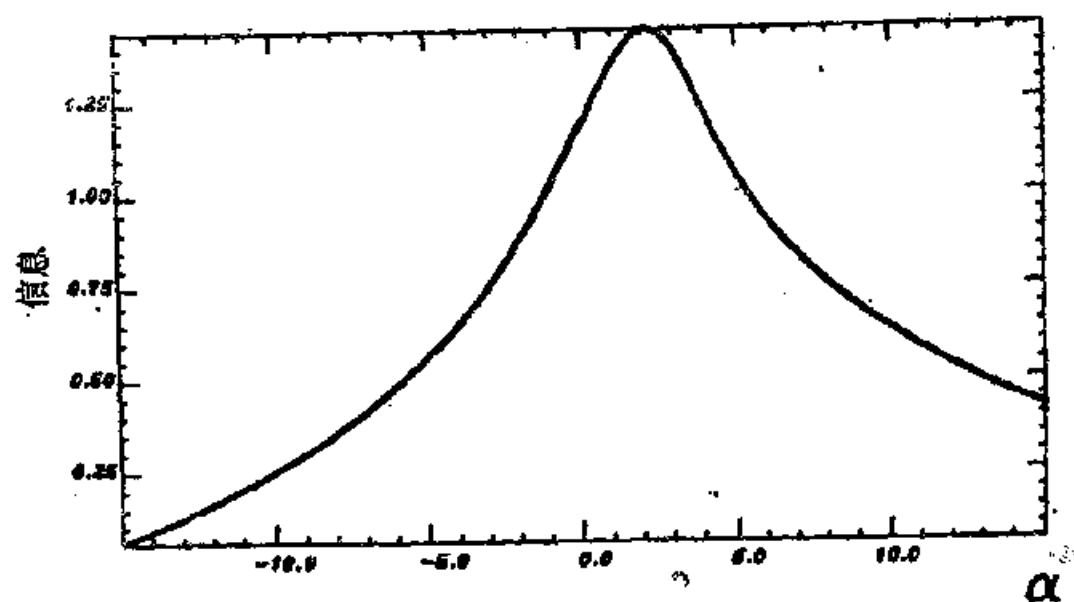
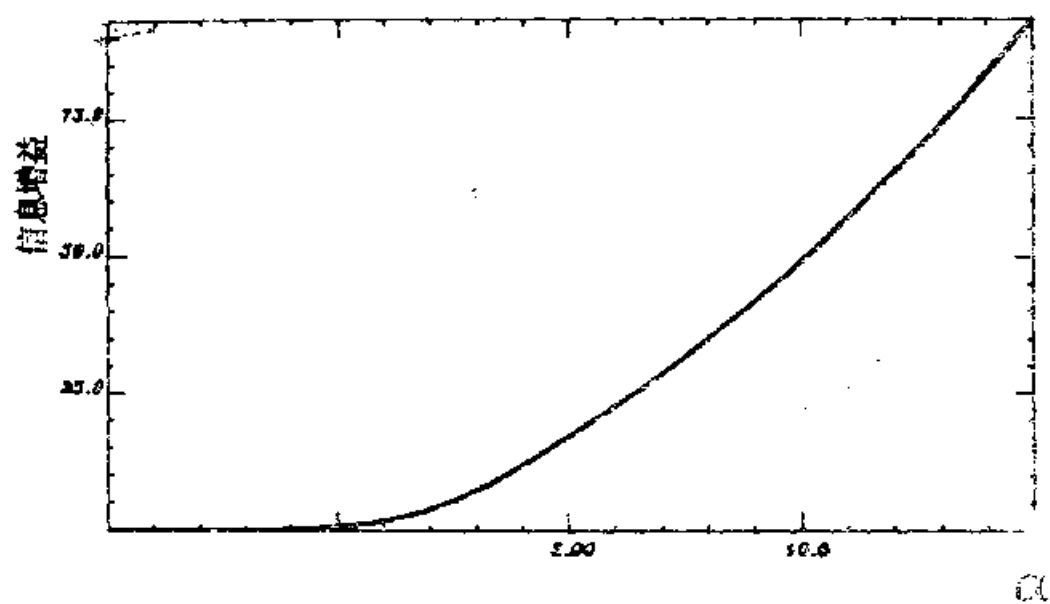


图7.2 相对于 $\beta = \varepsilon = 1$ 时，信息 (1.5) 随 $\alpha$ 的变化。



$$\alpha_0 = -5.00$$

图7.3 当 $\alpha_0 = -5.00$ 时， $\alpha$ 在-5到+15区间内的信息增益。

的定义是

$$K = \int_{-\infty}^{+\infty} d\xi f_{\alpha} \ln \left( f_{\alpha} / f_{\alpha_0} \right) \quad (7.88)$$

其中的 $f_{\alpha}$ 由下式给出：

$$f_{\alpha} = N(\alpha) \exp(\alpha \xi^2 - \beta \xi^4) \quad (7.89)$$

为了在阈值附近计算式 (7.88)，我们取 $\beta = 1$ ，在 $\alpha_0$ 固定而 $\alpha$ 从-5到15取值的情况下，我们计算了 $K$ （图7.3—图7.5）。

#### D. 讨论

以上，我们对一个用单序参量描述并经受二类相变的自组织系统，具体计算了其信息变化。如我们所见，虽然信息的变化很小，但与我们依据物理和信息理论所预期的完全一致。只要我们可以利用对称破缺实现两个可能状态之一，系统就可以存储 1 比特信息。

这就是当我们把式 (7.55) 应用到远离热平衡的系统时，与熵的概念相比，我们更愿使用信息的原因之一。另一方面，阈值附近的行为表明，由于临界涨落信息会显著增大。在这种情况下，人们会倾向于喜欢“熵”这个术语。但是，还有进一步的原因使我不愿意使用术语“熵”。正如在第五章中所见，使热平衡系统和远离热平衡的系统的信息熵最大所要求的约束是极不相同的。热平衡系统的约束之一总是能量，使用最大熵原理时，相应的拉格朗日乘子是温度的倒数。做些进一步的评论可能是合适的。

在非平衡相变（比如激光和流体动力学）中，我们熟知，在 $\alpha > 0$ 的区域将出现有序结构。图7.2表示的结果告诉我们，在这种有序状态中，信息熵比无序状态时大。这个结果是违背直觉的，而且确实，有些作者已经断言，不管是平衡还是非平衡情形，

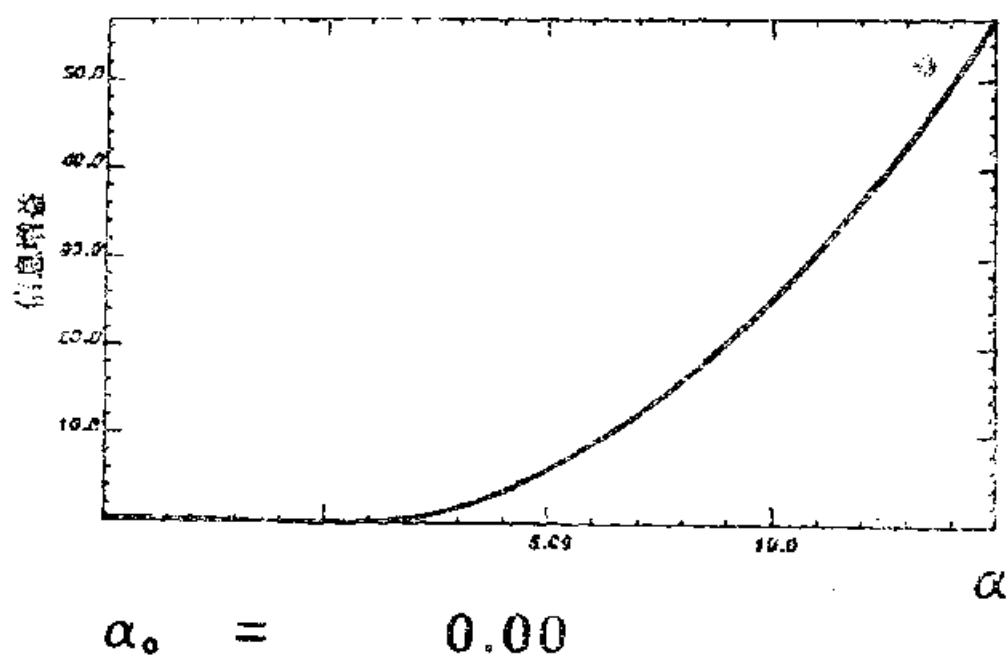


图7.4 与图7.3相同, 但 $\alpha_0 = 0$ 。

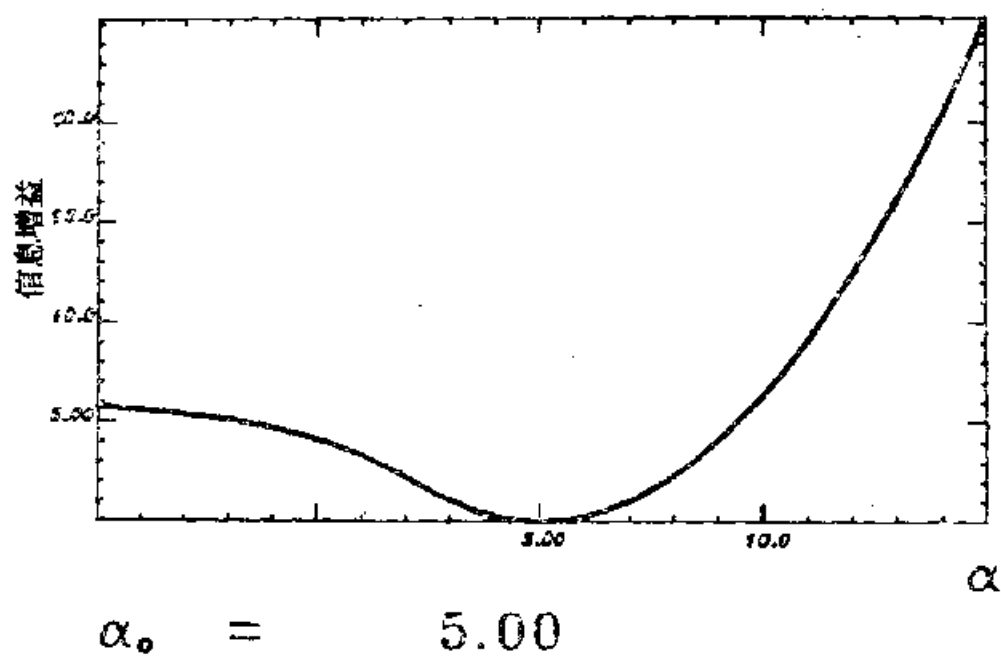


图7.5 与图7.3相同, 但  $\alpha_0 = 5$ 。

从无序转变为有序时，均有 $dS < 0$ 。然而，这一结论从未被证明过。更仔细地考查热力学，这个问题就明白了。根据热力学，封闭系统的熵及无序是只增不减的。因此，当我们维持整个系统的能量为常数而比较两种状态时，高熵状态具有较大的无序。然而在上述的开放系统中，我们没有在相同能量下对系统的状态进行比较。显然，当我们切断系统（如激光器）的能量输入和输出，使能量为一常数时，系统就会重新排布，增长它的熵而废弃它的有序状态。一般来说，如果我们只考虑（信息）熵，是不能对开放系统做出无序或有序的结论的。从形式上看，熵和信息在解释上的差别在于封闭或开放系统所采用的约束不同。

我们假定已经得到在约束下的开放系统的信息熵。然后，我们用分布函数计算平均能量。当我们把这个能量作为新的约束时，用该开放系统的分布函数不能使熵最大，而另一个只有能量是给定约束的函数可使熵最大。我们将在7.10节再从另一种观点来阐明这个结果。

## 7.9 单序参量的信息、信息增益和效率的严格解析结果

本节我们将给出信息 $I_i$ 、信息增益 $K_i$ 和效率 $W$ 的精确解析式，所有这些，都将用分布函数

$$f(\xi) = N \exp [\alpha \xi^2 - \beta \xi^4] \quad (7.90)$$

进行计算。我们简要地回顾一下这几个量。

$$I_i = -\ln N - \alpha \langle \xi^2 \rangle + \beta \langle \xi^4 \rangle \quad (7.91)$$

其中，分布函数的 $n$ 阶矩是

$$\langle \xi^n \rangle = \int_{-\infty}^{+\infty} f(\xi) \xi^n d\xi \quad (7.92)$$

$$K_f \equiv K_{f f_0} = \ln N - \ln N_0 + (\alpha - \alpha_0) \langle \xi^2 \rangle_0. \quad (7.93)$$

最后, 由于二阶矩  $\langle \xi^2 \rangle_0$  是辐射强度 (因此也是功) 的量度, 可以推得系统的效率是

$$W = \frac{d\langle \xi^2 \rangle_\alpha}{d\alpha} \quad (7.94)$$

利用式 (7.90) 来计算式 (7.94), 其结果为 (参见式 (7.36))

$$W = \langle \xi^4 \rangle_\alpha - \langle \xi^2 \rangle_\alpha^2 \quad (7.95)$$

导出  $I_f$ ,  $K_f$  和  $W$  以及控制参数  $\alpha$ ,  $\beta$  之间的关系式是非常有趣的。正如上面看到的那样, 按照  $\alpha$  值之不同可分为几个区域,  $\alpha \approx 0$ , 为不稳定区,  $|\alpha| \gg 0$ , 为稳定区。这里所需的主要数学工具是积分

$$\begin{aligned} & \int_0^\infty x^n \exp(\alpha x^{2m} - \beta x^{4m}) dx \\ &= (2m)^{-1} (2\beta)^{-\frac{n+1}{4m}} \Gamma\left(\frac{n+1}{2m}\right) D_{-\frac{n+1}{2m}}(\lambda) \exp(\lambda^2) \end{aligned} \quad (7.96)$$

其中,  $\lambda = \sqrt{\frac{-\alpha}{2\beta}}$ ,  $\Gamma$  是  $\Gamma$ -函数,  $D$  是抛物柱函数。由于

$$N^{-1} = \int_{-\infty}^{+\infty} \exp(\alpha \xi^2 - \beta \xi^4) d\xi \quad (7.97)$$

将式 (7.96) 简单应用到式 (7.97), 可得归一化常数  $N$ ,

$$N = (2\beta)^{\frac{1}{4}} \pi^{-\frac{1}{2}} \exp(-\lambda^2) [D_{-\frac{1}{2}}(\lambda)]^{-1} \quad (7.98)$$

同样, 利用式 (7.92) 和式 (7.96), 我们易于计算  $f(\xi)$  的二



阶矩和四阶矩，它们分别为

$$\langle \xi^2 \rangle = (2\beta)^{-\frac{1}{2}} D_{-\frac{3}{2}}(\lambda) / 2 D_{-\frac{1}{2}}(\lambda) \quad (7.99)$$

和

$$\langle \xi^4 \rangle = (8\beta)^{-1} 3 D_{-\frac{5}{2}}(\lambda) / D_{-\frac{1}{2}}(\lambda) \quad (7.100)$$

得到只用控制参数 $\alpha$ ， $\beta$ 表示的 $N$ ， $\langle \xi^2 \rangle$ 和 $\langle \xi^4 \rangle$ 之后，我们可以简单地把这些式子代入 $I_t$ ， $K_t$ 和 $W$ 的公式——也就是式(7.91)、式(7.93)和式(7.95)。于是，分布函数为 $f(\xi)$ 的系统所包含的信息为

$$\begin{aligned} I_t = & 0.3991 - \left(\frac{1}{4}\right) \ln \beta + \lambda^2 + \ln D_{-\frac{1}{2}}(\lambda) \\ & + [(\lambda/2) D_{-\frac{3}{2}}(\lambda) + \left(\frac{3}{8}\right) D_{-\frac{5}{2}}(\lambda)] / D_{-\frac{1}{2}}(\lambda) \end{aligned} \quad (7.101)$$

同样，可求出 $f(\xi)$ 中的参数 $\alpha$ 从 $\alpha_0$ 变到 $\alpha$ 时，信息增益为

$$K_{t,0} = (\alpha_0^2 - \alpha^2) / 2\beta + \ln [D_{-\frac{1}{2}}(\lambda_0) / D_{-\frac{1}{2}}(\lambda)] \quad (7.102)$$

其中， $\lambda_0 = \alpha_0 / \sqrt{2\beta}$ 。最后，从式(7.95)、式(7.99)和公式(7.100)得到系统之效率为

$$\begin{aligned} W = & \{3 D_{-\frac{5}{2}}(\lambda) D_{-\frac{1}{2}}(\lambda) - [D_{-\frac{3}{2}}(\lambda)]^2\} / \{8\beta \\ & \times [D_{-\frac{1}{2}}(\lambda)]^2\} \end{aligned} \quad (7.103)$$

值得强调的是，如果 $\beta \neq 0$ ， $\alpha < \infty$ ，则本节所有公式，即式(7.98)一式(7.103)可表示为控制参数 $\alpha$ 和 $\beta$ 的解析函数。我们特别感兴趣的量 $I_t$ ， $K_t$ 和 $W$ 十分紧密地依赖于 $\lambda$ 。由于抛物柱函数 $D$ 是特殊函数，一般来说，表示式(7.101)一式(7.103)

不能再进行简化，除了我们将各个 $\beta$ 值列成表。有三种特殊情况颇有研究价值。第一，不稳定点本身出现在 $\alpha=0$ 处，即 $\lambda=0$ 处；第二， $\alpha \rightarrow 0$ 和 $\lambda \rightarrow 0$ 表征不稳定点的紧邻；第三， $\lambda \gg 1$ 表征稳定区域。我们将详细分析每一种特殊情况。

### A. 不稳定点

抛物柱函数与韦伯函数的关系是

$$D_{-\nu-\frac{1}{2}}(\lambda) = U(P, \lambda) \quad (7.104)$$

在不稳定点， $\alpha=0$ ，利用性质

$$U(P, 0) = \sqrt{\pi} / 2^{\frac{a}{2} + \frac{1}{4}} \Gamma\left(\frac{a}{2} + \frac{3}{4}\right) \quad (7.105)$$

求得归一化常数 $N$ 的极大值

$$N_c = 0.5516\beta^{\frac{1}{4}} \quad (7.106)$$

二阶矩和四阶矩分别为

$$\langle \xi^2 \rangle_c = 0.3380\beta^{-\frac{1}{2}} \quad (7.107)$$

和

$$\langle \xi^4 \rangle_c = 0.2500\beta^{-1} \quad (7.108)$$

与此相应，不稳定点的信息 $I_c$ 等于

$$I_c = 0.8449 - \frac{1}{4} \ln \beta \quad (7.109)$$

同样，从分布 $f_c(\alpha \approx 0)$ 到 $f$ ，信息增益是

$$K_{ff_c} = 0.5949 - \frac{1}{4} \ln \beta + \ln N + \alpha \langle \xi^2 \rangle_c \quad (7.110)$$

最后，可以方便地求得不稳定点的效率，其表达式为

$$W_c = 0.1358\beta^{-1} \quad (7.111)$$

显然，这些值仅依赖于 $\beta$ （除信息增益外），并代表它们的函数对 $\alpha$ 的极大值。我们的兴趣是探索 $\alpha \rightarrow 0$ 时趋于这些极大值的途径。

## B. 研究不稳定性的途径

为了研究不稳定点的紧邻，下面我们使用韦伯函数的渐近展开式

$$U(P, \mp |\lambda|) \approx \left[ \sqrt{\pi} / 2^{\frac{p}{2} + \frac{1}{4}} \cdot \Gamma\left(\frac{P}{2} + \frac{1}{4}\right) \right] \times \exp(\mp P\lambda) \quad (7.112)$$

其中， $\lambda \approx 0$ ，并因此忽略了指数函数中 $\lambda$ 的高次幂。借此，我们得到趋于 $\lambda = 0$ 时， $N$ ， $\langle \xi^2 \rangle$ 和 $\langle \xi^4 \rangle$ 起主要作用的部分：

$$N \simeq N_c \exp(-\lambda^2) \quad (7.113)$$

$$\langle \xi^2 \rangle \simeq \langle \xi^2 \rangle_c \exp(\mp \lambda) \quad (7.114)$$

和

$$\langle \xi^4 \rangle \simeq \langle \xi^4 \rangle_c \exp(\mp 2\lambda) \quad (7.115)$$

其中， $N_c$ ， $\langle \xi^2 \rangle_c$ 和 $\langle \xi^4 \rangle_c$ 分别由式(7.106)、式(7.107)和式(7.108)给出。在式(7.114)和式(7.115)中，负号对应于 $\lambda > 0$  ( $\alpha < 0$ )，而正号对应于 $\lambda < 0$  ( $\alpha > 0$ )。因此，在靠近 $\alpha = 0$ 时， $N$ ， $\langle \xi^2 \rangle$ 和 $\langle \xi^4 \rangle$ 的值按指数形式增大，并分别趋于它们的临界值 $N_c$ ， $\langle \xi^2 \rangle_c$ 和 $\langle \xi^4 \rangle_c$ 。把这些结果代入 $I_t$ 中，得出

$$I_t \simeq I_c + \lambda^2 - 0.478\lambda \exp(\mp \lambda) + 0.250 [\exp(\mp 2\lambda) - 1] \quad (7.116)$$

它按照 $\lambda$ 的指数函数的组合趋于其临界值 $I_c$ 。当分布函数从 $f_0$ 变到 $f$ （二者皆距 $f_c$ 很近），可以算出取得的信息增益为

$$K_{f|f_0} \simeq K_{f|f_c} + \frac{\alpha_0^2}{2\beta} - \alpha_0 \langle \xi^2 \rangle_c \quad (7.117)$$

因此， $K_{f|f_0}$ 按照 $\alpha_0$ 的线性项与 $\alpha_0$ 的平方项的组合而趋于 $K_{f|f_c}$ 。

显然， $\alpha_0 \rightarrow 0$ 。令 $\lambda \simeq 0$ ，可进一步化简式(7.117)，得

$$K_{\text{eff}} \simeq \lambda_0^2 - \lambda^2 + 0.478(\lambda - \lambda_0) \exp(\mp \lambda) \quad (7.118)$$

最后，我们可以计算出趋于不稳定点的效率函数

$$W \simeq W_0 \exp(-2\lambda) \quad (7.119)$$

我们得到如下结论，我们感兴趣的所有量（除 $K$ 外）都以指数函数或指数函数的组合形式趋于它们的临界值。将会看到，这与它们在稳定区域的行为形成鲜明的对照。

### C. 稳定区域

另一个极限情况是 $\alpha \rightarrow \infty$ ，它可以由以下途径实现：增大 $\alpha$ 而使 $\beta$ 保持不变，或者（象平均场近似一样）让 $\alpha$ 和 $\beta$ 以相同的速度趋于无穷大，比如，通过增大系统的体积来实现。于是我们便渐近地离开了不稳定点。韦伯函数的相应展开式是

$$U(P, \lambda) \simeq e^{-\lambda^2/4} \lambda^{-P-\frac{1}{2}} \quad (7.120)$$

这里已略去 $1/\lambda$ 的高次幂。把式（7.120）应用到式（7.98）中，便得出归一化常数

$$N \simeq (\lambda/\pi)^{\frac{1}{2}} \exp(\lambda^2/4) \quad (7.121)$$

它显然要求 $\alpha < 0$ 。类似地得到二阶矩

$$\langle \xi^2 \rangle \simeq (4\lambda)^{-1} \quad (7.122)$$

以及四阶矩

$$\langle \xi^4 \rangle = \frac{3}{4} \lambda^2 \quad (7.123)$$

将这些表示式代入式（7.91），得信息

$$I_r \simeq 0.5724 - \left(\frac{1}{2}\right) \ln \lambda - \lambda^2/4 + (\beta/2)^{\frac{1}{2}}/2 + (3/4)\lambda^{-2} \quad (7.124)$$

仅当 $\alpha < 0$ 时 $I_r$ 这个公式方适用，它明显表示了 $\lambda \rightarrow \infty$ 时 $I_r$ 以幂律减少到零。当然， $\lambda$ 有一最大值，超过此最大值，式（7.124）就

不能使用，否则就会得出负的信息。此外，在这种情况下，还必须计及那些变得很重要的伺服模的影响。

下面，我们计算从 $f_0$ 到 $f_0$ 的信息增益

$$K \simeq \frac{1}{2} \ln \left( \frac{\alpha}{\alpha_0} \right) + (\alpha^2 - \alpha_0^2)/8\beta + (1 - \frac{\alpha_0}{\alpha}) \frac{\sqrt{\beta}}{2} \sqrt{2} \quad (7.125)$$

最后，求出系统的效率

$$W = (11/16)\lambda^{-2} \quad (7.126)$$

当 $\lambda$ 增大到无穷时，效率单调下降。因而我们看到，随着 $\lambda$ 的增大，相关的量以不同的幂律减小。

#### D. 注入的信号

如果不以耦合项 $-\sigma\xi$ 引入与 $\xi$ 耦合的外场，最可能的 $\xi$ 值是双重简并的，也就是两个值 $\xi = +(\alpha/2\beta)^{\frac{1}{2}}$ 和 $\xi = -(\alpha/2\beta)^{\frac{1}{2}}$ 是等概率的。将振幅为 $\sigma$ 的信号注入，会消除这种简并性，而且能优先选择与 $\sigma$ 有相同符号的 $\xi$ 值。在这种情况下，我们应当把分布函数修改为

$$f_0(\xi) = N \exp(-\sigma\xi + \alpha\xi^2 - \beta\xi^4) \quad (7.127)$$

我们今后将用下标“ $\sigma$ ”表示有信号注入的情况，而用下标“0”表示没有信号注入的情况。对式(7.127)给出的 $f_0(\xi)$ ，我们无法推导出严格的结果，因为式(7.96)不适用于它。作为替换，在式(7.90)中的 $f_0$ 附近，我们将把 $f_0(\xi)$ 展开为 $\sigma$ 的幂级数。由于已经假定，同 $\alpha$ 和 $\beta$ 相比， $\sigma$ 很小，我们仅保留最低阶项，并且一开始就将 $N_0$ 近似表示为式(7.98)的 $N_0$ 。将式(7.96)简单应用于 $f_0(\xi)$ 的级数展开式，在最低阶近似下得到

$$\langle \xi^2 \rangle_{\sigma} \simeq \langle \xi^2 \rangle_0 \left[ 1 + \frac{3}{4} \sqrt{2\beta^{-}} (D_{-\frac{5}{2}}(\lambda) / D_{-\frac{3}{2}}(\lambda)) \sigma^2 + \dots \right] \quad (7.128)$$

和

$$\langle \xi^2 \rangle_{\sigma} \simeq \langle \xi^4 \rangle_0 \left[ 1 + \frac{5}{4} \sqrt{2\beta^{-}} (D_{-\frac{7}{2}}(\lambda) / D_{-\frac{5}{2}}(\lambda)) \sigma^2 + \dots \right] \quad (7.129)$$

其中  $\langle \xi^2 \rangle_0$  和  $\langle \xi^4 \rangle_0$  分别由式 (7.99) 和式 (7.100) 给出。以上，注入信号的振幅  $\sigma$  通过修正项进入两个表达式，而且最低阶是平方项。给定  $\lambda$  的区域，就能详细分析这些项。在不稳定点的邻域，得到

$$\langle \xi^2 \rangle_{\sigma} \simeq \langle \xi^2 \rangle_0 \left[ 1 + 0.3698 \beta^{-\frac{1}{2}} \exp(\mp 0.4142\lambda) \sigma^2 \right] \quad (7.130)$$

和

$$\langle \xi^4 \rangle_{\sigma} \simeq \langle \xi^4 \rangle_0 \left[ 1 + 0.5070 \beta^{-\frac{1}{2}} \exp(\mp 0.3178\lambda) \sigma^2 \right] \quad (7.131)$$

$\sigma$  所产生的修正通过  $\lambda$  的一个指数函数实现，因此，当趋于不稳定点时，它变得越来越重要。我们可以进一步计算它对信息和效率的影响，这时

$$I_{\sigma} \simeq I_0 - 0.1768 \beta^{-\frac{1}{2}} \exp(\mp \sqrt{2\lambda}) \times [\lambda - 0.7172 \exp(\mp 0.9036\lambda)] \sigma^2 \quad (7.132)$$

和

$$W^{\sigma} \simeq W_0 + 0.0845 \beta^{-\frac{1}{2}} \exp(\pm 2.4142\lambda) \\ \times [1 + 1.5 \exp(\pm 0.0964\lambda)] \sigma^2 \quad (7.133)$$

