



# 统计物理选论

现代导引

陈童



Copyright © 2026 陈童

PUBLISHED BY PUBLISHER

下载网址: <http://newquanta.com/>

作者本人 (即陈童) 对本讲义保有所有版权, 目前读者可在上面的网站上自由下载, 也可以私自传播, 但仅限于学习或教学等非商业用途, 禁止私自以本讲义的部分或全部来获利, 一经发现, 作者将保留诉诸法律的权利。如果你有任何反馈, 或者发现书中的任何错误, 请给作者发邮件, [tongchen@ecut.edu.cn](mailto:tongchen@ecut.edu.cn)

2026 年 4 月


# Contents

<b>1</b>	<b>杨李相变理论</b> .....	<b>9</b>
1.1	液气相变简介	9
1.2	统计物理和相变	11
1.3	杨李理论基础	13
1.3.1	杨李理论的基本框架 .....	13
1.3.2	相变的物理图像 .....	17
1.3.3	对相变的进一步刻画 .....	19
1.4	伊辛模型与格子模型	20
1.5	单位圆定理	23
1.5.1	证明思路 .....	24
1.5.2	物理讨论 .....	25
<b>2</b>	<b>物质相的朗道范式</b> .....	<b>27</b>
2.1	引言	27
2.2	相与相变的基本概念	28
2.3	对称性自发破缺与序的扩张	29
2.3.1	相与对称性 .....	29
2.3.2	对称性自发破缺的机制 .....	30
2.3.3	相图的结构与子群格 .....	32

<b>2.4</b>	<b>连续对称性自发破缺</b>	<b>34</b>
2.4.1	序参量的涨落	34
2.4.2	铁磁相变	36
2.4.3	拓扑缺陷	37
2.4.4	物质相的朗道范式	39
<b>2.5</b>	<b>液晶的物理学</b>	<b>39</b>
2.5.1	朗道-de Gennes 理论	40
2.5.2	取向轴的形变	42
<b>2.6</b>	<b>结语：范式的辉煌与边界</b>	<b>43</b>
<b>3</b>	<b>重整化群理论</b>	<b>45</b>
<b>3.1</b>	<b>临界指数与普适性</b>	<b>46</b>
3.1.1	反常量纲	46
3.1.2	临界指数与普适性	48
<b>3.2</b>	<b>统计力学回顾</b>	<b>52</b>
<b>3.3</b>	<b>卡丹诺夫的物理图像</b>	<b>54</b>
<b>3.4</b>	<b>威尔逊重整化群理论</b>	<b>58</b>
3.4.1	重整化群变换的构建	58
3.4.2	流、不动点、普适性与标度	60
3.4.3	算符乘积展开	63
3.4.4	固定点附近的重整化群流	64
3.4.5	重整化群方程	66
<b>3.5</b>	<b>临界指数的计算</b>	<b>68</b>
3.5.1	经典平均场理论	68
3.5.2	积去小尺度自由度	71
3.5.3	从重整化群的物理实质到理解生命现象	74
<b>4</b>	<b>大 <math>N</math> 极限：从费曼图到随机矩阵理论</b>	<b>77</b>
<b>4.1</b>	<b>引言</b>	<b>77</b>
<b>4.2</b>	<b>中心极限定理与随机游走</b>	<b>78</b>
4.2.1	高尔顿钉板实验	78
4.2.2	随机变量的特征	78
4.2.3	中心极限定理的证明与随机游走	79
<b>4.3</b>	<b>扰动高斯分布与费曼图</b>	<b>80</b>
4.3.1	高斯分布	81
4.3.2	受扰动的高斯分布	82
4.3.3	圈展开	85

4.3.4	复变量 .....	86
4.3.5	微扰统计场论 .....	86
<b>4.4</b>	<b>大 N 展开与随机矩阵理论</b>	<b>87</b>
4.4.1	魏格纳半圆定律 .....	88
4.4.2	戴森气体 .....	91
<b>4.5</b>	<b>总结</b>	<b>94</b>
<b>5</b>	<b>KT 相变 .....</b>	<b>95</b>
<b>5.1</b>	<b>引言</b>	<b>95</b>
<b>5.2</b>	<b>两维没有连续对称性自发破缺</b>	<b>96</b>
<b>5.3</b>	<b>涡旋算符与 BKT 相变的正弦-戈登描述</b>	<b>98</b>
5.3.1	BKT 相变的连续场论描述 .....	98
5.3.2	阿贝尔对偶性 .....	99
5.3.3	涡旋算符与正弦-戈登描述 .....	100
<b>5.4</b>	<b>关于 BKT 相变的讨论</b>	<b>101</b>
<b>5.5</b>	<b>总结</b>	<b>103</b>
<b>6</b>	<b>两维伊辛模型的精确解 .....</b>	<b>107</b>
<b>6.1</b>	<b>模型的提出和伊辛的“错误证明”</b>	<b>107</b>
<b>6.2</b>	<b>两维伊辛模型, Kramers-Wannier 对偶</b>	<b>109</b>
<b>6.3</b>	<b>费米子方法</b>	<b>113</b>
<b>6.4</b>	<b>Grassmann 变量积分方法</b>	<b>123</b>





## 前言

本讲义是讲述一些统计物理专题，各章内容基本上相互独立。

统计物理学是一门令人敬畏的学科。它的目标极为宏大：从微观粒子那浩如烟海、看似混乱的自由度中，提炼出宏观世界简洁而普适的规律。然而，任何一个认真研习过统计物理的人大概都会承认，这也是一门对初学者极不友好的学问。它不像经典力学或电动力学那样，有着清晰可见的、从第一性原理出发的演绎路径。在这里，概念往往先于严密，直觉常常引领逻辑。我们谈论“系综”，却很难直观地“看到”它；我们使用“配分函数”，却常常忘记它背后那令人望而生畏的多重求和与积分。

在相当长的时期内，统计物理的教学陷入了一种略显尴尬的境地：要么停留在本科普物水平的热力学与初级统计介绍，要么直接跳入高深的场论方法，中间横亘着一道令人生畏的鸿沟。许多极具物理品味的学生，正是在这道鸿沟前丧失了继续探索的勇气。而事实上，统计物理学在二十世纪下半叶经历了数场深刻的革命——杨振宁和李政道的相变理论、朗道的对称性破缺范式、威尔逊的重整化群、Kosterlitz 和 Thouless 的拓扑相变——这些革命不仅重塑了我们对物质世界的理解，其思想脉络本身，也构成了一部惊心动魄的概念史诗。

我有幸在多年研读中，逐渐被这部史诗的恢弘与精妙所震撼。然而，当我想将这些心得分享给学生时，却发现很难找到一本合适的教材——它们要么太过专门，要么太过简略，鲜有作品致力于在“标准教材”与“前沿专题”之间架设一座平缓而坚实的桥梁。这本《统计物理选论：现代导引》，便是这种困境下的一份尝试。

本书取名“选论”，意在表明它并非一部面面俱到的教科书，而是一条精心选择的登山路径。它试图带领读者穿越统计物理现代发展中几处最壮丽的风景：从杨-李相变理论对相变本质的深刻洞察，到朗道范式以对称性破缺和序参量为支柱的宏大建构；从重整化群

对多尺度问题的优雅征服，到大  $N$  极限下图解方法与随机矩阵的奇妙交汇；从 KT 相变揭示的拓扑秩序，到二维伊辛模型精确解所展现的数学之美。这些章节虽各自独立，却共享着一种精神气质——它们都在追问同一个问题：宏观秩序如何从微观混沌中涌现？

在写作风格上，我选择了一条与标准教材有所不同的道路。我不追求事无巨细的推导（尽管关键步骤都有呈现），更不妄图罗列所有的知识点。我所渴望的，是捕捉那些伟大思想诞生时的原始冲动，还原那些关键转折处的物理直觉。本书中的许多段落，与其说是“讲解”，不如说是一种“重演”——我邀请读者与我一起，站在那些物理学巨匠的肩膀上，重新凝视他们当年面对的问题，感受那“山重水复疑无路，柳暗花明又一村”的智识喜悦。

本书默认读者已经掌握了热力学与初等统计物理的基本知识（如系综理论、理想气体、玻色-爱因斯坦与费米-狄拉克统计等），并对量子力学有一定基础。它不是为零起点者准备的，但我努力让每一个具有上述背景的物理系高年级本科生或低年级研究生，都能够循着书中的脉络，一步步进入统计物理现代理论的核心地带。对于那些已然熟悉这些专题的研究者，我也希望本书独特的叙事视角和某些技术细节的处理方式，能够带来新的启发。

我还想特别说明一点。在第三章的末尾，我用了少量篇幅讨论重整化群思想与生命现象的可能关联。这部分内容带有相当程度的思辨色彩，绝非成熟的物理理论。我将其纳入，并非声称找到了什么答案，而是想传递一种信念：统计物理处理多尺度复杂系统的思想框架，其力量或许远超出我们目前已经探明的疆域。在凝聚态物理和量子场论之外，是否还存在着这片思想大陆的“新世界”？这本身就是一个值得青年学子投身的迷人问题。

本书得以完成，离不开许多前辈和同行的杰出工作，我从他们的论文、专著和讲义中获益良多。各章末尾均已列出核心参考文献，此处不再重复。我还要特别感谢互联网时代那些愿意分享知识的无名者们——正是他们让我相信，在这个信息泛滥的时代，精心组织的、有深度的知识仍然有着不可替代的价值。本书的电子版已在网站上免费发布，欢迎读者下载、传播（仅限非商业用途）。如果你在阅读中发现任何错误，或是有任何建议，请不吝致信我的邮箱。一本好的讲义，永远是在作者与读者的对话中逐渐完善的。

最后，我想对正在阅读这本书的你说：统计物理是一门需要耐心，但绝不缺乏回报的学问。当你真正理解了威尔逊如何用“放大镜”看透临界现象的本质，或是明白了 Kosterlitz 和 Thouless 如何用“涡旋”破解二维世界的秩序之谜，你收获的将不仅是知识，更是一种看待复杂世界的全新眼光。愿这本书，能成为你这段智识旅程的忠实旅伴。

陈童

2026 年

# 1. 杨李相变理论

## 1.1 液气相变简介

水可以凝固成冰，也可以气化成水蒸气，气、水、冰本质都是由  $H_2O$  分子组成的物质，然而它们的物理性质截然不同。这种在环境变化无穷小时，物质的物理性质发生的不连续的改变，就是**相变**。气、水、冰就是宏观数量的  $H_2O$  分子的三种不同形态，或者说不同的相 (phase)。这里值得注意的是，相变不是组成物质的分子的微观改变，也不是少量分子的改变，而是宏观数量的分子宏观物理性质的改变。

通常有不止一个变量可以影响物质的相。例如可以通过降低温度或者增加压强来使得蒸气凝结。我们使用相图来显示平衡相与温度和压强的关系（注意压强的单位， $1\text{bar} = 10^5\text{Pa}$ ，Pa 表示帕斯卡，1bar 大约就是一个标准大气压 (atm)，实际上 1atm 约为 1.01325bar）。

为了确定相图，19 世纪人们就对气液相变进行了系统的研究。其中一种办法是将一定量的液体（如水、乙醚）封在容器中，缓慢加热，测定压力随温度的变化曲线。只要容器中同时存在液体和它的蒸气，气液两相就在共存，因此这样测得的  $P-T$  曲线，就是气液的相界线，它也叫气化线。

如图 (1.1) 所示，如果保持压强为  $P_0$ ，由液相开始加热，状态将从图中的  $L$  点沿水平线向右移动，与气化线交于  $Q$  点，一部分液体开始气化，这时尽管在继续加热，温度却不会升高，一直保持在  $Q$  点的  $T_0$  值，加热的热量全部被吸收来用于气化了。这种在相变时才表现出来的气化热，是一种相变潜热。日常生活中，手上沾了水会有清凉感，就是因为水气化时从手里吸收了潜热。

液体全部气化之后，温度才继续沿着图中的  $QG$  线继续升高。值得注意的是，沿着  $LQG$  线做实验，为了保持压强不变，要不断改变系统的体积，因此这样的实验不能用前面提到的密封容器来做。

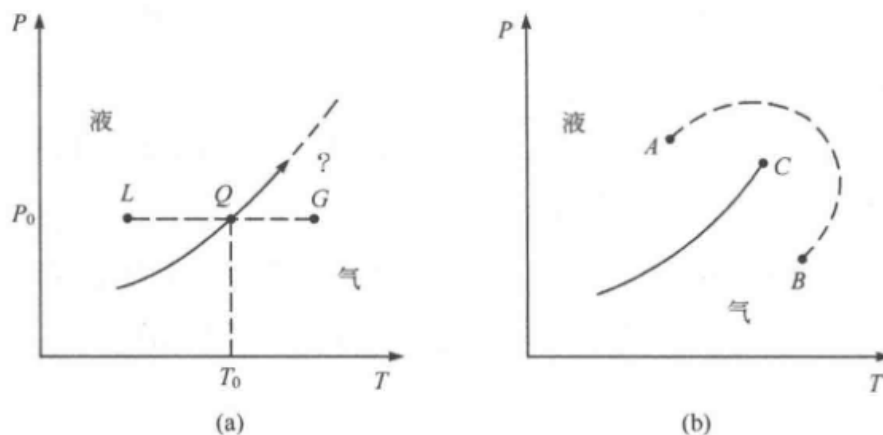


Figure 1.1:  $P-T$  相图上的气液相界线

一个自然的问题是，气化线是会随着温度和压强的不断增加而无限延伸，还是会有一个截止的端点？1869年英国物理学家安德鲁斯作了一个题为“论物质液态和气态的连续性”的报告，明确地回答了这个问题。他精确地测量了二氧化碳在液态与气态时的密度差，发现在31摄氏度附近，两者的差别消失了！这意味着气液相界线有一个明确的终点，安德鲁斯称之为“临界点”。其他物理学家也观察到这个现象，发现在这个点，相变潜热等于零！

这其实不难理解，关键是要注意到液气相界线总是有正的斜率，因此对于共存的液相和气相，如果要提高温度，就必须同时施加更高的压强，然而，随着压强的增加，气体将变得更加稠密，因此液体和气体之间的密度差会逐渐缩小，直至消失，也就是达到临界点。对于  $H_2O$ ，这个临界点在  $T_c = 374$  摄氏度、 $P_c = 221\text{bar}$ 。在临界点上，就不再有任何不连续的从液相到气相的变化了。

由于临界点的存在，液相其实可以连续地变成气相，而不需要经过任何不连续的相变。方法是沿着图 (1.1)b 中所示的虚线变化温度和压强。

1873年，荷兰青年范德瓦尔斯利用当时刚刚发展起来的分子运动论观念，对液气相变提出了理论解释。他的论文题目与安德鲁斯的报告类似，也叫“论气态和液态的连续性”。

为了描写气液相变，范德瓦尔斯建议计入气体分子间的两种相互作用：第一，每个分子具有一定的体积  $b$  (在小于  $b$  的范围之内，气体分子间会非常强烈地相互排斥)，因此气体活动的有效体积缩小为  $V - Nb$ 。第二，气体分子间在较远的距离上有微弱的相互吸引力，相当于补充了一点“内压力”，使  $P$  增加到  $P + a(\frac{N}{V})^2$ 。这里“内压力”之所以能写成  $a(\frac{N}{V})^2$ ，是因为，首先它来自于分子之间的微弱吸引，既然是分子之间的，那就应该正比于  $N \times N = N^2$ 。其次，作为压强，“内压力”应该是强度量，所以就只能正比于  $(\frac{N}{V})^2$  了。根

据这些考虑，于是范德瓦尔斯把理想气体的物态方程  $PV = Nk_B T$  修改为

$$\left(P + a\left(\frac{N}{V}\right)^2\right)(V - Nb) = Nk_B T. \quad (1.1)$$

我们可以在  $P-V$  图上画出范德瓦尔斯方程的等温线，如图 (1.2) 所示。在足够高的温度下，减小体积会导致压强平稳地上升到无穷，同时体积逼近  $Nb$ 。然而，在较低的温度下，等温线的行为要复杂得多，随着  $V$  的减小，等温线上升，下降，然后再次上升。似乎暗示着，对于某些状态来说，压缩液体会导致压强下降，即等温压缩系数  $\kappa_T \equiv -\frac{1}{V} \frac{\partial V}{\partial P} \Big|_T < 0$ ，但是，热力学稳定性的考虑告诉我们， $\kappa_T$  一定是大于零的。这就说明，为了正确地描写液气相变，范德瓦尔斯方程还需要一些修正。