修正项的主要部分又是 $\sigma$ 的平方项,并引入了 $\lambda$ 的一个指数函数。

另一方面,我们求出远离不稳定点的二阶矩和四阶矩

$$\langle \xi^2 \rangle_0 \simeq \langle \xi^2 \rangle_0 (1 + 0.5303 \lambda^{-1} \beta^{-\frac{1}{2}} \sigma^2) \quad (7.134)$$

$$\langle \xi^4 \rangle_0 \simeq \langle \xi^4 \rangle_0 (1 + 0.8839 \lambda^{-1} \beta^{-\frac{1}{2}} \sigma^2) \quad (7.135)$$

两式都有 $\sigma^2$ 的修正项,并随 $\lambda^{-1}$ 线性地减小。其结果是,可求出的信息为

$$I_0 \simeq I_0 - 0.1326 \beta^{-\frac{1}{2}} \lambda^{-2} (\alpha^2 - 5\beta \lambda^{-1}) \quad (7.136)$$

而效率近似表示为

$$W_0 \simeq W_0 - 0.5966 \beta^{-\frac{1}{2}} \lambda^{-3} \sigma^2 \quad (7.137)$$

### E. 结论

在不稳定点的邻域,自组织系统的信息和效率随输入功率 $\alpha$ 指数增长,于是伺服模的贡献被掩盖。另一方面,远离不稳定点时,系统的信息和效率按输入功率的二次幂减小。因而,在此区域伺服模的作用将重新显现。有趣的是,量 $\langle \xi^2 \rangle^2 / \langle \xi^4 \rangle$ 在两个极限情况下均为常数。在不稳定邻域中,该常数值等于0.4570,它标志着一个远比稳定区域大的信号强度。而在稳定区域,这个常数仅为0.083。我们已给出了 $\langle \xi^2 \rangle$ ,  $\langle \xi^4 \rangle$ ,  $I_0$ ,  $K_{ff0}$ 和 $W$ 的一般公式,它们适用于 $\alpha$ 和 $\beta$ 的各种情况,且是解析函数,我们预料能够计算出在 $\lambda$ 从0到 $\infty$ 的中间区域内介于0.083和0.4570间的所有

值。我们发现，注入的信号的作用是对未微扰结果引入了修正，其主要贡献与注入信号振幅的平方成正比。

## 7.10 柯里蒙托维奇 (Klimontovich) 的S-定理

本节，我们打算做两件事。首先要证明5.2节用宏观方法导出的激光分布函数也能够从微观理论求出。然后用这个结果去说明柯里蒙托维奇的一个定理，也就是所谓的S-定理，它能重新阐明我们的结果——尽管远离平衡的系统进入一种更有序的状态，但其熵却可以增加。

我们先从微观理论展开讨论。考虑一个复数序参量 $\xi$ ，我们研究它遵从下述方程的有关情况。将公式

$$\dot{\xi} = \lambda\xi - \beta\xi |\xi|^2 + F(t) \quad (7.138)$$

中的 $\xi$ 分解为实部和虚部，即

$$\xi = q_1 + iq_2 \quad (7.139)$$

类似地，涨落力 $F$ 可分解为

$$F = F_1 + iF_2 \quad (7.140)$$

上面的 $F_1$ 和 $F_2$ 是实数。于是，可把序参量方程 (7.138) 分解为实量 $q_1$ 和 $q_2$ 的方程

$$\dot{q}_j = \lambda q_j - \beta q_j (q_1^2 + q_2^2) + F_j, \quad j = 1, 2 \quad (7.141)$$

也可以把它们写成简洁的形式

$$\dot{q}_j = K_j(q_1, q_2) + F_j \quad (7.142)$$

详细的分析证明，涨落力具有下述性质

$$\langle F_j(t) \rangle = 0 \quad (7.143)$$



$$\langle F_1(t)F_1(t') \rangle = \delta_{11}Q\delta(t-t') \quad (7.144)$$

我们用2.3节的规定可以直接建立式 (7.141) 或式 (7.138)的福克-普朗克方程,

$$\dot{f} = -\frac{\partial}{\partial q_1}(K_1 f) - \frac{\partial}{\partial q_2}(K_2 f) + \frac{Q}{2}\left(\frac{\partial^2 f}{\partial q_1^2} + \frac{\partial^2 f}{\partial q_2^2}\right) \quad (7.145)$$

我们很快就会确信, 它满足细致平衡条件。因此, 对于稳定状态

$$\dot{f} = 0 \quad (7.146)$$

可直接求出

$$f(q_1, q_2) = N \exp \left[ \left\{ \lambda(q_1^2 + q_2^2) - \frac{\beta}{2}(q_1^4 + q_2^4) \right\} Q^{-1} \right] \quad (7.147)$$

为给以下研究带来方便, 我们把分布函数变换到一新坐标系。为此, 考虑概率函数

$$f(q_1, q_2) dq_1 dq_2 \quad (7.148)$$

我们现在引入代替 $q_1$ 和 $q_2$ 的新坐标系, 即

$$q_1 = r \cos \varphi \quad (7.149)$$

$$q_2 = r \sin \varphi \quad (7.150)$$

或等价地表为

$$q = r \exp [i\varphi] \quad (7.151)$$

在 $r, \varphi$ 这个新坐标中, 可以把体积元写为

$$dV = dq_1 dq_2 = r dr d\varphi \quad (7.152)$$

因而, 当我们使用新坐标 $r$ 和 $\varphi$ 时, 可以把式 (7.148) 具体写成

$$f(r)r dr d\varphi = N \exp \left[ \left\{ \lambda r^2 - \beta r^4/2 \right\} Q^{-1} \right] r dr d\varphi \quad (7.153)$$

现在, 用下式引入量 $u$ ,

$$r^2 = u \quad (7.154)$$

在物理学的许多情况中, 可以把 $u$ 解释为势能。我们从式(7.153)和式(7.154)直接得到

$$f(u)du = N \exp [\{\lambda u - \beta u^2/2\}Q^{-1}]du \quad (7.155)$$

在这里, 我们略去了公共因子 $d\varphi$ 。下面, 我们要求 $f(u)$ 归一化,

$$\int_0^\infty f(u)du = 1 \quad (7.156)$$

为了说明 $S$ -定理, 我们讨论信息熵

$$i = - \int_0^\infty f(u) \ln f(u) du \quad (7.157)$$

以及平均能量, 它的定义是

$$\langle u \rangle = \int_0^\infty u f(u) du \quad (7.158)$$

为了能使式(7.157)和式(7.158)具体化, 我们讨论激光物理中大家熟知的三种极限情况, 它们也可在其它领域找到。

(i) 在激光阈值以下

在 $\lambda < 0$ 的情况下, 激光振荡只受到弱激励, 于是它的能量遵守不等式

$$\frac{\beta}{2} u^2 \ll |\lambda| u \quad (7.159)$$

在这个假定下, 与线性项相比, 可以忽略分布函数(7.147)中的平方项。这样, 可简单地计算式(7.158), 得

$$\langle u \rangle = \frac{Q}{|\lambda|} \quad (7.160)$$

我们可用式 (7.160) 将不等式 (7.159) 化为

$$\frac{\beta Q}{2|\lambda|} \ll 1 \quad (7.161)$$

这种情况下，分布函数具体写为

$$f_1(u) = \frac{|\lambda|}{Q} \exp \left[ -\frac{|\lambda|}{Q} u \right] \quad (7.162)$$

这里已经包含了精确的归一化因子。现在，易于计算式 (7.157) 和式 (7.158)，我们有

$$i_1 = \ln \frac{Q}{|\lambda|} + 1 \quad (7.163)$$

和

$$\langle u \rangle_1 = \frac{Q}{|\lambda|} \quad (7.164)$$

由于我们现在在区域 1（即阈值以下）计算式 (7.157) 和公式 (7.158)，所以  $i$  和  $\langle u \rangle$  都附有脚标 1。当我们增加泵浦时，即增加泵浦功率时， $\lambda$  下降。于是式 (7.164) 告诉我们，平均能量增大。现在我们讨论激光处于阈值的情形。

(ii) 在阈值处

在这种情况下，控制参数  $\lambda$  满足

$$\lambda = 0 \quad (7.165)$$

相应的分布函数可写为

$$f_2(u) = \sqrt{\frac{2\beta}{\pi Q}} \exp \left[ -\frac{\beta u^2}{2Q} \right] \quad (7.166)$$

我们用脚标 2 表示阈值。信息熵和平均能量的计算仍然十分简单，

$$i_2 = -\frac{1}{2} \ln \frac{\pi Q}{2\beta} + \frac{1}{2} \quad (7.167)$$

$$\langle u \rangle_2 = \sqrt{\frac{2Q}{\pi\beta}} \quad (7.168)$$

(iii) 远在阈值以上

这里的  $\lambda > 0$ 。我们讨论不等式

$$\frac{Q\beta}{\lambda^2} \ll 1 \quad (7.169)$$

成立的区域。在这种情况下，用

$$f_3(u) = N \exp \left[ -\frac{\beta}{2Q} \left( u - \frac{\lambda}{\beta} \right)^2 \right] \quad (7.170)$$

可以很好地近似表示分布函数  $f$ ，我们用脚标 3 表示这种情形，并方便地计算出信息熵和平均能量，

$$i_3 = \frac{1}{2} \ln \frac{2\pi Q}{\beta} + \frac{1}{2} \quad (7.171)$$

$$\langle u \rangle_3 = \frac{\lambda}{\beta} \quad (7.172)$$

现在可以比较区域 1、2、3 的信息熵，也可以比较这三个区域中的平均能量。我们可以验证，它们遵守不等式

$$i_1 < i_2 < i_3 \quad (7.173)$$

和

$$\langle u \rangle_1 < \langle u \rangle_2 < \langle u \rangle_3 \quad (7.174)$$

信息熵随着泵浦的增大而增加，我们在 7.9 节已经对一个实数序参量导出了这一结果。在该节，我们已经观察到，仅当系统能量相同，才能对熵进行相应的比较。因此，按照柯里蒙托维奇的想法，我们现在选择涨落强度，以使各能量相同。换句话说，我们将改变区域 1 和 2 中的涨落力，从而得到

$$\langle u \rangle_{1,r} = \langle u \rangle_{2,r} = \langle u \rangle_{3,r} \quad (7.175)$$

我们将从阈值以下的区域着手，仔细研究其后果。

为使 $\lambda$ 减小时，能量保持不变，我们必须与 $\lambda$ 成比例地降低噪声水平 $Q$ 。通过激光物理学来讨论相应结果是很有趣的。减小 $\lambda$ ，也就是增大泵浦强度，这时出现了受激辐射过程，它意味着放大了光场的各个波列。因此，平均能量增大。另一方面，当我们降低噪声水平时，自发辐射的作用变小。依照放大因子来降低噪声水平或自发辐射的产生率，易于找到能量保持为常数的情况。同时可发现，信息熵也保持常数，也就是，在所谓的线性放大区域，在刚才所述的条件下信息熵保持为常数。

柯里蒙托维奇作过另一对比，他使用了固定平均能量的等式(7.175)，在阈值以上，

$$\langle u \rangle_{s.r.} = \frac{\lambda}{\beta} \quad (7.176)$$

其中， $\lambda$ 和 $\beta$ 现在是给定的量。我们现在从

$$\langle u \rangle_{2.r.} = \sqrt{\frac{2Q}{\pi\beta}} \quad (7.177)$$

中选取 $Q = Q_2$ 。使关系式

$$\langle u \rangle_{2.r.} = \frac{\lambda}{\beta} \quad (7.178)$$

成立，我们便得

$$Q_2 = \frac{\pi}{2} \frac{\lambda^2}{\beta} \quad (7.179)$$

按照式(7.173)，将式(7.179)代入信息熵的公式(7.171)，可得到

$$i_{2.r.} = \frac{1}{2} \ln \frac{\pi Q_2}{2\beta} + \frac{1}{2} = \ln \frac{\pi \lambda}{2\beta} + \frac{1}{2} \quad (7.180)$$

在区域 1 用同样的步骤, 进行计算, 从

$$\langle u \rangle_{1,r} = \frac{Q}{|\lambda_1|} \quad (7.181)$$

中选取  $Q = Q_1$ , 使得

$$\langle u \rangle_{1,r} = \frac{\lambda}{\beta} \quad (7.182)$$

我们容易得到

$$Q_1 = \frac{\lambda |\lambda_1|}{\beta} \quad (7.183)$$

将此结果代入信息熵的公式 (7.163), 得

$$i_{1,r} = \ln \frac{Q_1}{|\lambda_1|} + 1 = \ln \frac{\lambda}{\beta} + 1 \quad (7.184)$$

回忆不等式

$$\frac{Q\beta}{\lambda^2} \ll 1 \quad (7.185)$$

我们得到下面的信息熵不等式

$$i_{1,r} > i_{2,r} > i_{3,r} \quad (7.186)$$

因此, 在调节噪声水平而保持平均能量不变的前提下, 增大控制参数使系统进入更有序的状态, 我们的确得到了熵减少的结果, 例如获得激光作用。在有序化或自组织出现时, 系统的信息熵便减少, 这就是柯里蒙托维奇的S-定理, 他也研究了在 1 和 3 之间的中间区域的一般情况。在S-定理的推导中, 他考虑了能量和信息熵随控制参数的变化而改变。

在柯里蒙托维奇的论述中, 只考虑了能量随控制参数的改变, 而没有同时考虑噪声水平的变化, 而噪声水平的变化是保持能量为常数所必须的。因此, 他的定理的推导要稍加改变, 然而

本书不打算对此加以介绍。

## 7.11 接近非平衡相变时伺服模对信息的贡献

7.2节已经证明,可以把整个系统的信息分解为序参量的信息和伺服模的信息。前几节已计算了序参量的信息。现在我们研究伺服模的信息。为此,应用伺服模原理至打头项。正如我们已在2.6和7.4节证明的那样,条件概率 $P$ 为

$$P(\xi_s | \underline{\xi}_s) = N_s \exp [-(\lambda_s \xi_s - g_s(\underline{\xi}_s))^2 / Q_s] \quad (7.187)$$

其中,归一化因子 $N_s$ 为

$$N_s = |\lambda_s| (\pi Q_s)^{-\frac{1}{2}} \quad (7.188)$$

下面,我们认为 $\lambda_s, Q_s$ 以及归一化因子 $N_s$ 是控制参数 $\alpha$ 的函数,即

$$\lambda_s = \lambda_s(\alpha) \quad (7.189)$$

$$Q_s = Q_s(\alpha) \quad (7.190)$$

$$N_s = N_s(\alpha) \quad (7.191)$$

我们可进一步假定, $g_s$ 依赖于控制参数 $\alpha$ ,且

$$g_s(0) = 0 \quad (7.192)$$

成立。对于给定的控制参数和序参量 $\underline{\xi}_s$ ,伺服模 $s$ 之信息为

$$I_s(\underline{\xi}_s, \alpha) = - \int P(\xi_s | \underline{\xi}_s) \ln P(\xi_s | \underline{\xi}_s) d\xi_s \quad (7.193)$$

注意,积分被延拓到从负无限大到正无限大。我们现在利用下面关系引入一个新变量 $\xi'$ 来代替 $\xi_s$ ,

$$\xi_s - g_s / \lambda_s = \xi' \quad (7.194)$$

将式(7.187)和式(7.194)代入式(7.193),得

$$I_s = - \int N_s(\alpha) \exp [ - \lambda_s^2 \xi'^2 / Q_s ]$$

$$\times \ln [N_s(\alpha) \exp [-\lambda_s^2 \xi'^2 / Q_s]] d\xi' \quad (7.195)$$

为了尽可能消去 $I_s$ 对 $\alpha$ 的依赖关系,我们作如下变换

$$|\lambda_s(\alpha)| Q_s(\alpha)^{-\frac{1}{2}} \xi' = |\lambda_s(0)| Q_s(0)^{-\frac{1}{2}} \eta \quad (7.196)$$

或者再解出 $\xi'$ ,

$$\xi' = \frac{|\lambda_s(0)|}{|\lambda_s(\alpha)|} \left( \frac{Q_s(0)}{Q_s(\alpha)} \right)^{-\frac{1}{2}} \eta \quad (7.197)$$

把 $I_s$ 看作序参量 $\xi_s$ 和控制参数 $\alpha$ 的函数,得

$$\begin{aligned} I_s(0, \alpha) = & - \int \frac{N_s(\alpha)}{N_s(0)} N_s(0) \exp [-\lambda_s^2(0) \eta^2 / Q_s(0)] \\ & \times \ln \left\{ \frac{N_s(\alpha)}{N_s(0)} N_s(0) \exp [-\lambda_s^2(0) \eta^2 / Q_s(0)] \right\} \\ & \times \frac{|\lambda_s(0)|}{|\lambda_s(\alpha)|} \left( \frac{Q_s(0)}{Q_s(\alpha)} \right)^{-\frac{1}{2}} d\eta \quad (7.198) \end{aligned}$$

注意,这个表达式现在与序参量 $\xi_s$ 无关,这点已在 $I_s$ 中用零自变量表示了。利用关系式(7.188)可以看出,式(7.198)中的一些因子消掉了,因而式(7.198)可写为较简单的形式

$$I_s(0, \alpha) = -N_s(0) \int \exp[-\lambda_s^2(0) \eta^2 / Q_s(0)] \{A + B\} d\eta \quad (7.199)$$

其中 $A$ 是

$$A = \ln \left[ \frac{N_s(\alpha)}{N_s(0)} \right] \quad (7.200)$$

而 $B$ 是

$$B = \ln \{N_s(0) \exp[-\lambda_s^2(0) \eta^2 / Q_s(0)]\} \quad (7.201)$$



由于 $A$ 是一个常数,对高斯函数的积分可以立即算出,其结果与归一化因子之结合为1。我们容易确信,含有项 $B$ 的积分恰巧是控制参数 $\alpha=0$ 时信息 $I_s$ 的定义。因此,我们可以将式(7.199)化为

$$I_s(0, \alpha) = - \ln \frac{N_s(\alpha)}{N_s(0)} + I_s(0, 0) \quad (7.202)$$

使用式(7.188)所示的归一化因子 $N_s$ 的明确定义,也可将式(7.202)表示为

$$I_s(0, \alpha) = - \ln \left[ - \frac{|\lambda_s(\alpha)|}{|\lambda_s(0)|} \cdot \left( \frac{Q_s(\lambda)}{Q_s(0)} \right)^{-\frac{1}{2}} \right] + I_s(0, 0) \quad (7.203)$$

我们现在回忆一下,整个系统的信息 $i$ 一般可分解为序参量的信息 $I_r$ 和伺服模的信息 $I_s$ 。由于 $I_s$ 不依赖于 $\xi_s$ ,可以直接对序参量进行积分,并得到因子1。因此,最后结果是

$$i = I_r + \sum_s I_s(0) - \sum_s \ln \frac{N_s(\alpha)}{N_s(0)} \quad (7.204)$$

下面对两个不同控制参数值 $\alpha_2, \alpha_1$ 的信息进行比较,由式(7.204)可得

$$i(\alpha_2) - i(\alpha_1) = I_r(\alpha_2) - I_r(\alpha_1) - \sum_s \ln \frac{N_s(\alpha_2)}{N_s(\alpha_1)} \quad (7.205)$$

注意, $N^{-1}$ 相当于热力学的配分函数,根据

$$\ln N = -\beta \quad (7.206)$$

$\ln N$ 相当于自由能。显然,式(7.204)使热力学概念得到推广。当控制参数 $\alpha$ 在临界点取 $\alpha=0$ 时,我们通过求微商便可计算在临界点邻域中 $i_s$ 的变化。这就得到

$$\left. \frac{di_s}{d\alpha} \right|_{\alpha=0} = - \frac{N'_s(0)}{N_s(0)} \quad (7.207)$$

### 一个具体例子：单模激光

现在通过一个具体例子研究，接近非平衡相变时伺服模信息的行为。这方面的最简单的一个例子就是单模激光。因而，我们想使读者简短回顾一下所涉及的理论。激光由单一的场模构成，它具有无量纲振幅 $B(t)$ 。可以把 $|B|^2$ 解释为激光中出现的光子平均数目。激光还由它的激活原子组成，假定激活原子仅为二能级系统。用无量纲变量 $\alpha_\mu$ 表示原子的偶极矩。最后，每个原子的反转，也就是高能级的占据数和低能级占据数之差用 $d_\mu$ 表示。于是激光方程有如下形式：

$$\dot{b} = (-i\omega - \kappa)b - ig \sum_{\mu} \alpha_{\mu} \quad (7.208)$$

$$\dot{\alpha}_{\mu} = (-i\omega - \gamma)\alpha_{\mu} + igbd_{\mu} \quad (7.209)$$

$$\dot{d}_{\mu} = \gamma_{11}(d_0 - d_{\mu}) + 2ig(\alpha_{\mu}b^* - \alpha_{\mu}^*b) \quad (7.210)$$

其中， $\omega$ 是激光原子的光学跃迁频率， $\kappa$ 是场模的衰减常数，而 $g$ 是场模和原子偶极矩间的耦合常数。 $\gamma$ 是原子偶极矩的衰减常数。 $\gamma_{11}$ 是原子反转的弛豫常数。 $d_0$ 是由原子受激泵浦强度以及非相干弛豫过程给出的反转。

为消去常数 $\omega$ ，我们假定

$$b = Be^{-i\omega t}, \quad \alpha_{\mu} = A_{\mu}e^{-i\omega t} \quad (7.211)$$

其中， $B$ 和 $A_{\mu}$ 是含时变量。将式(7.211)代入式(7.208)一式(7.210)，容易得到方程

$$\dot{B} = -\kappa B - ig \sum_{\mu} A_{\mu} \quad (7.212)$$

$$\dot{A}_{\mu} = -\gamma A_{\mu} + ig b d_{\mu} \quad (7.213)$$

$$\dot{d}_\mu = \gamma_{11}(d_0 - d_\mu) + 2ig(A_\mu B^* - A_\mu^* B) \quad (7.214)$$

$d_0$ 等价于控制参数。大家知道, 反转 $d_0$ 小, 则不产生激光作用, 我们先研究这种情况。此种情况下, 场消失且原子的偶极矩也等于0, 而反转 $d_\mu$ 达到由泵浦过程和非相干弛豫过程描述的反转 $d_0$ 的值。因此, 我们得到

$$B_0 = 0, \quad A_{\mu,0} = 0, \quad d_{\mu,0} = d_0 \quad (7.215)$$

现在研究 $d_0$ 为何值, 解(7.215)才是稳定的, 而什么样的临界值 $d_0$ , 使解(7.215)失去稳定性。为此, 我们使用2.6节介绍的—般的稳定性分析, 即, 我们假定

$$B = B_0 + \delta B, \quad A_\mu = A_{\mu,0} + \delta A_\mu \quad (7.216)$$

将式(7.216)代入式(7.212)—式(7.214), 利用式(7.216)的关系, 我们在一级近似下(也就是取到线性项时)得到方程

$$\dot{\delta B} = -\kappa \delta B - ig \sum_{\mu} \delta A_{\mu} \quad (7.217)$$

$$\dot{\delta A}_{\mu} = -\gamma \delta A_{\mu} + ig \delta B d_0 \quad (7.218)$$

并能说明, 式(7.214)未产生 $d_\mu$ 的一阶变化。注意, 方程式(7.217)和(7.218)与2.6节的方程(2.72)具有相同形式, 在那里我们已详细阐明导出的线性方程组。为了求解方程(7.217)和(7.218), 我们通常假设

$$\delta B(t) = \delta B(0) e^{\lambda t} \quad (7.219)$$

$$\delta A_{\mu}(t) = \delta A_{\mu}(0) e^{\lambda t} \quad (7.220)$$

这里非含时振幅和 $\lambda$ 仍是未知量。将式(7.219)和式(7.220)代入式(7.217)及式(7.218), 我们便得到线性代数方程

$$\delta B(0)(\lambda + \kappa) = -ig \sum_s \delta A_s(0) \quad (7.221)$$

$$\delta A_s(0)(\lambda + \gamma) = ig \delta B(0) d_0 \quad (7.222)$$

相应于式 (7.221), 式 (7.222) 的行列式可写作

$$\begin{vmatrix} -\kappa - \lambda & -ig & -ig & \dots\dots -ig \\ igd_0 & -\gamma - \lambda & 0 & 0 \\ igd_0 & 0 & -\gamma - \lambda & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \end{vmatrix}$$

这个式子很容易加以计算, 从而得到

$$-(\kappa + \gamma)(\gamma + \lambda)^N + g^2 d_0 N(\gamma + \lambda)^{N-1} = 0 \quad (7.223)$$

其本征值 $\lambda$ ,

(i) 对于稳定模,  $\lambda_s$ 为:

$$\lambda_1 = -\frac{\kappa + \gamma}{2} - \frac{1}{2} \sqrt{(\kappa - \gamma)^2 + 4g^2 D_0}, D_0 = 4d_0 \quad (7.224)$$

(单模)

而对于 $s = 2, \dots, N$ 的情形

$$\lambda_s = -\gamma \quad (7.225)$$

为 $(N-1)$ 重的。

(ii) 对于不稳定模,  $\lambda_u$ 为:

$$\lambda_u = -\frac{\kappa + \gamma}{2} + \frac{1}{2} \sqrt{(\kappa + \gamma)^2 + 4g^2 D_0} \quad (7.226)$$

它也可以化为如下形式

$$\lambda = -\frac{\kappa + \gamma}{2} + \frac{1}{2} \sqrt{(\kappa - \lambda)^2 + 4g^2 D_0} \quad (7.227)$$

这样一来由不稳定条件便得出了临界反转 $D_c$ ，即

$$\lambda_s \geq 0 : D_c = \frac{\kappa \gamma}{g^2} \quad (7.228)$$

现在我们想研究一下，接近不稳点时何服模信息的改变。为此我们令

$$D_0 = D_c + \alpha' = \frac{\kappa \gamma}{g^2} + \alpha' \quad (7.229)$$

这里的 $\alpha'$ 起控制参数的作用。

按照通常的激光理论，我们现在假定，涨落力不依赖于控制参数。在该假定下，根据式(7.188)，则关系式(7.207)化简为

$$\left. \frac{dI_s}{d\alpha'} \right|_{\alpha'=0} = - \frac{\lambda'_s(\alpha'=0)}{\lambda_s(\alpha'=0)}, s=1, \dots, N \quad (7.230)$$

我们容易得到，对于 $s=2, \dots, N$ ，(7.230) = 0，同时

$$\frac{\lambda'_1(0)}{\lambda_1(0)} = -\frac{g^2}{(\kappa + \gamma)^2} \approx -\frac{g^2}{\gamma^2} = -\frac{\kappa}{\gamma} \cdot \frac{1}{D_c} \quad (7.231)$$

其中，我们专门研究了在

$$\kappa \ll \gamma \quad (7.232)$$

成立时的结果，这是大多数激光器中发生的情况。最后结果可表示为

$$\frac{dI_s}{d\alpha'} \cdot \alpha' = -\frac{\kappa}{\gamma} \frac{\alpha'}{D_c} \quad (\text{对于 } s=1) \quad (7.233)$$

$$\frac{dI_s}{d\alpha'} \cdot \alpha' = 0 \quad (\text{其它情况})$$

它告诉我们，随泵浦率 $\alpha'$ 增加，信息减少。只要我们回顾一下伺服模概率分布的一般形式(7.187)，就很容易解释这个结果。若 $Q_0$ 是常数，按照式(7.194)变换，去掉 $g$ 以后，则伺服模的信息依赖于高斯分布函数，当 $\lambda_0$ 的绝对值增大（即 $\alpha'$ 增加）时，高斯分布函数的宽度就变窄。然而较窄的分布函数意味着一种较大的确定性，同时信息减小。

因此，在伺服原理的一级近似中我们看到当较强烈地泵浦激光器时，激光原子的信息熵减少。

最后来建立式(7.229)引入的控制参数 $\alpha'$ 与序参量分布函数

$$f = N \exp [\alpha \xi^2 - \beta \xi^4] \quad (7.234)$$

中的控制参数 $\alpha$ 的关系式。为此，我们必须回顾一下在激光理论中与特殊情况(7.232)相应的某些结果，以及微观理论的某些基本方程。为了说明问题，我们暂时假定，场振幅 $b$ 是一个实变量，我们令它等价于式(7.234)中的 $\xi$ 。在激光理论中已经证明， $B \equiv \xi$ 遵从式(2.1)与式(2.3)型的朗之万方程。这种朗之万方程可以变换为福克-普朗克方程，它的稳定解就是刚才给出的式(7.233)，我们可以认为

$$\alpha = \lambda_0(\alpha')/Q_{00} \quad (7.235)$$

$Q_{00}$ 是作用于激光场的整个涨落力，而 $\lambda_0(\alpha')$ 正是

$$\lambda_0 = -\frac{\kappa + \gamma}{2} + \frac{1}{2} \sqrt{(\kappa + \gamma)^2 + 4g^2\alpha'} \quad (7.236)$$

其中，我们使用了式(7.226)和式(7.229)。可以方便地将式(7.236)展开为 $\alpha'$ 的一些幂次，其首项是

$$\lambda_0 \approx [g^2/(\kappa + \gamma)] \alpha' \quad (7.237)$$

对照一下式(7.235)和式(7.237)，我们得到了两种控制参数

$\alpha$  和  $\alpha'$  之间的关系式

$$\alpha \approx \frac{g^2}{(\kappa + \gamma)Q_{101}} \alpha' \quad (7.238)$$

根据激光理论，下式应当成立，

$$Q_{101} \approx 2\kappa N_{2s}/D_c \quad (7.239)$$

这里的  $N_{2s}$  是受激的原子数，我们可把式(7.238)化成形式

$$\alpha = \frac{1}{2N_{2s}} \alpha' \quad (7.240)$$

在这里曾用了关系式 (7.229) 和 (7.232)。我们可令

$$N_{2s} \approx N/2 \quad (7.241)$$

这是一个很好的近似。这里的  $N$  是激光激活原子的全部数目。

借助于式 (7.240)、式 (7.241) 我们可把式(7.231)变换为

$$\begin{aligned} \text{及} \quad \frac{d\lambda_1}{d\alpha} / \lambda_1 &= -\frac{\kappa}{\gamma} \frac{N}{D_c} \\ \frac{d\lambda_s}{d\alpha} &= 0 \quad (s \geq 2) \end{aligned} \quad (7.242)$$

因此，信息的变化 (7.230) 即变换为

$$\sum_{s=1}^N \left. \frac{dI_s}{d\alpha} \right|_{\alpha=0} = -\frac{\kappa}{\gamma} \frac{N}{D_c} \quad (7.243)$$

上式中的  $D_c/N$  可解释为每个原子的临界反转。由氦-氖激光的典型数据我们可以求得，对复激光强度而言

$$\sum_s \left. \frac{dI_s}{d\alpha} \right|_{\alpha=0} \approx 10^6, \quad \left. \frac{dI_n}{d\alpha} \right|_{\alpha=0} \approx 10^8 \quad (7.244)$$





## 第八章 拉氏乘子的直接确定

### 8.1 临界点之上和之下系统的信息熵

众所周知, 根据非平衡相变理论, 当系统处于远离临界点之上或之下时, 序参量的分布函数可以简化。这时, 临界点附近的临界涨落让位于序参量的高斯分布的正规涨落。另外, 在临界点之下的无序相, 序参量相当小, 使我们可令式

$$P_1(\xi_1 | \xi_0) = N \exp \left[ -(\xi_1 - g_1(\xi_0) / |\lambda_1|)^2 / Q_1 \right] \quad (8.1)$$

中  $\xi_0 = 0$ 。

在临界点之上, 系统通过对称性破缺, 定位于  $\xi_0 = \xi_0 \neq 0$  的态, 这时涨落仍然很小, 因而可在 (8.1) 中将  $\xi_0$  换为  $\xi_0$ 。因此, 联合分布函数成为高斯型。在本节我们将证明怎样可明显地确定拉氏乘子, 在其中只有第一、二两个矩, 而较高阶矩将被唯一确定。将如此计算和实验上确定的较高阶矩进行比较, 可以检验高斯近似的正确性。

让我们将分布函数写成通常的形式

$$f(\lambda, \underline{q}) = \exp \left[ -\lambda - \sum_j \lambda_j q_j - \sum_{jk} \lambda_{jk} q_j q_k \right] \quad (8.2)$$

其中的拉氏乘子是我们现在想要决定的。让我们将它与下式的分布函数比较：

$$\begin{aligned} \tilde{f}(\underline{q}) &= (2\pi)^{-\frac{n}{2}} |Q|^{-\frac{1}{2}} \\ &\times \exp \left[ -\frac{1}{2} (\underline{q} - \underline{m})^T Q^{-1} (\underline{q} - \underline{m}) \right] \end{aligned} \quad (8.3)$$

其中我们用了矢量记号。上标 $T$ 意指转置矢量， $Q = (Q_{jk})$ 是一矩阵。 $|Q|$ 是行列式。函数(8.3)是在 $\underline{q}$ 空间归一化的，这可简单加以证明。

还可相当简单地证明

$$m_j = \langle q_j \rangle \quad (8.4)$$

$$Q_{jk} = \langle (q_j - m_j)(q_k - m_k) \rangle \equiv \langle q_j q_k \rangle - \langle q_j \rangle \langle q_k \rangle \quad (8.5)$$

这意味着，一旦矩 $\langle q_j \rangle$ 和 $\langle q_j q_k \rangle$ 已知（或给定），我们就可决定(8.3)中的 $\underline{m}$ 和 $Q$ 。

因此，为了确定(8.2)中的拉氏乘子，我们必须一方面在 $\underline{m}$ 和 $Q$ 之间，另一方面在 $\lambda$ ， $\lambda_j$ ， $\lambda_{jk}$ 之间建立关系。这些步骤很简单，其结果是

$$\begin{aligned} \exp [-\lambda] &= (2\pi)^{-\frac{n}{2}} |Q|^{-\frac{1}{2}} \\ &\times \exp \left[ -\frac{1}{2} \underline{m}^T Q^{-1} \underline{m} \right] \end{aligned} \quad (8.6)$$

$$\lambda_{jk} = \frac{1}{2} [Q^{-1}]_{jk} \quad (8.7)$$

$$\lambda_j = -\frac{1}{2} \sum_k m_k \{ [Q^{-1}]_{kj} + [Q^{-1}]_{jk} \} \quad (8.8)$$

现在我们可以应用 (3.47)，并将其写为

$$i = \lambda + \sum_j \lambda_j \langle q_j \rangle + \sum_{jk} \lambda_{jk} \langle q_j q_k \rangle \quad (8.9)$$

将式 (8.6) — 式 (8.8) 代入式 (8.9)，并利用式 (8.4) 和式 (8.5)，我们得到

$$i = \frac{n}{2} \ln(2\pi) + \frac{1}{2} \ln |Q| + \frac{1}{2} \sum_{jk} [Q^{-1}]_{jk} (\langle q_j q_k \rangle - \langle q_j \rangle \langle q_k \rangle) \quad (8.10)$$

$$Q_{kj}$$

但矩阵关系

$$\sum_{jk} [Q^{-1}]_{jk} Q_{kj} = \sum_j (Q^{-1}Q)_{jj} = \sum_j 1 = n \quad (8.11)$$

使我们得到最后结果，即在约束 (8.4) 和 (8.5) 下与分布函数 (8.2) 相应的信息，

$$i = \frac{n}{2} \ln 2\pi + \frac{1}{2} \ln |Q| + \frac{n}{2} \quad (8.12)$$

显然，一阶矩  $\langle q_i \rangle$  已被丢掉。

与式 (7.61) 比较后可知，式 (8.12) 可以解释为宽度为  $1/\alpha_k$  的独立高斯分布的几个信息熵的求和，其中

$$\ln |Q| = - \sum_k \ln(2\alpha_k) \quad (8.13)$$

练习：检验以上叙述。

提示：归一化因子 (8.4) 和 (8.5)：通过变换  $\tilde{q} = A \tilde{\xi}$  将  $Q$  对角化。

关系式 (8.6) — (8.8)：去掉式 (8.3) 中括号后，将式 (8.2) 和式 (8.3) 逐项比较。

## 8.2. 临界点之上和之下拉氏乘子的直接确定

本节，我们要证明，怎样才能不计算积分而确定最大信息原理所需要的拉氏乘子。我们必须付出的代价是：我们将不得不假定，我们在处理具有一种特点的相变，例如象7.4节到7.9节中的软模相变。为此，分两步进行，并从下面形式的约束出发，

$$f_j = \langle q_j \rangle, \quad f_{jj'} = \langle q_j q_{j'} \rangle, \dots \quad (8.14)$$

其中，我们考虑到四阶矩。我们将变量 $q_i$ 变换为一组新变量并决定相应的变换。然后，考虑软模不稳定的特殊情形，证明如何样决定未知的拉氏乘子。同时，将推导出约束(8.14)的自治关系。通过这些条件，我们可以检验所考虑的系统是否象我们原先假设的那样经受了一个软模相变。

为了计算方便，我们引入以下记号：利用式(8.14)的 $f_{jj'}$ ，我们可以构成一矩阵，这个矩阵可以认为是矢量 $\underline{q}$ 与自身的直积

$$\langle \underline{\tilde{q}} \otimes \underline{\tilde{q}} \rangle = ( \langle q_i q_{j'} \rangle ) \quad (8.15)$$

以类似的方式，我们可以构成

$$\langle \underline{\tilde{q}} \otimes \underline{\tilde{q}} \otimes \underline{\tilde{q}} \rangle = ( \langle q_i q_{j'} q_{j''} \rangle ) \quad (8.16)$$

其中右边意味着我们必须取所有 $j, j', j''$ 的组合。四阶矩可以类似地处理。现在我们看一看微观方法，其中将态矢量表示为下面的叠加形式：

$$\underline{\tilde{q}} = \underline{\tilde{q}}_0 + \sum_k \xi_k \underline{\tilde{v}}_k \quad (8.17)$$

其中 $\underline{\tilde{q}}_0$ 是一个态矢量，这里令其为零，即

$$\underline{\tilde{q}}_0 = 0 \quad (8.18)$$

然而, 考虑  $\underline{\tilde{q}}_0 \neq 0$  的情形也不困难,  $\xi_k$  是随机变量, 而  $\underline{\tilde{v}}_k$  是常矢量, 它仍需以后确定, 我们引入伴随矢量

$$\underline{\tilde{v}} = (v_{k1}^*, v_{k2}^*, \dots, v_{kn}^*) \quad (8.19)$$

并假定正交关系

$$\underline{\tilde{v}}_k \cdot \underline{v}_k = \delta_{kk}, \quad (8.20)$$

对式 (8.17) 取平均, 我们立即得到

$$\langle \underline{\tilde{q}} \rangle = \sum_k \langle \xi_k \rangle \underline{\tilde{v}}_k \quad (8.21)$$

用伴随矢量  $\underline{\tilde{v}}_k$  乘以式 (8.21) 得到

$$\underline{\tilde{v}}_k \cdot \langle \underline{\tilde{q}} \rangle = \langle \xi_k \rangle \quad (8.22)$$

其中用了正交关系 (8.20)。引入缩写

$$C_k \equiv \underline{\tilde{v}}_k \cdot \langle \underline{\tilde{q}} \rangle \quad (8.23)$$

我们可将 (8.22) 写为

$$\langle \xi_k \rangle = C_k \quad (8.24)$$

以适当的方式用伴随矢量  $\underline{\tilde{v}}_k$  乘以式 (8.15), 并利用式 (8.20), 我们立即得到

$$\langle \xi_k \xi_{k'} \rangle = \langle \underline{\tilde{v}}_k \cdot \underline{\tilde{q}} \cdot \underline{\tilde{v}}_{k'} \cdot \underline{q} \rangle \equiv \bar{C}_{kk'}, \quad (8.25)$$

其中最后一个等式是  $C_{kk}$  的定义, 而中间部分更明显地写为

$$\langle \underline{\tilde{v}}_k \cdot \underline{\tilde{q}} \cdot \underline{\tilde{v}}_{k'} \cdot \underline{q} \rangle = \sum_i v_{ki}^* q_i \cdot \sum_j v_{k'j}^* q_j \quad (8.26)$$

以非常相同的方式，我们发现三、四阶矩为

$$\langle \xi_k \xi_k, \xi_k, \dots \rangle = \langle \tilde{v}_k q \cdot \tilde{v}_k q \cdot \tilde{v}_k \dots q \rangle = C_{kk, k, \dots} \quad (8.27)$$

以及

$$\begin{aligned} \langle \xi_k \xi_k, \xi_k, \dots, \xi_k, \dots \rangle &= \langle \prod_{v=1}^4 \tilde{v}_k q \rangle \\ &= C_{kk, k, \dots, k, \dots} \end{aligned} \quad (8.28)$$

到现在为止，我们已将关于 $q$ 的旧矩变换为关于 $\xi_k$ 的新矩。下面，我们假定式(8.14)的旧矩由实验给出，而新矩

$$\langle \xi_k \xi_k, \dots \rangle \quad (8.29)$$

以及矢量 $\tilde{v}_k$ 仍需决定。为此，我们假定我们所处理的现象由式(8.14)的矩所描述，且它具有软模不稳定性。为继续进行研究，我们提醒读者注意一下微观方法，在阈值之下，系统可由关于所有变量的高斯分布来很好的近似。这时，只处理一、二阶矩就足够了。为了找出合适的模矢量，我们将

$$\langle \tilde{q} \circ \tilde{q} \rangle \quad (8.30)$$

对角化，借此，我们得到代表模的二阶矩的实本征值。现在我们知道，当我们接近临界点时，由于软模不稳定性，一个模将显示出临界涨落，因此，可以预期一个本征值将变得很大，所有其他的将仍然很小。这个很大的本征值将表示不稳模，而其他则仍属于稳定模。这样，我们可区分出软模和稳定模。

然而，与数学的线性化相反，在那里软模的本征值将发散，而在这里的情形中，本征值将仍然有限，因为在式(8.30)中我们处理的是实验提供的资料，其中非线性将阻碍本征值变成无穷大。

现在我们区分出稳定模和不稳定模

$$\underline{v}_j \rightarrow \underline{V}_0, \underline{V}_j \quad (8.31)$$

进一步, 我们考虑典型的软模相变, 由于对称性的原因, 其中方程

$$\langle \xi_0 \rangle = 0 \quad (8.32)$$

成立。除此之外, 我们以下还假设

$$\langle \xi_0 \xi_j \rangle = 0 \quad (8.33)$$

成立。为了更明显起见, 我们假定进行平均的联合概率分布函数形式为

$$P(\xi_0, \xi_j) = f(\xi_0) \prod_j P_j(\xi_j | \xi_0) \quad (8.34)$$

按照第六章, 它可从微观方法或宏观方法得到。其中的  $f$  是序参量的分布函数, 而  $P_j$  是伺服模幅度  $\xi_j$  的条件概率, 它由下式给出:

$$P_j = N_j \exp [ - (\xi_j - g_j(\xi_0))^2 / Q_j ] \quad (8.35)$$

以下我们考虑一特殊情形(这一情形是经常得到满足的)。即  $g_j$  是序参量的偶函数, 这种情况通过变换

$$\xi_j = \bar{\xi}_j + g_j(\xi_0) \quad (8.36)$$

很易计算出平均值, 例如

$$\langle \xi_j \rangle = \langle \bar{\xi}_j + g_j(\xi_0) \rangle = \langle \bar{\xi}_j \rangle + \langle g_j(\xi_0) \rangle \quad (8.37)$$

由于式 (8.35), 关系式

$$\langle \xi_0 \rangle = 0 \quad (8.38)$$

成立。以下我们采用简写, 将  $\xi_j$  用  $s$  来代替, 即

$$\xi_j \rightarrow s \quad (8.39)$$

用新的记号可将式 (8.37) 写为

$$\langle s \rangle = \langle \bar{s} + g_s \rangle = \langle \bar{s} \rangle + \langle g_s \rangle \quad (8.40)$$

现在，我们用特定分布函数的形式，即 (8.34) 和式 (8.35) 来研究表示式 (8.24)、(8.25)、(8.27) 和式 (8.28)。从式 (8.32)、(8.23) 和式 (8.24)，我们很容易得到

$$\bar{\psi}_u \langle q \rangle = 0 \quad (8.41)$$

(此式现在是对  $\bar{q}$  的平均值的约束) 以及

$$\langle \xi_u \rangle = \langle g_u(\xi_u) \rangle = C_u \quad (8.42)$$

如我们将要证明的那样，在合理的假设下，式 (8.42) 将确定  $g_u$ 。据式 (8.25)，令  $k=u$ ， $k'=u$ ，我们得到

$$\langle \xi_u^2 \rangle = C_{uu} \quad (8.43)$$

令  $k=u$ ， $k'=s$ ，利用式 (8.34) 和式 (8.35) 我们得到

$$\langle \xi_u \xi_s \rangle = 0 = C_{us} \quad (8.44)$$

这是对  $C_{us}$  的一个约束。最后令  $k=s$ ， $k'=s'$ ，得到

$$\langle ss' \rangle = \langle \bar{s}^2 \rangle \delta_{ss'} + \langle g_s g_{s'} \rangle = C_{ss'} \quad (8.45)$$

利用

$$\langle \bar{s}^2 \rangle = Q_s \quad (8.46)$$

式 (8.45) 可写为

$$Q_s \delta_{ss'} + \langle g_s g_{s'} \rangle = C_{ss'} \quad (8.47)$$

以后我们将看到，这个方程可用来确定  $Q_s$ ，现在我们研究关系式

$$\langle \xi_k \xi_{k'} \xi_{k''} \rangle = C_{kk'k''} \quad (8.48)$$

导致的后果。

(i) 首先我们令

$$k=u, \quad k'=s', \quad k''=s'' \quad (8.49)$$



以上关系中，它们可以置换。我们易得到

$$C_{\alpha\beta\gamma\delta} = 0 \quad (8.50)$$

以及表示式

$$\langle g_{\alpha} g_{\beta} g_{\gamma} \rangle \quad (8.60)$$

此式暂时还不能简化。将式 (8.57) 一式 (8.60) 各项以及它们的适当的置换集中起来, 我们得到以下关系

$$Q_{\alpha} \delta_{\alpha\beta} \langle g_{\gamma} \rangle + Q_{\beta} \delta_{\alpha\beta} \langle g_{\alpha} \rangle + Q_{\gamma} \delta_{\alpha\beta} \langle g_{\beta} \rangle + \langle g_{\alpha} g_{\beta} g_{\gamma} \rangle = C_{\alpha\beta\gamma}, \quad (8.61)$$

这可由式 (8.48) 得到。为了完整起见, 我们写出由

$$\langle \xi_{\alpha} \xi_{\beta} \xi_{\gamma} \xi_{\delta} \rangle = Q_{\alpha\beta\gamma\delta} \quad (8.62)$$

所产生的结果, 但性急的读者可以立即跳到以下的最终结果式 (8.83)。

根据指标结合  $k, k', k'', k'''$ , 我们区分出各种情形。

情况 (i)

$$u, u', u'', u''' \quad (8.63)$$

我们得到

$$\langle \xi_u^4 \rangle = C_{uuuu} \quad (8.64)$$

情况 (ii)

$$u, u, u, s'' \quad (8.65)$$

我们得到

$$C_{uuus} = 0 \quad (8.66)$$

情况 (iii)

$$u, u, s'', s''' \quad (8.67)$$

对式 (8.36) 进行分解, 得到

$$\langle \xi_u^2 (\bar{s}'' + g_{\alpha}) (\bar{s}''' + g_{\beta}) \rangle \quad (8.68)$$

考虑每一项, 我们得到

$$\langle \xi_u^2 \bar{s}'' \bar{s}''' \rangle = \langle \xi_u^2 \rangle Q_{\dots} \delta_{\dots} \quad (8.69)$$

$$\langle \xi_u^2 \bar{s}'' g_{\dots} \rangle = 0 \quad (8.70)$$

以及

$$\langle \xi_u^2 g_{\dots} g_{\dots} \rangle \neq 0 \quad (8.71)$$

情况 (iv)

$$u, s', s'', s''' \quad (8.72)$$

根据对称性, 我们容易得到

$$C_{u, \dots, \dots, \dots} = 0 \quad (8.73)$$

情况 (v)

$$s, s', s'', s''' \quad (8.74)$$

再次进行适当的分解, 我们必须考虑表示式

$$\langle (\bar{s} + g_s)(\bar{s}' + g_{s'}) (\bar{s}'' + g_{s''}) (\bar{s}''' + g_{s'''}) \rangle \quad (8.75)$$

然后, 可考虑单个项。为了计算

$$\langle \bar{s} \bar{s}' \bar{s}'' \bar{s}''' \rangle \quad (8.76)$$

我们必须考虑那些给出非零平均值的指标  $s$  的组合, 有一些组合, 其中两部分指标重合, 即我们必须使关系式

$$\begin{aligned} s &= s', & s'' &= s''' \\ s &= s'', & s' &= s''' \\ s &= s''', & s' &= s'' \end{aligned} \quad (8.77)$$

满足。代替式 (8.76), 我们容易得到表达式

$$\langle \bar{s} \bar{s}' \bar{s}'' \bar{s}''' \rangle = \langle \bar{s}^2 \rangle \langle \bar{s}''^2 \rangle \delta_{s, s''} \delta_{s', s'''} \dots$$

$$\begin{aligned}
& + \langle \bar{s}^2 \rangle \langle \bar{s}'^2 \rangle \delta_{s,s'} \delta_{s',s''} \dots \\
& + \langle \bar{s}^2 \rangle \langle \bar{s}'^2 \rangle \delta_{s,s'} \delta_{s',s''} \dots \quad (8.78)
\end{aligned}$$

要考虑的下一项具有

$$\langle \bar{s} \bar{s}' \bar{s}'' g_{s'} \dots \rangle = 0 \quad (8.79)$$

的形式，且其 $g_s$ 在不同位置可代替 $\bar{s}$ 。另一项是

$$\begin{aligned}
\langle \bar{s} \bar{s}' g_{s'} g_{s'} \dots \rangle & = \langle \bar{s}^2 \rangle \delta_{s,s'} \langle g_{s'} g_{s'} \dots \rangle \\
& = Q_s \delta_{s,s'} \langle g_{s'} g_{s'} \dots \rangle \quad (8.80)
\end{aligned}$$

其中，为求出非零结果必须选择 $\bar{s} = \bar{s}'$ 。通过变换指标，我们还求出了这种形式的其他项。对于公式

$$\langle \bar{s} g_{s'} g_{s'} g_{s'} \dots \rangle = 0 \quad (8.81)$$

我们得到已指出的结果。最后，出现了

$$\langle g_s g_{s'} g_{s'} g_{s'} \dots \rangle \quad (8.82)$$

将式(8.64)到式(8.82)的所有项集合起来，我们得到最终结果

$$\begin{aligned}
& Q_s Q_{s'} \delta_{s,s'} \delta_{s',s''} \dots + Q_s Q_{s'} \delta_{s,s'} \delta_{s',s''} \dots \\
& + Q_s Q_{s'} \delta_{s,s'} \delta_{s',s''} \dots \\
& + Q_s [\delta_{s,s'} \langle g_{s'} g_{s'} \dots \rangle + \delta_{s,s'} \langle g_s g_{s'} \dots \rangle \\
& + \delta_{s,s'} \langle g_{s'} g_{s'} \dots \rangle] \\
& + Q_{s'} [\delta_{s',s''} \langle g_s g_{s'} \dots \rangle + \delta_{s',s''} \langle g_s g_{s'} \dots \rangle] \\
& + Q_{s'} \delta_{s',s''} \dots \langle g_s g_{s'} \dots \rangle \\
& + \langle g_s g_{s'} g_{s'} g_{s'} \dots \rangle = C_{s,s',s'',\dots} \quad (8.83)
\end{aligned}$$

此结果来自式(8.62)，其中所有指标皆为s型。

为了确定出现在分布函数(8.34)及(8.35)中的未知量拉氏乘子，让我们看一看我们的结果。为此，将我们的分析限于阈值之上不太远的区域，在这种情况下，我们可用打头项来近似表

示 $g$ , 即

$$g_s(\xi_s) \approx \alpha_s \xi_s^2 \quad (8.84)$$

我们进一步做一些关于序参量分布函数形式的假设。即

$$f(\xi_s) = N \exp [\alpha \xi_s^2 - \beta \xi_s^4] \quad (8.85)$$

这与我们以前从微观和宏观理论得到的结果一致。因此, 我们看到, 现在下列常数需要确定:  $\alpha$ ,  $\beta$ ,  $Q_s$ ,  $a_s$ 。  $\alpha$ ,  $\beta$  和归一化常数通过式 (8.43) 和式 (8.64) 来确定, 即通过关系式

$$\langle \xi_s^2 \rangle = C_{ss} \quad (8.86)$$

$$\langle \xi_s^4 \rangle = C_{ssss} \quad (8.87)$$

来确定。  $a_s$  可由式 (8.42), 即

$$a_s \langle \xi_s^2 \rangle = C_s \quad (8.88)$$

来确定。最后,  $Q_s$  可通过式 (8.47) (取其中  $s = s'$ ) 和关系式

$$\langle g_s^2 \rangle = a_s^2 \langle \xi_s^4 \rangle \quad (8.89)$$

来确定, 上式来自式 (8.84)。因此,  $Q_s$  由下式给出:

$$Q_s = C_{ss} - C_s^2 \langle \xi_s^2 \rangle^2 / \langle \xi_s^4 \rangle \quad (8.90)$$

这样, 我们能够确定所有常数, 读者可以检验, 这样, 我们便确定了出现在特殊形式 (8.34)、(8.35)、(8.84)、和式 (8.85) 中的拉氏乘子。应该注意到, 系统被过分地确定了, 这意味着方程数目多于未知量数目。因此, 上面建立的剩余的关系 (观测量  $C$  出现于其中) 是单软模非平衡相变出现的自洽条件。

将以上结果推广于更复杂的情形也是不困难的, 例如, 当式 (8.84) 还含有奇变量和 (或者)  $\xi_s$  的高次项时, 或有几个序参量出现的情况。



# 第九章 随机过程的无偏模型 化：怎样推测路径积分， 福克-普朗克方程和朗 之万-伊藤方程

---

## 9.1 一维状态矢量

前面的章节已经指出，最大信息原理怎样使我们通过对约束的适当选择，推导出非平衡系统的分布函数；在那些章节中，我们的注意力集中于定态分布函数上，而在本章，我们打算进一步研究时间相依分布函数。由吉尼斯（Jaynes）提出的最大定标原理可作为这种研究的一个合适出发点。他通过一个类熵泛函在所在的时-空区域上求极大值，推广了最大熵原理，得到了最大定标原理。吉尼斯把他的这个原理用来推导不可逆过程的一些重要热力学公式，以研究弛豫过程。在这里，我们想向读者指出，通过对约束的适当选择，就可以处理任何接近或远离平衡的物理系统，或其它系统。如果假定起主导作用的过程是连续马尔可夫过程，用我们这种方法得出的表述形式和公式体系将是特别华丽

的。在这种情况下,我们发现了一种十分简捷的方法,用这种方法可以求得路径积分和福克-普朗克方程,其漂移和扩散系数可显式地加以确定。这样,我们的这种方法就为从实验数据推算基本过程的决定性力和涨落力提供了一种切实可行的工具。

考虑一时间序列  $t_0, t_1, \dots, t_N$ , 并在这些时刻对体系进行观测, 在  $t_i$  时刻所测得的态矢量之值为  $\underline{q}_i$ 。我们希望对联合分布函数

$$P_N = P(\underline{q}_N, t_N; \underline{q}_{N-1}, t_{N-1}; \dots; \underline{q}_0, t_0) \quad (9.1)$$

进行无偏估算。为此,我们将信息量  $i$ , 或吉尼斯公式中的标度

$$i = - \int Dq P \ln P \quad (9.2)$$

在给定约束条件下对  $P$  求极大值。式中之  $Dq$  是由所有态矢量  $\underline{q}_i$  所张成的态空间上的积分体积元。

众所周知,求解这类问题的关键在于适当地选择约束。如果假定导致式 (9.1) 的基础过程是一马尔可夫过程,则该选择将大为简化,在该假设下,式 (9.1) 可分解为。

$$P_N = \prod_{i=0}^{N-1} P(\underline{q}_{i+1}, t_{i+1} / \underline{q}_i, t_i) \cdot P_0(\underline{q}_0, t_0) \quad (9.3)$$

为了书写简便,我们将略去时间  $t_i$ , 因而式 (9.3) 中的各单个因子有如下形式,

$$P(\underline{q}_{i+1} | \underline{q}_i) \quad (9.4)$$

式 (9.4) 表示条件概率。在本章的第一部分中,我们将讨论平稳随机过程。

为了揭示出其本质,我们将采用单变量  $q$  代替状态矢量加以讨论。现在引入约束,

$$f_1 = \langle q_{i+1} \rangle q_i \quad (9.5)$$



及

$$f_2 = \langle q_{i+1}^2 \rangle q_i \quad (9.6)$$

其定义为:

$$f_2 = \int P(q_{i+1}/q_i) q_{i+1}^2 dq_i \quad (9.7)$$

按这种方式我们讨论的是, [在时刻  $t_i$  测得值为  $q_i$  的条件下,  $q_i$  在  $t_{i+1}$  时刻的一阶和二阶矩。把最大信息原理运用于式 (9.4), 并考虑到约束 (9.5) 与 (9.6), 立即可得:

$$P(q_{i+1} | q_i) = \exp \{ \lambda + \lambda_1 q_{i+1} + \lambda_2 q_{i+1}^2 \} \quad (9.8)$$