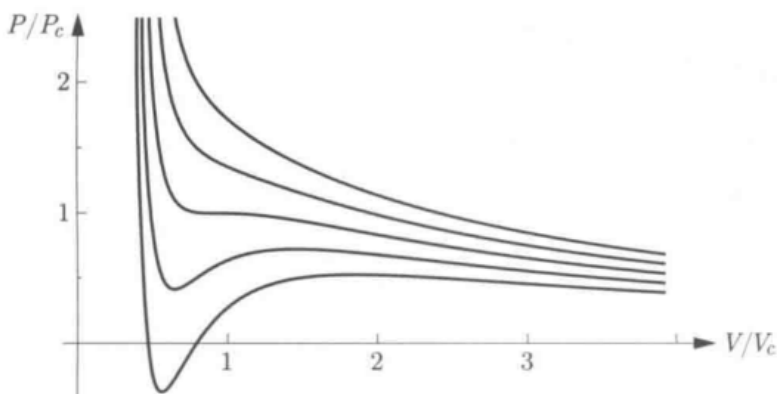


Figure 1.2: 范德瓦尔斯方程的等温线。

为了进行这样的修正，麦克斯韦提出了所谓的**麦克斯韦构造法** (细节我们这里不做讨论)。图 (1.3) 给出了在不同温度下重复麦克斯韦构造法得到的结果。对于每个温度，都有一个确定的压强 (对应阴影部分的水平直线)，称为蒸气压，在该温度和压强下会发生液气转变，换言之，在一定的温度之下， $P-V$  图上的等温曲线有一个平台区，它描写的就是液气两相共存的情况。这也符合实验，因为实验表明，两相共存时压强是确定的。平台区的左侧，随着体积的减小压强迅速增大，对应液相，平台区的右侧，体积减小导致的压强增加相对比较小，对应气相。

但是，范德瓦尔斯方程毕竟是一个唯象的理论，麦克斯韦构造法更是一个人为的东西。因此，一个自然的问题就是，能不能从一个基本理论出发，比如从统计物理出发，得到能同时描写液气两相的物态方程呢？比如推导出如图 (1.3) 那样的  $P-V$  图等温曲线呢？尤其是，我们是通过人为的麦克斯韦构造法得到等温曲线上描写两相共存的平台区的，能不能从一个基本理论出发，自动得到这些平台区呢？在很长时间之内，这都是未解的难题，杨振宁和李政道的相变理论正是解决了这个难题。

## 1.2 统计物理和相变

为了看清如何从微观的统计物理出发理解宏观的液气相变，我们不妨将问题一般化为，如何从统计物理出发理解相变。从热力学的角度来看，所谓的相变，就是当我们连续

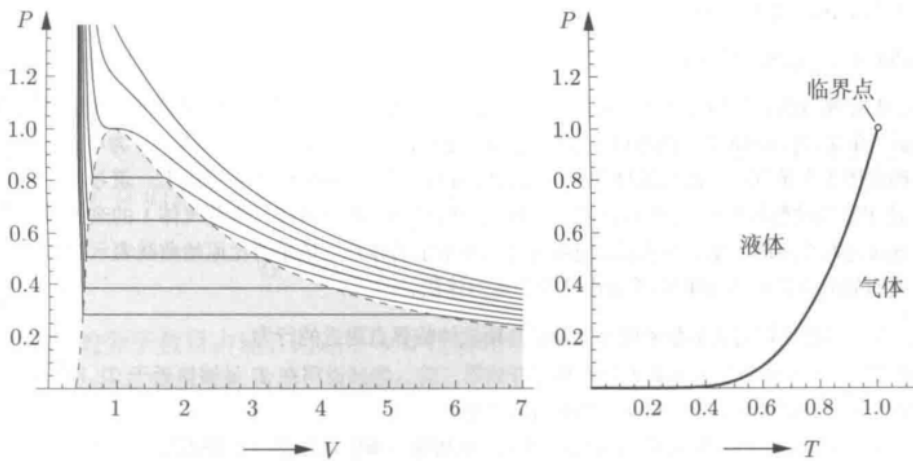


Figure 1.3: 范德瓦尔斯模型预测的相图。阴影部分的稳定态是两相共存。右图是相应  $P-T$  图上的液气相界线。

变化系统的控制参数 (例如温度) 时, 某些宏观物理量 (比如比热、磁化率等等) 表现出来的对控制参数的不连续依赖。根据热力学, 各种宏观物理量都可以从热力学势函数的偏导计算中得出, 因此, 相变就是某些热力学势函数偏导对控制参数的不连续性, 或者说, 非解析性。

但是根据统计物理, 热力学势函数正比于配分函数的对数, 而配分函数的计算就是以  $e^{-\beta H}$  ( $H$  为系统哈密顿量) 为权重对系统位形进行求和。但是, 权重因子  $e^{-\beta H}$  对控制参数都是解析依赖的。因此, 如果这个位形求和只涉及有限项, 那有限个解析函数的和必定依然是解析函数, 那就必然不会有相变。因此, 结论就是, 一个有限系统从严格意义上来说是没有相变的。现实中的系统当然都是有限的, 但是, 它们也是宏观的, 也就是涉及超过  $10^{23}$  量级的粒子数, 同时系统的体积  $V$  也是宏观的。对于这样的宏观系统而言, 我们可以近似认为其粒子数  $N \rightarrow \infty$ , 体积  $V \rightarrow \infty$ , 同时保持粒子数密度  $n = N/V$  固定, 这就是所谓的热力学极限。可以近似认为宏观系统是满足热力学极限的, 而热力学极限下, 位形求和就是一个无穷求和, 无穷多个解析函数的和是有可能不解析的, 换言之, 在热力学极限下, 相变可以发生。

但是, 为了保证热力学极限的存在性, 通常需要对粒子间的相互作用势能  $u(r)$  ( $r$  是两粒子间距) 进行一些假设:

- 第一, 随着粒子间距大到一定程度,  $u(r)$  随着  $r$  的增加必须下降得足够快, 通常来说要求随着  $r \rightarrow \infty$ ,  $|u(r)| \leq C/r^{d+\epsilon}$ , 式中  $C > 0$ ,  $d$  为空间维数,  $\epsilon$  为一个正常数。这种系统也就是所谓的短程相互作用的系统。之所以短程相互作用的系统才能保证热力学极限存在, 是因为, 有限系统必定有一个空间边界, 边界条件可能影响系统体内的行为, 如果相互作用是短程的, 那么边界条件就只能影响边界附近薄薄的一层区域, 对真正体内的影响可以忽略, 从而在  $V \rightarrow \infty$  的热力学极限下, 整个边界都可以忽略。相反, 如果相互作用是长程的, 那边界条件的影响就可以穿透整个体内, 那

$V \rightarrow \infty$  下的行为将严重依赖边界条件，从而无法定义热力学极限。

- 第二，当粒子间距足够小时， $u(r)$  必须有一个排斥势部分，它随着距离的减小势能迅速上升到无穷。这是为了防止整个系统坍缩到一个点上（那粒子数密度  $n$  就趋于无穷了）。这里我们不妨假设当  $r < r_0$  时， $u(r) = \infty$ ，也就是说，每个粒子都有一个有限大的硬核。根据这个假设，给定一个有限的体积  $V$ ，整个系统内能放下的粒子总数必定有一个上限，不妨记这个上限为  $M$ 。在热力学极限之下， $M/V$  是固定的。
- 第三，无论粒子间距为多少，势能  $u(r)$  都有下界！换言之， $u(r) \geq -u_0$ ，式中  $u_0$  是一个正常数。这一条是为了保证整个系统的总能量有下界，因为否则系统是不可能稳定存在的。

杨李相变理论是以上面这三条假设为基础的。考虑一个体积为  $V$  的盒子，内部是一盒子气体分子，分子之间的相互作用势能为  $u(r)$ 。盒子处于温度为  $T$  的环境中，并与一个化学势为  $\mu$  的粒子源交换气体分子。则系统的巨配分函数为

$$\Xi_V(T, z) = \sum_{N=0}^M \mathcal{Z}_N(T) z^N, \quad (1.2)$$

式中  $\mathcal{Z}_N(T)$  为  $N$  个粒子的正则系综配分函数，

$$z = e^{\mu/k_B T} \quad (1.3)$$

为逸度。根据配分函数的定义，显然  $\mathcal{Z}_N(T)$  是正实数， $\Xi_V(T, z)$  就是一个以  $\mathcal{Z}_N(T)$  为系数的关于  $z$  的多项式。杨李理论的关键想法是，将这个多项式的定义域从正实轴延拓到整个复  $z$  平面！

有了巨配分函数以后，我们就可以定义有限体积的压强  $P_V$  和有限体积的粒子数密度  $n_V$  如下（式中  $\log$  表示以  $e$  为底的对数）

$$\begin{aligned} P_V(z) &= k_B T \frac{1}{V} \log \Xi_V(T, z), \\ n_V(z) &= \frac{\partial}{\partial \log(z)} \frac{1}{V} \log \Xi_V(T, z). \end{aligned} \quad (1.4)$$

并且人们还可以把这个定义从  $z$  的正实轴往整个复平面延拓，当然延拓以后，除正实轴之外的复  $P_V(z)$  和复  $n_V(z)$  就只有数学上的意义。取热力学极限之后，就得到我们关心的压强和粒子数密度

$$P(z) = \lim_{V \rightarrow \infty} P_V(z), \quad n(z) = \lim_{V \rightarrow \infty} n_V(z). \quad (1.5)$$

通过将  $P(z)$  和  $n(z)$  消去逸度  $z$ ，就可以得到系统的物态方程。问题是，上述  $P(z)$  和  $n(z)$  是否对整个复  $z$  平面都有定义？对于液气相变而言，我们怎么知道这样得到的物态方程既可以描写液相，同时又可以描写气相，以及如何刻画两相之间的相变？

## 1.3 杨李理论基础

### 1.3.1 杨李理论的基本框架

注意到  $\Xi_V(T, z)$  是一个关于  $z$  的  $M$  次多项式，杨李的关键观察是，除了这个多项式在复  $z$  平面上的零点处之外， $P_V(z)$  和  $n_V(z)$  关于  $z$  均是解析的，因此相变只可能和  $\Xi_V(T, z)$

关于  $z$  的零点有关。注意到多项式  $\Xi_V(T, z) = \sum_{N=0}^M \mathcal{Z}_N(T) z^N$  各项系数都是正实数，所以  $\Xi_V(T, z)$  在  $z$  的正实轴上必定没有零点，但是， $z$  的正实轴才是物理的逸度，所以，这就说明，对于有限的  $V$ ，系统必定没有相变。

类似的，由于多项式  $\Xi_V(T, z) = \sum_{N=0}^M \mathcal{Z}_N(T) z^N$  各项系数都是实数，所以它的零点必定成对出现，每一个零点都有一个相应的复共轭零点。进而，我们可以把这个多项式因式分解成

$$\Xi_V(T, z) = \Xi_0(T) \prod_{i=1}^M \left[ 1 - \frac{z}{z_i(T)} \right], \quad (1.6)$$

式中  $z_i(T)$  表示第  $i$  个零点位置， $\Xi_0(T)$  是与  $z$  无关的常数。注意，这些零点的分布与温度  $T$  有关，随着温度的改变，零点分布也会改变。同样，零点分布还依赖于体积  $V$ 。特别的，因为零点数目  $M$  是给定体积下粒子数目的上限，所以，随着  $V \rightarrow \infty$ ，必定同步的也有  $M \rightarrow \infty$ 。

根据 (1.4) 式，即可以得到

$$\begin{aligned} P_V(z) &= P_0(T) + k_B T \frac{1}{V} \sum_{i=1}^M \log \left[ 1 - \frac{z}{z_i(T)} \right], \\ n_V(z) &= \frac{\partial}{\partial \log(z)} P_V(z) / k_B T = \frac{1}{V} \sum_{i=1}^M \frac{z}{z - z_i}. \end{aligned} \quad (1.7)$$

式中  $P_0(T) \equiv k_B T (\log \Xi_0(T)) / V$ ，这一项对  $P_V(z)$  的贡献其实不重要，因为我们关心的是  $P_V$  作为  $z$  的函数。注意，由上式不一定能推出  $n(z) = \frac{\partial}{\partial \log(z)} P(z) / k_B T$ ，因为正如后文会分析的，虽然  $P(z)$  在整个复  $z$  平面上都有定义，但是  $n(z)$  却并非如此。

对于物理的  $z$ ，也即是正实轴，根据统计物理的标准知识，有

$$\frac{\partial}{\partial \log z} \log \Xi_V(z) = \langle N \rangle, \quad (1.8)$$

也即是粒子数  $N$  的统计期望值。进一步也有

$$\frac{\partial^2}{(\partial \log z)^2} \log \Xi_V(z) = \langle N^2 \rangle - \langle N \rangle^2 = \langle (\Delta N)^2 \rangle > 0, \quad (1.9)$$

式中  $\Delta N \equiv N - \langle N \rangle$ 。即  $\log \Xi_V$  对  $\log(z)$  的二阶偏导大于零，由于  $n_V(z) = \frac{\partial}{\partial \log(z)} \frac{1}{V} \log \Xi_V(z)$  正比于  $\log \Xi_V$  对  $\log(z)$  的一阶偏导，这就说明，对于物理的  $z$ ， $n_V(z)$  是  $\log(z)$  的单调递增函数，并且很显然，根据 (1.7) 给出的  $n_V(z)$  表达式， $n_V(z)$  在正实轴上有一个有限的上界  $M/V$ 。进一步，同样根据这个  $n_V(z)$  的表达式，不难看出  $n_V(0) = 0$ ，结合  $n_V(z)$  是  $z$  的单调递增函数即可知，在正实轴上， $n_V(z) > 0$  且有上界。也即， $\log \Xi_V$  对  $\log(z)$  的一阶偏导在  $z$  的正实轴上大于零且有上界，进而可知，在  $z$  的正实轴上， $\log \Xi_V$  是  $\log(z)$  的单调增函数。显然这说明，在  $z$  的正实轴上， $P_V(z)$  是  $\log(z)$  的单调增函数。取热力学极限，即有  $P(z), n(z)$  在正实轴上都是单调递增函数（对于有定义的  $z$ ），且  $n(z)$  在正实轴上有上界  $M/V$ 。

注意到除了  $\Xi_V(z)$  的零点之外， $P_V(z)$  在整个复平面上均解析，换言之，给定复  $z$  平面上的某个任意点  $z_0$ ， $P_V$  在  $z_0$  处解析的充要条件是：围绕  $z_0$  取一个任意小的邻域，此邻

域内不含有  $\Xi_V(z)$  的零点。换言之，随着  $V \rightarrow \infty$ ，只要  $\Xi_V(z)$  的零点不以  $z_0$  为聚点，则  $P(z)$  在  $z_0$  处就一定解析。同样的道理也适用于  $n(z)$ 。这种解析性得以满足的地方当然和相变没有关系。因此，相变只可能发生在  $\Xi_V(z)$  零点的聚点处（随着  $V \rightarrow \infty$ ）。进一步，由于真正的物理量只对正实数的  $z$  有定义，所以，相变对应于  $\Xi_V(z)$  的零点在正实轴上的聚点。前面说过，对于有限的  $V$ ， $\Xi_V(z)$  在正实轴上没有零点，但，这并不表示，随着  $V \rightarrow \infty$ ，复  $z$  平面上的零点不会在正实轴上形成聚点。

同时，由  $n_V(z)$  的表达式可知，对于有限的  $V$ ，复  $z$  平面上这些零点是  $n_V(z)$  的极点，但是，这些零点的聚点，就会成为  $n(z)$  的本性奇点，所以，在这些零点的聚点处， $n(z)$  一般来说会有无穷多个取值（或者说  $n(z)$  在该处没有定义）。类似的分析也适用于  $\Xi_V(z)$  对  $\log(z)$  的高阶偏导。问题是， $P(z)$  在这些聚点处有没有定义呢？这个问题我们稍后再做讨论。

现在我们来改写一下方程 (1.7)。设复平面上的坐标  $z = x + iy$ ，零点  $z_j = x_j + iy_j$ ，进而定义零点密度函数  $\rho(z)$  为

$$\rho(z) \equiv \frac{1}{V} \sum_{i=1}^M \delta(x - x_i) \delta(y - y_i), \quad (1.10)$$

由于零点是成对出现的，每一对中的两个零点互为对方的复共轭，所以显然有

$$\rho(z) = \rho(\bar{z}), \quad (1.11)$$

式中  $\bar{z}$  表示  $z$  的复共轭。假设所有的零点均分布在复平面上的区域  $\mathcal{D}$  之内（ $\mathcal{D}$  当然依赖于温度  $T$ ），则很显然

$$\int_{\mathcal{D}} dz \rho(z) = \frac{M}{V}, \quad (1.12)$$

式中  $dz \equiv dx dy$ 。可以设想，随着  $V \rightarrow \infty$ ， $M \rightarrow \infty$  的热力学极限，复平面上的零点可能在区域  $\mathcal{D}$  内汇聚成连续的一片，即， $\mathcal{D}$  内的每一个点都是聚点，并使得  $\rho(z) = \rho(x, y)$  成为  $x, y$  的连续函数（在热力学极限之下）。后文我们默认所说的  $\rho(z)$  都是指此热力学极限之下的零点密度函数。