注意到  $f_1$  和  $f_2$  是  $q_i$  的函数, 所以, 一般而论, 上式中的  $\lambda_1, \lambda_2, \lambda_3$  至少也将成为  $q_i$  的函数。考虑到我们取极限  $\tau \rightarrow 0$ , 下面将用  $q(i+\tau)$  代替  $q(i+1)$ 。现在, 我们将把一通常的归一化条件加诸于式 (9.8) 上, 即:

$$\int P(q_{i+\tau} | q_i) dq_{i+\tau} = 1 \quad (9.9)$$

同时, 还必须要求

$$\tau \rightarrow 0: \quad P(q_{i+\tau} | q_i) \rightarrow \delta(q_{i+\tau} - q_i) \quad (9.10)$$

上式右端的  $\delta$  函数将按一种特定的方式加以解释, 即, 若我们对  $q_{i+\tau}$  求积分,  $P$  的行为与一  $\delta$  函数相似。

为了进一步探讨条件式 (9.9) 及 (9.10), 我们把式 (9.8) 右边改写成:

$$\exp \left\{ - \frac{1}{2} \frac{\lambda_1^2}{\lambda_2} \left( q_{i+\tau} - \frac{\lambda_1}{\lambda_2} \right)^2 - \frac{\lambda_1^2}{4 \lambda_2} + \lambda \right\} \quad (9.11)$$

因为

$$\int \exp \{ - |\lambda_2| q^2 \} dq = \sqrt{\pi / |\lambda_2|} \quad (9.12)$$

归一化条件 (9.9) 可写成

$$\exp \{ \lambda - \lambda_1^2 / (4 |\lambda_2|) \} = \sqrt{|\lambda_2|} / \pi \quad (9.13)$$

这就使  $P$  具有如下形式:

$$P(q_{i+\tau} | q_i) = \sqrt{|\lambda_2(q_i)|} \pi \\ \cdot \exp \left\{ - |\lambda_2| \left( q_{i+\tau} - \frac{\lambda_1}{2 |\lambda_2|} \right)^2 \right\} \quad (9.14)$$

为了使  $\tau \rightarrow 0$  时, 式 (9.14) 变成  $\delta$  函数, 还必须要求

$$\tau \rightarrow 0, \quad |\lambda_2| = G/\tau \quad (9.15)$$

(这是  $\tau$  的另一个函数, 当  $\tau \rightarrow 0$  时它趋于零。这恰恰意味着另一个时间标度的存在, 这个标度便是时间本身的函数。)

从式 (9.10) 中, 我们可进一步导出, 当  $\tau \rightarrow 0$  时,

$$q_{i+\tau} - \lambda_1 / (2 |\lambda_2|) \rightarrow q_{i+\tau} - q_i \quad (9.16)$$

或者, 换句话说,

$$\tau \rightarrow 0, \quad \lambda_1 / (2 |\lambda_2|) \rightarrow q_i \quad (9.17)$$

对有限而微小的时间  $\tau$ , 显然可把式 (9.17) 推广为

$$\lambda_1 / (2 |\lambda_2|) = q_i + \tau K(q_i) + \tau^2 H(q_i) + \dots \quad (9.18)$$

显而易见, 我们已可以说, 式 (9.17) 左边的表达式, 当  $\tau$  足够小时, 应表现为  $\tau$  的一种解析函数。考虑到式 (9.14) 和式 (9.18), 我们便得到本节的一个主要结果, 即在短时间极限情况下, 条件概率为:

$$P(q_{i+\tau} | q_i) = \sqrt{|\lambda_2(q_i)|} \pi \\ \cdot \exp \left\{ - \frac{G}{\tau} (q_{i+\tau} - q_i - \tau K(q_i))^2 \right\} \quad (9.19)$$

注意,  $G$  可能仍为  $q_i$  的函数。

## 9.2 对多维状态矢量的推广

在本节中, 我们想在平稳马尔可夫过程假设的前提下, 对出现在式 (9.3) 中的条件概率进行无偏估算。我们把在时刻  $i + \tau$  和  $i$  的状态矢量分量表示为:

$$q_1(i + \tau); q_k(i) \quad (9.20)$$

为了把单变量之第一矩及第二矩的表示式 (9.5) 和式 (9.6) 加以推广, 我们引入下列第一和第二条件矩:

$$f_{1,1} = \langle q_1(i + \tau) \rangle_{q(i)} \quad (9.21)$$

$$f_{2,1,k} = \langle q_1(i + \tau) q_k(i + \tau) \rangle_{q(i)} \quad (9.22)$$

把最大熵原理用于条件概率, 我们得到:

$$P(\underline{q}(i + \tau) | \underline{q}(i)) = \exp \left\{ \lambda + \sum_i \lambda_{1i} q_{1i}(i + \tau) + \sum_{k,i} \lambda_{k1} q_k(i + \tau) q_{1i}(i + \tau) \right\} \quad (9.23)$$

其中, 拉格朗日乘子仍是  $\underline{q}(i)$  的函数。在下面的讨论中, 我们将假设问题是非退化的。亦即

$$\text{Det} \lambda_{k1} \neq 0 \quad (9.24)$$

通过简单的代数运算, 可以将式 (9.23) 改写成:

$$P = \exp \left\{ \tilde{\lambda} + \sum_{i=1}^n (q_{1i}(i + \tau) - h_{1i}) \lambda_{1i} (q_{1i}(i + \tau) - h_{1i}) \right\} \quad (9.25)$$

其中,  $h_{1i}$  定义为

$$h_{1i} = \sum_m A_{1m} \lambda_m(\underline{q}(i)) \quad (9.26)$$

我们用  $A$  表示由元素  $\lambda_{k1}$  组成的矩阵。式 (9.26) 中的  $A$  定义为

$$A = 1/2 A^{-1} \quad (9.27)$$

注意到矩阵  $\Delta$  是对称的, 而且出现在式 (9.25) 的  $\tilde{\lambda}$  为

$$\tilde{\lambda} = \lambda - \sum_{k=1}^n h_k \lambda_{k1} h_k \quad (9.28)$$

为了把条件式 (9.10) 推广, 要求在对  $q_{i+1}(\tau)$  进行积分时,  $P$  对  $\tau \rightarrow 0$  表现为一  $\delta$  的函数

$$P \rightarrow \delta(q(i+\tau) - q(i)) \quad (9.29)$$

如果令

$$\lambda_{k1} = -\frac{1}{\tau} G_{k1}(q(i)) \quad (9.30)$$

便可体现出此  $\delta$  函数的奇异性。为了推广式 (9.18), 我们可将  $h_k$  展成  $\tau$  的幂级数, 并保留第一项:

$$h_k = q_k(i) + \tau K_k(q(i)) + \dots \quad (9.31)$$

合并式 (9.25)、(9.28)、(9.30) 及 (9.31), 我们就可把条件概率改写成如下形式:

$$P = \exp\{\tilde{\lambda}(q_i)\} \exp\left\{-\frac{1}{\tau} \sum_{k=1}^n (q_k(i+\tau) - q_k(i) - \tau K_k) G_{k1}(q_1(i+t) - q_1(i) - \tau K_1)\right\} \quad (9.32)$$

将式 (9.32) 中的第一个因子缩写为  $N(q_i)$  式 (9.32) 的结果可表示为如下形式:

$$P = N(q_i) \exp L_i \quad (9.33)$$

把式 (9.30) 代入式 (9.3), 其中  $P_0$  为一可设定的分布函数, 若需要, 可通过对  $q_i$  之矩设定某一数值而加以估算。这样一来, 即可得出我们期望的最终结果, 即确定联合分布函数式 (9.1)。

在实际应用中, 人们主要感兴趣的是条件概率, 即是在  $T=0$ , 且状态矢量  $q$  已知的条件下,  $T$  时刻  $q$  值的分布函数。此概率

可采用在各中间时刻对一切的 $\underline{q}$ 求积分的办法求得。令

$$Nr = T \quad (9.34)$$

借助式(9.33)和式(9.32)确定 $L$ ，我们把所期望的结果写为：

$$P(\underline{q}(T) | \underline{q}(0)) = \int Dq \exp \left\{ \sum_{i=0}^{N-1} L_i \right\} \quad (9.35)$$

其中， $N(\underline{q}_i)$ 包含在 $Dq$ 之中。

在 $r \rightarrow 0$ 的极限情况下，对 $i$ 的求和转换为积分。式(9.35)就是人们熟知的 $P$ 的路径积分表达式。从2.5节对福克-普朗克方程的路径积分解的阐述可知，式(9.35)，(9.33)及(9.32)显然是漂移系数为 $K_i$ 的一个福克-普朗克方程的解。其扩散矩阵由矩阵元为 $G_{ij}$ 的矩阵之逆给定。

在我们的论述中，只要让读者知道，式(9.35)是具有刚才提及的漂移及扩散系数的一个福克-普朗克方程的解就足够了(与9.4节比较)。

在结束本节时应该指出，这种方法很容易推广至时间相依的连续马尔可夫过程。在这种情形下，一般而论，必须使用如下代换：

$$\underline{K}(\underline{q}(i)) \rightarrow \underline{K}(\underline{q}(i); t_i) \quad (9.36)$$

$$G_{ij}(\underline{q}(i)) \rightarrow G_{ij}(\underline{q}(i); t_i) \quad (9.37)$$

### 9.3 作为约束的关联函数

在前几节中，我们使用了条件矩(当 $q_i$ 已被量度时，变量 $q_{i+1}$ 之矩)的概念。在众多的实际应用中，我们也许无法得知 $q_i$ ，但我们很可能有办法测得如下类型的关联函数：

$$\langle q_{i+1} \cdot q_i \rangle \quad (9.38)$$

根据

$$\langle q_{i+\tau}, q_i \rangle = \int \int dq_{i+\tau} dq_i q_{i+\tau} q_i P(q_{i+\tau}, q_i) \quad (9.39)$$

它可被解释成联合分布函数的均值。

我们的任务就是对出现在式 (9.39) 中的联合分布函数, 而不是对以前所研究过的条件概率作出无偏估计。为了简化这个任务, 假设测量是在定态条件下进行的。在这种情形下, 我们可把联合概率表示成条件概率与定态概率分布函数的乘积:

$$P(q_{i+\tau}, q_i) = P(q_{i+\tau} | q_i) P_{st}(q_i) \quad (9.40)$$

因为, 我们没有关于  $q_i$  的精确资料, 仅能通过关联函数获得一点信息, 所以, 现在需要更多的约束。我们将要使用的一些约束是:

$$\langle q_i^\mu \rangle, \quad \mu = 1, \dots, m \quad (9.41)$$

$$\langle q_{i+\tau} q_i^\nu \rangle, \quad \nu = 0, \dots, n \quad (9.42)$$

$$\langle q_{i+\tau}^2 q_i^\kappa \rangle, \quad \kappa = 0, \dots, k \quad (9.43)$$

这里, 数字  $m, n$ , 和  $k$  遵从将在下面讨论的某些自洽条件, 而且必须经过选择, 这样, 数据的合理收敛性才可能得以保证。我们预计, 在接近非平衡相变点时,  $n$  和  $k$  可能会限于  $n = k = 3$ 。本节与前面的章节有一个根本的差别, 应该注意现在的乘子  $\lambda_i$  是与  $q_i$  无关的。由于最大信息原理, 或更确切地说, 最大定标原理涉及到两个量  $q_{i+\tau}$  和  $q_i$ , 所以最大信息原理派生出了

$$P_{st} = \exp \left\{ \lambda_{st} + \sum_{\mu=1}^m \lambda_\mu q_i^\mu \right\} \quad (9.44)$$

而最大定标原理则为我们提供了下列联合概率表达式:

$$P(q_{i+\tau}, q_i) = \exp \{ \lambda_0 + a(q_i) + q_{i+\tau} b(q_i) + q_{i+\tau}^2 \\ \times C(q_i) \} \quad (9.45)$$

这里，我们使用了缩写符号：

$$a = \sum_{\mu=1}^m \lambda_{\mu}^{(1)} q_i^{\mu} \quad (9.46)$$

$$b = \sum_{\nu=0}^n \lambda_{\nu}^{(2)} q_i^{\nu} \quad (9.47)$$

$$C(q_i) = \sum_{\kappa=0}^k \lambda_{\kappa}^{(3)} q_i^{\kappa} \quad (9.48)$$

根据式 (9.40)，我们可以确定条件概率，其表达式为：

$$P(q_{i+\tau} | q_i) = \exp \{ \hat{\lambda} + \hat{a}(q_i) + q_{i+\tau} b(q_i) + q_{i+\tau}^2 c(q_i) \} \quad (9.49)$$

其中，又使用了缩写符号

$$\hat{\lambda} = \lambda_0 - \lambda_{11} \quad (9.50)$$

和

$$\hat{a} = a - \sum_{\mu=1}^m \lambda_{\mu} q_i^{\mu} \quad (9.51)$$

按这种方式，我们发现了条件概率的一种形式，可直接类推至式 (9.8)。从现在起，我们借助下面的等价代换，就可采用与 9.1 节完全相同的方法来处理问题了：

$$\lambda(q_i) \Longleftrightarrow \hat{\lambda} + \hat{a}(q_i) \quad (9.52)$$

$$\lambda_1(q_i) \Longleftrightarrow b(q_i) \quad (9.53)$$

$$\lambda_2(q_i) \Longleftrightarrow c(q_i) \quad (9.54)$$

应该注意到，由于  $\lambda$  与  $\lambda_1$  和  $\lambda_2$  有关，式 (9.52) — (9.54) 右端所涉及的一些自洽的要求必须满足（参见式 (9.13)）。若  $\lambda_2$  是  $q_i$  的函数，至少在原则上，式 (9.13) 要求  $q_i$  的无限高次项，或者说要求无限多个矩作为约束条件，则  $\lambda$  对  $\lambda_1$  和  $\lambda_2$  的这种依赖

性就会多少变得复杂起来。

然而在一些实际事例中，可望有很少几个矩阵作为约束就足够了。另一方面，若 $\lambda_2$ 与 $q_1$ 无关，则它为一 $m=2n$ 阶的多项式。现在，如何才能把本节中与给定时刻的单变量有关的结果，推广至给定时刻状态矢量的范畴，结论已经十分明显，为了节省篇幅，我们将不再显式地写出相应的结果，

本节已经指出，如何从简单的矩或作为约束的关联函数，推演出统计物理的基本表达式，即联合概率及路径积分，它们又可被解释为是基础福克-普朗克方程的解，这样建立的方法使我们可通过对实验数据的观测，对潜在基础过程的机制加以推测。即是说，我们可以确定作为决定性力的漂移系数和作为涨落力量度的扩散系数。在实际事例中，并非总要假设基础过程是马尔可夫过程或连续的马尔可夫过程，但在这种情况下，我们的方法却体现了对基础的决定性力和涨落力的一种估计。十分清楚，我们建立的方法可以从不同角度加以推广，例如，可用以讨论各高阶矩或关联函数，也可以对较长时间间隔的联合分布函数或条件概率进行估计，但在这种情形下，却至少会失去这种方法的华丽性。

#### 9.4 相应于短时传播子(9.35)的福克-普朗克方程

相应于式(9.35)的福克-普朗克方程的推导程序是众所周知的。但为了以显式表明哪类福克-普朗克方程(伊藤或斯特拉特洛维奇方程)适合于式(9.32)，我们将扼要阐明其主要步骤。由于福克-普朗克方程涉及到无穷小的时间间隔，因此，只需考虑两个相邻的时刻 $i$ 和 $(i+\tau)$ 就足够了。

相应的概率分布函数之间的关系为



$$P(\underline{q}_{i+\tau}) = \int P(\underline{q}_{i+\tau} | \underline{q}_i) P(\underline{q}_i) d^N \underline{q}_i \quad (9.55)$$

其中的条件概率已由式 (9.32) 给出。我们期望推导出一个  $P(\underline{q}_{i+\tau})$  的微分方程，为此我们用一任意函数  $g(\underline{q}_{i+\tau})$  去乘式 (9.55)，从而得到

$$\begin{aligned} A &\equiv \int g(\underline{q}_{i+\tau}) P(\underline{q}_{i+\tau}) d^N \underline{q}_{i+\tau} \\ &= \int d^N \underline{q}_i P(\underline{q}_{i+\tau}) \underbrace{\int d^N \underline{q}_{i+\tau} g(\underline{q}_{i+\tau}) P(\underline{q}_{i+\tau} | \underline{q}_i)}_{F(\underline{q}_i)} \end{aligned} \quad (9.56)$$

为了求出  $F(\underline{q}_i)$ ，我们引入一新的变量矢

$$\underline{q}_{i+\tau} - \underline{q}_i - \tau \underline{K}(\underline{q}_i) = \underline{\xi} \quad (9.57)$$

则有：

$$F = \int d^N \underline{\xi} N(\underline{q}_i) \exp\left\{-\underline{\xi} \frac{G}{\tau} \underline{\xi}\right\} g(\underline{q}_i + \tau \underline{K}(\underline{q}_i) + \underline{\xi}) \quad (9.58)$$

为了求出到  $\tau$  的各线性项的积分值，把式 (9.58) 中的  $g$  展成  $\underline{\xi}$  的幂级数并保留到第二阶，同时展成  $\tau$  的幂级数并保留到一阶项，

由于

$$\int N(\underline{q}_i) \exp\left\{-\underline{\xi} \frac{G}{\tau} \underline{\xi}\right\} d^N \underline{\xi} = 1 \quad (9.59)$$

及

$$\int N(\underline{q}_i) \exp\left\{-\underline{\xi} \frac{G}{\tau} \underline{\xi}\right\} \xi_k \xi_l d^N \underline{\xi} = \frac{\tau}{2} (G^{-1})_{kl} \quad (9.60)$$

我们知道对  $\underline{\xi}$  之奇次幂积分为零，于是很容易得到

$$\begin{aligned} F(\underline{q}_i) &= g(\underline{q}_i) + \tau \underline{K}(\underline{q}_i) \nabla_{\underline{q}_i} g(\underline{q}_i) \\ &\quad + \tau \cdot \frac{1}{4} \sum_{kl} (G^{-1})_{kl} \partial^2 g | \partial q_{ik} \partial q_{il} \end{aligned} \quad (9.61)$$

把式 (9.61) 代入式 (9.56), 并进行分步积分, 则代替式 (9.56) 可得到:

$$A = \int d^N q_i g(\underline{q}_i) \{1 - \tau \nabla_{\underline{q}_i} \cdot \underline{K}(\underline{q}_i) + \frac{\tau}{4} \sum_{kl} \partial^2 / \partial q_{ik} \partial q_{il} (G^{-1}(\underline{q}_i))_{kl} \} P(\underline{q}_i) \quad (9.62)$$

现在, 引入记号

$$P(\underline{q}_i, \tau) \Rightarrow f(\underline{q}_i, t + \tau) \quad (9.63)$$

$$P(\underline{q}_i) \Rightarrow f(\underline{q}_i, t) \quad (9.64)$$

出现在式 (9.56) 中部及式 (9.62) 右端的  $g$  是一任意函数, 相当的表达式即使不积分也必定相等, 这就导致了如下的关系式:

$$f(\underline{q}_i, t + \tau) = f(\underline{q}_i, t) + \tau L f(\underline{q}_i, t) \quad (9.65)$$

其中,  $L$  是福克-普朗克算子。

$$L f = -\nabla_{\underline{q}_i} \cdot (\underline{K}(\underline{q}_i) f) + \frac{1}{4} \sum_{kl} \partial^2 / \partial q_{ik} \partial q_{il} \{ (G^{-1}(\underline{q}_i))_{kl} f \} \quad (9.66)$$

它出现在伊藤积分中。

这样, 我们确认已找到了伊藤-福克-普朗克方程的短时传播子。

## 9.5 从实验数据能推导出牛顿定律吗

在本节中, 我们打算讨论 9.1, 9.2 和 9.4 节中的程序, 是否有可能使我们根据与一种特定的运动有关的实验数据 (如行星运动) 及导致该运动的力  $\underline{K}$ , 推导出牛顿定律

$$\ddot{\underline{P}} = \underline{K}(\underline{q}) \quad (9.67)$$

显然, 首先必须研讨噪音为零的极限情况, 此时, 条件概率为一

$\delta$ -函数:

$$\underline{P}(\underline{q}_{i+\tau} | \underline{q}_i) = \delta(\langle \underline{q}_{i+\tau} - \underline{q}_i \rangle | \tau - K(\underline{q}_i)) \quad (9.68)$$

或表示为其多维推广形式:

$$\underline{P}(\underline{q}_{i+\tau} | \underline{q}_i) = \delta(\langle \underline{q}_{i+\tau} - \underline{q}_i \rangle | \tau - \underline{K}(\underline{q}_i)) \quad (9.69)$$

乍看起来,按这种方式我们将得不到式(9.67)所表达的牛顿定律,而只能推得:

$$\dot{\underline{q}} = \underline{K}(\underline{q}) \quad (9.70)$$

牛顿力学诞生之前,这个定律曾在科学中使用过,并被用来描述阻尼运动。但是,如果我们在讨论时,不仅把粒子(或天体)的位置,而且也把它们动量(或速度)纳入可观测变量组 $\underline{q}$ 中加以考查,那也容易把牛顿定律再推导出来!这样一来,(当我们采用牛顿力学)去检验条件矩表示式(9.5)和(9.6)会表现为甚么形式,并确证它们的确是式(9.67)的结果,就只不过是一个简单的练习罢了。

把我们的方法和步骤“翻译”为量子力学所述的办法,这样一来,很可能有希望推导出基本粒子所遵从的种种基本定律。

练习:从牛顿定律 $\dot{P} = K(q)$ 入手,计算式(9.21)及(9.22)的条件矩

$$\langle q_{l,i+\tau} \rangle_{q_{k,i}}, \quad \langle q_{l,i+\tau}^2 \rangle_{q_{k,i}}, \quad l=1,2; \quad k=1,2$$

在 $\tau$ 为无穷小时之值。上式中

$$q_1 \equiv q, \quad q_2 \equiv P$$

然后,再使用上面的公式去确定式(9.69)!

## Introduction

The purpose of this study is to investigate the effects of a new educational program on the learning outcomes of students in a secondary school.

The study was conducted in a secondary school in the city of Istanbul, Turkey, during the 2022-2023 academic year.

The research was carried out by a team of researchers from the Department of Educational Sciences, Istanbul University.

The study was designed as a quasi-experimental study, with the new educational program being the experimental group and the traditional program being the control group.

The data were collected through a series of tests and questionnaires administered to the students in both groups.

The results of the study show that the new educational program had a positive effect on the learning outcomes of the students in the experimental group.

The findings of this study suggest that the new educational program is a promising approach for improving the learning outcomes of students in secondary schools.

The study also highlights the need for further research to explore the long-term effects of the new educational program on students' learning outcomes.

The results of this study can be used by educators and policymakers to make informed decisions about the implementation of the new educational program in secondary schools.

The study was funded by the Ministry of National Education, Turkey, and the results will be published in a peer-reviewed journal.

The authors would like to thank the students and teachers of the participating school for their cooperation and support during the study.

The authors also thank the reviewers for their valuable comments and suggestions, which helped to improve the quality of the study.

The authors declare that they have no conflict of interest and that the study was conducted in accordance with the ethical standards of the Ministry of National Education, Turkey.

## 第十章 应用于一些物理系统

---

目前,为数众多的非平衡相变的事例已被人们知晓,它们可产生于截然不同的系统之中。5.2节和5.3节中的单模与多模激光器(无相位关系)可作为说明我们新观点的一个明显的例子。本节首先对激光器进一步详加阐述,并处理各种更为复杂的情况。然后,我们将把注意力转向流体动力学,去讨论对流不稳定性问题。

### 10.1 有相位关系的多模激光器

让我们来考查一环型激光器,其中的电场强度 $E$ 可表示为:

$$E(x,t) = \sum_i B_i(t) \exp [ik_i x - i\omega_i t] + \text{c.c.} \quad (10.1)$$

我们假设波动仅沿一个方向传播,同时, $k$ 和 $\omega$ 分别在比 $k_0$ 及 $\omega_0$ 为小的某一区间取值,表示为:

$$\begin{aligned} k_i &\text{ 属于 } k_0 \pm \Delta k \\ \omega_i &\text{ 属于 } \omega_0 \pm \Delta \omega \end{aligned} \quad (10.2)$$

下面,我们讨论 $E$ 及其幂次的空间和时间平均值。时间平均值已

在第五章中加以定义，于是很容易得到：

$$\bar{E} = 0 \quad (10.3)$$

$$\overline{E_1(x, t) E_{1'}(x, t)} = 2\delta_{11'} B_1 B_{1'} \quad (10.4)$$

$$\overline{E_1 E_{1'} E_{1''} E_{1'''}} = 0 \quad (10.5)$$

对平面波积分，并利用k-选择规则，不难得到下面的结果：

$$\begin{aligned} \overline{E_1 E_{1'} E_{1''} E_{1'''}} &= C \sum_{\pm} \delta(\pm k_1 \pm k_{1'} \pm k_{1''} \pm k_{1'''}) \\ &\times B_1^{\pm} B_{1'}^{\pm} B_{1''}^{\pm} B_{1'''}^{\pm} \end{aligned} \quad (10.6)$$

其中  $B^+ = B^*$ ,  $B^- = B$ 。

关于3.3节中的讨论，现在可作如下等价替换：

$P_i$  的下标

$$i = (B_1, B_1^*, B_2, B_2^*, \dots, B_M, B_M^*) \quad (10.7)$$

约束

$$f_1 = \langle B_1^* B_1 \rangle \quad (10.8)$$

$$\begin{aligned} f_{1,1',1'',1'''} &= \langle B_1^* B_{1'}^* B_{1''} B_{1'''} \rangle \delta(k_1 + k_{1'} - k_{1''} \\ &\quad - k_{1'''}) \end{aligned} \quad (10.9)$$

再应用最大信息熵原理，最后可以得到：

$$\begin{aligned} P(\underline{B}, \underline{B}^*) &= \exp \left[ -\lambda - \sum_i \lambda_i B_i^* B_i - \sum_{1,1',1'',1'''} \lambda_{1,1',1'',1'''} \right. \\ &\quad \left. \cdot B_1^* B_{1'}^* B_{1''} B_{1'''} \delta(k_1 + k_{1'} - k_{1''} - k_{1'''}) \right] \end{aligned} \quad (10.10)$$

此关系式就是在细致平衡原理成立的情况下，曾经从微观理论导出的结果。

## 10.2 包含极化和反转的单模激光器

现在我们处理单模激光器问题，其中原子系统的极化和反转是作为动力学变量来考虑的。为了简单起见，我们讨论行波，因而有如下的一些变量：

$$E(x, t) = B(t)e^{ikx - i\omega t} + \text{c.c.} \quad \text{场} \quad (10.11)$$

$$P(x, t) = P(t)e^{ikx - i\omega t} + \text{c.c.} \quad \text{极化} \quad (10.12)$$

$$D(x, t) \quad \text{慢变化反转} \quad (10.13)$$

其中  $B(t)$ ,  $P(t)$  和  $D$  是时间的慢变量，求其空间平均值，容易得到：

$$\bar{E} = \bar{P} = 0, \quad \bar{D} \neq 0 \quad (10.14)$$

$$\overline{E^*E} = B^*B, \quad \overline{E^*E^*E E} = B^*B^*B B \quad (10.15)$$

$$\overline{|P(x, t)|^2} = P^*(t)P(t), \quad \overline{P^*E} = P^*(t)B(t) \quad (10.16)$$

为了把前面的讨论加以推广，我们将保留直到第四阶的一切非零项。这时分布函数将取如下形式：

$$P = \exp [\lambda + \lambda_1 |E|^2 + \dots + \lambda_2 D^4] \quad (10.17)$$

不幸的是，还没有一种微观理论能确切计算目前这个问题中的分布函数。因此，我们只好再度进行绝热近似处理；然而，对于  $P$  和  $D$  我们将显式地保存其分布函数。我们仍然从激光方程入手（参见 7.10 节），

$$\dot{E} = -\kappa E + gP + F \quad (10.18)$$

$$\dot{P} = -\gamma P + gED + \Gamma \quad (10.19)$$

$$\dot{D} = \gamma_{11}(D_0 - D) - g(EP^* + E^*P) + \Gamma_D \quad (10.20)$$

同时，利用绝热近似

$$\dot{P} = \dot{D} = 0 \quad (10.21)$$

我们便可从式 (10.19) 得到如下关系:

$$\Gamma = \gamma P - g ED \quad (10.22)$$

由于涨落力  $\Gamma$  按高斯律分布, 我们可以断定表达式

$$\gamma P - g ED \quad (10.23)$$

也有相同的分布函数。同样, 由式 (10.20), 我们亦可得到该方程左端前两个括号之差的分布函数。最后, 在阈值附近绝热消去  $P$  和  $D$ , 代替式 (10.18), 我们得到:

$$\dot{E} = -\kappa E + GE - CE |E|^2 + F_{tot} \quad (10.24)$$

很容易把式 (10.24) 转变为福克-普朗克方程, 而且, 如我们以前所见, 该方程的定态解易于求得。 $E$ ,  $P$  和  $D$  的总的分布函数是  $E$ ,  $P$  和  $D$  分别遵从的分布函数之积, 因而可表示为:

$$\begin{aligned} f = & e^{\lambda} \exp [a |E|^2 - b |E|^4 \\ & \cdot \exp [-(\gamma P - gED)^2/Q) \\ & - (g(EP^* + E^*P) + \gamma_{11}(D - D_0))^2/Q'] \end{aligned} \quad (10.25)$$

比较分布函数 (10.25) 与表示推测的函数 (10.17), 发现许多拉格朗日参数  $\lambda$  都恒等于零, 如表 10.1 所示。但是, 这些  $\lambda$  参数为零, 究竟是我们采用绝热消去手段的后果, 抑或甚至在一般性方程 (10.18) — (10.20) 的场合也仍然成立, 目前还是一个悬而未决的问题。

### 10.3 流体动力学: 对流不稳定性

首先, 让我们和读者一起简要地回顾一下对流不稳定性。试考虑从下方加热的矩形容器内的液体。液体若较下层与较上表层



表10.1

	1	$ E ^2$	$ P ^2$	$EP^*$	$E^*P$	$D$	$D^2$
1	X	X	X	X	X	X	X
$ E ^2$	/	X	X				X
$ P ^2$	/	/					
$EP^*$	/	/	/	X	X	X	
$E^*P$	/	/	/	/	X	X	
$D$	/	/	/	/	/	X	
$D^2$	/	/	/	/	/	/	

温差很小，则仅能以传导的方式传热，无宏观运动可见。但是，当温差超过某一临界值，便有涡旋状结构型的宏观运动出现。它以波长  $\lambda$  为涡旋直径两倍的垂直速度场为特征。由于垂直速度场的正弦式变化代表着演化结构，这本身就提醒我们，应采用相应的正弦波（或平面波）的振幅作为序参量。取前四阶矩，并考虑到对称性因素，奇数阶矩必定趋于零，这样，我们立即可得到分布函数：

$$f(\xi) = N \exp [\lambda \xi^2 - \beta \xi^4] \quad (10.26)$$

这与微观理论推出的结果一致。

如果假定速度场的前四个空间关联函数已被测出，我们便可利用第六章的方法去确定联合分布函数  $P(\underline{\xi}_0, \underline{\xi}_1)$ ，其形式为

$$P = f(\underline{\xi}_0) \prod_i P(\xi_i | \underline{\xi}_0) \quad (10.27)$$

这种处理稍嫌冗长，且并不能给我们以新的启示，因此我们不准

备进一步加以讨论。

但是，我们要提出另一种现象。例如，在一圆形容器内，一些特殊的流体可能形成六角形结构。这些波纹结构可解释为波矢量形成等边三角形的诸平面波 $\xi_i \exp[i k x]$ 的叠加。所以，我们将用这些平面波的复振幅 $\xi$ 作为序参量。为了便于区分，我们将根据其在流体表平面上的方位，分别注以下标 $0, \frac{\pi}{3}$ 或 $\frac{2}{3}\pi$ 。采用类似于10.1节中为讨论多模激光器而导出选择定则那样的论证方法，我们便可证明不少的矩皆为零，而只有一些特殊的矩及其拉格朗日乘子存在。这些不为零的矩为：

$$\langle |\xi_0|^2 \rangle, \langle |\xi_{\pi/3}|^2 \rangle, \langle |\xi_{2\pi/3}|^2 \rangle \quad (10.28)$$

$$\langle \xi_0^* \xi_{\pi/3} \xi_{2\pi/3} \rangle$$

及其轮换排列

$$\langle |\xi_i|^2 |\xi_k|^2 \rangle \quad (10.29)$$

其中， $j, k$ 为 $0, \frac{\pi}{3}, \frac{2\pi}{3}$ 中之任一个。这样，如何去构筑分布函数，就变得一目了然，其结果仍与微观理论一致。

# 第十一章 生物学中的行为模式转变·一个例子:手的运动

---

在本章中, 我们想指出如何才能把前几章中建立起的方法运用于生物体系。正是在这里, 事实将会毫不含糊地表明, 对于复杂体系行为和性能的模拟, 我们所创立的方法将是多么强有力的工具。

## 11.1 一些实验事实

在研究两只食指的随意摆动时, 凯尔索 (Kelso) 观察到一种有趣现象。若指令受试者加快手指不同相的、反对称的运动 (牵涉到屈肌和伸肌同时活动) 的频率, 接受实验者的手指运动将出人意料地变为一种同相的对称运动模式, 这涉及到同源肌肉簇的同时激活作用 (图11.1)。这一发现不仅仅局限于手指运动。后来, 在不少的实验中, 使用类似的处理方法, 也观察到了手的绕腕运动中之模式转变: 两手间反对称位相关系被对称的位

相关系所取代。而且还进一步发现，尽管不同受试者的手运动发生相转变的频率迥然不同，但这种转变是可以预言的。如果用各特定频率为单位来表示转变频率，即采用一种彼此独立的速度单位来度量它，单位的选择随每一接受实验者的意愿而定，即以“仿佛他(或她)惯于以此快慢，终日作转手运动”的那种速度为准，所得的无量纲比或“临界值”，对一切受试者都是同一常数。引入摩擦阻力后，各个接受实验者的特定频率和转变频率都将改变，但对于所有接受实验者，临界值却都不会变化。

这些简单实验最富于戏剧性的方面在于，肌肉簇在排列顺序及位相关系上那种突然的完全不由自主的变化，竟发生在一个决定于其本性的临界频率处，在这一点上，手的运动数据正是运动步调转变的表象。例如，席克(Shik)等人曾证明，对切除大脑的猫的中脑区给以电刺激，并稳定地增大电流，不仅可以导致其运动速率增大，而且，当电流超过一定值时，也可以观察到步调的转换。正象在手的实验中，运动步率较高时，往往会发生从一种运动模式到另一种运动模式的“跳跃”一样，他们也指出存在着一个不稳定区域，此时猫可从小跑进入狂奔状态，然后再回到原态。尽管手运动的数据以及四足动物运动速率增大的发现使人强烈地感到，这些协调的变化可能受制于一个单参数的变化，但对作为这运动位相转变之基础的神经系统过程，我们仍然知之甚少。

总之，上述实验之主要特征是：(i)手的运动状态只有两种稳定的位相(或称“吸引子”)存在(观察到哪一种取决于体系如何设定，即取决于两手以异相位或同相位模式运动的指令)；

(ii) 从一吸引子态向另一吸引子态的突变发生在临界转动频率

处；(iii) 突变后，只可观察到一种运动模式（对称同位相的）；(iv) 当转动频率降低时，体系仍滞留在对称运动模式中，即体系并不返回其初始设定态——这一结果使人想到对称及反对称模式吸引子区共存，而且其中之一已耗尽。

## 11.2 怎样将转变模型化

现在我们试图用前面介绍的方法，把这些发现模型化。如图11.1所示，我们以 $q_1$ 和 $q_2$ 表示两手指的伸张度，因而可将其表示为

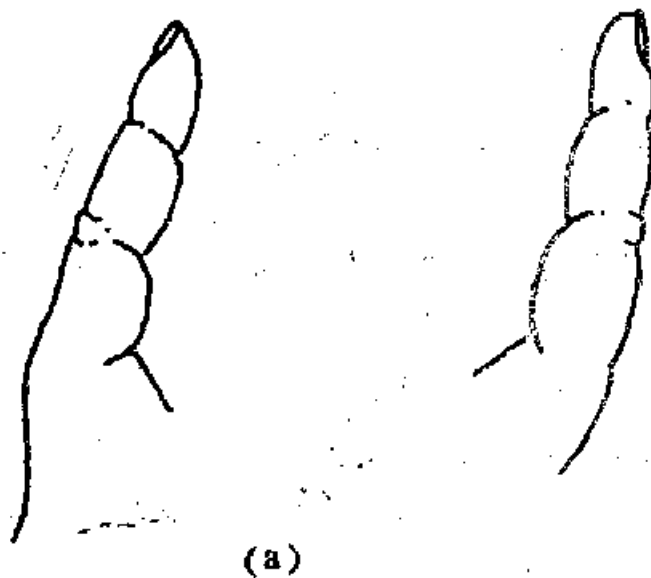
$$q_1 = r_1 \cos(\omega t + \phi_1) \quad (11.1)$$

$$q_2 = r_2 \cos(\omega t + \phi_2) \quad (11.2)$$

其中 $r_1$ 和 $r_2$ 为振幅，根据实验结果，可认为它们与时间无关，而 $q_1$ 及 $q_2$ 只取决于预先设定的手指运动角频率 $\omega$ 。相位 $\phi_1$ 和 $\phi_2$ 仍可视作时间慢变化函数。

在以下的讨论中，我们令

$$r_1 = r_2 = r \quad (11.3)$$



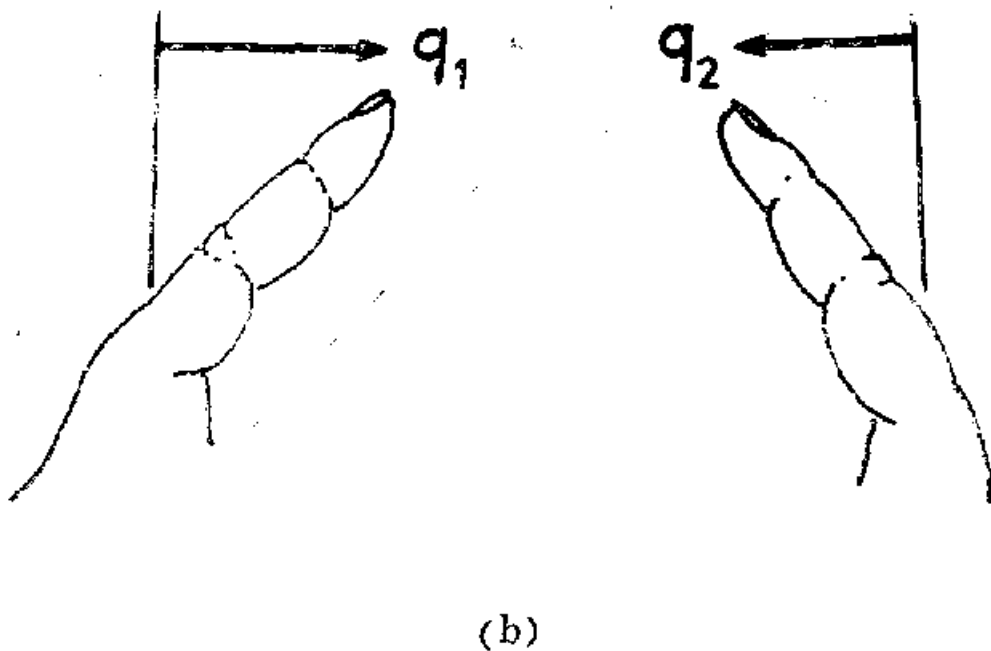


图11.1 手指的平行运动 ((a)图)不自觉地变成一种反平行对称运动((b)图)

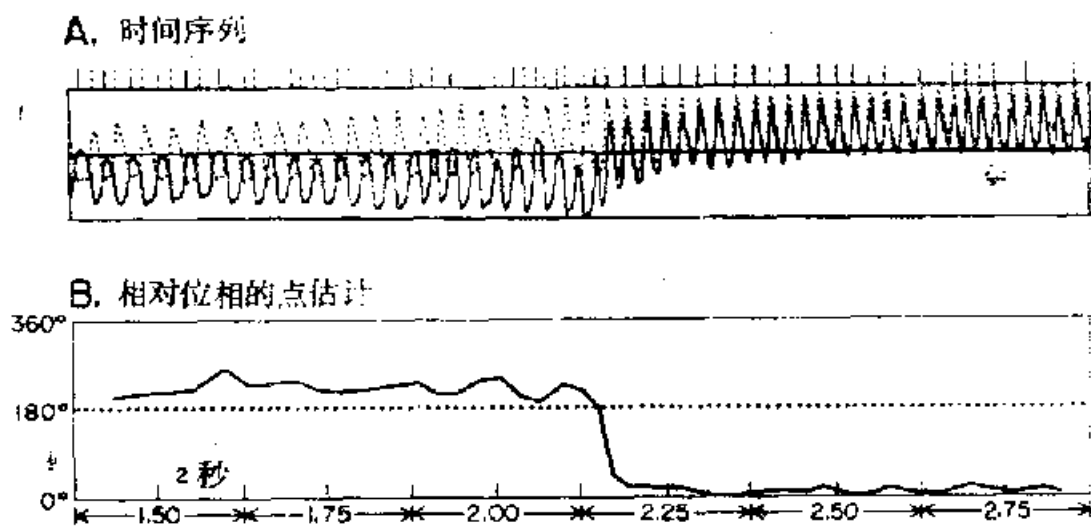


图11.2 上图:左手(实线)及右手(虚线)随时间的位移。接受实验者依照实验者的口令不断增加转动频率。下图:两手间的位相关系。一支手运动的位移峰值用作“指标”线,其相位置按另一“参比”线从峰值到峰值的周期间隔连续地加以计算。图象重复显示了相位曲线,因而可注意到相位的滞后和超前。

为了使用我们的一般方法，可将 $q_1$ 及 $q_2$ 视为观测量，它们的矩是可以测定的，如

$$\langle q_i \rangle \dots \quad (11.4)$$

根据以前建立的方法，我们考虑四阶以下的矩。类似于对激光的讨论，我们感兴趣的是慢变化的位相 $\phi_1, \phi_2$ ，而不是相对快速的振荡。为此，首先考虑时间平均值，定义

$$\langle q_i \rangle_t = \frac{1}{T} \int_0^T q_i dt \quad (11.5)$$

其中 $T$ 由下式给定：

$$T = \frac{2\pi}{\omega} \quad (11.6)$$

根据实验，我们假定 $\phi_1, \phi_2$ 为时间慢变化函数，因而可使用如下关系：

$$\frac{1}{T} \int_0^T \cos(\omega t + \phi_i) dt = 0 \quad (11.7)$$

或者，当余弦函数可分解成指数函数时，采用等价形式

$$\frac{1}{T} \int_0^T \exp [i \omega t + i \phi_i] dt = 0 \quad (11.8)$$

由式(11.8)，我们甚至可以要求

$$\frac{1}{T} \int_0^T \exp [in(\omega t + \phi_i)] dt = 0 \quad (11.9)$$

其中 $n$ 是整数。

又由式(11.7)，我们得到

$$\frac{1}{T} \int_0^T q_i dt = 0 \quad (11.10)$$

现在，考虑第二矩

$$\frac{1}{T} \int_0^T q_i q_k dt = 0 \quad (11.11)$$

这里，我们将使用分解式

$$q_i = \frac{r}{2} \{ \exp [i(\omega t + \phi_i)] + \exp [-i(\omega t + \phi_i)] \} \quad (11.12)$$

让我们讨论，将式 (11.12) 代入式 (11.11) 时所得到的各种项。首先考虑项

$$\frac{1}{T} \int_0^T \exp [2i\omega t + i\phi_i + i\phi_k] dt = 0 \quad (11.13)$$

由于式 (11.9)，它显然等于零。另一方面，由于在下式中快速振荡部分相消，

$$\frac{1}{T} \int_0^T \exp [i(\phi_i - \phi_k)] dt = \exp [i(\phi_i - \phi_k)] \quad (11.14)$$

故可用其右端代换左端之平均值，合并式 (11.13) 及式 (11.14)，可立即得到

$$\frac{1}{T} \int_0^T q_i q_k dt = \frac{r^2}{2} \cos(\phi_i - \phi_k) \quad (11.15)$$

现在让我们来区分两种情况。

第一种情况， $j = k$ 。

由 (11.15)，我们易于获得一个与时间无关的常数，使

$$(11.15) = \text{常数} \quad (11.16)$$

既然此时再没有变量出现，在以后的讨论中完全可以忽略此项。

第二种情况， $j \neq k$ 。

式 (11.15) 的结果启示我们引入简写记号，



$$\phi_2 - \phi_1 = \phi \quad (11.17)$$

这样, 该结果可写为

$$\langle 11.15 \rangle = \frac{r^2}{2} \cos \phi \quad (11.18)$$

现在, 讨论三阶矩

$$\frac{1}{T} \int_0^T q_{j_1} q_{j_2} q_{j_3} dt = 0 \quad (11.19)$$

实际上, 我们已直接写出了结果, 因为很容易弄清楚式 (11.19) 的左端仅可能出现如下形式的表示式或其复共轭,

$$\exp [i\omega t] ; \exp [3i\omega t] \quad (11.20)$$

根据式 (11.9), 它们的贡献为零。最后便轮到讨论四阶矩

$$\frac{1}{T} \int_0^T q_{j_1} q_{j_2} q_{j_3} q_{j_4} dt \quad (11.21)$$

将式 (11.12) 代入式 (11.21), 很容易得到含有

$$\exp [4i\omega t] ; \exp [2i\omega t] \quad (11.22)$$

或其复共轭的诸表达式。对整个时间间隔  $T$  求平均值, 这些表达式为零。这样, 我们得到快速振荡项不再出现的一些表达式, 所以, 最终我们只关心如下的表达式,

$$\exp [i(\phi_{j_1} + \phi_{j_2} - \phi_{j_3} - \phi_{j_4})] \quad (11.23)$$

或者将其足标  $j_1, \dots, j_4$  加以交换所得到的表达式。

现在我们研究, 给足标  $j_1, \dots, j_4$  赋以数值 1 或 2 时, 所得到的各类表达式。结果详见表 11.1。

使用表 11.1 的结果, 可把四阶矩写作

$$\langle q_{j_1} q_{j_2} q_{j_3} q_{j_4} \rangle = C_1 + C_2 \cos \phi + C_3 \cos(2\phi) \quad (11.24)$$

其中系数  $C_1, C_2$  和  $C_3$  与  $r^4$  成比例。

表11.1

$j_1$	$j_2$	$j_3$	$j_4$	指 数
1	1	1	1	0
2	2	2	2	0
1	2	1	2	0
1	2	2	1	0
2	1	2	1	0
2	1	1	2	0
1	1	2	1	$\phi_1 - \phi_2$
1	1	1	2	$\phi_1 - \phi_2$
2	1	2	2	$\phi_1 - \phi_2$
1	2	2	2	$\phi_1 - \phi_2$
				$= -\phi$
2	1	1	1	$\phi_2 - \phi_1$
1	2	1	1	$\phi_2 - \phi_1$
2	2	1	1	$\phi_2 - \phi_1$
2	2	2	1	$\phi_2 - \phi_1$
				$= -\phi$
1	1	2	2	$2(\phi_1 - \phi_2) = -2\phi$
2	2	1	1	$2(\phi_2 - \phi_1) = -2\phi$

$C_1$ 不含体系的任何变量，我们因此对 $C_1$ 项毫无兴趣。第二项在讨论二阶矩时已被考察过，所以可以从式(11.24)中划去。这样一来，相关约束仅包含于式(11.24)的最后一项中。

汇集以上结果，我们可将从第二阶到第四阶矩得到的约束表示为如下公式

$$r^2 \langle \cos \phi \rangle = f_1 \quad (11.25)$$

$$r^4 \langle \cos(2\phi) \rangle = f_2 \quad (11.26)$$

现在，面临的一个简单任务就是运用最大信息原理，得出如下形式的分布函数 $P$ ，

$$P(\phi) = \exp [-\lambda - \lambda_1 \cos \phi - \lambda_2 \cos(2\phi)] \quad (11.27)$$

应该注意，拉格朗日乘子 $\lambda$ ， $\lambda_1$ ， $\lambda_2$ 是 $\omega$ 的函数，这不仅是由于式(11.25)或式(11.26)前面存在着因子 $r^2$ 和 $r^4$ ，而且是由于实验已经表明括号所表示的平均值取决于 $\omega$ 。式(11.27)中的指数可用一势函数表示：

$$\hat{V}(\phi) = \lambda_1 \cos \phi + \lambda_2 \cos(2\phi) \quad (11.28)$$

这个势函数可决定体系的稳定和不稳定态，它随 $\lambda_1$ 与 $\lambda_2$ 比值的不同而变化（参看图11.3），显然，图中左上角的局域极小点，相当于手指的平行运动。当参数 $\lambda_2/\lambda_1$ 减小时，该极小区会变得越来越扁平，最后即消失。故，用来表示体系状态的小球最终将落入较低的极小点，这已意味着两个手指作反平行（即对称）运动。所以，根据我们的结果，很容易将上面描述的手指运动的转变模型化，唯一需要假定的是：比率 $\lambda_2/\lambda_1$ 取决于传动频率 $\omega$ 。但我们的模型的用处还不仅限于此，让我们讨论这样一个情况：传动频率足够大，以致只有手指对称运动的状态可以实现。因而，从图11.3很容易推断，当我们减小频率，小球，即体系将停留在 $\phi = 0$ 的状态。这意味着可以预料，即使在临界频率以下，实验者仍以一种对称方式运动其手指。这一预言已被凯尔索证实。

为了描述时间相依过程，我们现在试图对分布函数(11.27)所遵从的福克-普朗克方程加以推断。正如6.6节所示，该推断结果不会是唯一的，因为噪音源与各变量有关。现在作一个最简单的假设，即噪音与 $\phi$ 无关，于是，立即可确认式(11.27)为：

$$P(\phi) = N \exp [-2V(\phi)/Q] \quad (11.29)$$

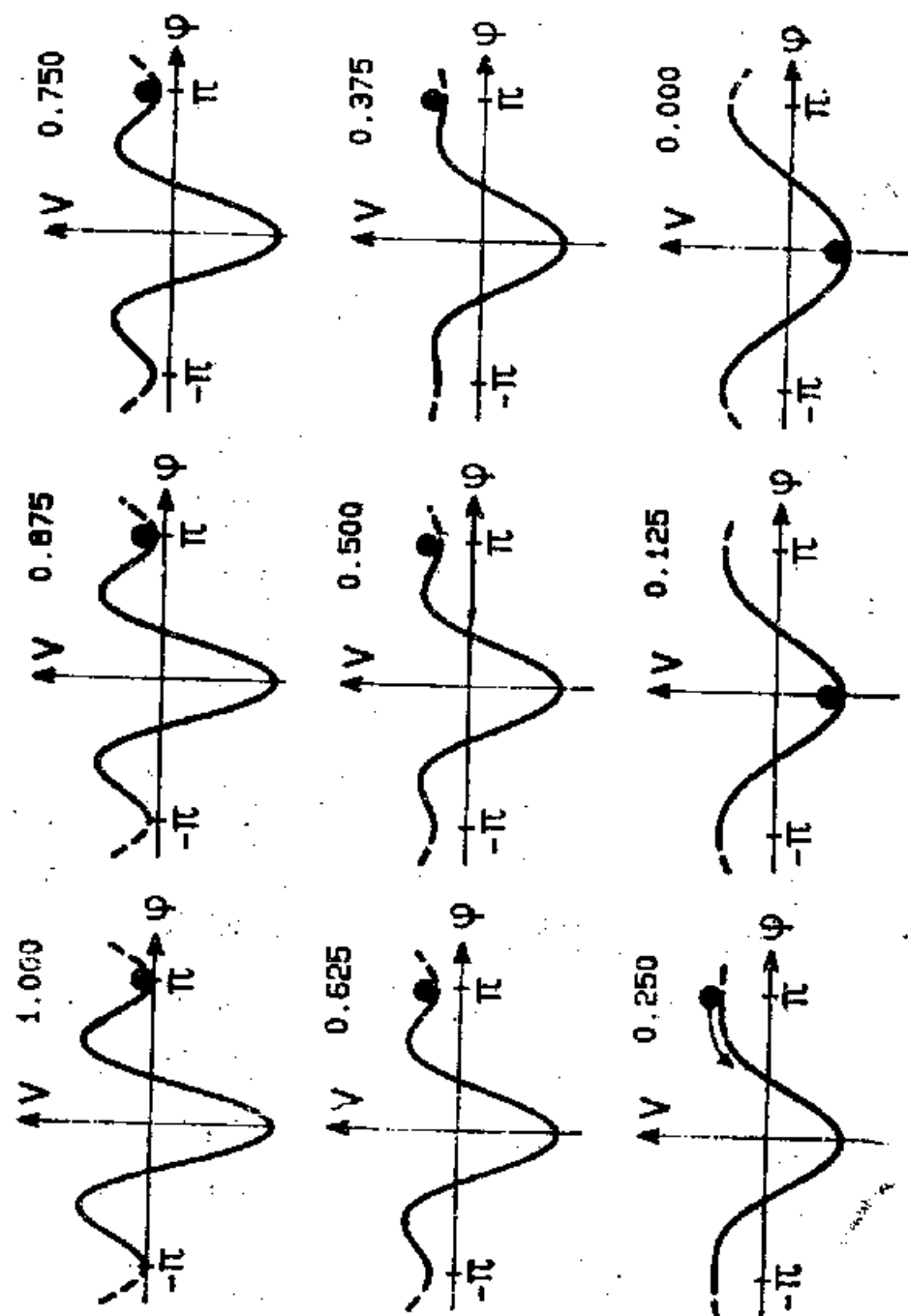


图11.3 势函数 $V$ 随值 $\lambda_2/\lambda_1$ 的变化, 数字表示比值 $\lambda_2/\lambda_1$ 。

其中，我们运用了缩写记号：

$$\hat{V} = \frac{2}{Q} V \quad (11.30)$$

这里Q是噪音源强度。根据第二章中的微观理论，福克-普朗克方程可写作

$$\dot{P}(\phi, t) = -\frac{\partial}{\partial \phi} (K(\phi) P) + \frac{Q}{2} \frac{\partial^2}{\partial \phi^2} P \quad (11.31)$$

上式中又应用了缩写记号

$$K(\phi) = -\frac{\partial V}{\partial \phi} \quad (11.32)$$

如我们在第二章所见，与福克-普朗克方程 (11.31) 相应的朗之万方程是

$$\dot{\phi} = K(\phi) + F(t) \quad (11.33)$$

应用式 (11.27), (11.28), (11.30), (11.32) 及 (11.33)，我们可显式地将上式之右端表示为

$$\dot{\phi} = a \sin \varphi + b \sin(2\varphi) + F(t) \quad (11.34)$$

其中  $a = Q\lambda_1/2$ ,  $b = Q\lambda_2$ 。

这个方程蕴含着对手指运动行为的一些新的预言。即是说，现在我们可以解式(11.34)，并可用这种方式算出各时刻  $\phi$  的关联函数。换言之，现在我们已可以从理论上研究瞬变现象，并与实验结果加以比较。

### 11.3 临界涨落

让我们从作为模型方程的式(11.34)入手进行讨论，这个方

程有三个参数,即 $a, b$ 以及隐含的涨落力 $F(t)$ 的强度,记为 $\sqrt{Q}\xi$ ,而 $\xi$ 具有下列性质:

$$\langle \xi_i \rangle = 0 \quad (11.35)$$

$$\langle \xi_i \xi_{i'} \rangle = \delta(t - t') \quad (11.36)$$

首先,研究以对称模式运动的位相涨落。由于此时体系处在 $\phi = 0$ 附近,我们可将朗之万方程(11.34),或等价地将福克-普朗克方程(11.31)线性化。例如,朗之万方程可写为如下形式

$$\dot{\phi} = -(4b + a)\phi + \sqrt{Q}\xi_i \quad (11.37)$$

求解线性化福克-普朗克方程是不难的,从而可轻而易举地算出绝对位相的平均值及其均方差。

$$\begin{aligned} \langle |\phi| \rangle_{\text{stat}} &= \int_{-\infty}^{+\infty} d\phi |\phi| P_{\text{stat}}(\phi) \\ &= \frac{1 - \exp(-\pi^2 d^2)}{\sqrt{\pi} d \operatorname{erf}\{\pi d\}} \end{aligned} \quad (11.38)$$

$$\begin{aligned} \sigma_{\text{stat}} &= \langle \phi^2 \rangle_{\text{stat}} - \langle |\phi| \rangle_{\text{stat}}^2 \\ &= \frac{1}{2d^2} - \frac{\sqrt{\pi} e^{-\pi^2 d^2}}{d \operatorname{erf}\{\pi d\}} \langle |\phi| \rangle_{\text{stat}}^2 \end{aligned} \quad (11.39)$$