根据 (1.7) 式不难看出，

$$\begin{aligned} P(z) &= P_0(T) + k_B T \int_{\mathcal{D}} dz' \rho(z') \log \left[ 1 - \frac{z}{z'} \right], \\ n(z) &= \frac{\partial}{\partial \log(z)} P(z) / k_B T = \int_{\mathcal{D}} dz' \rho(z') \frac{z}{z - z'}. \end{aligned} \quad (1.13)$$

当然， $P(z)$  是  $z$  的多值函数。很明显，只要知道了零点的分布函数  $\rho(z)$ ，那我们就能根据上式算出  $P(z)$  和  $n(z)$ ，对于物理的  $z$ （正实轴），通过消去它，就能得到系统的物态方程  $P = P(n, T)$ 。然而，给定一个系统，算出其零点分布函数  $\rho(z)$  是一个极端困难的事情，事实上，即使只是决定零点的分布区域  $\mathcal{D}$  都是一个难题。好在对于一大类系统（伊辛模型和格气模型），杨振宁和李政道的原始论文解决了后面这个难题，这就是著名的单位圆定义，它说的是，对于这一类系统，其零点均分布在复  $z$  平面的单位圆上，即  $\mathcal{D}$  是一个单位圆。后文我们再回到对单位圆定理的讨论和证明。

现在, 不妨将  $P(z) - P_0(T)$  的实部和虚部分开来考虑, 定义

$$[P(z) - P_0(T)]/k_B T \equiv \varphi(z) + i\psi(z), \quad (1.14)$$

式中  $\varphi(z)$ ,  $\psi(z)$  是两个实函数。很显然,

$$\varphi(z) = \int dz' \rho(z') \log |1 - \frac{z}{z'}|. \quad (1.15)$$

注意  $\frac{1}{2\pi} \log |z|$  是二维拉普拉斯算符  $\Delta = \partial^2/\partial x^2 + \partial^2/\partial y^2$  的格林函数。进而通过将拉普拉斯算符作用于上式, 不难看出

$$\Delta\varphi(z) = 2\pi\rho(z). \quad (1.16)$$

这意味着  $\varphi(z)$  是两维平面上电荷密度  $\rho(z)$  所对应的静电势! 根据电动力学的知识可知, 对于有限区域  $\mathcal{D}$  内的任何电荷密度  $\rho(z)$ , 相应的静电势  $\varphi(z)$  在整个两维平面上都有解, 且这个解是连续函数! 特别的, 这说明  $\varphi(z)$  在零点的聚点处有定义, 且在这些地方也是连续的!

不仅如此, 假设最终的零点汇聚在复平面上的一条封闭曲线上 (这是很典型的情况), 当然它也就是我们的电荷分布区域  $\mathcal{D}$ , 因此这样的曲线  $\mathcal{D}$  就将整个复平面分割成内外两个区域, 区域 1 和区域 2。在这两个区域内部, 由于电荷为零, 所以静电势  $\varphi(z)$  是解析函数, 分别记这两个解析函数为  $\varphi_1(z)$ ,  $\varphi_2(z)$ 。由于静电势在整个两维平面上连续, 所以必定有

$$\varphi_1(z)|_{\mathcal{D}} = \varphi_2(z)|_{\mathcal{D}}. \quad (1.17)$$

但是, 请注意, 在区域分割线  $\mathcal{D}$  上,  $\varphi(z)$  虽然连续, 但并不解析。

承接上一段的假设, 假设封闭曲线  $\mathcal{D}$  与  $z$  的正实轴相交, 那这些交点就对应相变发生的地方。同时, 由于在  $z$  的正实轴上,  $P(z) - P_0(T)$  是一个实函数, 它只有实部  $k_B T \varphi(z)$ 。因此上面关于  $\varphi(z)$  连续性的讨论就说明,  $P(z)$  在整个正实轴上都有定义, 且是连续函数。特别的, 对于零点在正实轴上的聚点处,  $P(z)$  有定义, 且在此处连续。总之, 对于任何物理的逸度  $z$ ,  $P(z)$  均有定义, 且是连续函数。

依然承接上面的假设, 由于  $\mathcal{D}$  是一条曲线, 所以相应的零点密度  $\rho(z)$  可以改成线密度  $\lambda(z)$ 。进一步假设  $z_p = x_p$  为  $\mathcal{D}$  与正实轴的某个交点。为了看清楚沿着正实轴从两侧趋于  $x_p$  时, 粒子数密度  $n(z)$  的跳变情况 (注意, 前文分析过, 在这样的聚点处,  $n(z)$  一般来说有无穷多个取值), 我们不妨把两维  $z$  平面逆向旋转 90 度, 使得实轴变成虚轴, 虚轴变成实轴, 从而旋转之后  $z \rightarrow iz = ix - y$ 。记  $x^\pm = x \pm \varepsilon$ , 式中  $\varepsilon$  为无穷小正实数, 则根据  $n(z)$  的计算公式 (1.13), 有 (注意  $\rho(z) \rightarrow \lambda(z)$  了)

$$n(x_p^-) = \int_{\mathcal{D}} dz' \lambda(z') \frac{iz_p}{iz_p - i\varepsilon - z'} = \int_{\mathcal{D}} dz' \lambda(z') \frac{-ix_p}{z' - iz_p + i\varepsilon}. \quad (1.18)$$

利用数学公式

$$\frac{1}{z - i\varepsilon} = \frac{1}{z} + i\pi\delta(z), \quad \frac{1}{z + i\varepsilon} = \frac{1}{z} - i\pi\delta(z). \quad (1.19)$$

可得

$$n(x_p^-) = -ix_p \mathcal{P} \int_{\mathcal{D}} dz' \frac{\lambda(z')}{z' - iz_p} - x_p \pi \lambda(iz_p). \quad (1.20)$$

式中  $\mathcal{P} \int$  表示主值积分。同理可得

$$n(x_p^+) = -ix_p \mathcal{P} \int_{\mathcal{D}} dz' \frac{\lambda(z')}{z' - iz_p} + x_p \pi \lambda(iz_p). \quad (1.21)$$

从而即有 (坐标轴旋转回来了)

$$n(x_p^+) - n(x_p^-) = 2\pi x_p \lambda(z_p) = 2\pi x_p \lambda(x_p). \quad (1.22)$$

所以, 只要零点密度在聚点  $x_p$  处不为零, 那么  $n(z)$  在  $z = x_p$  处就是不连续的, 而是会发生一个  $2\pi x_p \lambda(x_p)$  的正向的跳变。

### 1.3.2 相变的物理图像

根据以上分析, 我们可以得出如下物理图像。假设在热力学极限之下,  $z$  的正实轴  $x$  上没有聚点, 如图 (1.4) 所示。则, 压强随着  $\log(x)$  连续单调递增, 同时, 粒子数密度  $n$  随着  $\log(x)$  也连续单调上升, 且有一个上界。所以, 系统只有唯一的一个相, 没有相变。最终在  $P - 1/n$  ( $1/n$  为比体积) 图上, 等温线连续单调下降, 没有平台区。

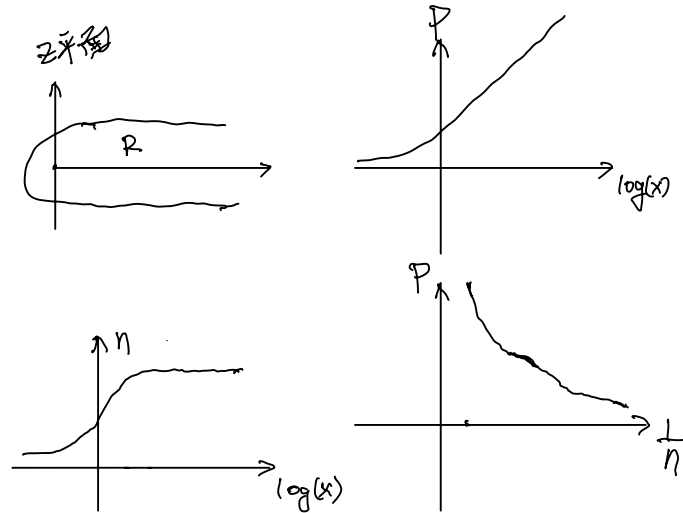


Figure 1.4: 在热力学极限之下, 左上角图中的区域  $R$  内没有零点, 则系统只有一个相, 在给定温度下, 其压强随着比体积  $1/n$  单调下降。

而假设在热力学极限之下,  $z$  的正实轴  $x$  上有比方说两个聚点  $x_p$  和  $x_q$ , 如图 (1.5)(a) 所示。则系统共有三个不同的相,  $x_p$  和  $x_q$  就是相变发生的地方。压强  $P$  随  $\log(x)$  连续单

调增,但在  $\log(x_p)$  和  $\log(x_q)$  处不解析,如图 (1.5)(b) 所示。粒子数密度  $n$  随  $\log(x)$  也单调增,同时在  $\log(x_p)$  和  $\log(x_q)$  处有正向跃变,如图 (1.5)(c) 所示。根据图 (b) 和图 (c) 可以画出最终压强和比体积  $1/n$  的曲线,如图 (1.5)(d) 所示,注意,在粒子数密度跳变的地方, $\log(x)$  是固定的,因此相应的压强  $P$  也固定,这就对应  $P-1/n$  曲线上的平台区。

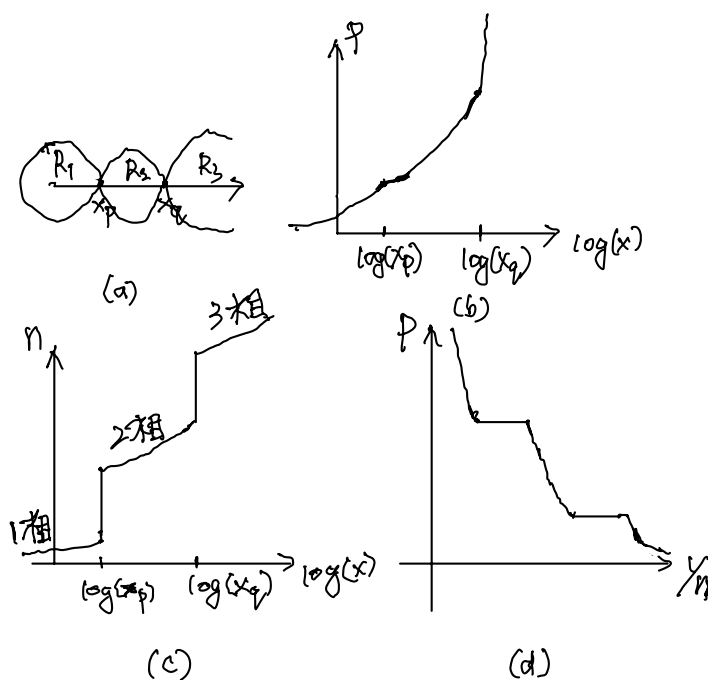


Figure 1.5: 图 (a) 表示正实轴上有两个聚点  $x_p$  和  $x_q$ , 区域  $R_1, R_2, R_3$  之内没有零点。图 (b) 是相应的  $P-\log(x)$  曲线,  $P$  随  $\log(x)$  连续单调增,但在  $\log(x_p)$  和  $\log(x_q)$  处不解析。图 (c) 是  $n-\log(x)$  曲线,  $n$  随  $\log(x)$  单调增,同时在  $\log(x_p)$  和  $\log(x_q)$  处有正向跃变。图 (d) 是  $P-1/n$  曲线, 注意在两相共存区自动出现了等温线的平台。

随着温度的变化,聚点  $x_p$  和  $x_q$  当然会在正实轴上移动。假设在某个温度  $T_c$  之下,  $x_p$  处的聚点消失了,那就说明 1 相和 2 相之间的相变消失了,那这个温度  $T_c$  就是一个所谓的临界温度。当然,也有可能比方说聚点  $x_p$  处,零点密度  $\lambda(x_p) = 0$ ,则根据 (1.22) 式可知,这时候,粒子数密度  $n$  没有跳变,而是连续的。因此这时候  $x_p$  处的相变就是所谓的连续相变。

很显然,如果正实轴上的聚点只有一个,那就是两个相,分析和上面完全类似。特别的,  $P-1/n$  等温线上会自动出现描述两相共存的平台区。很显然,这幅物理图像自然可以正确描述液气相变。但,这里的关键是,我们得出这幅物理图像并不需要任何唯象的或者手动的假设,而是从统计物理的基本理论中自动得出来的。特别的,从不同温度下的  $P-1/n$  等温曲线,我们原则上可以得到能够同时描写两个不同相的物态方程。所以,杨李的这个相变理论自然解决了如何从基本理论出发描述液气相变的难题。

### 1.3.3 对相变的进一步刻画

为了进一步刻画相变，依然让我们假设零点的分布区域  $\mathcal{D}$  是一条曲线，并假设它与正实轴交于  $z_0$  点，如图 (1.6) 所示。当然， $z_0$  就是相变点。不妨以  $s$  来表示这条曲线  $\mathcal{D}$  的弧长参数，其中弧长从  $z_0$  点开始量起，并以逆时针方向为  $s$  增加的方向。显然，这条曲线将复  $z$  平面分成了左右两个区域，记左侧为区域 1，右侧为区域 2。

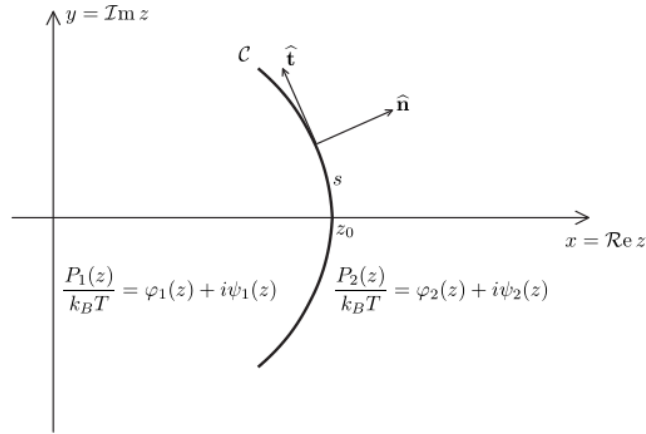


Figure 1.6: 零点分布区域  $\mathcal{D}$  是一条曲线的典型情况。图中此曲线与正实轴交于  $z_0$ ，当然，这就是相变点。

将方程  $\nabla^2 \varphi(z) = 2\pi\rho(z)$  类比于静电学的拉普拉斯方程，并在曲线  $\mathcal{D}$  附近应用静电场的高斯定理，即有

$$[\nabla\varphi_2(z) - \nabla\varphi_1(z)] \cdot \mathbf{n}|_{\mathcal{D}} = \lambda(s), \quad (1.23)$$

式中  $\mathbf{n}$  为曲线  $\mathcal{D}$  的单位法向量， $\lambda(s)$  是曲线上的零点线密度。由于  $P(z) - P_0(T) = k_B T[\varphi(z) + i\psi(z)]$  在左右两个区域均为  $z$  的解析函数，应用解析函数的柯西-黎曼方程，即有

$$\nabla\varphi_{1,2}(z) \cdot \mathbf{n}|_{\mathcal{D}} = \nabla\psi_{1,2}(z) \cdot \mathbf{t}|_{\mathcal{D}}, \quad (1.24)$$

式中  $\mathbf{t}$  为曲线  $\mathcal{D}$  的单位切向量。结合上面这两个方程，即有

$$\frac{d}{ds}[\psi_2(z) - \psi_1(z)]|_{\mathcal{D}} = 2\pi\lambda(s). \quad (1.25)$$

特别的，在相变点  $z_0$  处，我们有

$$\frac{d}{ds}[\psi_2(z) - \psi_1(z)]|_{z_0} = 2\pi\lambda(0). \quad (1.26)$$

这些结果，再结合下面的连续性条件就可以对相变构成进一步的刻画

$$\varphi_2(z)|_{\mathcal{D}} = \varphi_1(z)|_{\mathcal{D}}. \quad (1.27)$$

为此，我们把  $[P(z) - P_0(T)]/k_B T$  在  $z_0$  点的两侧分别做泰勒展开

$$[P_{1,2}(z) - P_0]/k_B T = C(z_0) + a_{1,2}(z - z_0) + b_{1,2}(z - z_0)^2 + \mathcal{O}((z - z_0)^3), \quad (1.28)$$