十分明显,式(11.38)和式(11.39)的结果取决于单参数 $d$ ,它由下式给定

$$d = \sqrt{\frac{4b + a}{Q}} \quad (11.40)$$

这些结论使我们可将相位涨落的实验结果与已建立的模型加以比较。为此,首先向读者介绍图11.4,它描绘出实验所得结果。

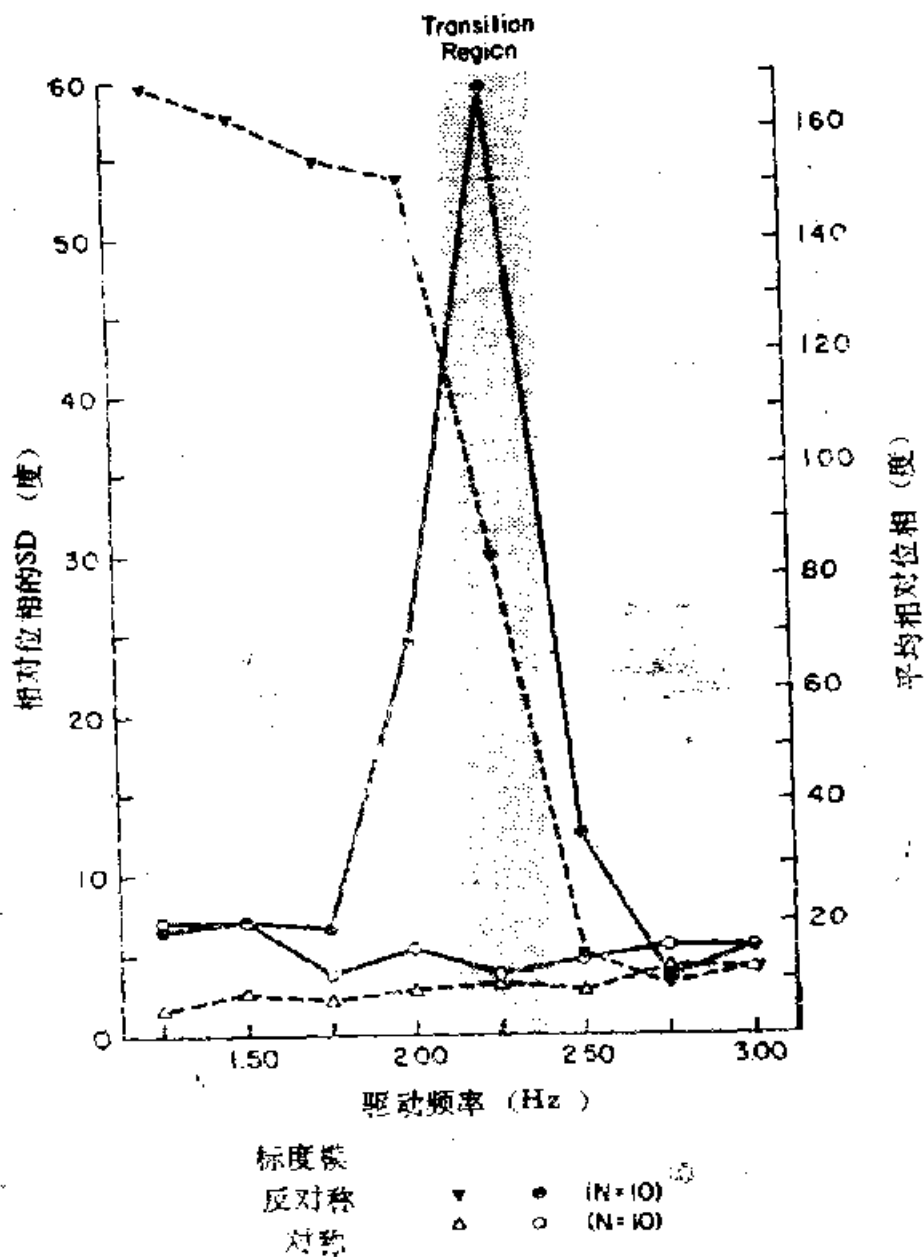


图11.4 在稳定极限下，在每一频率的平稳段(即每一频率的最后三秒钟)测定的平均相位模数 $\langle |\phi| \rangle$  (▼AMS, △SMS) 及其 $SD = [\langle \phi^2 \rangle - \langle \phi \rangle^2]^{\frac{1}{2}}$  (•AMS, ○SMS)。图上的每一点代表10次实验的平均值。AMS = 反对称标度模，SMS = 对称标度模 (引自Kelso及Scholz, 1985)。

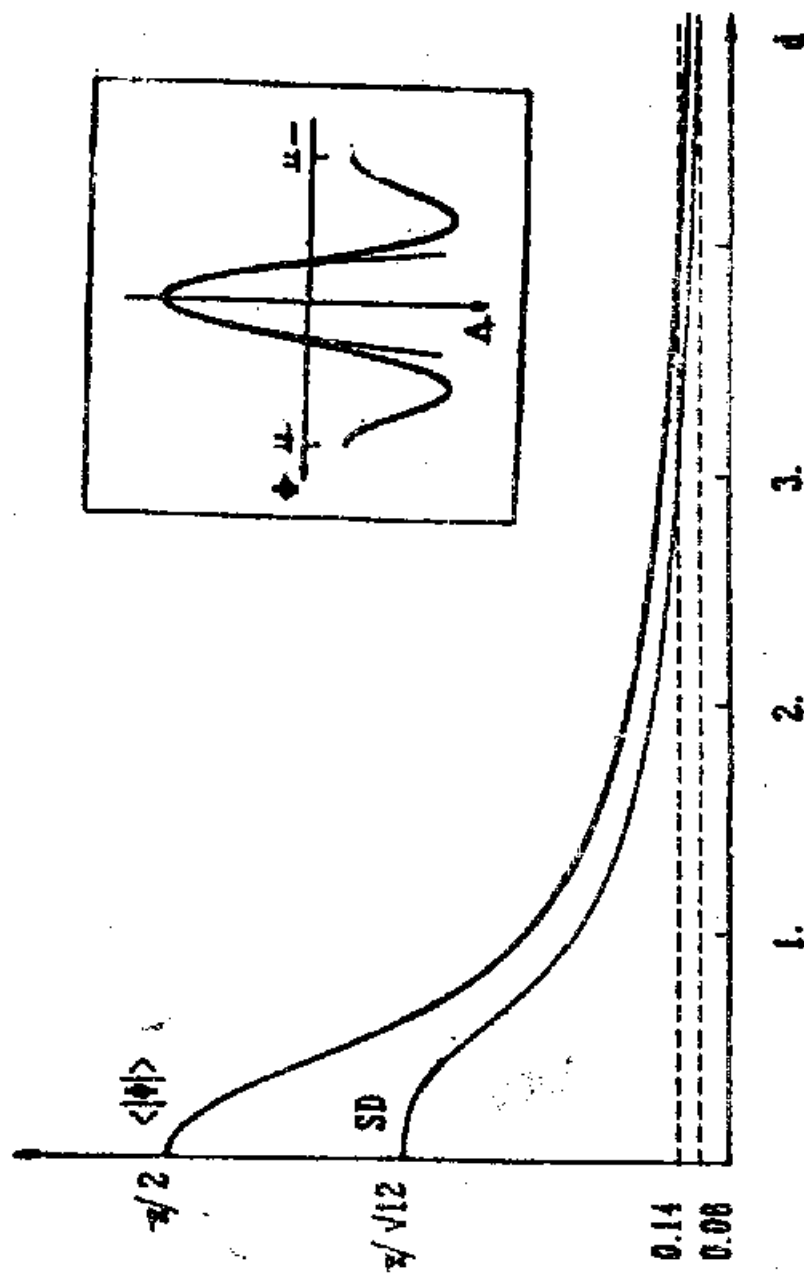


图11.5 根据对称模 $\phi_{0,1}=0$ 的局域模型计算的,  $d = [(a+4b)/Q]^{1/2}$  的函数  
平均相位模数及其SD值。两虚线表示均值( $=0.14$ )及SD( $=0.08$ )的实验值, 因而 $d$   
之实验值可以读出。插图表明了 $a=b=1$ 时势函数(粗线)及局域模型(细线)的情况。

图11.5中, 实线表示式(11.38)和式(11.39)对  $d$  的图象, 而实验值(可从图11.4中导出)的图象则用虚线表示。这样一来, 我们就



可以确定 $d$ ，即，根据实验值

$$\langle |\phi| \rangle \approx 8 \text{ deg} \quad (11.41)$$

可得

$$d = 3 \cdots 5 \quad (11.42)$$

同时从

$$\sigma_{\dots} \approx 4.5 \text{ deg} \quad (11.43)$$

并根据

$$d = 3 \cdots \quad (11.44)$$

我们可求得 $d$ 。显然，若假设

$$d \approx 4 \quad (11.45)$$

两实验结果可以协调。

在 $\phi = 0$ 附近，体系的弛豫时间，可通过求解决定论情况下（即无涨落力时）的线性朗之万方程得到。我们很容易求得

$$\phi(t) = \phi(0) \exp [-(4b + a)t] \quad (11.46)$$

指数中时间 $t$ 前的因子之倒数确定了弛豫时间。所以，利用式(11.40)，我们有

$$\tau_{\dots} = \frac{1}{4b + a} = \frac{1}{d^2 Q} \quad (11.47)$$

此弛豫时间出现在一些迥然不同的弛豫或反应现象中。例如，当红灯一亮，我们必须把脚从气门踏板上提起。 $\tau_{\dots}$ 的一个粗略估计值为0.25秒。由这个数值及 $d \approx 4$ ，再根据式(11.47)可算出

$$Q \approx 0.25 \text{ Hz} \quad (11.48)$$

考虑到涨落，并研究在两个不同时刻 $\phi$ 的关联函数，理论上我们仍可得到相同结果。同理，我们可讨论反对称模式。此时， $\phi$ 的中心在 $+\pi$ 或 $-\pi$ 附近，所以我们可通过下面的形式引入一个小

量 $\varepsilon$ ,

$$\varepsilon = \begin{cases} \phi - \pi, & \text{当 } 0 < \phi \leq \pi \\ \phi + \pi, & \text{当 } -\pi < \phi \leq 0 \end{cases} \quad (11.49)$$

使用 $\varepsilon$ , 并只保留朗之万方程(11.37)中的线性项, 立即可得

$$\dot{\varepsilon} = -(4b - a)\varepsilon + \sqrt{Q}\xi, \quad (11.50)$$

相应的线性化福克-普朗克方程也不难求得, 它的定态解为

$$P_{\text{stat}}(\varepsilon) = \frac{f}{\pi \operatorname{erf}\{\pi f\}} e^{-f^2 \varepsilon^2} \quad (11.51)$$

其中, 我们引入了参数

$$f = \sqrt{\frac{4b - a}{Q}} \quad (11.52)$$

沿引前述位相平均值及标准方差的定义, 并利用式(11.51), 立即可得

$$\begin{aligned} \langle |\phi| \rangle_{\text{stat}} &= \pi - \langle |\varepsilon| \rangle_{\text{stat}} \\ &= \pi - \frac{1 - \exp(-\pi^2 f^2)}{\sqrt{\pi} f \operatorname{erf}\{\pi f\}} \end{aligned} \quad (11.53)$$

$$\begin{aligned} \sigma_{\text{stat}} &= \langle \phi^2 \rangle_{\text{stat}} - \langle |\phi| \rangle_{\text{stat}}^2 \\ &= \langle \varepsilon^2 \rangle_{\text{stat}} - \langle |\varepsilon| \rangle_{\text{stat}}^2 \\ &= \frac{1}{2f^2} - \frac{\sqrt{\pi} \exp(-\pi^2 f^2)}{f \operatorname{erf}\{\pi f\}} - \langle |\varepsilon| \rangle_{\text{stat}}^2 \end{aligned} \quad (11.54)$$

式(11.53)和(11.54)之值对参数的图象见图11.6。根据图11.6, 我们可查出与每一 $f$ 值相应的 $\phi$ 的平均值和标准方差。反之, 若把这一对值纳入图11.6中, 也很容易找出相应的 $f$ 值。与式(11.47)类似, 弛豫时间可由下式给出,

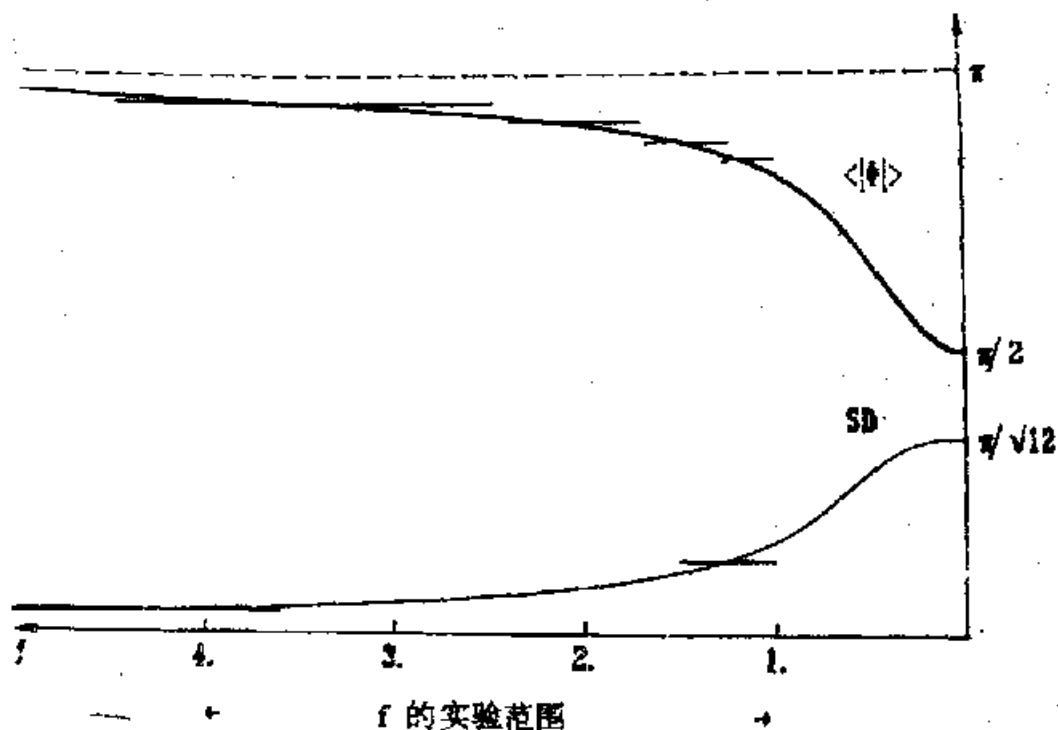


图11.6 反对称模的局域模型平均绝对相位及其SD值与  $f = [(4b-a)/Q]^{1/2}$  的函数关系。趋向转变点的途径相当于  $f \rightarrow 0$ 。  $f$  轴向左取向。曲线已用于根据均值及SD的实验值确定相应的  $f$  值，其范围如图所示。

$$\tau_{rel} = \frac{1}{4b-a} = \frac{1}{f^2 Q} \quad (11.55)$$

取  $\tau_{rel} = 0.25$  秒，并根据实验数据，让  $f$  在数域  $1 \cdots 4$  内取值，可得  $Q$  在数域  $0.25 \cdots 4$  内之值，其数量级与以前讨论对称模式所得结果一致。从式(11.40)及式(11.52)，不难得到下面的关系式，

$$4b + a = Qd^2, \quad 4b - a = Qf^2(\omega) \quad (11.56)$$

现在既然已知  $f$  是振荡频率  $\omega$  的函数，我们即可从式(11.56)出发，借助于

$$a(f) = \frac{Qd^2}{2} - \frac{Q}{2} f^2(\omega) \quad (11.57)$$

$$b(f) = \frac{Qd^2}{8} + \frac{Q}{8} f^2(\omega) \quad (11.58)$$

导出频率与 $a$ 及 $b$ 的依赖关系。其相应的结果已描绘于图11.7中。转变发生时，考察频率及相应的 $f$ 值，我们得到如下结果：当 $d \approx 4$ ， $Q = 0.25\text{Hz}$ 时， $a_c = 4b_c$ 。因而，近似地有

$$a_{c,1} = 2.0 \text{ Hz} \quad (11.59)$$

$$b_{c,1} = 0.5 \text{ Hz} \quad (11.60)$$

总而言之，我们可以说，我们已能以一种合理的方式去确定诸参数；并且在每一种情况下，运用三个模型参数，就能找出四个相互匹配的取决于实验的值，即对称和反对称模式的绝对位相平均值及标准方差。当然，我们的模型仍需改进；但是，我深信以上讨论已毋庸置疑地表明，我们建立的方法能够把复杂系统的行为和性能模型化。

我们因此能得出一些重要结论。

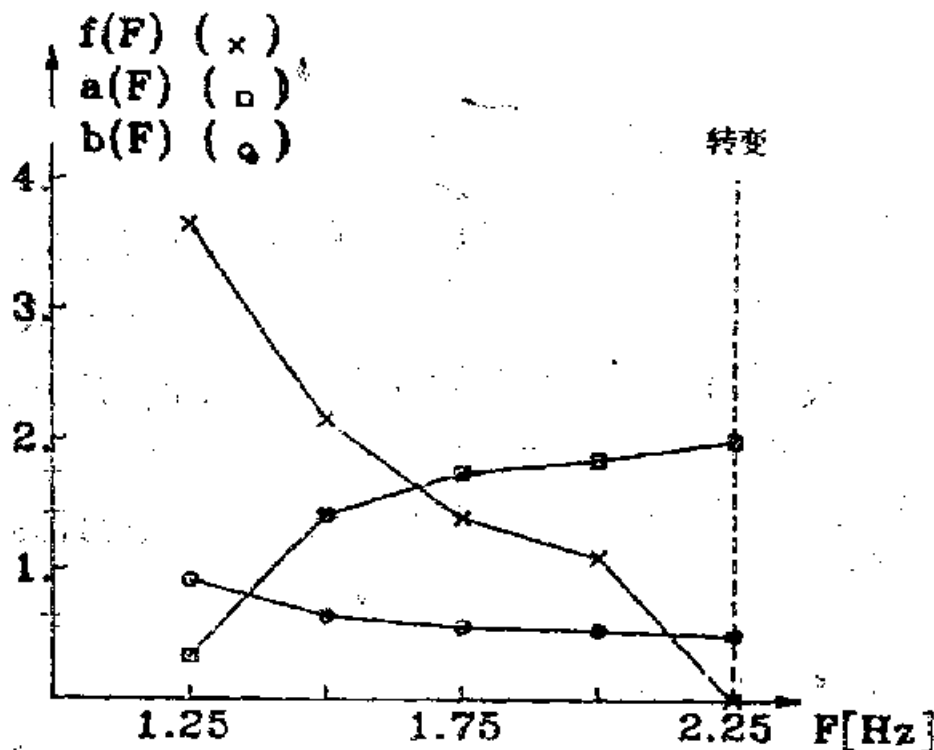


图11.7 实验控制参数——频率 $F \equiv \omega$ 之函数的模参数 $f$ ， $a$ 和 $b$ 。它们由均值与SD值以及弛豫时间估计值确定。

## 11.4 一些结论

### A. 运动原\*程序与自组织

为了解释肌肉的高度协调性,生物学常采用所谓运动原程序的概念。按照这一概念,中枢神经或多或少地类似于一台贮存着特定程序的计算机,一旦得到指令,便开始执行各自的动作。但是,在这种情形下,根本难以理解涨落的产生。首先,运动程序终究是一种固定程序,容不得任何涨落。另一方面,临界涨落是出现自组织时所产生的非平衡相变的典型特征。正是由于这个原因,我们相信,在这些实验中发现的涨落,为肌肉及神经形成了一个自组织系统这一观点提供了强有力的证据。这很可能与激光类似,在那里,当单控制变量改变时,也显示了不同行为间的相互转变。当然,根据这些实验事实,我们还不能肯定,自组织究竟产生于整个神经和肌肉系统,还是产生在神经的子系统之中。

### B. 信息的压缩

肌肉和神经高度的协调已日益为各种类型的特定宏观运动的出现所证实,在上述示例中,这种宏观运动可用单序参量描述。与这种单序参量的出现紧密相联的是一种容量极低的信息,与描述一切神经元和肌肉的各别状态所必须的信息量相比,这种信息是要少得多。

### C. 行为的形态发生学

众所周知,物理学和化学中的大量事例表明,自组织体系能

---

\*运动原指肌肉、运动神经或中枢。——译者注

产生一种特定的时间图样。在这里，我们又看到生物体系也可以产生一种特定的行为模式，如象手指的特殊运动。我们还发现，根据最大信息熵原理，我们可以找到合适的序参量，甚至还可求得相应的方程。可以预期，更为复杂的行为模式也有可能用少量的序参量来加以描述。

## 第十二章 模式识别

---

在本章中我们希望向读者展示，我们建立的一般方法是怎样使我们重新推导出了模式识别现代理论的关系式。假定读者并不熟悉模式识别方法，所以，本章的开始我们对其加以扼要的评述，但仅限于考虑数字模式识别。按照这种方法，一种模式被分解为所谓的特征元。以字母为例，让我们讨论一种特别简单的情形，即仅有两个字母X和O的情形（参看图12.1）。可赋予左开的括号以数字0，右开的括号以数字1的代号。使用这种编码，我们就可进一步把组合（0，1）解释为字母X，组合（1，0）解释为字母O。

更一般地讲，我们可以把一种模式分解为诸特征元，并赋予每一特征元以一分量 $q_i$ 的代号。用这种方式，整个模式就可用一个模式矢量 $\underline{q} = (q_1, q_2, \dots, q_n)$ 来描述（图12.2）。当然，在多维相空间中，一个矢量可用其终点的坐标表示。这样，我们就可把一个矢量，或相应的一个点解释为一种模式或构型。该矢量 $\underline{q}$ 的分量并不需要仅取值0或1，它们也可以是连续的。

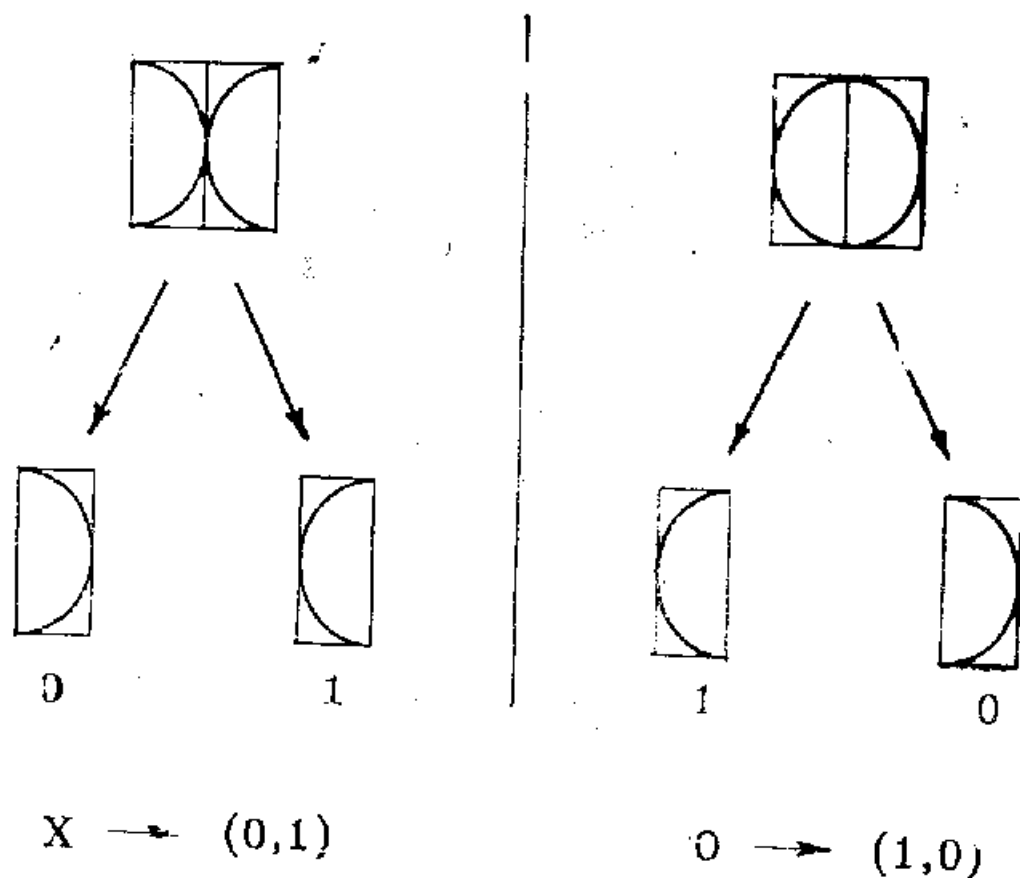


图12.1 把字母X和O分解为特征元的例子，各特征元被赋以数字0或1的代号。

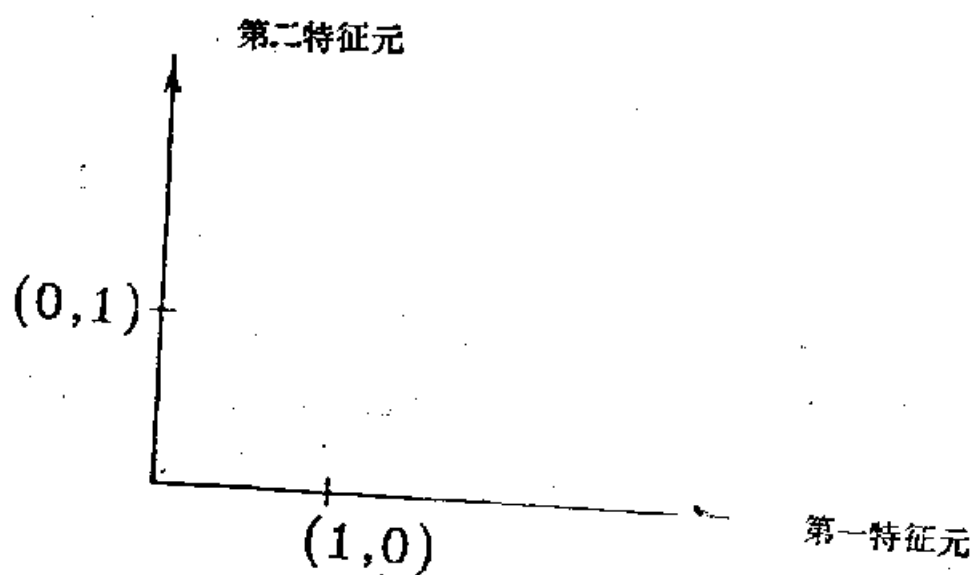


图12.2 适合于图12.1的符号的特征空间。图中点(1, 0)相当于字母O，而点(0, 1)相当于字母X。



例如，当我们把一种模式分解为单个的小单元，并把强度 $q_i$ 赋予每一点 $i$ ，就属于这种情况。因此，现阶段我们可以说，模式识别是分两步完成的。当特征元被选定以后，我们必须按一种具体的量度赋予诸特征元以矢量 $\underline{q}$ ；然后，还得查看该模式矢量所隶属的特征空间。