式中  $C(z_0)$  为泰勒展开的零阶项，由  $P(z)$  在  $z_0$  处的连续性，这个零阶项对于左右两侧都是一样的。同时，为了确保  $P_{1,2}(z)$  在正实轴上为实函数，上式所有的展开系数都必定是实数。记  $z = x + iy$ ，则根据连续性条件 (1.27)，即可以得到曲线  $\mathcal{D}$  在  $z_0$  附近的方程，为

$$(a_2 - a_1)(x - z_0) + (b_2 - b_1)[(x - z_0)^2 - y^2] + \text{Re}[\mathcal{O}((z - z_0)^3)] = 0. \quad (1.29)$$

下面分几种情况进行讨论：

- **一阶相变**：假如  $a_2 \neq a_1$ ，则复压强的一阶导数在  $z_0$  处不连续，相应的相变称之为**一阶相变**。假如进一步有  $b_2 \neq b_1$ ，则曲线  $\mathcal{D}$  在  $z_0$  附近是双曲线，方程为

$$y^2 = (x - z_0)^2 + \frac{a_2 - a_1}{b_2 - b_1}(x - z_0). \quad (1.30)$$

它在  $z_0$  处的切向量平行于虚轴。进而通过取方程 (1.28) 的虚部，并应用 (1.26)，即可以得到

$$\lambda(0) = \frac{a_2 - a_1}{2\pi}. \quad (1.31)$$

因此，对于一阶相变，相变点处的零点密度非零。这与上一小节的物理图像完全吻合。

- **二阶相变**：假如  $a_2 = a_1$ ，但是  $b_2 \neq b_1$ ，那这就是**二阶相变**。相应  $\mathcal{D}$  在  $z_0$  附近的方程为

$$y = \pm(x - z_0), \quad (1.32)$$

很显然，这说明  $\mathcal{D}$  在  $z_0$  处与虚轴成  $\pm\pi/4$  的夹角。进而通过取方程 (1.28) 的虚部，并应用 (1.25)，即可以得到

$$\lambda(s) = \frac{b_2 - b_1}{\pi}|s| + \mathcal{O}(s^2). \quad (1.33)$$

即零点密度在相变点处线性地趋于零。

- **高阶相变**：假如复压强的不连续性出现在更高阶的导数上，那这就是**高阶相变**。这时候，完全类似于上面的分析，也可以得到  $\mathcal{D}$  在相变点附近的方程，并能得出零点密度在相变点处也是趋于零的。实际上，假设相变是  $n$  阶的 ( $n \geq 3$ )，则具体分析表明  $\lambda(s) \sim |s|^{n-1}$ 。

## 1.4 伊辛模型与格子模型

为了考察具体的零点分布情况，杨和李考察了两类相互等价的模型，伊辛模型和格子模型，其中格子模型可以用来理解液气相变。下面我们从伊辛模型开始讲起。

考虑  $N$  个伊辛自旋  $\sigma_i = \pm 1 (i = 1, 2, \dots, N)$ , 式中  $\sigma_i = +1$  表示自旋向上状态,  $\sigma_i = -1$  表示自旋向下状态。(1). 这  $N$  个自旋放置在一个  $d$ -维格点上。(2). 不同的自旋两两结对, 每对的两个自旋之间存在铁磁相互作用, 即当这两个自旋平行时 (都向上或都向下) 能量比较低, 而反平行时 (一个向上一个向下) 则能量比较高。(3). 整个系统处于一个非均匀磁场中, 第  $i$  个自旋感受到的磁场为  $h_i$ 。

则系统的哈密顿量  $H_Y$  可以写成

$$H_Y = -\frac{1}{2} \sum_{i,j,i \neq j} J_{ij} \sigma_i \sigma_j - \sum_i h_i \sigma_i, \quad (1.34)$$

铁磁相互作用要求  $J_{ij} > 0$ , 且关于  $i, j$  指标对称。则系统的正则配分函数为

$$\mathcal{Z}(h_1, h_2, \dots, h_N) = \sum_{\{\sigma_i\}} e^{-H_Y/k_B T}. \quad (1.35)$$

很明显, 在磁场翻转之下, 即在所有  $h_i$  同步变换为  $-h_i$  时,  $\mathcal{Z}$  保持不变, 因为我们可以同时作翻转自旋的变换  $\sigma_i \rightarrow -\sigma_i$ , 从而可以保持系统的哈密顿量  $H_Y$  不变, 进而使得配分函数保持不变, 即有

$$\mathcal{Z}(h_1, h_2, \dots, h_N) = \mathcal{Z}(-h_1, -h_2, \dots, -h_N). \quad (1.36)$$

为了得到与之等价的格子模型, 我们进行如下变量代换

$$\sigma_i = 2n_i - 1, \quad (1.37)$$

显然  $n_i = 0, 1$ , 分别对应  $\sigma_i = -1, +1$ 。代入  $H_Y$  的表达式, 即有

$$H_Y = -2 \sum_{i,j,i \neq j} J_{ij} n_i n_j - \frac{1}{2} \sum_{i,j,i \neq j} J_{ij} + 2 \sum_i (-h_i + \sum_{j,j \neq i} J_{ij}) n_i + \sum_i h_i. \quad (1.38)$$

记  $\sum_{j,j \neq i} J_{ij} \equiv J_i$ ,  $\frac{1}{2} \sum_{i,j,i \neq j} J_{ij} = J$ , 则上式可以重写成

$$H_Y = -2 \sum_{i,j,i \neq j} J_{ij} n_i n_j - J + 2 \sum_i (-h_i + J_i) n_i + \sum_i h_i. \quad (1.39)$$

下面我们定义相应格子模型的哈密顿量  $H_G$  为

$$H_G \equiv -2 \sum_{i,j,i \neq j} J_{ij} n_i n_j - J. \quad (1.40)$$

之所以称这个哈密顿量描述的格点模型为格子模型, 原因在于我们可以把  $n_i = 1$  的格点想象成是某种气体原子占据的格点, 而  $n_i = 0$  的格点则是未被占据的格点, 且每个格点最多被一个原子占据。那哈密顿量  $H_G$  描述的就是这些格点气体原子之间以  $-4J_{ij} < 0$  的吸引型相互作用能进行相互作用的气体系统, 称之为格子系统。

则由 (1.39) 式和  $H_G$  的定义, 不难看出,

$$\mathcal{Z}(h_1, h_2, \dots, h_N) = \sum_{\{n_i\}} e^{-H_G/k_B T} \prod_i e^{-2n_i(J_i - h_i)/k_B T} \prod_i e^{-h_i/k_B T}. \quad (1.41)$$

定义局域逸度  $y_i$  为

$$y_i \equiv e^{-2J_i/k_B T} e^{2h_i/k_B T} \equiv a_i z_i, \quad \text{其中 } a_i \equiv e^{-2J_i/k_B T}, z_i \equiv e^{2h_i/k_B T}. \quad (1.42)$$

后文我们将保持  $a_i > 0$  为实数不变, 但是把  $z_i$  延拓到复平面上去 (相当于让局域磁场  $h_i$  可以取复数值)。进而即有

$$\mathcal{Z}(h_1, h_2, \dots, h_N) = (z_1 z_2 \cdots z_N)^{-1/2} \sum_{\{n_i\}} [e^{-H_G/k_B T} \prod_i y_i^{n_i}]. \quad (1.43)$$

很显然, 式中  $\sum_{\{n_i\}} e^{-H_G/k_B T} \prod_i y_i^{n_i}$  是一个以  $\{n_i\}$  为变量, 局域逸度为  $y_i = a_i z_i$  的格子系统的巨配分函数, 不妨记作  $\Xi(z_1, z_2, \dots, z_N)$ ,

$$\Xi(z_1, z_2, \dots, z_N) = \sum_{\{n_i\}} [e^{-H_G/k_B T} \prod_i y_i^{n_i}] = \sum_{\{n_i\}} [e^{-H_G/k_B T} \prod_i (a_i z_i)^{n_i}].$$

从而

$$\mathcal{Z}(h_1, h_2, \dots, h_N) = (z_1 z_2 \cdots z_N)^{-1/2} \Xi(z_1, z_2, \dots, z_N). \quad (1.44)$$

很明显, 磁场翻转  $h_i \rightarrow -h_i$  相应于变换  $z_i \rightarrow 1/z_i$  (根据  $z_i$  的定义)。进而结合式 (1.36) 与式 (1.44), 即可知

$$\Xi(1/z_1, 1/z_2, \dots, 1/z_N) = (z_1 z_2 \cdots z_N)^{-1} \Xi(z_1, z_2, \dots, z_N). \quad (1.45)$$

将  $H_G$  的定义式 (1.40) 代入巨配分函数  $\Xi(z_1, z_2, \dots, z_N)$  的计算公式, 即有 (结合  $a_i$  的定义式, 以及  $J_i$  的定义)

$$\begin{aligned} \Xi(z_1, z_2, \dots, z_N) &= \sum_{\{n_i\}} [(e^{-H_G/k_B T} \prod_i (a_i z_i)^{n_i})] \\ &= e^{J/k_B T} \sum_{\{n_i\}} \prod_{i,j,J \neq i} e^{2n_i n_j J_{ij}/k_B T - 2n_i J_{ij}/k_B T} z_i^{n_i}. \end{aligned} \quad (1.46)$$

由于  $J_{ij} > 0$ , 所以很显然  $2n_i n_j J_{ij}/k_B T - 2n_i J_{ij}/k_B T \leq 0$ , 所以

$$0 < e^{2n_i n_j J_{ij}/k_B T - 2n_i J_{ij}/k_B T} \leq 1, \quad (1.47)$$

并且很明显, 当  $n_i = n_j = 1$  时 (即  $i, j$  格点均被占据时), 或者  $n_i = 0$  时 (即  $i$  格点为空时), 有  $e^{2n_i n_j J_{ij}/k_B T - 2n_i J_{ij}/k_B T} = 1$ 。又请注意  $i$  格点为空时,  $z_i^{n_i} = 1$ , 而当  $i$  格点被占据时,  $z_i^{n_i} = z_i$ 。

不妨将所有被原子占据的格点 (即  $n_i = 1$ ) 的集合记作  $S \subseteq \{1, 2, \dots, N\}$ , 其补集 (空格点集合) 记作  $\bar{S}$ , 并记  $z^S \equiv \prod_{i \in S} z_i$ 。则不难将上面的巨配分函数按照被占据的格点泰勒展开成如下形式

$$\Xi(z_1, z_2, \dots, z_N) = e^{J/k_B T} \sum_S [z^S \prod_{i \in S, j \in \bar{S}} A_{ij}] \equiv e^{J/k_B T} Q_N(z_1, \dots, z_N). \quad (1.48)$$

式中  $0 < A_{ij} \leq 1$ , 且关于指标  $i, j$  对称。人们也常常记

$$\prod_{i \in S, j \in \bar{S}} A_{ij} \equiv A_S. \quad (1.49)$$

很显然,  $A_{\bar{S}} = A_S$ 。进而将多项式  $Q_N(z_1, \dots, z_N)$  简写成

$$Q_N(z_1, \dots, z_N) = \sum_S A_S z^S. \quad (1.50)$$

根据 (1.45) 式, 当然也有

$$Q_N(1/z_1, 1/z_2, \dots, 1/z_N) = (z_1 z_2 \cdots z_N)^{-1} Q_N(z_1, z_2, \dots, z_N). \quad (1.51)$$

这个结论也可以通过直接在 (1.50) 式中交换集合  $S$  与其补集  $\bar{S}$ , 并利用  $A_{\bar{S}} = A_S$  直接证明。

我们要研究的, 就是巨配分函数  $\Xi(z_1, z_2, \dots, z_N)$  关于所有复自变量  $z_i$  的零点, 很显然, 这等价于研究多项式  $Q_N(z_1, \dots, z_N)$  关于所有复自变量  $z_i$  的零点。

## 1.5 单位圆定理

### 定理叙述

先概括一下上一节关于多项式  $Q_N(z_1, \dots, z_N)$  的定义: 设  $(A_{ij})$  是实对称矩阵, 满足任意矩阵元  $0 < A_{ij} \leq 1$ 。对  $[N] = \{1, 2, \dots, N\}$  的子集  $S$ , 定义  $A_S = \prod_{i \in S} \prod_{j \notin S} A_{ij}$ , 其中  $A_\emptyset = A_{[N]} = 1$ 。定义  $N$  元多项式

$$Q_N(z_1, \dots, z_N) = \sum_{S \subseteq [N]} A_S z^S,$$

式中  $z^S = \prod_{i \in S} z_i$ 。容易证明, 这样的  $N$  元多项式满足

$$Q_N(1/z_1, 1/z_2, \dots, 1/z_N) = (z_1 z_2 \cdots z_N)^{-1} Q_N(z_1, z_2, \dots, z_N). \quad (1.52)$$

所为的**单位圆定理**, 说的就是: 如果  $|z_1| \leq 1, |z_2| \leq 1, \dots, |z_N| \leq 1$ , 则  $Q_N(z_1, \dots, z_N)$  在此范围内的零点, 全部位于  $N$  为复空间  $\mathbb{C}^N$  的如下单位环面  $\mathbf{T}^N$  上

$$\mathbf{T}^N : |z_1| = |z_2| = \cdots = |z_N| = 1. \quad (1.53)$$

不妨引入一个术语, 用  $\mathbb{D}$  表示 (开) 单位圆盘  $\{z : |z| < 1\}$ , 称一个多元多项式  $P(z_1, \dots, z_N)$  是  $\mathbb{D}$ -stable 的, 如果对任意  $z_1, \dots, z_N \in \mathbb{D}$ ,  $P(z_1, \dots, z_N) \neq 0$ 。称线性变换  $T$  是  $\mathbb{D}$ -stable 的, 如果  $T$  将  $\mathbb{D}$ -stable 多项式映为  $\mathbb{D}$ -stable 多项式或者 0。

很显然, 要证明单位圆定理, 只需证明,  $Q_N(z_1, \dots, z_N)$  是  $\mathbb{D}$ -stable 的。而且由于  $Q_N(z_1, \dots, z_N)$  满足反演关系 (1.52), 所以, 如果它对圆盘内部是  $\mathbb{D}$ -stable 的, 则它对圆盘外部也同样是  $\mathbb{D}$ -stable 的。特别的, 这就意味着  $N$  次多项式  $Q_N(z, \dots, z)$  的零点全部位于  $|z| = 1$  的单位圆之上。

### 1.5.1 证明思路

证明的策略是从一些基本的、容易验证的  $\mathbb{D}$ -stable 多项式出发, 通过一系列  $\mathbb{D}$ -stable 的变换得到  $Q_N(z_1, \dots, z_N)$ 。为此, 我们需要以下引理:

#### 引理 1

设复数  $|a| < 1$ , 则多项式  $g(z_1, z_2) = 1 + az_1 + \bar{a}z_2 + z_1z_2$  是  $\mathbb{D}$ -stable 的。证明: 若  $z_1, z_2 \in \mathbb{D}$ , 则

$$|1 + az_1|^2 - |\bar{a} + z_1|^2 = (1 - |a|^2)(1 - |z_1|^2) > 0,$$

$$\text{于是 } |1 + az_1| > |\bar{a} + z_1| \geq |\bar{a}z_2 + z_1z_2|,$$

$$\text{从而必有 } g(z_1, z_2) \neq 0.$$

#### 引理 2: (Asano Contraction)

考虑线性映射  $T_{zw}$  将  $P_1(z, w) = a + bz + cw + dzw$  映射成  $P_2(z) = a + dz$ , 那么  $T_{zw}$  是  $\mathbb{D}$ -stable 的。注意,  $T_{zw}$  相当于将变量  $z$  和  $w$  进行配对收缩, 一对变量收缩成一个 (记为  $z$ ), 如果某个变量没有相应的配对, 那它就收缩为零。人们也称  $T_{zw}$  为 Asano Contraction。

证明: 假设  $P_1(z, w) = a + bz + cw + dzw$  是  $\mathbb{D}$ -stable 的多项式, 下证  $|a| \geq |d|$ 。

反证, 假设  $|a| < |d|$ , 对任意  $|z| < 1$ , 由  $P_1(z, w) = a + bz + (c + dz)w$  可知  $|a + bz| \geq |c + dz|$ , 取极限可知该式对  $|z| = 1$  也成立。

取  $z = 0$  可得  $|a| \geq |c|$ 。若  $|a| \leq |b|$ , 取  $z = -a/b$  得矛盾, 从而必有  $|a| > |b|$ 。

取  $|z| = 1$  满足  $|c + dz| = |c| + |d|$ , 则由三角不等式有  $|a| + |b| \geq |c| + |d|$ , 由  $|a| < |d|$  可知  $|b| > |c|$ , 即  $|d| > |a| > |b| > |c|$ 。