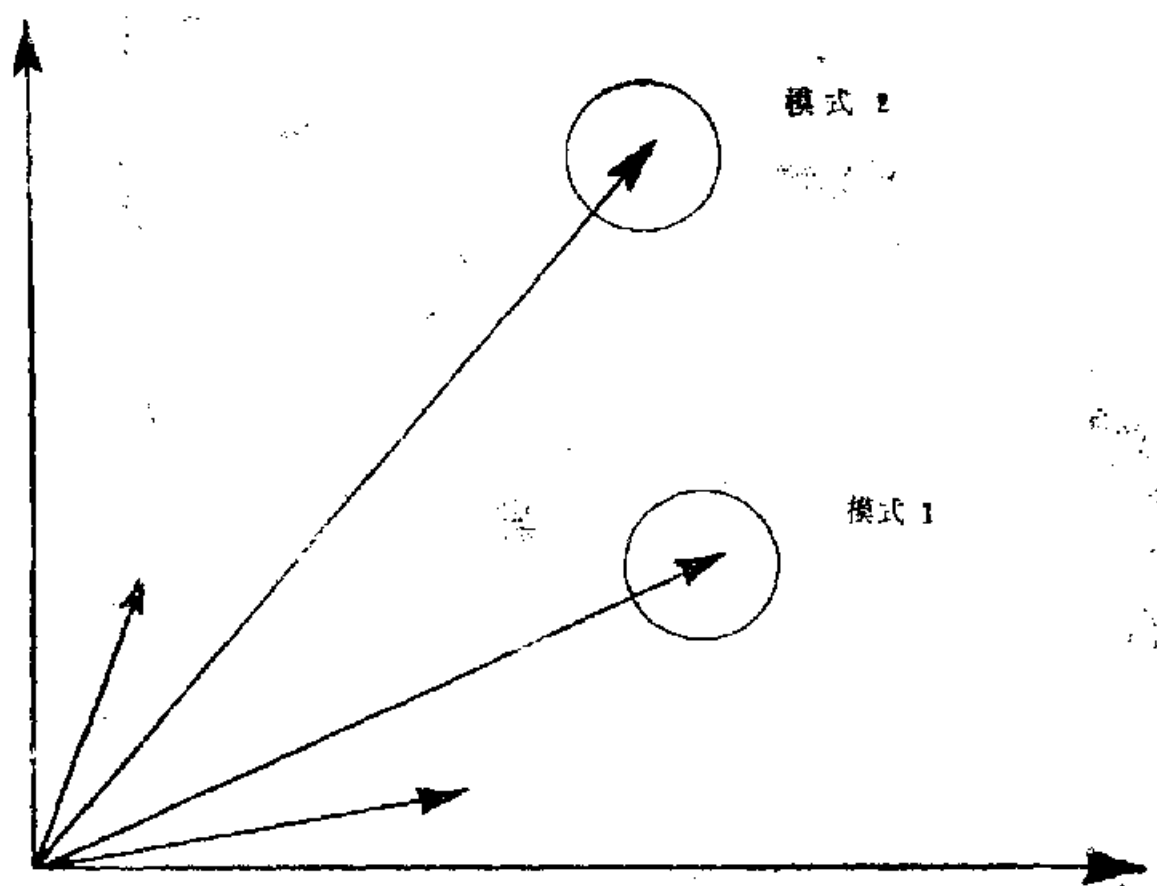


图12.3 广义特征空间的形象化表示。各矢量之终点表示一种典型模式，而它们的邻近区域则表示稍偏离于该典型的模式，因此可作为特定的模式加以识别。

但是，实际上存在着某些困难。首先，特征元可用各种方式选择。因此，人们面临着以适当的方式选择特征元的问题。这里，“适当的”一词意味着特征空间不应该太大，同时模式可清

楚地加以区分。小维度则意味着必须处理的信息不是太多。还有一个更深入一步的问题产生于测量或模式本身，这是由于测量可能引入随机噪声，而模式可能发生内涨落。因此，我们所考虑的非决定论矢量  $\underline{q}$ ，确切地说，是一些分量具有确定概率分布函数的随机矢量。

下面的讨论将涉及到两个问题：

第一，怎样才能根据实验确定概率分布函数？

第二，怎样才能抽取出各特征元，以使待处理的信息变得最小？

为了导出概率分布函数  $P(\underline{q})$  我们援引最大信息原理。假定我们已对打算进行分类或研究的模式系统测定了关联函数

$$\langle \underline{q}_i \underline{q}_j \rangle = Q_{ij} \quad (12.1)$$

我们可不失普遍性地再假定均值为零，即

$$\langle \underline{q} \rangle = 0 \quad (12.2)$$

否则，我们可采用简单的方法，从这些矢量中减去非零均值，并引入新的模式矢量。应用最大信息熵原理立即可以得到

$$P(\underline{q}) = N \exp [-\underline{q}_T M \underline{q}] \quad (12.3)$$

式中我们已将式(12.1)，(12.2)作为约束。拉格朗日乘子可安排在一矩阵  $M$  内，该矩阵通过下式与关联矩阵  $Q$  相联系，

$$M = Q^{-1} \quad (12.4)$$

而  $Q$  则由式 (12.1) 确定。

现在我们变换式 (12.3)，其方法与 6.3 节导出和临近非平衡相变的模相关的模式所采用的方法大致相同。我们先求适合于方程

$$Q\tilde{u}_i = \lambda_i \tilde{u}_i \quad (12.5)$$

的本征矢量 $\tilde{u}_i$ 及本征值 $\lambda_i$ 。这些本征矢量使矩阵 $Q$ 对角化。我们打算证明这些本征矢量也会使矩阵 $M$ 对角化。为此，我们把诸本征矢 $\tilde{u}_i$ 按

$$(\tilde{u}_1 \quad \tilde{u}_2 \cdots \tilde{u}_N) = U \quad (12.6)$$

汇集为矩阵 $U$ 。 $U$ 的矩阵元由

$$u_{ki} = U_{ki} \quad (12.7)$$

给定。指标 $k$ 用以区分每一矢量的各分量，而指标 $i$ 则用来区分不同的本征矢量。利用矩阵 $U$ ，可将式(12.5)所示的各个方程汇总为矩阵方程

$$QU = UA \quad (12.8)$$

式中 $A$ 具有分量

$$A_{ij} = \lambda_j \delta_{ij} \quad (12.9)$$

矩阵方程(12.8)可用分量表示为

$$\sum_k Q_{ik} U_{kj} = \sum_k u_{ki} A_{kj} = \lambda_j u_{ji} \quad (12.10)$$

在下面的讨论中我们将假定 $Q$ 为非奇异矩阵，这已隐含于式(12.4)中。现在我们把方程(12.8)左乘以 $Q^{-1}$ ，从而得到

$$U = Q^{-1}UA \quad (12.11)$$

再将该方程右乘以 $A^{-1}$ ，可得

$$UA^{-1} = Q^{-1}U \quad (12.12)$$

据式(12.4)，方程(12.2)可改写为如下形式，

$$MU = U\tilde{A} \quad (12.13)$$

上式中使用了缩写符号

$$\tilde{A} = A^{-1} \quad (12.14)$$

Q为一对称矩阵, 故

$$U^T U = 1 \quad (12.15)$$

成立。式中 $U^T$ 为 $U$ 的转置矩阵。现在我们把一般矢量 $\underline{q}$ 展开成本征矢量组 $\underline{q}_j$ 。这种分解可用如下的普遍形式表示,

$$\underline{q} = U \underline{\xi} \quad (12.16)$$

或写为更一般的显式

$$\underline{q} = \sum_j \xi_j \underline{u}_j \quad (12.17)$$

应当注意, 矢量 $\underline{q}$ 为随机变量, 而本征矢量 $\underline{u}_j$ 却是决定论的。因此, 式(12.16)或式(12.17)中的矢量 $\underline{\xi}$ 就必须具有随机变量的分量。现在我们把式(12.3)中的指数, 即

$$\underline{q}^T M \underline{q} \quad (12.18)$$

变换为新变量 $\underline{\xi}$ 的表达式, 为此可采用式(12.17)的转置矢量

$$\underline{q}^T = \underline{\xi}^T U^T \quad (12.19)$$

这样, 我们就可得到

$$\underline{q}^T M \underline{q} = \underline{\xi}^T U^T M U \underline{\xi} \quad (12.20)$$

再利用式(12.13)及式(12.14)把式(12.20)转换为

$$M U = \underline{\xi}^T U^T U \Lambda^{-1} \underline{\xi} \quad (12.21)$$

据式(12.15)及式(12.14), 上式还可化简为

$$\underline{q}^T M \underline{q} = \underline{\xi}^T \tilde{A} \underline{\xi} \quad (12.22)$$

众所周知, 根据概率分布函数的变换理论, 我们不仅必须变换概率分布函数的指数, 还必须变换其体积元。后者的变换因子由雅可比行列式给出, 在这种情况下, 可以证明

$$\text{Det } U = 1 \quad (12.23)$$

因此，该变换使体积元不变。用这种方式代替式(12.3)，可得到新的概率分布函数，即

$$P(\underline{\xi}) = N \exp \left[ - \underline{\xi}^T \tilde{A} \underline{\xi} \right] \quad (12.24)$$

根据 $\tilde{A}$ 为对角矩阵这一事实，我们可将式(12.14)写成如下形式。

$$P(\underline{\xi}) = N \exp \left[ - \sum_i \tilde{\lambda}_i \xi_i^2 \right] \quad (12.25)$$

其右端为一乘积

$$P(\underline{\xi}) = \prod_i P_i(\xi_i) \quad (12.26)$$

式中每一个因子由下式给定，

$$P_i(\xi_i) = N_i \exp \left[ - \tilde{\lambda}_i \xi_i^2 \right] \quad (12.27)$$

选择适当的归一化因子 $N_i$ ，使 $P_i$ 归一。把特定形式的关系式(12.26)与式(12.27)联合起来，可得关联函数

$$\langle \xi_i \xi_k \rangle = \lambda_i \delta_{ik} \quad (12.28)$$

即是说关联矩阵现在已对角化了。进一步，我们还要注意，最小的 $\tilde{\lambda}_i$ 对应于最大的 $\lambda_i$ 。

现在我们研究，把 $\underline{q}$ 变换为与信息有关的变量 $\underline{\xi}$ ，我们究竟得到些什么。信息可由下式极普遍地给出，

$$i = - \int d^N q P(\underline{q}) \ln P(\underline{q}) \quad (12.29)$$

由于上面陈述的理由，这个式子可变换为如下的形式，

$$i = - \int d^N \xi \tilde{P}(\underline{\xi}) \ln \tilde{P}(\underline{\xi}) \quad (12.30)$$

由式(12.25)，式(12.30)可改写为各种单个贡献之和，即

$$i = - \sum_j \int d\xi_j P_j(\xi_j) \ln P_j(\xi_j) \quad (12.31)$$

或表示为缩写形式

$$i = \sum_j i_j \quad (12.32)$$

再利用式 (12.27), 可直接计算信息量并得到用本征值  $\lambda_j$  来表示的关系式,

$$i_j = -\frac{1}{2} \ln \lambda_j + \frac{1}{2} \ln \pi + \frac{1}{2} \quad (12.33)$$

或者, 由于关系式 (12.14) 成立, 我们还可把信息量  $i_j$  改写为

$$i_j = \frac{1}{2} \ln \lambda_j + \frac{1}{2} \ln \pi + \frac{1}{2} \quad (12.34)$$

根据式 (12.33) 及式 (12.34), 可以认识到, 最大信息即是  $\lambda_j$  取最大值的  $i_j$ , 而总信息则表现为无关联的信息之和。现在让我们来讨论这些结果对于模式识别意味着什么。

我们从一组具体选择的特征元开始, 给它赋予一个矢量  $\underline{q}$ , 借助于形如式 (12.17) 的特定的线性组合, 现在作出了一种不同的特征元组选择, 或者说抽取了一类新的特征元。这些新的特征元具有非常有趣的性质, 即信息为式 (12.34) 所表示的各项贡献之和。这样我们就可和数字模式识别中所用的方法相联系了, 展式 (12.17) 中之  $\underline{u}$  按式 (12.5) 加以选择时称卡若伦-洛弗 (Karhunen-Loeve) 展开式。就这个意义上说,  $i$  称为总体熵。我们得到的结果可按下面的方式加以解释:

卡若伦-洛弗展开为我们提供了把模式分解为无关联分量 (特征元) 的一种方法。正如下面将讨论的那样, 用这种方式减少特征分量数是可以办到的。适于卡若伦-洛弗展开的坐标系  $\underline{u}$  是决定性的。为构筑模式所需的一切信息都包含在这些分量之

中，更确切地说包含在这些系数之中。有最大 $\lambda_i$ 值的系数包含着与模式矢量 $\underline{q}$ 有关的最多的信息。因此，为了识别一种模式，我们只需要考虑具有最大本征值的系数，而具有小方差（恒等于本征值）的系数所转移掉的信息的量是可以忽略的。换言之，我们可以说， $\underline{q}$ 变换为 $K$ - $L$ 坐标系，产生了一种信息压缩。一旦 $K$ - $L$ 坐标系被确定以后，它对我们想加以分类的一切模式都是等效的。

由此可见，一切可加以辨认的信息必定被展式的系数 $\xi_i$ 所携带，因为绝大多数信息已被压缩到少数几个这类系数之中。在维数较低的特征元空间中，执行相应的决定，从计算上来看，问题的复杂程度较低。信息的使用还另有其极微妙的特点。诸本征值 $\lambda_i$ 的离差越大，相应的熵 $i$ 就越小，而信息的压缩就越大。反过来说，当 $\lambda$ 的值均一分布时，不确定性就最大，熵也最大；而对一个系数 $\xi_i$ 的观测导致的不确定性或熵的减少只能是很小的。

用非平衡相变的观点去解释这种方法是饶有兴味的。在这方面我们已经看到，临近相变点时，序参量的涨落的大幅度增长，与其它一切模相比，已达到相当大的程度。回顾这种解释，我们可以说，具有最大方差的，或者具有最大 $\lambda_i$ 值的诸 $\xi_i$ 实际上与序参量对应，并决定着模式。

对于导致模式形成的动力学过程，我们可用一种自然的方式去区分支配模（即序参量与伺服模）；然而，在一般情况下，我们究竟能在多大程度上进行类似的区分，即是否存在一些受制于支配性特征元的非支配性特征元，还是一个未解决的问题。或许，存在着一些类型，这种论述对其成立；但也可能对另外一些例子，该论述又变得不适用了。





## 第十三章 量子系统

---

### 13.1 为什么要研究量子信息理论

迄今为止，我们已研究过了具有宏观维度的经典系统。在本章中，我们想讨论一下必须借助量子理论来描述的各种体系。作为例证先考查一些量子效应起作用的体系。第一个例子就是激光器，其单个原子都必须用量子力学加以处理，而且其中的场也是由自发或受激辐射之类的量子力学事件产生的。因此，关于激光器的严格理论必须建立在量子力学的基础上。在前面的章节中，基于所谓的半经典处理，我们建立的激光理论，实际上完全可以从彻底的量子力学处理出发，利用所谓的量子经典对应方法导出。对此，我们不加以详细讨论。

另一个更重要的例子是计算机元件。这些元件的制作已日趋小型化，量子效应将不可避免地变得极其重要。因而，我们必须研讨捕陷在所谓量子势阱中的一个或几个电子的运动情况，从而可面临耿氏振荡器中那种电子流振荡，或者半导体中那种由电子

形成的丝极之类的情况。在各种捕陷状态中的电子或空穴可贮存信息，而电子则可能形成时间或空间结构。生物分子也是可以贮存信息的，我们可以预料，有可能制作一些元件，这些元件能将生物分子和半导体结合起来，贮存和处理信息。因此，我们必须研究，究竟能否将以前得到的结果推广到在这些微观元件中进行的量子力学过程中去。

下面我们将会看到，在很大程度上，这一目标是可以达到的。我们假定读者对量子理论的基本概念是熟悉的，而且，作为开端我们将和读者一起去回顾量子理论的一些最显著的特征，然后逐渐进入论题。为了从经典物理学过渡到量子力学，我们必须把诸如粒子的位置和动量这些观测量 $q_i$ 用相应的算符代替。在下面的讨论中，我们仍将用 $q_i$ 这个符号（参看表13.1）。以这种方式可以将经典力学，甚至电磁场理论转换为量子力学。为此，必须用一个算符 $E(x)$ 来替换位置为 $x$ 处的电场的幅值 $E(x)$ 。当我们讨论薛定格方程的解——波函数 $\psi(x)$ 时，由于相应的场是一种经典场，也须代之以算符 $\hat{\psi}(x)$ ，这就是所谓的二次量子化程序，电子密度从而变成了一个算符（对照表13.1）。

表13.1

经典的	量子力学的
观测量 $q_i$	算符 $\hat{q}_i$
在位置 $x$ 处的场振幅 $E(x)$	$E(x)$
电子波函数 $\psi(x)$	$\hat{\psi}(x)$
电子密度 $\tilde{\rho}(x) = \psi^*(x)\psi(x)$	$\tilde{\rho}(x) = \hat{\psi}^+(x)\hat{\psi}(x)$
分布函数 $f(q)$	密度矩阵 $\rho$
矩	

$$f_i^{(1)} = \langle q_i \rangle = \int f(q) q_i d^N q \quad \langle q_i \rangle = \text{tr}(\hat{q}_i \rho)$$

续 表

$$f_{ij}^{(2)} = \langle q_i, q_j \rangle = \int f(q) q_i q_j d^N q \quad \text{tr}(q_i q_j \rho)$$

$$i = - \int f \ln f d^N q \quad i = - \text{tr}(\rho \ln \rho)$$

现在让我们再向前跨出更重要的一步，去考虑一下统计力学。在第二章中我们曾讨论过分布函数  $f(q)$ ，在量子理论中该分布函数必须用密度矩阵  $\rho$  代替。分布函数遵从福克-普朗克方程，而密度矩阵却遵从所谓的主方程。下面我们将致力于用宏观方法推导密度矩阵  $\rho$  的显式表达式，这与过去进行的分布函数  $f(q)$  的“宏观”推导十分类似。因此，在此我们不打算详述密度矩阵方程或主方程的形式，而宁愿去推敲另外一种与矩有关的类推法。经典理论中的矩的两个示例列在表13.1左边倒数第三行及倒数第二行中。

表13.1右端的各相应表达式为我们指出了量子理论中相应的矩的定义。缩写符号  $\text{tr}$  表示迹，这是矩阵论中一个众所周知的概念。如果有一个矩阵  $A$ ，矩阵元为  $a_{ii}$ ，则该矩阵之迹定义为

$$\text{tr } A = \sum_i a_{ii} \quad (13.1)$$

当我们企图把经典物理学的结果转换为量子力学的表述时，困难在于，至少一般地说来算符  $q_i$  是不对易的，亦即一般而论下面的关系是成立的

$$q_i q_j - q_j q_i \neq 0, \quad i \neq j \quad (13.2)$$

大家从量子力学知道，算符  $q_i$  最终是要作用在一个波函数上的。但是，由于诸算符在同一波函数的作用顺序不同，最后将得到不同的波函数。正因为如此，小心对待各算符出现的顺序就显得特别重要了。但是，应该注意，在以下大多数的问题中，我们使

用的是算符 $q_i, q_j$ 的对称积,因而各算符的顺序已不再重要。鉴于经典力学和量子力学表达式形式上的类似,人们立即会联想到在量子论中应如何定义信息,这个定义已列在表13.1的最后一行中。

作了这些预备以后,现在我们就可以转向对最大信息原理的量子力学表述的讨论了。

## 13.2 最大信息原理

本节我们将使用表13.1中所示的转换方案,并以字母 $S$ 表示信息或熵。与最大信息原理有关的进一步的转换步骤总结在表13.2中。根据表13.1的末行所示的公式,要求在经典的或量子力学的情况下, $S$ 的相应表达式为给定约束下的极大值。给定的约束条件列在表13.2的第二行中,而归一化将如第三行的公式所示。为了求出极大值,我们将再次使用拉格朗日乘子。为此,将第二行中的表达式乘以 $\lambda_k$ ,同时把末行之表达式乘以 $\lambda-1$ ;然后从 $S$ 中减去所得的诸表达式,让最后所得的表达式变分为零,即得

$$\delta\{-\text{tr}(\rho \ln \rho) - (\lambda-1)\text{tr}\rho - \sum \lambda_k \text{tr}(g^{(k)})\}\rho = 0 \quad (13.3)$$

表13.2 最大信息原理

经典的	量子力学的
$S = - \int f \ln f d^N q = \text{极大!}$	$S = - \text{tr}(\rho \ln \rho) = \text{极大!}$
约束	
$f^{(k)} = \int f g^{(k)}(\underline{q}) dV$	$f^{(k)} = \text{tr}(g^{(k)}(\underline{q})\rho)$
归一化	
$\int f d^N q = 1$	$\text{tr} \rho = 1$

必须小心地进行变分运算。因为一般说来，变分 $\delta\rho$ 与 $\rho$ 是不可对易的，

$$[\rho, \delta\rho] \equiv \rho\delta\rho - \delta\rho\rho \neq 0 \quad (13.4)$$

另一方面，我们还必须使用迹的一个重要性质，即其轮换性

$$\text{tr}(AB) = \text{tr}(BA) \quad (13.5)$$

为了进行式(13.3)中的变分运算，可逐一讨论其所含之各项。

我们首先从变分

$$\delta\text{tr}\rho \quad (13.6)$$

开始运算。进行这个变分运算意味着，密度矩阵取值 $\rho$ 和 $\rho + \delta\rho$ 时，将对应的 $\text{tr}\rho$ 值加以比较，这里 $\delta\rho$ 很小。因而式(13.6)的变分决定于

$$\delta\text{tr}\rho = \text{tr}(\rho + \delta\rho) - \text{tr}(\rho) = \text{tr}\delta\rho \quad (13.7)$$

在上面的连等式中，为了得到最后一个等式，我们利用了求迹运算的线性特征。用类似的方式立即可以求得

$$\delta\text{tr}(g^{(k)}\rho) = \text{tr}(g^{(k)}\delta\rho) \quad (13.8)$$

而式(13.3)中第一项变分，由于式(13.4)所示的不可对易性，计算时需加小心。根据定义，迹的变分是对 $\rho + \delta\rho$ 及 $\rho$ 之迹求差，因而我们可得

$$\delta\text{tr}(\rho \ln \rho) = \text{tr}\{(\rho + \delta\rho) \ln(\rho + \delta\rho) - \rho \ln \rho\} \quad (13.9)$$

用先减去再加上一项的办法，式(13.9)可转换为

$$\begin{aligned} & \text{tr}\{(\rho + \delta\rho) \ln(\rho + \delta\rho) - \rho \ln(\rho + \delta\rho)\} \\ & + \text{tr}\{\rho \ln(\rho + \delta\rho) - \rho \ln \rho\} \end{aligned} \quad (13.10)$$

对第一个差值加以计算，仅保留 $\delta\rho$ 的线性项，立即可给出

$$\text{tr}\{\delta\rho \ln(\rho + \delta\rho)\} = \text{tr}(\delta\rho \ln \rho) \quad (13.11)$$

为计算第二个差值，我们可采用一种技巧，把式(13.12)左端写成

一种很特殊的形式，这样就可用通常的办法把对数展成泰勒级数

$$\ln \rho = \ln \langle 1 + \rho - 1 \rangle = \sum_{v=0}^{\infty} C_v (\rho - 1)^v \quad (13.12)$$

用完全相同的方式我们可得

$$\ln (\rho + \delta \rho) = \sum_{v=0}^{\infty} C_v (\rho - 1 + \delta \rho)^v \quad (13.13)$$

现在让我们来讨论上面的和式中的一个单项，从  $v = 1$  开始，即有

$$\rho - 1 + \delta \rho \quad (13.14)$$

对  $v = 2$ ，我们很容易得到

$$(\rho - 1)^2 + (\rho - 1)\delta \rho + \delta \rho(\rho - 1) + (\delta \rho)^2 \quad (13.15)$$

在上式中我们已注意了算符  $\rho - 1$  及  $\delta \rho$  的严格顺序。一般项  $v$ ，当保持其主要项时，可写为

$$\begin{aligned} & (\rho - 1)^v + (\rho - 1)^{v-1}\delta \rho + (\rho - 1)^{v-2}\delta \rho(\rho - 1) \\ & + (\rho - 1)^{v-3}\delta \rho(\rho - 1)^2 + \dots \\ & \dots + \delta \rho(\rho - 1)^{v-1} \\ & + (\delta \rho)^2 + \dots \end{aligned} \quad (13.16)$$

在下面的讨论中，我们仅保留与  $\delta \rho$  无关的或线性相关的各项。现在将公项 (13.16) 这种近似形式左乘以  $\rho$  并求其迹。为揭示问题的本质，我们取式 (13.16) 之任意项为代表，由此得

$$\text{tr}\{\rho(\rho - 1)^{v-1}\delta \rho(\rho - 1)^v\} \quad (13.17)$$

利用式 (13.5) 所示的轮换性，进一步可得

$$(13.17) = \text{tr}\{\delta \rho (\rho - 1)^v \rho(\rho - 1)^{v-1}\} \quad (13.18)$$

由于  $\rho$  可与  $\rho - 1$  交换，最后我们即可把式 (13.17) 的结果写为如

下形式:

$$\text{tr}\{\delta\rho \rho^{<\rho-1>^{-1}}\} \quad (13.19)$$

上面的转换对我们有如下的启示: 求迹运算会得到如象 $\rho$ 与 $\delta\rho$ 对易时那样的结果。因此, 从现在起, 在计算

$$\text{tr}\{\rho \ln(\rho + \delta\rho) - \rho \ln\rho\} \quad (13.20)$$

时, 我们将略去一切代数细节而仅仅假定 $\rho$ 与 $\delta\rho$ 在运算前可以交换, 换句话说, 为了计算式(13.20)我们可以把 $\rho$ 及 $\delta\rho$ 当成 $c$ -数来处理。利用通常的对数性质就可以代替式(13.20)而得到

$$\text{tr}\{\rho \ln(1 + \delta\rho\rho^{-1})\} \quad (13.21)$$

或者只保留最主要的项

$$(13.20) = \text{tr} \delta\rho \quad (13.22)$$

根据式(13.7), 式(13.8)及式(13.20), 式(13.22)的诸结果我们即可计算式(13.3), 从而得到

$$\{-\text{tr}(\delta\rho \ln \rho) - \lambda \text{tr} \delta\rho - \sum_k \lambda_k \text{tr}(\delta\rho g^{(k)})\} = 0 \quad (13.23)$$

为具体地计算变分 $\delta\rho$ , 我们将 $\rho$ 的任何一种表象, 对应于一组可构成分离或连续谱的本征态。使用这样一种表象就可以把 $\rho$ ,  $\delta\rho$ 及 $g^{(k)}$ 表示为矩阵。

现在利用迹的定义式(13.1)写出相应的各种结果。逐项写出式(13.23)结果, 可得

$$\sum_j (\delta\rho \ln\rho)_{jj} = \sum_{jj} (\delta\rho)_{jj} (\ln\rho)_{jj} \quad (13.24)$$

$$\sum_j (\delta\rho)_{jj} \quad (13.25)$$

$$\sum_{jj} (\delta\rho)_{jj} (g^{(k)})_{jj} \quad (13.26)$$

把各相应的表达式代入式(13.23), 我们即可得

$$\sum_j (\delta\rho)_{ji} \{ (\ln\rho)_{ij} + \lambda\delta_{ij} + \sum_k \lambda_k (g^{(k)})_{ij} \} = 0 \quad (13.27)$$

在上式中使用了柯氏符号, 当  $l = j$  时,  $\delta_{ij} = 1$ , 而当  $l \neq j$  时,  $\delta_{ij} = 0$ 。由于运用了拉格朗日乘子我们还可假定, 各变分  $(\delta\rho)_{ij}$  是彼此独立的。其结果, 只有当花括号内之运算对每一对  $l, j$  皆为零时, 方程 (13.27) 才可能成立。可认为整个括号是标号为  $l, j$  的一矩阵元, 因而式 (13.27) 中的花括号为零与如下的矩阵方程等价,

$$\ln\rho + \lambda I + \sum_k \lambda_k g^{(k)} = 0 \quad (13.28)$$

这里,  $I$  为单位矩阵。该矩阵或算子方程具有如下的解,

$$\rho = \exp\{-\lambda - \sum_k \lambda_k g^{(k)}\} \quad (13.29)$$

现在, 拉格朗日乘子  $\lambda$  和  $\lambda_k$  至少在原则上可根据归一化条件

$$\text{tr}\rho = 1 \quad (13.30)$$

及约束

$$\text{tr}(g^{(k)}\rho) = f^{(k)} \quad (13.31)$$

加以确定。把式 (13.29) 代入式 (13.30), 可得

$$\text{tr}\rho = e^{-\lambda} \text{tr}\{\exp[-\sum_k \lambda_k g^{(k)}]\} = 1 \quad (13.32)$$

这就立即导致了下面的关系式成立,

$$e^{\lambda} = Z = \text{tr}\{\exp[-\sum_k \lambda_k g^{(k)}]\} \quad (13.33)$$

这个式子也包含了量子力学配分函数  $Z$  的定义。我们在 3.3 节中导出的很多关系式都可以转换为量子力学的表达式, 但正如读者将会注意到的那样, 我们决不能掉以轻心, 而必须采用某些窍门。在这里我们将不对转换过程的全貌进行表述, 而只希望给读