取  $|z| = 1$  满足  $|a + bz| = |a| - |b|$  可得矛盾。

于是必有  $|a| \geq |d|$ , 从而  $P_2(z) = a + dz$  是  $\mathbb{D}$ -stable 的。

#### 定义: (Schur-Hadamard 乘积)

设多项式

$$f(z_1, \dots, z_N) = \sum_{S \subseteq [N]} a_S z^S, \quad g(z_1, \dots, z_N) = \sum_{S \subseteq [N]} b_S z^S,$$

定义其 Schur-Hadamard 乘积为

$$(f \cdot g)(z) = \sum_{S \subseteq [N]} a_S b_S z^S.$$

#### 引理 3: (Hinkkanen's composition theorem)

沿用上面的定义, 若  $f, g$  是  $\mathbb{D}$ -stable 的, 则  $f \cdot g$  也是  $\mathbb{D}$ -stable 的。

**证明：**考虑  $2N$  元多项式  $\phi(z, w) = f(z)g(w) = \sum_{S \subseteq [N]} a_S z^S g(w)$ 。将 Asano Contraction  $T_{z_k, w_k}$  简记为  $T_k$ ，不难看出 (请大家想象一下配对收缩的定义)

$$T_N T_{N-1} \cdots T_1 \phi(z, w) = \sum_{S \subseteq [N]} a_S b_S z^S = f \cdot g$$

显然  $\phi$  关于  $z, w$  是  $\mathbb{D}$ -stable 的，故我们依次作 Asano Contraction  $T_1, T_2, \dots, T_N$  后得到的多项式  $f \cdot g$  也是  $\mathbb{D}$ -stable 的。

最后，考虑多项式

$$f_{i,j}(z) = (1 + A_{ij}z_i + A_{ji}z_j + z_i z_j) \prod_{k \neq i,j} (1 + z_k)$$

对所有  $i < j$  求 Schur-Hadamard 乘积，计算可得

$$Q_N(z_1, \dots, z_N) = f_{12} \cdot f_{13} \cdots f_{N-1,N}$$

由引理 1,  $f_{i,j}$  是  $\mathbb{D}$ -stable 的多项式。于是不断使用引理 3 可得  $Q_N(z_1, \dots, z_N)$  是  $\mathbb{D}$ -stable 的，证毕。

### 1.5.2 物理讨论

回到我们的物理问题，如果取  $z_1 = z_2 = \cdots = z_N = z$ ，即取所有的局域复磁场  $h_i$  均相等，也就是相当于对系统施加的是一个整体的磁场  $h$ 。相应伊辛模型的哈密顿量成为

$$H_Y = -\frac{1}{2} \sum_{i,j,i \neq j} J_{ij} \sigma_i \sigma_j - h \sum_i \sigma_i. \quad (1.54)$$

这里， $z \equiv e^{2h/k_B T}$ 。

这时候格气模型的巨配分函数  $\Xi(z, \dots, z)$  可以简记为  $\Xi(z)$ ，很显然

$$\Xi(z) = \sum_{\{n_i\}} [e^{-H_G/k_B T} \prod_i y_i^{n_i}] = \sum_{\{n_i\}} [(e^{-H_G/k_B T} \prod_j a_j^{n_j}) z^{\sum_i n_i}].$$

注意  $\sum_i n_i$  其实就是总粒子数。所以这就是一个比较标准的以  $z$  为逸度的巨配分函数，它是关于  $z$  的一个  $N$  次多项式。

这时候相应的，也可以将多项式  $Q_N(z, \dots, z)$  简记为  $Q_N(z)$ ，它由一个  $N$  元多项式变成了一个  $N$  次多项式。则由 (1.48) 式知

$$\Xi(z) = e^{J/k_B T} Q_N(z). \quad (1.55)$$

引用单位圆定理，即有， $\Xi(z)$  的零点均分布在复  $z$  空间的单位圆  $|z| = 1$  上。

注意，在格气模型中，格点的总数  $N$  就相当于总体积  $V$ ，也相当于最大粒子数  $M$ ，所以热力学极限其实就是  $N \rightarrow \infty$  的极限。如果这个极限存在 (这对相互作用  $J_{ij}$  有要求，一般要求它是一种短程相互作用)，那根据上面的讨论可知， $\Xi(z)$  在热力学极限之下的零点依然分布在单位圆上，即零点分布区域  $\mathcal{D}$  为复  $z$  平面上的单位圆。

## 参考文献

两篇原始文献:


1. Statistical Theory of Equations of State and Phase Transitions. I. Theory of Condensation, C. N. Yang and T. D. Lee
2. Statistical Theory of Equations of State and Phase Transitions. II. Lattice Gas and Ising Model, T. D. Lee and C. N. Yang

两篇综述:

3. Statistical Mechanics of Equilibrium and Nonequilibrium Phase Transitions: The Yang-Lee Formalism, Ioana Bena, Michel Droz, Adam Lipowski, arXiv:cond-mat/0510278
4. 知乎文章: 李-杨单位圆定理, YoiHong  
<https://zhuanlan.zhihu.com/p/1963625182053762633>

李新征老师组的两篇综述文章:

5. 李-杨相变理论: 历史与新进展, 王方成, 叶麒俊, 李新征
6. 物质的态与相, 叶麒俊, 欧阳霄宇, 李新征



## 2. 物质相的朗道范式

朗道立相变之论，以对称破缺为本，序参量为纲。高温相具至大之对称，低温相则自发性亏其群，由是序以扩张。相界分居里之线，相变有连续之殊。此范式统摄物态虽广，然于分数量子霍尔效应等奇相，犹未尽其妙也。——题记

### 2.1 引言

人们常说物质有三种状态（三种相）：固态、液态和气态。中国古代哲学则有五行学说，大致是说物质有五种状态：金、木、水、火、土，这在今天已经被认为是一种迷信。

稍微了解一点科普的人可能知道，至少还有一种“等离子体”态，它构成了恒星的大部分。最初的那三种状态，如今已演变成了众多类型。从重要且精确的意义上说，磁体是一种独特的物质形态。金属与绝缘体不同。超导体和超流体则是引人注目的新物质状态。你电脑的液晶屏幕是众多不同液晶物质状态中的一种（向列相、胆甾相、蓝相、蓝相和蓝雾相、近晶相 A、B、C、C\*、D、I 等）。存在超过 200 种具有质的差异的晶体类型，更不用说还有准晶体了。还有像自旋玻璃这样的无序物质状态，以及像分数量子霍尔效应这样具有电荷为  $e/3$  的激发态（类似夸克）的物质状态。

总之，物质状态多种多样，因此，我们就必须建立一套体系，它要能一般性地分类物质的不同相，正确描述不同相的不同物理性质，甚至还要能描述不同相之间的相变。这个理解物态的体系主要是由朗道建立起来的（L. D. Landau, On the theory of phase transitions, Zh. Eksp. Teor. Fiz. 7, 19 (1937).）。介绍朗道关于物质相的这个范式就是本文的核心目标。

不过，朗道这个体系并不能解释诸如分数量子霍尔效应这样的新物态，这些新物态的发现曾经令人们大伤脑筋，虽然最终人们弄明白了它们为何会有那样的表现，并且由此拓展了我们理解物态的体系。不过，如果你想进一步理解这些新的拓展，那就需要在本文之外进一步去接触科研的前沿了。

## 2.2 相与相变的基本概念

相变不是组成物质的分子的微观改变，也不是少量分子的变化，而是宏观数量的分子宏观物理性质的改变。通常有不止一个变量可以影响物质的相。例如可以通过降低温度或者增加压强来使得蒸气凝结。我们使用相图来显示平衡相与温度和压强的关系，这就是所谓的  $T-p$  相图。在  $T-p$  图上的不同区域标示出物质相应的相，不同相之间的分界线，就是所谓的相界线。为了明晰起见，我们主要讨论  $T-p$  相图，显然，这相应于等温等压的系统。

问题是，给定一个压强和温度，如何判断系统会处于哪个相呢？

为了回答这个问题，考虑某个和温度为  $T$  的热库接触，体积可变，但是压强与热库一样为  $p$  的系统。假设经过某个过程，系统的体积发生了改变，同时从热库中吸收了热量  $Q$ ，由于热库很大，其放热过程可以看作是可逆的，因此热库的熵变为  $dS_R = -Q/T$ 。假设系统的熵变为  $dS$ ，则系统和热库一起的总熵变就是  $dS + dS_R$ ，由于这两者一起构成了一个大的孤立系统，所以根据熵增加原理，必有

$$dS + dS_R \geq 0. \quad (2.1)$$

对系统应用热力学第一定律，即有  $Q = dE + pdV$ ， $dE$  是上述过程中系统的能量增加量， $dV$  是其体积增加量，从而即有

$$dS + dS_R = dS - \frac{1}{T}(dE + pdV) = -\frac{1}{T}(dE - TdS + pdV) = -\frac{1}{T}d\mathcal{G} \geq 0. \quad (2.2)$$

式中  $\mathcal{G} = E - TS + pV$  为系统的吉布斯自由能。因此熵增加原理告诉我们，等温等压系统的任何自发过程都必然会导致吉布斯自由能减少，即

$$d\mathcal{G} \leq 0. \quad (2.3)$$

特别的，当系统达至热平衡时，吉布斯自由能取极小值。

将上述原理用到相变问题上，即有：在给定的温度和压强下，稳定的相永远是拥有更低吉布斯自由能的相！其自然推论是，在两相之间的相界线上，两个相的吉布斯自由能必定相等！

简单起见，这里只讨论液体和气体的相界线，类似的讨论当然也适用于任何相界线。我们考虑一定量的物质，比如 1mol。我们已经知道，在相界线上，液体和气体的吉布斯自由能（分别记为  $\mathcal{G}_l$  和  $\mathcal{G}_g$ ）必定相等，即

$$\mathcal{G}_l = \mathcal{G}_g. \quad (\text{在相界线上}) \quad (2.4)$$

沿着相界线吉布斯自由能保持相等的结论也可以说成是：沿着相界线，两相的化学势相等。因为根据 Gibbs-Duhem 关系式有  $\mathcal{G} = \mu N$ ，而不管物质处于哪一相，其粒子数肯定是一样的，所以吉布斯自由能相等就等价于化学势相等，即

$$\mu_l = \mu_g. \quad (2.5)$$

通常称这个条件为**相平衡条件**，它也可以用另一种方法来得到，即考虑共存的两相，把它们看成一个孤立系统的两个子系统，两相热平衡的必然推论就是：**两相的温度相等；两相的压强相等；同时，两相的化学势也相等。**

由于相变前后，吉布斯自由能保持相等，也就是吉布斯自由能在相变前后是连续的。但是，如果相变前后熵和体积都发生了不连续的突变，而熵和体积都是吉布斯自由能的一阶偏导（分别对  $T$  和  $p$ ），那这就说明，对于这种相变，相变前后吉布斯自由能的一阶偏导是不连续的！这样的相变就称之为**一级相变**，也就是吉布斯自由能连续但是其一阶偏导不连续的相变。

但是，还有一种相变，系统越过相界线从一种相进入另一种相时，相变潜热等于零，或者说熵和体积也同样没有突变。换言之，对于这种相变，不仅吉布斯自由能是连续的，熵也是连续的，类似的体积也是连续的。像这样的，不仅吉布斯自由能连续，而且其一阶偏导也连续的相变就称作**连续相变**。值得注意的是，连续相变也会产生物理量的不连续或者无穷尖峰的奇异行为，但是这些物理量都是吉布斯自由能的二阶以上的偏导，是诸如压缩率、磁化率这样的物理量。

## 2.3 对称性自发破缺与序的扩张

在最初的固液气三相中，液相和气相之间的相变相对比较好理解。而液体与晶体的本质区别在于，液体是各向同性的，而晶体是各向异性的。从晶体到液体，或到具有不同对称性的晶体的每一次相变，都与某些对称元素的消失或出现相关联。但对称元素要么存在，要么不存在，不可能存在中间情况，所以与液气之间的区别不同，液体和晶体之间的区别更具本质性。

注意，液体与晶体之间或对称性不同的晶体之间的正常相变是有相变潜热的，其中物体的状态（特别是能量）要发生不连续变化，这种相变虽然重要，但并非本文的关注重点。我们主要感兴趣的是这样一种转变：即使对称性发生不连续变化，物体的状态（特别是能量）却是连续变化的，这就是前面说过的连续相变。注意，它们与液-气转变的连续性意义有所不同，这里在任何时刻，我们都能判断物体具有这种或那种对称性。

### 2.3.1 相与对称性

基础于对液体与晶体之间区别的观察，同时也基础于不同对称性的晶体之间的区别，朗道提出：物质的每一个相都对应一个对称群  $H$ （请不要与哈密顿量混淆），不妨记这个相 (Phase) 为  $P^H$ 。假设系统微观相互作用的对称群为  $G$ ，那么  $H$  一定是  $G$  的子群， $H \subseteq G$ ，因为  $H$  反映的是物质宏观态的对称性，它当然是微观对称性在宏观上的反映，但只会比

微观相互作用的对称性小。对称性  $H$  不同的相之间会被相界线相区隔，也就是会发生相变。

在高温时，组成物质的分子热运动的能量远远超过分子之间相互作用的能量，因此，这些分子各自为政，每个分子的运动状态都随机地改变，各种运动状态完全对称，整个系统是对称而且无序的。这时候系统的宏观状态具有微观允许的最大对称性  $G$ ，相应的相  $P^G$  也是最无序的。换言之，高温相的对称性是微观相互作用所能允许在宏观上表现出来的最大对称性。

随着温度的下降，分子之间的相互作用效应开始超过无规则热运动的效应，这时候会发生一件神奇的事情：这些分子开始自发且协调一致地行动，形成一种规律性的宏观模式。换言之，系统从无序的热运动中自发产生了有序。但是，这种规律性的宏观模式一般来说不能保持最大的对称性  $G$ ，而只能保持它的一个子群  $H$ 。朗道称这种由微观相互作用的对称性  $G$  到宏观状态的对称性  $H$  之间的转变为对称性自发破缺。例如晶体自发破缺了空间平移对称性，铁磁体自发破缺了自旋旋转对称性。随着对称性的自发破缺  $G \rightarrow H$ ，系统也就发生了相变  $P^G \rightarrow P^H$ 。由于高温相  $P^G$  是最无序的，对称性破缺以后的相  $P^H$  是在无序中产生了有序，所以我们也称  $P^G \rightarrow P^H$  的相变为**序的扩张**，记作  $P^G \subset P^H$ ，这个符号的含义不是说  $P^G$  相包含在  $P^H$  相之中，而是说  $P^H$  相比  $P^G$  更有序，或者说序更大。注意，对称性的包含关系  $H \subset G$  和序的包含关系  $P^H \supset P^G$  是相反的。

事实上，在朗道之前，皮埃尔·居里在研究磁性时，已经强调了群论和对称性破缺的重要性 (Curie, P., Sur la symétrie dans les phénomènes physiques, symétrie d'un champ électrique et d'un champ magnétique, J. Phys. Theor. Appl. 3, 393 (1894))。问题是，对称性自发破缺的机制是什么呢？以及，从这个对称性的角度来看，连续相变和一阶相变有什么不同呢？

### 2.3.2 对称性自发破缺的机制

现在让我们来研究系统如何从具有最大对称性  $G$  的高温相自发破缺到具有对称性  $H$  的相  $P^H$ 。设高温相  $P^G$  由取值为  $\phi_G$  的宏观状态量描述，由于这时候  $G$  也是系统的宏观对称性，所以  $G$  作为系统微观相互作用的对称性作用在  $\phi_G$  上时，应该要保持它不变，记作

$$g \cdot \phi_G = \phi_G, \quad g \in G \quad (2.6)$$

式中  $g \cdot \phi_G$  表示群元  $g$  作用在  $\phi_G$  上。在高温相时系统的吉布斯自由能当然是  $\phi_G$  的函数，记作  $\mathcal{G}(\phi_G) \equiv \mathcal{G}_G$ ，当然  $\mathcal{G}_G$  同时是温度  $T$  和压强  $p$  的函数， $\mathcal{G}_G(T, p)$ 。

当系统对称性自发破缺到  $H$  时，相变发生了，系统的序也扩张了  $P^G \subset P^H$ 。为了描述  $P^H$  相中多出来的秩序，我们引入**序参量**  $\phi_i$ ，由于它是  $P^G$  相之外的秩序，所以  $\phi_i$  当然不是  $G$  不变的，不妨假设它构成了群  $G$  的线性不可约表示，指标  $i$  表示的就是线性矢量的第  $i$  分量，记作

$$g \cdot \phi_i \equiv \sum_j R(g)_{ij} \phi_j, \quad g \in G, \quad (2.7)$$

式中  $R(g)$  表示群元  $g$  的表示矩阵。同时，由于  $\phi_i$  描述的是  $P^H$  相的序，所以它得具有  $H$

对称性，即在  $H$  的作用下保持不变，记作

$$h \cdot \phi_i = \phi_i, \quad h \in H. \quad (2.8)$$

不过， $\phi_i$  我们是看作一个变量，它具体取何值待定，可以认为，正是  $\phi_i$  最终的具体取值才**通过上式确定了低温相的对称性  $H$** 。我们需要解释的，正是  $\phi_i$  是如何从高温相时取值为零，到穿过相界线以后取值非零，从而使得系统对称性自发破缺的。

为此，我们需要同时将自由能看作  $\phi_i$  的函数，记作  $\mathcal{G}(\phi_G, \phi_i)$ ，然后通过自由能极小的条件来确定  $\phi_i$  的具体取值。由于  $G$  是系统的微观对称性，所以我们还得假设，虽然  $\phi_i$  构成  $G$  的表示，但自由能  $\mathcal{G}$  本身得是  $G$  不变的。另外，由于高温时  $\phi_i = 0$ ，所以为了研究相变如何发生，我们可以把  $\phi_i$  当作小量，然后将自由能  $\mathcal{G}(\phi_G, \phi_i)$  对  $\phi_i$  进行级数展开。

展开的零阶当然就是  $\mathcal{G}_G$ ，而由于由不可约表示无法构造一阶的  $G$  不变量，因此展开的一阶项是缺失的，由不可约表示构造的二阶  $G$  不变量就是二次型  $\sum_i \phi_i^2$ ，所以我们有

$$\mathcal{G}(\phi_G, \phi_i) = \mathcal{G}_G(T, p) + A(T, p) \sum_i \phi_i^2 + \dots \quad (2.9)$$

为了确保对于高温相  $P^G$ ，自由能的极小值对应  $\phi_i = 0$ ，就必须要求在高温相  $A(T, p) > 0$ 。类似的，为了确保在  $P^H$  相，自由能的极小值对应非零的  $\phi_i$ ，就得要求在  $P^H$  相， $A(T, p) < 0$ 。所以， $P^G$  相和  $P^H$  相之间的相界线由下式定义

$$A(T, p) = 0. \quad (2.10)$$

但是，为了确保这条线上的点真的以  $\phi_i = 0$  为自由能的极小值位置，还需要展开的三阶项等于零，否则这个三阶项就可以导致一个负的非零  $\phi_i$  值。为此，我们假设对于我们所考虑的特定不可约表示  $\phi_i$ ，**不存在三阶  $G$  不变量**。虽然不是所有情况下这条假设均成立，但一方面假设成立的情况很常见（比如，如果在  $G$  的作用之下， $\phi_i$  可以改变正负号，那所有奇数阶的  $G$  不变量都将不存在），另一方面，这样假设有助于简化对朗道相变理论的分析。

与此同时，如果自由能级数展开的四阶项恒正，则  $P^G$  相和  $P^H$  相之间的界线就由 (2.10) 式完全确定。不难看出，这定义的是  $T-p$  相图上的一条曲线，通常称作**居里线**或者**临界线**。对称性自发破缺的机制正在于，穿过居里线之后， $A(T, p)$  由正变成了负，导致  $\phi_i$  自动获得了一个非零取值，这个非零取值当然不是  $G$  不变的，但根据序参量  $\phi_i$  的定义，它一定是  $H$  不变的，从而使得系统宏观状态的对称性自发破缺到了  $H$ 。

为了看得更清楚一点，不妨定义

$$\sum_i \phi_i^2 \equiv \phi^2, \quad \phi_i/\phi \equiv \gamma_i. \quad (2.11)$$

则自由能可以展开为

$$\mathcal{G} = \mathcal{G}_G + A\phi^2 + B\phi^4 + \dots, \quad (2.12)$$

其中所有的系数都是  $T, p$  的函数，且假设  $B > 0$  恒成立。一般来说， $B$  还会依赖于  $\gamma_i$ ，这时候求自由能的极小值就要求先求出  $B(\gamma_i)$  对  $\gamma_i$  的极小值位置。不过，依然是为了简化分

析(同时也是很常见的情况),我们假设  $B$  与  $\gamma_i$  无关,这相当于说四阶的  $G$  不变量有且仅有  $(\sum_i \phi_i^2)^2 = \phi^4$  这一个。

如此一来,在居里线以上(高温相),  $A > 0, B > 0$ , 自由能的极小值对应  $\phi = 0$ , 系统的宏观状态具有对称性  $G$ 。而在居里线之下(穿过居里线),  $A < 0, B > 0$ , 从极小值条件  $\frac{\partial \mathcal{G}}{\partial \phi} = 0$  可以得到

$$A + 2B\phi^2 = 0 \Rightarrow \phi^2 = -A/(2B) \equiv \phi_H^2. \quad (2.13)$$

从而  $\phi$  获得了一个非零取值  $\phi_H$ , 系统的对称性自发破缺到了子群  $H$ 。进一步不难看出,这时候自由能的值为

$$\mathcal{G} = \mathcal{G}_0 - A^2/(4B). \quad (2.14)$$

由于在相界线上  $A = 0$ , 不难看出穿过相界线的时候,自由能是连续的,类似的,也不难看出自由能对  $T, p$  的一阶偏导也是连续的,从而,这样的相变为连续相变。所以, **对称性自发破缺对应的常常是连续相变。**

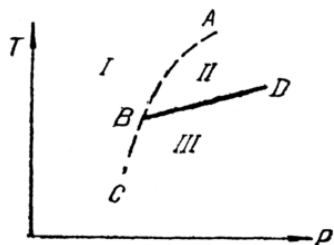
不难求出从下面的  $P^H$  相无限趋近于居里线时的定压热容  $c_p$ , 为

$$c_p = -T \frac{\partial^2 \mathcal{G}}{\partial T^2} = c_G + \frac{T}{2B} \left( \frac{\partial A}{\partial T} \right)^2, \quad (2.15)$$

式中  $c_G$  为  $\mathcal{G}_G$  项贡献的热容。注意,虽然在居里线上  $A = 0$ , 但是  $\frac{\partial A}{\partial T}$  一般来说可不等于零,这就说明,从高对称性相  $P^G$  穿过居里线到达低对称性相  $P^H$  时,系统的热容发生了一个不连续的跳变,这正是相变在物理上的反映。

### 2.3.3 相图的结构与子群格

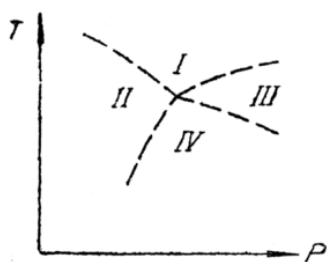
在更具一般性的情形中,四阶系数  $B$  可能依赖于  $\gamma_i$ , 成为  $B(\gamma_i)$ , 这时候自由能的极小值要求同时确定  $B(\gamma_i)$  关于  $\gamma_i$  的极小值位置。由于  $H$  是由  $\phi_i$  的具体取值所保持的对称性所确定的,从而也就是由  $\gamma_i$  的这个取值(这个取值是依赖于  $T, p$  的)所保持的对称性来确定的。这时候就可能发生这样的事情,即居里线的不同部分确定了不同的低温对称性  $H$ 。这种情况的相图将会类似于图(2.3.3)。



图(2.3.3)中的虚线就是居里线,  $I$  区域就是高对称的  $P^G$  相 ( $\phi_i = 0$ ),  $II$ 、 $III$  两个区域是两个不同的低对称相,它们的  $\phi_i$  非零取值所保持的对称性不同。假设  $II$ 、 $III$  区域的对称群不是彼此的子群,从而在两者之间无法发生上述对称性自发破缺。那这两个区域之

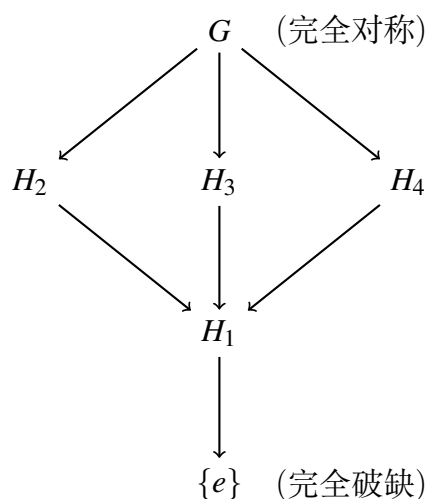
间的相界线，即图中的  $BD$  线，就不可能是居里线。从而，这两个区域之间的相变一般来说就不会是连续相变，而更可能是一阶相变。

当然，还可能发生这样的事情，即一次对称性自发破缺没有把高温对称性  $G$  破缺光，剩余的对称性可能再次发生对称性自发破缺。这时候就可能出现图 (2.3.3) 这样的居里线相交的情形。在这幅图中， $I$  区域是高对称相， $II$ 、 $III$  区域的对称性就低一些，它们的对

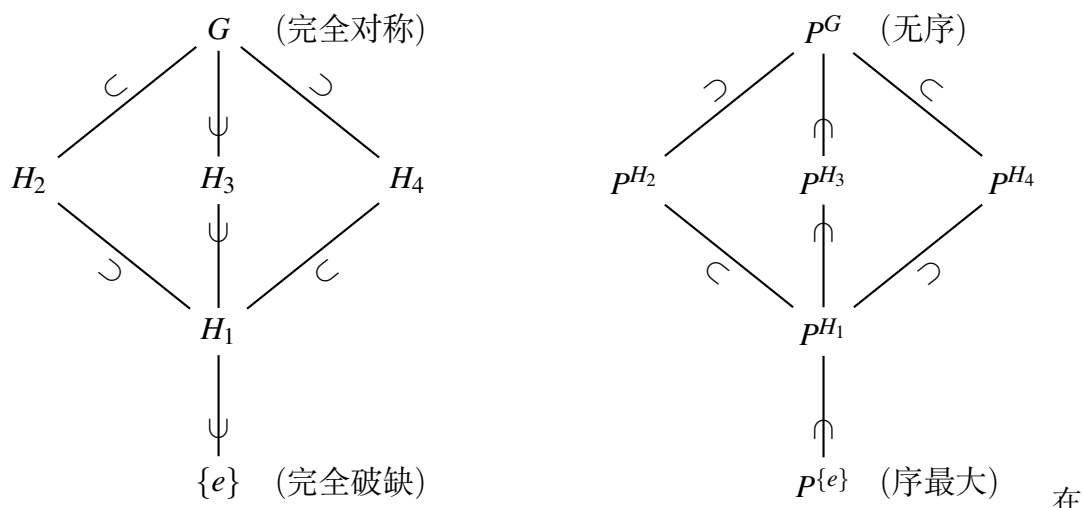


称群分别是  $I$  区域对称群的子群。 $IV$  区域的对称性则再低一些，它的对称群既是  $II$  区域对称群的子群，也是  $III$  区域对称群的子群。从  $I$  区域穿过居里线到  $II$  区域会发生对称性自发破缺，系统的序扩张了，从  $II$  区域到  $IV$  区域再次发生对称性自发破缺，系统的序再次扩张。

我们可以用类似于下图的子群格来刻画整个相图的总体结构。图中的箭头表示对称性自发破缺的方向，每个箭头都从一个更大的群指向其对称性自发破缺之后的子群。当然，比方说  $H_2$  也可能是  $H_3$  的子群，但图中没有从  $H_3$  指向  $H_2$  的箭头就说明没有发生从  $H_3$  到  $H_2$  的对称性自发破缺。



更明确的，我们也可以把对称性自发破缺导致的子群包含关系以及对称性自发破缺导致序的不断扩张画成如下图示。注意，在图中序的扩张方向和子群的包含关系刚好是相反的。



在  $T-p$  相图上，能够通过对称性自发破缺联系起来的两个相之间的相界线是居里线，而像上图中  $P^{H_2}$  和  $P^{H_3}$  这样不能通过对称性自发破缺联系起来的两个相，它们之间的相界线就不是居里线，而往往是一阶相变线。

## 2.4 连续对称性自发破缺

如果高温对称性  $G$  和自发破缺以后剩余的对称性  $H$  都是连续对称性，那么还会有一些新的物理效应。为了讲清楚这个效应，我们限制于考虑 (2.3.2) 小节中假设的情形，即假设四阶的  $G$  不变量有且仅有  $(\sum_i \phi_i^2)^2 = \phi^4$  这一个。

### 2.4.1 序参量的涨落

根据 (2.3.2) 小节中的讨论，这时候在低温相，序参量的取值得满足下式

$$\sum_i \phi_i^2 = \phi^2 = \phi_H^2, \quad (2.16)$$

式中  $\phi_H$  是给定的。但很显然，这个式子并不能唯一确定  $\phi_i$ ，任何一个满足上式的特定  $\phi_i$  值相当于在整个系统的全空间选取一个特定的  $\phi$ —“方向”，从而把对称性从高温相的  $G$  自发破缺到低温相的  $H$ 。由于每一个  $\phi_i$  都是  $H$  不变的，很显然所有满足 (2.16) 式的  $\phi_i$  的集合刚好构成陪集空间

$$G/H. \quad (2.17)$$

但是，由于满足 (2.16) 式的  $\phi_i$  有很多，所以也完全可能发生在空间不同区域  $\phi_i$  取不同“方向”的事情，甚至不同空间局域上  $\phi_i$  的取值在 (2.16) 式允许的所有可能性中随机选取，我们称之为**序参量的涨落**。它使得  $\phi_i$  是空间点  $\mathbf{x}$  的函数，从而使得 (2.16) 式变成

$$\sum_i \phi_i^2(\mathbf{x}) = \phi_H^2. \quad (2.18)$$

因此，对称性自发破缺以后，如果考虑到序参量的“方向”可以在整个陪集空间  $G/H$  上涨落，对于连续对称性，那低温有序相的序参量实际上就是从三维空间到  $G/H$  上的连续

映射，即

$$\mathbf{R}^3 \rightarrow G/H. \quad (2.19)$$

低温有序相的物理性质就取决于序参量的这种涨落。所以为了描述低温有序相，就需要写出低温时自由能作为满足 (2.18) 式的序参量场  $\phi_i(\mathbf{x})$  的泛函。不过，由于序参量的“方向”作一个整体的改变结果在物理上完全等价，也即是说，序参量“方向”的整体改变完全没有物理效应，从而可以知道，自由能对序参量的级数展开中除了 (2.3.2) 小节中写出来的整体贡献部分之外，序参量涨落引起的额外贡献  $\mathcal{G}_e$  必定只依赖于序参量场随空间的变化情况，也就是序参量场的空间梯度。这里的一个推论是，序参量场涨落所贡献的自由能可以任意地小，只需要让序参量场足够平滑，随空间的变化足够缓慢。这个推论被称为戈德斯通定理。

另外， $\mathcal{G}_e$  当然不依赖于空间  $G/H$  的局部坐标选取。因此对  $\mathcal{G}_e$  最简单的贡献只能由陪集流形  $G/H$  上的  $G$  不变线元

$$ds^2 = \sum_{\alpha, \beta} g_{\alpha\beta}(\psi) d\psi^\alpha d\psi^\beta. \quad (2.20)$$

其中  $\psi^\alpha$  为流形  $G/H$  上的局部坐标，指标  $\alpha = 1, 2, \dots, m$ ,  $m$  为空间  $G/H$  的维数。因此，在此局部坐标下， $\psi^\alpha(\mathbf{x})$  作为序参量场就定义了映射  $\mathbf{R}^3 \rightarrow G/H$ 。那么结合这两段的讨论，比如结合  $\mathcal{G}_e$  不依赖于  $G/H$  上局部坐标的具体选取的要求，可以知道  $\mathcal{G}_e$  的最低阶贡献必定可以写成

$$\mathcal{G}_e = \int d^3\mathbf{x} \rho_s \sum_{\alpha, \beta} g_{\alpha\beta}(\psi) \nabla\psi^\alpha \cdot \nabla\psi^\beta + \dots \quad (2.21)$$

其中我们已经考虑到了三维空间的旋转不变性，式中的  $\rho_s$  是某个比特常数。式中这一项是  $(\nabla\psi)^2$  阶项，省略号表示诸如  $(\nabla\psi)^4$  这样的具有更多个偏导的高阶修正项。这些更多偏导的项之所以能被忽略，是因为我们主要考察随空间变化比较平缓的序参量涨落，它的空间偏导是很小的，因此某项中包含的偏导个数越多就越能够被忽略，这个论证也适用于后文的相关地方，之后我们不再重复。