者指出一些重要步骤，并写出一重要结果，为了推导关系式 (3.44) 及 (3.45) 中的量子力学模拟式，我们必须学会如何对含有不可对易算子的指数函数进行微分。该节余下的部分略微偏重于数学技巧，对这些技术细节不感兴趣的读者可以只注意最后得到的公式 (13.43)，然后直接转入下一节就行了。

现在明确一下我们的目的，我们希望计算的是

$$-\frac{\partial}{\partial \lambda_i} \text{tr} \rho \quad (13.34)$$

亦即，我们想探寻如何对求迹运算符号下的密度矩阵  $\rho$  微分。为此，我们得讨论出现在下式左端的指数函数

$$\exp [A + \lambda' B] = T \exp \left[ \int_0^1 dt (A_t + \lambda' B_t) \right] \quad (13.35)$$

该函数含有一个参数  $\lambda'$ ，我们希望将指数函数对这一参数微分。但是，由于已假定  $A$  和  $B$  是不可对易的，该微分运算不可能以通常的方式来完成。更确切地说，我们必须使用一种特殊的技巧，即引入所谓时序算符。为此我们将以标号  $t$  来标记  $A$  和  $B$ ，同时在方程 (13.35) 中，以右端的表达式代替左端的表达式。 $T$  意味着时序。当我们计算指数函数时，具有较小  $t$  指标的算符必须先于具有较大  $t$  值的算符参与运算，在这种规定下，就可以把  $A_t$  和  $B_t$  当成好似可以对易的一样加以处理。因此我们可把式 (13.35) 写成如下形式，

$$T \exp \left[ \int_0^1 dt A_t \right] \exp \left[ \int_0^1 \lambda' dt B_t \right] \quad (13.36)$$

现在将式 (13.35)，或等价地将式 (13.36) 对  $\lambda'$  微分，即

$$\frac{d}{d\lambda'} T \exp \left[ \int_0^1 dt (A_t + \lambda' B_t) \right] \quad (13.37)$$

于是得到

$$T \int_0^1 B_{i1} dt' \exp \left[ \int_0^1 dt (A + \lambda' B)_{i1} \right] \quad (13.38)$$

根据约定的时序，出现在指数函数中的一些与  $t < t'$  对应的算符必须先于  $B_{i1}$  使用；而另一方面，那些与  $t > t'$  对应的算符则必须后于  $B_{i1}$  运算。这样，我们便可用分裂指数函数的办法具体地按时间顺序排列，从而得到

$$\begin{aligned} & \int_0^1 dt' \exp \left[ \int_{t'}^1 dt (A + \lambda' B)_{i1} \right] B_{i1} \\ & \times \exp \left[ \int_0^{t'} dt (A + \lambda' B)_{i1} \right] \end{aligned} \quad (13.39)$$

既然单时间次序的排列已经给定，我们即可再用没有时间标号  $t$  的算符去代换算符  $A_{i1}$  和  $B_{i1}$ ，并进行积分。这就得到

$$\int_0^1 dt' \exp [(1-t')(A + \lambda' B)] B \exp [t'(A + \lambda' B)] \quad (13.40)$$

现在取式 (13.40) 之迹，同时利用轮换性 (13.5)，把式 (13.40) 中的第一个指数函数移到右端。这两个指数函数可汇集为一单指数函数，而不再与  $t'$  有关。这样，我们便可立即对  $t'$  积分，而得到一。进一步注意到微分运算和求迹运算可以相互交换，可得如下的最后结果，

$$\frac{d}{d\lambda'} \text{tr} \{ \exp[A + \lambda' B] \} = \text{tr} \{ B \exp[A + \lambda' B] \} \quad (13.41)$$

为了达到我们最初的目的，需要将  $\rho$  的迹对  $\lambda_i$  微分。为此，可以认为  $B$  即是  $g^{(i)}$ ，而  $\rho$  的指数函数的其余部分是  $A$ ，从而可以很快得到所希望的结果

$$\frac{\partial}{\partial \lambda_k} \text{tr} \rho = \text{tr}(g^{(k)} \rho) \quad (13.42)$$

再利用  $\rho$  的显式及  $Z$  的定义，很容易求出最终的结果

$$f_k = \text{tr}\{g^{(k)} \rho\} = - \frac{\partial}{\partial \lambda_k} \ln Z \quad (13.43)$$

这个式子与 (3.45) 完全类似。当我们打算求  $Z$  对  $\lambda_k$  的二阶微商时，可以再次使用上面用过的某些特殊技巧，但最终结果看起来比 3.3 节相应的结果稍为复杂。

作为本节的结论，我们要提出，第四章对热力学的论述结果可以直接转换到量子力学问题中去。例如，当我们将体系能量作为约束时，可得

$$\rho = Z^{-1} \exp[-H/(kT)] \quad (13.44)$$

式中  $H$  为哈密顿算符， $k$  为玻尔兹曼常数，而  $T$  是绝对温度。配分函数  $Z$  的定义为

$$Z = \text{tr}\{\exp[-H/(kT)]\} \quad (13.45)$$

### 13.3 序参量、伺服模及构型

本节我们想究研一下，如何将第六章的结果转换到量子力学中去。为了达到这个目的，我们特别再次选用下面的约束

$$f_k = \text{tr}(g^{(k)} \rho) \quad (13.46)$$

即是说，我们将以  $f_k$  去代换对称化矩

$$\langle q_i \rangle \quad (13.47)$$

$$\langle q_i q_k + q_k q_i \rangle \quad (13.48)$$

等。在下面的讨论中，我们将一直考虑到第四阶的各种矩。因此，使信息最大的密度矩阵由下式给出，

$$\rho = \exp \left[ -\lambda - \sum_i \lambda_i q_i - \sum_{ij} \lambda_{ij} q_i q_j - \dots \right] \quad (13.49)$$

式中的指数含有关于 $q_i$ 的一直到四阶的式子。以后可将指数缩写为

$$V = [\dots] \quad (13.50)$$

我们将尽可能按与第六章相近的步骤讨论。为此，我们希望除掉式(13.49)，(13.50)中的线性项，因而作如下的假定，

$$q_i = c_i + \bar{q}_i \quad (13.51)$$

这里 $c_i$ 为 $c$ -数。把式(13.51)代入式(13.49)便会得到若干关于 $\bar{q}_i$ 的线性项，现在需要令它们的系数为零，即

$$\lambda_i + \sum_j (\lambda_{ij} c_j + \lambda_{ji} c_i) + \dots = 0 \quad (13.52)$$

很容易明白，如果我们要求

$$\frac{\partial V}{\partial q_i} = 0 \quad (13.53)$$

对 $q_i^0$ 等于零，从而可把 $q_i$ 当作 $c$ -数处理，式(13.52)即可成立。

我们作如下的等价替换

$$q_i^0 \longleftrightarrow \bar{q}_i = 0 \quad (13.54)$$

从而 $c_i$ 决定于

$$c_i = q_i^0 \quad (13.55)$$

用这种方式我们已把式(13.49)转换成为

$$\exp [-\bar{V}(\bar{q})] \quad (13.56)$$

此处，缩写式(13.50)用下式代替，

$$\bar{V}(\bar{q}) = \bar{\lambda} + \sum_{ij} \bar{\lambda}_{ij} \bar{q}_i \bar{q}_j + \dots \quad (13.57)$$

由于我们选取的是对称化矩，所以十分明显下式应当成立，

$$\lambda_{ji} = \lambda_{ij} \quad (13.58)$$

现在作如下假设，

$$q_i = \sum_k \xi_k v_{ki} \quad (13.59)$$

式中  $v_{ki}$  为  $c$ -数，而  $\xi_k$  是算符。将假设 (13.59) 代入式 (13.57)，可知，对线性项

$$\sum_{k,l} \xi_k \xi_l \left\{ \sum_{ij} \lambda_{ij} v_{ki} v_{lj} \right\} \quad (13.60)$$

成立。现在我们对  $v_{ki}$  的选择使下式成立，

$$\sum_{ij} \lambda_{ij} v_{ki} v_{lj} = \hat{\lambda}_k \delta_{kl} \quad (13.61)$$

$\hat{\lambda}_{ij}$  为一对称矩阵，使我们可将  $v_{ki}$  选为一实矩阵，同时使  $\lambda_k$  之值亦为实数。在这些转换条件下，式 (13.49) 由下式所代替，

$$\rho(\underline{\xi}_+, \underline{\xi}_-) = \exp \left[ -\lambda - \sum_+ \hat{\lambda}_+ \xi_+^2 - \sum_- \hat{\lambda}_- \xi_-^2 + V_+(\underline{\xi}_+) + V_-(\underline{\xi}_+, \underline{\xi}_-) \right] \quad (13.62)$$

式中我们已再度将正  $\lambda_k$  与负  $\lambda_k$  加以区分，这样一来就可以把序参量与伺服模清楚地加以区别。

现在以下面的方式引入与序参量  $\xi_k$  有关的密度矩阵，

$$\rho_+(\underline{\xi}_+) = \text{tr}_- \rho(\underline{\xi}_+, \underline{\xi}_-) \quad (13.63)$$

由于总迹是归一的，所以下式应当成立，

$$\text{tr } \rho_+ = 1 \quad (13.64)$$

用与第六章中类似的处理程序, 我们以下面的方式把与序参量算符和伺服模算符相关的密度矩 $\rho$ , 分裂为密度矩阵 (13.63) 与一种条件密度矩阵之积,

$$\rho(\underline{\xi}_n, \underline{\xi}_n) = \rho_n(\underline{\xi}_n) \rho_n(\underline{\xi}_n | \underline{\xi}_n) \quad (13.65)$$

当我们将上式两端对与算符 $\xi_n$ 有关的诸变量求迹, 根据定义, 左端即得到式 (13.63); 而右端则得到

$$\rho_n(\underline{\xi}_n) \text{tr}_n \rho_n(\underline{\xi}_n | \underline{\xi}_n) \quad (13.66)$$

把左右端加以比较, 立即可得如下的关系,

$$\text{tr}_n \rho_n(\underline{\xi}_n | \underline{\xi}_n) = 1 \quad (13.67)$$

现在我们希望通过变换把式 (13.62) 写成式 (13.65) 的形式, 由于诸算符 $\xi$ 不可对易, 这决不是一件无足轻重的事。为了完成这一任务我们将式 (13.62) 写为如下形式,

$$\rho(\underline{\xi}_n, \underline{\xi}_n) = \exp [\hat{V}_n(\underline{\xi}_n) + \hat{V}_n(\underline{\xi}_n, \underline{\xi}_n)] \quad (13.68)$$

上式的意义是十分明显的, 但究竟把 $\bar{\lambda}$ 放入 $V_n$ 还是 $V_n$ 之内却是未决的。

现在我们加上再减去同一个算符函数, 即可得

$$(13.68) = \exp [\hat{V}_n + h(\underline{\xi}_n) - h(\underline{\xi}_n) + \hat{V}_n(\underline{\xi}_n, \underline{\xi}_n)] \quad (13.69)$$

我们的目的是把式 (13.69) 改写为式 (13.65) 的形式, 为此需要一种辅助定理。考虑如下的表达式,

$$T \exp \left[ \int_0^1 dt (A_t + B_t) \right] \quad (13.70)$$

它是在式 (13.35) 中令 $\lambda = 1$ 而得到的。 $T$ 还是时序算符,  $A_t$ 及 $B_t$

是必须按时间次序排列的算符。现在我们利用费曼 (Feynman) 的分辨定理, 将式 (13.70) 改写为下面的形式,

$$\exp AT \exp \tilde{B} \quad (13.71)$$

式中

$$\tilde{B} = \int_0^1 dt (\exp[-At] B \exp[At]), \quad (13.72)$$

现在作如下的等价代换,

$$\hat{V}_n + h = A, \quad -h + \hat{V}_1 = B \quad (13.73)$$

这就使我们可立即导出

$$\rho_n(\xi_n) = \exp [\hat{V}_n + h] \quad (13.74)$$

而条件密度矩阵可以按下式确定,

$$\rho_n(\xi_n | \xi_1) = T \exp \tilde{B} \quad (13.75)$$

但仍然未知的算符函数  $h$  则必须以

$$\text{tr}_n \rho(\xi_n | \xi_1) = 1 \quad (13.76)$$

成立为前提才能确定。

让我们总结一下本节得到的结果。从一至四阶的矩算符出发, 借助于最大信息熵原理我们构筑了密度矩阵。因此, 用把双线性部分对角化的办法就可确认与序参量和伺服模对应的算符。最后我们可把  $\xi_n$  及  $\xi_1$  的联合密度矩阵改写为乘积, 其因子之一  $\rho_n$  是只与序参量有关的密度矩阵, 另一因子  $\rho_1$  则可被解释为条件密度矩阵。现在的情况与经典的情况完全类似, 例如我们可以假定,  $q_i$  为空间点  $i$  处的强度, 而将  $v_k$  理解为属于模  $k$  的空间点  $j$  处的构型。十分值得注意的是,  $v$  为一经典量, 它意味着出现的构型可用经典量加以描述, 因而毋需任何量子力学波函数的知识就

可加以解释。

在结束本章的时候,我们想证明一下,信息可分为只涉及序参量的部分和与条件密度矩阵有关的部分(即与伺服模有关的部分)。

#### 13.4 序参量与伺服模的信息

让我们从整个体系的信息开始讨论。整个体系的信息表达式为

$$i = \text{tr}(\rho \ln \rho) \quad (13.77)$$

将式(13.65)代入,该式中的因子分别按式(13.74)及式(13.75)来确定。另外,再利用关系式(13.69),我们即可得

$$(13.77) = \text{tr} \{ \rho_s \rho, [(\hat{V}_s + h) + (-h + \hat{V}_s)] \} \quad (13.78)$$

这个式子可分解为

$$\text{tr}_s \{ \rho_s (\text{tr}_s \rho_s) (\hat{V}_s + h) \} \quad (13.79)$$

及

$$\text{tr}_s \{ \rho_s \text{tr}_s \rho_s (-h + \hat{V}_s) \} \quad (13.80)$$

之和。由于式(13.67),表达式(13.79)可进一步简化为

$$\text{tr}_s \{ \rho_s (\hat{V}_s + h) \} \quad (13.81)$$

同时,可认为式(13.80)中s的迹是对应于伺服模s的信息。因为它决定于算符 $\xi_s$ ,所以这是一个信息的算符表达式。式(13.81)仅与序参量有关,我们将其缩写为 $i_s$ ,就可得到下面的最后结果,

$$i(13.77) = i_s + \text{tr}_s \{ \rho_s i_{s,s} (\xi_s) \} \quad (13.82)$$



这样，和经典的情况完全类似，把整个体系的信息分解为序参量的信息与涉及到伺服模的信息就是可能的了，正因为如此，就需要用序参量的分布函数或密度矩阵对后者加以平均。

作为本章的结论我们可以宣称，我们已能把在经典力学框架内获得的很多结果推广到使用量子力学算符的场合。特别重要的是，本章建立的方法使我们可借助于矩阵 $v_{ij}$ 去辨别量子力学系统中演化着的模式，从而在量子系统中识别出经典的构型。



## 第十四章 结语和展望

---

在结尾的这一篇章中，将论述本书中我们取得了哪些成果，在哪些方面还存在着悬而未决的问题。

第一章讨论了复杂系统的本性。为彻底认识它们，我们作了种种努力，涉及到它的许多方面，发现其最显著的特征似乎是本质上的不可穷尽性。因此，我们只好选择一些适合我们的目的或兴趣的特殊方面加以研究。根据这一观念，本书中选择了一个颇为一般的方面加以研讨，亦即把我们的注意力集中在使复杂系统宏观性能发生质变的方面，以图驾驭它们。更确切地说，我们研究了那些发生了所谓非平衡相变的情况。这种方法不仅使我们可以研讨诸如激光器和流体之类的物理系统，而且可以讨论象生物系统那样更复杂的系统。毋庸置疑，这种方法还有多种多样的进一步的应用。我们建立的方法也可用于振荡现象的研究，这已为有关激光的例子所证实，但是，如何向混沌系统推广，这个问题仍未解决。因此，混沌系统显然是一个值得进一步研究的领域。对于我们的研究，信息是一个媒介手段。正如我们在书中所见，

信息这个字眼在十分不同的意义下被使用着，因而我们不得不讨论这个词义的各个方面。特别是我们已看到，在申农意义上，信息的概念似乎在各种情况下都比熵的概念更为合理。这是基于如下事实：至少，在物理学中熵有着十分特殊的意义，而且总是与热平衡系统联系在一起。但当我们讨论远离平衡的物理系统或生物系统时，就不得不去重新解释一些结果，同时还必须采用新的约束使信息或（统计）熵极大化。

特别是，在某种意义上，我们已看到，一种新型的信息已经出现，它涉及到集体变量或序参量。这使人感到，应该把涉及序参量并且反映着体系集体性质的那部分信息称为“协同学信息”。同时，序参量也得到了“信息子”这样一种新的意义。

根据我们在研讨变量矩的基础上建立的一般方法，对通常意义上的那种信息，究竟可以揭示出哪一些类型。这是很重要的一个问题，我们已经看到，用这种方式可以确定序参量及其定态分布函数。但这类信息充分与否则决定于我们的目的。以激光器为例，定态分布函数完全足以表征与激光强度及涨落有关的定态性质。但是，该分布函数却并不能告诉我们有关激光的一个重要性质——时间相干性的任何一点内容。为此，人们必须非常细致地去研究动力学问题。在我们所建立的方法框架中，这类动力学特征可用与时间有关的矩及转移概率来推断。根据最大信息熵原理，我们可由它们推出所需的结果。一般地说，只要特定宏观数据的平均值已知，这一整套方法可使我们对微观事件及微观过程作出应有的估计。

这使我们得到以下很有吸引力的结论，在远离热平衡的系统中，甚至在非物理系统中，（申农）信息，与在热平衡或接近热

平衡系统中的熵的作用相同，可作为过程的起因。由于下述原因我不愿勉强附合这种解释：首先我们已认识到，分析的结果取决于适当选择的约束。我们已可能对以非平衡相变为特征的一类现象确定适宜的约束，同时也可对定态分布函数作出推断。就这方面而论，可以说我们能定义出一种势函数，体系在它的驱使下向特定的定态演化。但是，该定态势并不能给我们唯一的指示，以单一地导出决定动力学特征的福克-普朗克方程，这就是R.朗道尔（R. Landauer）屡次强调的一个方面。因此，信息（或熵），或者说得更恰当些，信息的极大化似乎并非是以唯一方式驱动体系的一条基本定律。另一方面，在发生分支的情况下，如何去构筑随机过程的势函数，格拉哈姆（Graham）及铁尔（Tel）等人已作过非常值得注意的研究。在此，我们并不打算将我们与他们的方法加以比较，遗留下的这一问题可供进一步研究之用。这些讨论对生物的进化或发育，换句话说对生物的种系发生及个体发育等问题都有根本性的重要意义。当然，问题是进化和发育究竟是否受制于极值原理（尤其是与象熵或信息这样的单个函数相联系的极值原理）。我们得到的结果可能会给人们一种启发：如果这样一种函数是存在的，我们付出的代价至少是对动力学特征的一无所知或知之甚少。不过，这样一种函数究竟是否真正存在，仍然还是一个没有解决的问题。从另一方面来说，选择矩作为非平衡相变中信息的适当约束这一事实，正是我们的方法的精髓所在。矩的行为反映了系统发生相干的趋势。因此，我们在本书中所研究的，正是复杂系统中通过自组织出现的相干性和宏观秩序。但是，是否可把生物学完全置于这种观点之下来加以考察，也还是一个悬而未决的问题。正如我们在前面说过的那样，复杂

系统的本性似乎是无穷尽的。

正因为如此，虽然一个深刻而基本的问题仍未解决，但我们在本书中所概述的方法，的确是一种以统一的观点去研究邻近非平衡相变的复杂系统的强有力工具。

## 参 考 文 献

---

### 复杂系统的挑战

#### 1.6 普适性的追求

##### 1. 热力学

##### 2. 统计物理学

- [1] Landau-Lifshitz, In Coures of Theoretical Physics, Vo1.5, Statistical Physics (Pergamon Press, London-Paris 1952) .
- [2] R. Becker, Theory of Heat (Springer, Berlin, Heidelberg, New York 1967) .
- [3] A. Münster, Statistical Thermodynamics, Vo1. 1 (Springer, Berlin-Heidelberg-New York 1969) .
- [4] H.B. Callen, Thermodynamics (Wiley, New York 1960) .
- [5] P.T. Landsberg, Thermodynamics (Wiley, New York 1961) .

- [6] R.Kubo, Thermodynamics (North Holland, Amsterdam 1968) .

### 3.协同学

- [7] H. Haken, Synergetics, An Introduction 3rd ed., Springer, Berlin, Heidelberg, New York 1983.
- [8] H. Haken, Advanced Synergetics, 2nd. corr. Print., Springer, Berlin, Heidelberg, New York 1987.

### 1.7 信息

#### 1.申农信息：意义的消除

- [9] C.E. Shannon, A Mathematical Theory of Communication, *Bell System Techn.*, I. **27**, 370—423, 623—656 (1948) .
- [10] C.E. Shannon, *Bell System Techn.*, I. **30**, 50 (1951) .
- [11] C.E. Shannon, W. Weaver, The Mathematical Theory of Communication (üniv. of Illin. Press, Urbana 1949) .
- [12] L. Boltzmann, Vorlesungen über Gastheorie, 2 Vols. (Leipzig 1896, 1898) .

#### 2.信息的效应

- [13] H. Haken, in Thermodynamics and Regulation of Biological Processes, I. Lamprecht and A. I. Zotin, eds, walter de Gruyter Co, Berlin, New York 1984.



### 3. 意义的自创生

- [14] H. Haken, paper presented at international conference, Tokyo, 1986.

也可参阅

- [15] H. Haken, in *Ordnung aus dem Chaos. Prinzipien der Selbstorganisation und Evolution des Lebens*, Bernd-Olaf Küppers (Hrsg.), Piper, München Zürich 1987.

### 从微观世界到宏观世界

- [1] H. Haken, *Synergetics, An Introduction*, 3rd., ed., Springer, Berlin, Heidelberg, New York (1983) .
- [2] H. Haken, *Advanced Synergetics*, 2nd, corr. print., Springer, Berlin, Heidelberg, New York 1987.

### 最大信息原理 (MIP)

#### 3.2 信息增益

- [1] S. Kulback, *Ann. Math. Statist.* **22**, 79 (1951) .
- [2] S. Kulback, *Information Theory and Statistics* (Wiley, New York 1951) .

#### 3.3 信息熵和约束条件

- [3] E.T. Jaynes, *Phys. Rev.*, **106**, 4,620 (1957) ; *Phys. Rev.*, **108**, 171 (1957) .
- [4] E.T. Jaynes, In *Delaware Seminar in the Foundations of Physics* (Springer, Berlin,

Heidelberg, New York 1967) .

- [5] E.T. Jaynes, *Am. Jour. Phys.*, **33**, 391—398  
(1965) .

- [6] E.T. Jaynes, In *The Maximum Entropy Formalism*,  
R.D. Levine and M. Tribus, Editors, MIT  
Press, Cambridge, Mass. (1978) .

物理学中的一个例子：热力学

最大信息原理在自组织系统中的应用

在英伽顿和他的合作者以及其它一些人的许多论文中，  
对吉尼斯原理在“几个温度”的系统中的早期应用加以  
了阐述。例如，可参见：

- [1] R.S. Ingarden, *Bull. Acad. Polon. Sci., Ser. Sci.  
Math. Astron. Phys.*, **11**, 541 (1963) .
- [2] R.S. Ingarden, *Fortschr. Phys.*, **12**, 567 (1964b) .
- [3] R.S. Ingarden, *Fortschr. Phys.*, **13**, 755 (1965a) .
- [4] R.S. Ingarden, *Acta Phys. Polon.*, **27**, 179  
(1965b) .
- [5] R.S. Ingarden, *Ann. Inst. Henri Poincare*, **8**, 1  
(1968a) .
- [6] R.S. Ingarden, *Bull. Acad. Polon. Sci., Ser.  
Sci. Math. Astron. Phys.*, **16**, 57 (1968b) .
- [7] R.S. Ingarden, *Acta Physica Polonica Vol.  
XXXVI, Fasc. 5 (11)*, 855 (1969)

本章和下一章主要的参考文献如下：

- [8] H. Haken, *Z. Physik B-Condensed Matter* **61**, 335

(1985)

- [9] H. Haken, Z. Physik B-Condensed Matter 63, 487

(1986)

### 5.2 自组织系统中的应用：单模激光

激光分布函数由以下文献导出：

- [10] H. Risken, Z. Phys., 186, 85 (1965) .

- [11] H. Haken, Laser Theory, Hdb. Phys., Vol. 25/2c, Springer, Berlin 1970, corr. Print. 1984.

### 5.3 没有位相关系的多模激光

- [12] H. Haken, Laser Theory, Hdb. phys., Vol. 25/2c, Springer Berlin 1970, corr. print. 1984.

非平衡相变的最大信息原理：序参量、伺服模与模式的确  
定

参看最大信息原理在自组织系统中的应用这一章的参考  
资料

自组织系统在接近不稳定点时的信息、信息增益和效率

7.1—7.8节的主要文献是：

- [1] H. Haken, Z. Physik B-Condensed Matter, 61, 329  
(1985)

- [2] H. Haken, Z. Physik, B-Condensed Matter, 62, 255  
(1986)

7.9节的参考文献是：

- [3] J. A. Tuszynski, Z. Phys. B-Condensed Matter  
65, 375 (1987) .

- [4] W. Witschel, Chemical Physics 50 (1980) 265,

North-Holland Publishing Company.

- [5] W. Witschel, J. Bohmann, *J. Phys. A: Math. Gen.*, **13** (1980) 2735 (Great Britain) .

#### 7.10 柯里蒙托维奇的S-定理

- [6] Yu.L. Klimontovich, *Pisma v JTP* **9**, 1412(1983)  
(in Russian) .
- [7] Yu.L. Klimontovich, *Physica A*, **142**, 390(1987)
- [8] Yu.L. Klimontovich, *Z. Phys. B-Condensed Matter*, **65**, 125 (1987) .

#### 拉氏乘子的直接确定

8.2节的参考文献是:

H. Haken and J. Wehberuss, 尚未发表的结果。

随机过程的无偏模型化: 怎样推测路径积分, 福克-普朗克方程  
和朗之万-伊藤方程

H. Haken, *Z. Physik B-Condensed Matter* **63**,  
505 (1986) .

#### 应用于一些物理系统

##### 10.1 有相位关系的多模激光器

- [1] H. Haken, in *Frontiers in Quantum Optics*,  
E.R. Pike, S. Sarkar, eds. Adam Hilger, Bristol  
and Boston 1986.
- [2] H. Haken, *Laser Theory, Hdb. phys.*, Vol.  
25/2c, Springer, Berlin 1970, corr. print. 1984.

##### 10.2 包含极化和反转的单模激光器

- [3] H. Haken, in *Frontiers in quantum Optics*, E.R.

pike, S. Sarkar, eds. Adam Hilger, Bristol and Boston 1986.

### 生物学中的行为模式转变·一个例子：手的运动

实验部分可参见：

- [1] J.A.S. Kelso, *Bull. Psychon. Soc.*, **18**, 63(1981),
- [2] J.A.S. Kelso, *Am. J. Phys. Reg. Integr. Comp.*, **15**, R1000—1004 (1984) .

以下文献发展了借助于决定论方程或随机方程的理论

- [3] H. Haken, J.A.S. kelso, H.H. Bunz, *Biol. Cybern.*, **51**, 347 (1985) .
- [4] G. Schöner, H. Haken, J.A.S. Kelso, *Biol. Cybern.*, **53**, 247 (1986) .

### 模式识别

P.A. Devijver and J. Kittler, *Pattern Recognition, A Statistical Approach* Prentice Hall International, Englewood Cliffs, New Jersey, 1982.

### 量子系统

### 结语和展望

- [1] R. Graham, T. Tel, *Phys. Rev.*, **35A**, 1328 (1987) .
- [2] R. Graham, and T. Tel, *Phys. Rev. Lett.* **52**, 9 (1984) .
- [3] R. Graham, and T. Tel, *J. Stat. Phys.* **35**, 729 (1984) ; **37**, 709 (1984) .

- [4] R. Graham, T. Tel, *Phys. Rev.* **A31**, 1109  
(1985) .
- [5] R. Graham, D. Roekaerts, and T. Tel, *Phys.*  
*Rev.* **A31**, 3364 (1985) .
- [6] M.I. Freidlin, and A.D. Wentzell, *Random*  
*Perturbations of Dynamical Systems* (Springer,  
New York, 1984) .
- [7] R. Landauer, *Relative Hability in the Dissipative*  
*State*, preprint 1978.