很显然， $\mathcal{G}_e$  是场变量  $\psi(\mathbf{x})$  的泛函，而且不是平方类型的泛函，而是真正非线性的，所以它也称作非线性西格玛模型。注意，它描写的是低温有序相的序参量涨落。这样的非线性西格玛模型在离散对称性自发破缺中是不存在的，它是连续对称性自发破缺特有的物理。

利用低温时的这个  $\mathcal{G}_e$ ，我们就可以计算低温时序参量涨落所贡献的配分函数，公式是

$$\mathcal{Z} = \int \mathcal{D}\psi(\mathbf{x}) e^{-\beta\mathcal{G}_e[\psi(\mathbf{x})]}, \quad (2.22)$$

式中  $\psi(\mathbf{x})$  示意性地指代所有  $\psi^\alpha(\mathbf{x})$ 。当然，计算这个泛函积分并不是一件简单的事情，感兴趣的读者请自行阅读统计场论的相关书籍。

### 2.4.2 铁磁相变

不妨举铁磁相变为例。铁磁体可以看作是由一些自旋组成的，在微观上具有作用在自旋上的三维空间旋转对称性，通常称之为  $O(3)$  对称性。邻近的自旋之间存在相互作用，它使得当邻近自旋指向一致时，系统的能量取值比较低。高温时，热运动效应占据了上风，自旋间的相互作用相比不重要，这时候每个自旋都随机指向，整个系统在宏观上表现出各向同性，因此具有  $G = O(3)$  的三维自旋旋转对称性。低温时，自旋间的相互作用占了上风，它使得所有自旋趋向于指向同一个空间方向，系统的对称性自发破缺到绕这个特定方向的旋转对称性，即  $H = O(2)$ 。这个低温相就是所谓的铁磁相。

为了描写铁磁相变的这种对称性自发破缺，我们可以将自旋磁矩在物理小的局部区域上取平均，进而定义磁化强度，现在它同样是三维矢量，记作  $\mathbf{m}(\mathbf{x})$ 。 $\mathbf{m}(\mathbf{x})$  是空间上的矢量场，可以想象成是每个空间点上放一个小箭头（表示  $\mathbf{m}$  的方向和大小），只不过，每一点上的小箭头方向都可以进行任意的三维旋转。在那个高温的具有  $O(3)$  对称性的相，由于每一个自旋都随机取向，因此  $\mathbf{m}$  的取值为零，没有哪个方向特殊。

在临界温度以下，每一个空间点附近的物理小区域内的自旋就会因为相互作用而协调起来，进而都指向某一个特定的方向，产生了宏观上的秩序，这种秩序就表现为磁化强度  $\mathbf{m}(\mathbf{x})$  不再是零了，系统出现了自发磁化！因此  $\mathbf{m}(\mathbf{x})$  从取值为零变成非零就标志着系统从高温无序相转变成了低温铁磁相，因此  $\mathbf{m}(\mathbf{x})$  就是序参量。

人们完全可以仿照朗道的办法，将自由能对序参量进行泰勒展开，进行建立上述铁磁相变的物理模型。很显然，在高温相，自由能的极小要求  $\mathbf{m} = 0$ 。类似于前文的讨论，在低温相，自由能的极小要求  $\mathbf{m}(\mathbf{x})$  为常矢量，且

$$|\mathbf{m}(\mathbf{x})| = m_H, \quad (2.23)$$

式中  $m_H > 0$  是常数。也即是说，在居里线之下，相互作用会协调空间每一点上的小箭头，使得它们长度都为  $m_H$ ，且方向都一样。

我们已经看到，如果不考虑序参量的涨落，那么居里线以下，空间点上这些长度一样的小箭头就会同步地选取一个特定的空间指向  $\mathbf{n}$ （这里  $\mathbf{n}$  是一个不依赖于空间点  $\mathbf{x}$  的单位矢量，满足  $|\mathbf{n}| = 1$ ），进而破缺掉三维旋转对称性。换言之，居里线之下，序参量可以写成  $\mathbf{m}(\mathbf{x}) = m_H \mathbf{n}$ 。当然，小箭头们具体选哪个指向是随机的，任何一个方向都同样可能。因此在居里线之下， $O(3)$  对称性自发破缺了！不过，破缺之后其实还剩余一种对称性，那就是，选定了空间指向的小箭头们还可以绕着各自的轴线旋转任意一个角度，通常称这为  $O(2)$  对称性，也就是二维平面绕着某个竖直轴的旋转对称性。所以，在居里线之下，铁磁体的  $O(3)$  对称性自发破缺到了  $O(2)$  对称性。

所有不同指向但是地位平等的单位向量  $\mathbf{n}$  的集合构成了三维空间中的二维单位球面  $S^2$ 。在绕  $\mathbf{n}$  自身轴旋转的  $O(2)$  对称性的作用下，这个  $S^2$  上的点是保持不动的，但在模去  $O(2)$  的  $O(3)$  旋转作用下， $\mathbf{n}$  会改变方向，从而会把  $S^2$  上的一点变换到另一点。正是在这个意义上，我们说这个  $S^2$  可以等同于破缺前的对称性  $O(3)$  模去破缺后剩余的对称性  $O(2)$ ，记作

$$S^2 = O(3)/O(2). \quad (2.24)$$

依然假定在居里线之下。不过，如果考虑到序参量的涨落，那方向  $\mathbf{n}$  就可以依赖于空间点  $\mathbf{x}$ ，也即是说不同点的小箭头可以指向不同的方向了。这时候的序参量就是一个不均匀的场分布  $\mathbf{m}(\mathbf{x}) = m_H \mathbf{n}(\mathbf{x})$ ,  $|\mathbf{n}(\mathbf{x})| = 1$ 。很显然， $\mathbf{n}(\mathbf{x})$  定义了一个三维空间到  $S^2$  的连续映射，

$$\mathbf{n}(\mathbf{x}) : \mathbf{R}^3 \rightarrow S^2, \quad (2.25)$$

这个映射决定了低温铁磁相序参量场  $\mathbf{m}(\mathbf{x})$  的空间分布。低温铁磁相的物理性质显然由这些序参量的涨落决定。

我们可以设所有的小箭头都同步指向某个特定方向时，序参量涨落额外贡献的自由能  $\mathcal{G}_e = 0$ 。但是，由于所有的方向都是等价的，所以如果把这些指向某个特定方向的小箭头同步地转过一个任意的方向，结果自由能的额外贡献将依然为零，即依然有  $\mathcal{G}_e = 0$ 。由此我们可以推断， $\mathcal{G}_e$  的非零贡献只能来自于  $\mathbf{m}(\mathbf{x})$  随不同空间点的变化（因为同步的方向变化不贡献能量），也就是依赖于  $\mathbf{m}(\mathbf{x})$  的空间梯度，进而也就是依赖于  $\mathbf{n}(\mathbf{x})$  的空间梯度。

由于  $\mathbf{n}(\mathbf{x})$  就是一个  $\mathbf{R}^3 \rightarrow S^2$  的连续映射，如果想把这个映射写得更清楚的话，就需要在  $S^2$  上引入局部坐标系，比如球坐标。这里的要点是， $S^2$  上的局部坐标完全是人为的，因此不止可以取球坐标，实际上是可以有无数多种局部坐标，但是，自由能泛函  $\mathcal{G}_e[\mathbf{m}(\mathbf{x})]$  当然不依赖于  $S^2$  上局部坐标的选取方式。因此，为了写出自由能泛函，我们首先得问， $S^2$  上什么东西是不依赖于局部坐标的，回答可能很多，但最简单的一定是，两个无限靠近的点在  $S^2$  上的距离平方不依赖于局部坐标的选取！这个距离平方很显然是  $ds^2 = (dn_x)^2 + (dn_y)^2 + (dn_z)^2$ ，其中  $\mathbf{n} = (n_x, n_y, n_z)$ ，且满足约束  $n_x^2 + n_y^2 + n_z^2 = 1$ 。

出于一些纯数学上的方便，可以在  $S^2$  上引入复的局部坐标  $(\psi, \bar{\psi})$  ( $\bar{\psi}$  表示  $\psi$  的复数共轭)，其定义如下

$$n_x + in_y = \frac{2\psi}{1 + |\psi|^2}, \quad n_x - in_y = \frac{2\bar{\psi}}{1 + |\psi|^2}, \quad n_z = \frac{1 - |\psi|^2}{1 + |\psi|^2}. \quad (2.26)$$

利用这个局部坐标，经过一番计算即可以将  $S^2$  上的  $ds^2$  写成

$$ds^2 = \frac{4d\psi d\bar{\psi}}{(1 + |\psi|^2)^2}. \quad (2.27)$$

注意这个式子虽然是在特定局部坐标下写出来的，但  $ds^2$  本身实际上并不依赖于局部坐标的选取。

$\mathbf{R}^3 \rightarrow S^2$  的连续映射当然也可以看作是由  $\psi(\mathbf{x})$  和  $\bar{\psi}(\mathbf{x})$  决定。而自由能泛函  $\mathcal{G}_e$  只依赖于  $\mathbf{n}(\mathbf{x})$  的空间梯度也就是只依赖于  $\psi(\mathbf{x})$  和  $\bar{\psi}(\mathbf{x})$  的空间梯度。结合  $\mathcal{G}_e$  不依赖于  $S^2$  上局部坐标的具体选取的要求（因此由  $S^2$  上的  $ds^2$  决定），可知，对  $\mathcal{G}_e$  的最简单贡献必定可以写成

$$\mathcal{G}_e = \int d^3\mathbf{x} \rho_s \frac{4\nabla\psi \cdot \nabla\bar{\psi}}{(1 + |\psi|^2)^2} + \dots \quad (2.28)$$

### 2.4.3 拓扑缺陷

在铁磁相，序参量是  $\mathbf{R}^3 \rightarrow S^2$  的映射。通常这个映射可能就如水面上的小涟漪。但是，有时候也可以出现更奇特的映射。比如如图 (2.1) 所示的那样，磁化强度从一点向四周发

散出去，形成类似于刺猬那样的序参量场位形。这样的场位形是所谓拓扑缺陷的一种。所谓**拓扑缺陷**，是指在某些物理系统中，由于整体结构的“拓扑性质”限制，而无法通过平滑连续的微小变化消除的、稳定的不规则性或者“瑕疵”。

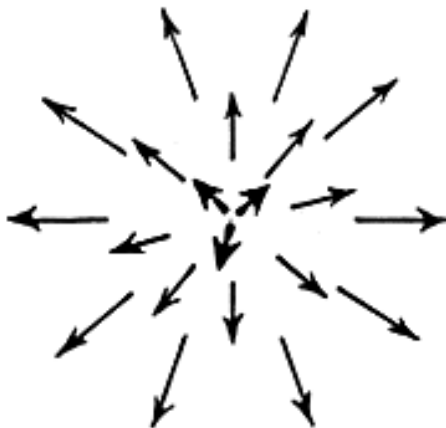


Figure 2.1: 铁磁相刺猬形的拓扑缺陷

那么限制如图 (2.1) 所示刺猬形拓扑缺陷的整体结构是什么呢？为了回答这个问题，不妨设想取一个二维球面包围刺猬中心的那一个点，因为刺猬的每一根刺都对应序参量空间  $S^2$  上的一个点，因此这球面上的那些“刺”就定义了一个  $S^2 \rightarrow S^2$  的连续映射，后一个  $S^2$  表示的是序参量空间，也就是  $O(3)/O(2)$ 。而且，具体对于图 (2.1) 中画的刺猬，很显然，这个映射正好是用前一个  $S^2$  包裹后一个  $S^2$  一次。这种包裹当然无法通过连续的微小变化消除，因此是限制刺猬形拓扑缺陷的整体结构。

然而，人们也可以想象映射  $S^2 \rightarrow S^2$  不止包裹后者一次，而是可以包裹任意整数次，其中负整数表示内外翻转的包裹。很显然，不同的包裹次数之间不能连续地过渡，因此包裹次数的不同就表示不同的拓扑缺陷，包裹次数分类了铁磁相的刺猬形拓扑缺陷。包裹次数只能是整数，称作刺猬形拓扑缺陷的**拓扑量子数**。所有不同次数的包裹集合在一起就构成整数集  $\mathbf{Z}$ ，数学家通常称之为序参量空间  $O(3)/O(2)$  的第二同伦群，记作

$$\pi_2(O(3)/O(2)) = \pi_2(S^2) = \mathbf{Z}. \quad (2.29)$$

也即是说，铁磁相的刺猬形拓扑缺陷由  $\pi_2(O(3)/O(2))$  分类。

假设把两个拓扑量子数分别为  $n_1$  和  $n_2$  的刺猬形拓扑缺陷相互靠近，最终它们就可能结合成一个“更大”的刺猬形拓扑缺陷，这个“更大”的刺猬形拓扑缺陷的拓扑量子数必定是  $n_1 + n_2$ 。也即是说，在拓扑缺陷的结合过程前后，其拓扑量子数是守恒的。同样，一个“大”的刺猬形拓扑缺陷也可能分裂成两个“小”的拓扑缺陷，分裂前后拓扑量子数也是守恒的。特别的，假设取一个拓扑量子数为  $n$  的拓扑缺陷，和一个拓扑量子数为  $-n$  的拓扑缺陷，让它们结合，结果就会得到拓扑量子数 0，也就是不存在拓扑缺陷了，也即是说，这样的两个拓扑缺陷相互湮灭了！这就和正反粒子对会相互湮灭一样。

值得注意的是，拓扑缺陷的自由能一定是有限的，不能任意小。因此，你必须耗费一定的有限的能量才能激发出拓扑缺陷。

推广到 (2.4.1) 小节讨论的一般情况，可知，如果输入的能量足够，那么就有可能在低温有序相中激发起拓扑缺陷，比方说类似于铁磁体情形的那种刺猬状拓扑缺陷，类比于铁磁体情形的相关讨论可以知道，这种拓扑缺陷由  $G/H$  的二阶同伦群分类，也即由下式分类

$$\pi_2(G/H). \quad (2.30)$$

不过，在有些例子中，由于对称性  $G$  和  $H$  不够凑巧，导致  $\pi_2(G/H) = 0$ ，那就说明在这些例子中不存在这种刺猬状拓扑缺陷。

#### 2.4.4 物质相的朗道范式

下面我们可以概括一下主要物质相的朗道范式：

对于相互作用保持某种对称性（构成一个群  $G$ ）的系统，相由对称性如何自发破缺来区分，不同的相由未被破缺的对称性子群  $H$  标记，由不同  $H$  标记的相总是被相变分隔开，具有兼容对称性  $H_1 \subset H_2$  的相之间可以存在连续相变。这种连续相变导致系统序的扩张，它由序参量的涨落驱动，该序参量在  $H_1$  下变换平凡，但在  $H_2$  下变换非平凡。对于连续对称性情形，序参量的每一种涨落对应一个  $\mathbf{R}^3 \rightarrow H_2/H_1$  的连续映射。

而对于相互作用不保持任何对称性的系统，相图中不同部分之间没有尖锐的区分，只存在一个相。

## 2.5 液晶的物理学

不少棒状或扁盘状分子在一定温度区间会处于某种非液非固的中间相——液晶态。它们不是固体，其分子质心位置完全无序，表现为液体；但它们也不是通常的液体。其分子取向有明显的各向异性，能产生光的双折射等晶体中特有的现象。虽然早在一百多年前就发现了液晶，但系统的物理研究还是近几十年的事。

液晶物理学是一个跨越基础科学与应用技术的典范领域：在理论层面，它提供了研究对称性破缺、拓扑缺陷和非平衡统计物理的理想平台；在应用层面，它支撑了价值千亿美元的显示屏产业，并持续向软体机器人、生物医学等新兴领域拓展。液晶屏的普及更是一个关乎中国制造业崛起的传奇故事，一代又一代中国工程师通过工艺简化、材料创新和规模扩张的三重奏，将曾经昂贵得堪比“液体黄金”的液晶面板，一步步送进了寻常百姓家。

液晶有很多相，比如向列相、胆甾相、蓝相、蓝相和蓝雾相、近晶相 A、B、C、C\*、D、I 等。在向列相的液晶中，棒状分子有一个平均取向，用指向矢  $\mathbf{n}$  代表（见图 (2.2)），而质心位置完全无序，表现为液体。向列相液晶具有单轴各向异性，没有外场时， $\mathbf{n}$  轴的取向是任意的。本节我们就以向列相液晶为例，讨论一下液晶的物理学。

开始理论讨论之前，不妨提一个技术应用上的大难题：液晶是液体，里面的分子像一群懒散的学生，东倒西歪。要让它们整齐排列才能显示图像，传统方法是在玻璃上刻很多



Figure 2.2: 向列相液晶

“小沟槽”，逼分子排队。但刻沟槽很费事，而且一旦刻不好，屏幕就会出现“水波纹”，颜色不均匀。

华星的工程师想了个聪明办法：在液晶里掺入一种特殊“胶水”（聚合物），然后用紫外线一照，这胶水就在分子两头形成了看不见的“小钉子”，把液晶分子固定成预想的倾斜角度。这就好比不用在地上画白线，而是给每个学生脚上绑个隐形支架，让他们自然站成整齐的军姿。好处立竿见影：1. 不用刻复杂的沟槽，工艺简单了。2. 液晶排列更整齐，透光率提高了 15%，意味着同样亮度需要的灯泡更少，省电又省成本。3. 屏幕按压时不会出现难看的水波纹。

京东方则走了另一条路，开发了一种叫“ADS”的硬屏技术。原理类似，也是通过优化电极设计，让液晶听话排列。结果是，屏幕更耐用，颜色更鲜艳，成本更低。

### 2.5.1 朗道-de Gennes 理论

本节我们讨论液晶材料如何由高温时的各向同性液体转变为低温时的向列相 (Nematic Phase) 的相变，也就是所谓的  $IN$  转变 (Isotropic-Nematic 转变)。假设液晶材料由棒状分子组成，每个分子的长轴 (沿着棒) 沿着单位矢量  $\vec{v}$ ，由于在所考虑的情况下棒状分子没有极性，其头尾是等价的，因此  $\vec{v}$  和  $-\vec{v}$  等价。高温时，每个分子的方向  $\vec{v}$  是随机的，整个材料在宏观上表现出各向同性，也就是具有对称性  $G = O(3)$ 。低温时，棒状分子有一个平均取向  $\mathbf{n}$  (当然它也要满足  $\mathbf{n}$  和  $-\mathbf{n}$  等价)，原来的对称性自发破缺了，整个材料相变到向列相。

为了描述这种  $IN$  相变，我们需要定义一个合适的序参量，它在高温时取值为零，低温时非零。一个自然的想法是取  $\vec{v}$  在一个宏观看很小微观看很大的区域内的平均值  $\langle \vec{v} \rangle$  (即将这个区域内每个分子的  $\vec{v}$  相加再除以这个区域的总分子数) 为序参量，但这样无法处理  $\vec{v}$  和  $-\vec{v}$  的等价。一个解决办法是引入如下二阶对称无迹张量  $Q_{ij}$

$$Q_{ij} = \langle v_i v_j \rangle - \frac{1}{3} \delta_{ij}, \quad (2.31)$$

式中的平均依然是在一个宏观看很小微观看很大的区域内进行， $v_i$ ,  $i = 1, 2, 3$  是单位矢量  $\vec{v}$  的三个空间分量。由于  $\langle 1 \rangle = 1$ ，很显然  $Q_{ij}$  是对称无迹的。很显然， $Q_{ij}$  在等价关系  $v_i \rightarrow -v_i$  的作用下是不变的。高温时，由于每个分子的方向  $\vec{v}$  随机，所以很显然

$\langle v_i v_j \rangle = \langle \vec{v}^2 \rangle \frac{1}{3} \delta_{ij} = \frac{1}{3} \delta_{ij}$ , 所以  $Q_{ij} = 0$ 。而且  $Q_{ij}$  构成了旋转群  $O(3)$  的一个不可约张量表示。所以  $Q_{ij}$  作为序参量是合理的。

作为对称张量,  $Q_{ij}$  (相应的  $3 \times 3$  矩阵记作  $\mathbf{Q}$ ) 可以在一个合适的三维正交坐标轴中对角化, 对角元是它的三个本征值, 不过, 无迹条件告诉我们, 这三个本征值中只有两个是独立的, 通常可以作如下选取

$$\mathbf{Q} = \begin{pmatrix} -\frac{1}{3}S_1 + S_2 & 0 & 0 \\ 0 & -\frac{1}{3}S_1 - S_2 & 0 \\ 0 & 0 & \frac{2}{3}S_1 \end{pmatrix}. \quad (2.32)$$

把三个正交轴的方向一般化, 则有

$$\mathbf{Q} = S_1(\mathbf{nn} - \frac{1}{3}I) + S_2(\mathbf{mm} - \mathbf{ll}) \quad (2.33)$$

式中  $I$  为  $3 \times 3$  单位矩阵,  $\mathbf{n}, \mathbf{m}, \mathbf{l}$  分别为三个正交轴的单位矢量  $\mathbf{l} = \mathbf{n} \times \mathbf{m}$ , 不难看出其中  $S_1 = \frac{3}{2} \langle (\vec{v} \cdot \mathbf{n})^2 \rangle - \frac{1}{2}$ 。如果  $S_2 = 0$ , 就称系统为单轴的, 因为这时候只需要主排列轴  $\mathbf{n}$  就能完全确定  $\mathbf{Q}$ 。

根据朗道的理论, 为了构造自由能展开, 我们需要利用  $\mathbf{Q}$  来构造  $G = O(3)$  的不变量, 很显然这种不变量必定具有如下形式

$$\text{Tr}(\mathbf{Q}^n). \quad (2.34)$$

但是,  $3 \times 3$  对称无迹矩阵有一个数学性质, 即它的任何一个  $\text{Tr}(\mathbf{Q}^n)$  都可以表示成  $\text{Tr}(\mathbf{Q}^2)$  以及  $\text{Tr}(\mathbf{Q}^3)$  的多项式。所以, 真正独立的  $O(3)$  不变量只有下面两个

$$\text{Tr}(\mathbf{Q}^2), \quad \text{Tr}(\mathbf{Q}^3). \quad (2.35)$$

注意, 这个例子存在三阶不变量。因此, 系统的自由能可以展开成

$$\mathcal{G} = \mathcal{G}_G + \frac{1}{2}A\text{Tr}(\mathbf{Q}^2) + \frac{1}{3}B\text{Tr}(\mathbf{Q}^3) + \frac{1}{4}C[\text{Tr}(\mathbf{Q}^2)]^2 + \dots \quad (2.36)$$

式中系数  $A, B, C$  均是  $T, p$  的函数。

注意, (2.36) 式允许存在三阶项, 正如我们将要看到的, 这导致向列相-各向同性相转变成为一级相变。各向同性相-向列相转变发生在  $A = 0$  附近。  $A = 0$  确定了  $T - p$  相图上的一条曲线  $T^*(p)$ , 给定压强  $p$ , 我们可以把  $A$  在  $T^*$  附近展开为  $A = a(T - T^*)$ , 其中  $a$  是一个正常数。由于  $IN$  转变发生在  $T^*$  附近, 因此为了考察这种转变, 我们可以合理的假设  $B, C$  两个系数均为常数 (给定压强)。

利用  $\mathbf{Q}$  的对角形式, 不难算出

$$\text{Tr}(\mathbf{Q}^2) = \frac{2}{3}S_1^2 + 2S_2^2, \quad \text{Tr}(\mathbf{Q}^3) = \frac{2}{9}S_1^3 - 2S_1S_2^2. \quad (2.37)$$

代入自由能的展开式 (2.36), 即得

$$\mathcal{G} = \mathcal{G}_G + \frac{1}{3}AS_1^2 + \frac{2}{27}BS_1^3 + \frac{1}{9}CS_1^4 + S_2^2[A - \frac{2}{3}BS_1 + \frac{2}{3}CS_1^2] + CS_2^4 + \dots$$

为了找到稳定的相，原则上我们必须求出  $\mathcal{G}$  同时关于  $S_1, S_2$  的极小值。不过，这里我们只关心  $S_2 = 0$  的**单轴相**，记这时候的  $\mathcal{G} = \mathcal{G}_G + \mathcal{G}_1$

$$\mathcal{G}_1 = \frac{1}{3}AS_1^2 + \frac{2}{27}BS_1^3 + \frac{1}{9}CS_1^4 + \dots \quad (2.38)$$

要使最小值在有限的  $S_1 = S^*$  处， $C$  必须为正。根据后面给出的计算可知， $B$  的符号与  $S^*$  的符号相反。因此，对于传统的棒状分子向列相液晶 ( $S^* > 0$ )， $B < 0$ ；而对于盘状向列相液晶， $B > 0$ 。如果  $T > T^*$  (注意， $T^*$  并不是相变温度)，从而  $A > 0$ ，则  $S_1 = S^*$  的这个值将确保原来的自由能展开中  $S_2^2$  项的系数为正，从而保证了这时候  $S_2 = 0$  的确是  $\mathcal{G}$  关于  $S_2$  的极小值位置。这也正是我们关心的单轴情形。

$S^*$  可以由  $\frac{\partial \mathcal{G}_1}{\partial S_1} = 0$  给出，

$$S^* = \begin{cases} 0 \\ \frac{1}{4C} [-B \pm \sqrt{B^2 - 24AC}] \end{cases} \quad (2.39)$$

真正的  $S^*$  当然是这三个值中使得  $\mathcal{G}_1$  的值最低的那个。为了找到相变的临界温度，我们改变温度 (从而改变  $A$ )，进而改变  $S^*$ ，使得当  $S^* = S_{IN}$  时，自由能  $\mathcal{G}_1 = 0$ ，此时系统发生一阶相变进入向列相，相应的相变温度记作  $T_{IN}$ 。通过繁琐的代数计算可以得到

$$S_{IN} = -\frac{B}{6C}, \quad T_{IN} - T^* = \frac{B^2}{27aC} > 0. \quad (2.40)$$

注意  $S_{IN} \neq 0$ ，通过公式  $S_1 = \frac{3}{2} \langle (\vec{v} \cdot \mathbf{n})^2 \rangle - \frac{1}{2}$ ，这就在三维空间给出了一个取向轴  $\mathbf{n}$ ，显然它也满足  $\mathbf{n}$  与  $-\mathbf{n}$  等价。

类似的考虑可以确定另一个温度  $T^I > T_{IN}$ 。(1)  $T > T^I$  时，自由能最小值对应各向同性相， $S_1 = S^* = 0$ ；(2)  $T_{IN} < T < T^I$  时，最小值对应各向同性相，此外存在一个局部最小值，对应可能存在的过热向列态；(3)  $T^* < T < T_{IN}$ ，最小值对应向列相，同时存在一个局部最小值，对应可能存在的过冷各向同性态；(4) 当  $T < T^*$  时，最小值对应向列相。

在单轴向列相中，系统有绕取向轴  $\mathbf{n}$  旋转任意角度的对称性，有空间反演对称性，有关于垂直于取向轴  $\mathbf{n}$  的水平镜面的镜像对称性  $\sigma_h$ ，还有关于过取向轴的任意垂直镜面的镜像对称性  $\sigma_v$ ，这些剩余对称性构成群  $D_{\infty h}$ 。当然，它是高温  $O(3)$  对称性的子群， $D_{\infty h} \subset O(3)$ 。

后来人们还发现，单轴向列相可以进一步发生对称性自发破缺，破缺掉镜像对称  $\sigma_h$  和空间反演，进而得到一个极性向列相，这时候取向轴是有方向的，不能头尾等价，因为系统有一个非零的沿着取向轴的极化矢量。极性向列相的对称性是  $C_{\infty v}$ ，当然也有  $C_{\infty v} \subset D_{\infty h}$ 。

极性向列相的对称性还可以进一步破缺，进入层状相，比如近晶 C 相，它的对称性是  $C_{2v}$ 。所以，从高温时的各向同性液体开始，可以发生一系列的对称性破缺，它们由子群链  $O(3) \supset D_{\infty h} \supset C_{\infty v} \supset C_{2v}$  刻画。更多的讨论留给读者自行查阅相关文献。

### 2.5.2 取向轴的形变

单轴向列相完全由取向矢  $\mathbf{n}$  刻画，并且它满足等价关系  $\mathbf{n} \equiv -\mathbf{n}$ 。但是取向矢本身也可以涨落，或者取向轴可以在外力作用下发生形变，从而在液晶中引入了一个取向场  $\mathbf{n}(\mathbf{x})$ 。

取向轴的形变当然对自由能有额外贡献，记这个贡献为  $\mathcal{G}_e$ 。很明显， $\mathcal{G}_e$  的展开式只可能含有矢量  $\mathbf{n}$  的分量及其导数的标量组合。总共只存在两种关于一阶导数的线性标量组合：真标量  $\nabla \cdot \mathbf{n}$  和赝标量  $\mathbf{n} \cdot (\nabla \times \mathbf{n})$ 。其中前者在对体积积分时变为对物体表面的积分，因此，在考虑材料的体性质时并不重要。另外，空间反演对称或者镜像对称又否定了赝标量项的贡献。

呈一阶导数二次的真标量可以通过写下四阶张量

$$\frac{\partial n_k}{\partial x_i} \frac{\partial n_l}{\partial x_m} \quad (2.41)$$

并由它缩并一对指标或者与矢量  $n_i$  进行合适缩并的方法来得到。这时还应该考虑到  $\mathbf{n}$  为单位矢量，从而

$$0 = \frac{\partial \mathbf{n}^2}{\partial x_i} = 2n_k \frac{\partial n_k}{\partial x_i}, \quad (2.42)$$

式中我们已经默认了求和约定，即对表达式中的重复指标默认求和（下同）。

用这种方法可以求得如下标量

$$[(\mathbf{n} \cdot \nabla) \mathbf{n}]^2, \quad \frac{\partial n_k}{\partial x_i} \frac{\partial n_k}{\partial x_i}, \quad (\nabla \cdot \mathbf{n})^2, \quad \frac{\partial n_k}{\partial x_i} \frac{\partial n_i}{\partial x_k}. \quad (2.43)$$

但是，最后两式彼此只差一个散度

$$\frac{\partial n_i}{\partial x_i} \frac{\partial n_k}{\partial x_k} - \frac{\partial n_k}{\partial x_i} \frac{\partial n_i}{\partial x_k} = \frac{\partial}{\partial x_i} (n_i \frac{\partial n_k}{\partial x_k} - n_k \frac{\partial n_i}{\partial x_k}), \quad (2.44)$$

所以它们对  $\mathcal{G}_e$  的贡献只相差意义不大的表面积分项。又由于

$$\frac{\partial n_k}{\partial x_i} \frac{\partial n_k}{\partial x_i} = [\mathbf{n} \cdot (\nabla \times \mathbf{n})]^2 + (\nabla \cdot \mathbf{n})^2. \quad (2.45)$$

所以这样构造出来的独立项只有下面三个

$$[(\mathbf{n} \cdot \nabla) \mathbf{n}]^2, \quad [\mathbf{n} \cdot (\nabla \times \mathbf{n})]^2, \quad (\nabla \cdot \mathbf{n})^2. \quad (2.46)$$

而且这三项均在  $\mathbf{n} \rightarrow -\mathbf{n}$  的等价变换下保持不变。所以，单轴向列液晶取向轴的形变（或者涨落）贡献的自由能可以写成

$$\mathcal{G}_e = \int d^3 \mathbf{x} \left[ \frac{1}{2} K_1 (\nabla \cdot \mathbf{n})^2 + \frac{1}{2} K_2 [\mathbf{n} \cdot (\nabla \times \mathbf{n})]^2 + \frac{1}{2} K_3 [(\mathbf{n} \cdot \nabla) \mathbf{n}]^2 \right] \quad (2.47)$$

这称作 Frank-Oseen 自由能，式中  $K_1$  称作展曲 (Splay) 弹性常数， $K_2$  称作扭曲 (Twist) 弹性常数， $K_3$  称作弯曲 (Bend) 弹性常数。

## 2.6 结语：范式的辉煌与边界

朗道范式以对称性破缺为魂，以序参量为骨，为纷繁复杂的物质相世界绘制了一幅统一而优美的图景。从磁性到超导，从液晶到超流，这一范式不仅成功分类了数以千计的物相，更深刻揭示了相变本质——序的扩张与对称性的隐匿。朗道理论的深邃洞见在于，它

将相变这一宏观现象与微观对称性结构精妙联结，为凝聚态物理学奠定了坚实的概念基石。

具体到朗道范式在液晶物理学上的应用。从莱尼茨尔试管中的浑浊液体，到万亿级市场的显示帝国，液晶的百年历程见证了科学发现如何经由技术创新，最终惠泽人类生活。朗道-de Gennes 理论在其中扮演了不可替代的角色：它不仅是理解液晶物理的概念框架，更是指导材料设计与工艺优化的实用工具。今天，当人们透过高清屏幕观看世界时，或许未曾想到，那变幻的光影背后，凝聚着一代代物理学家、工程师和产业工人对自然秩序的深刻洞察与不懈求索。

然而，任何伟大的理论都有其边界。朗道范式在临界点附近遭遇了深刻的挑战。作为一种平均场理论，它忽略了临界涨落的长程关联，导致对临界指数的预言与实验及严格解存在系统偏离。这一失效的根源在于，当趋近临界点时，关联长度发散，涨落的幅度与尺度使得平均场近似崩溃。正是这一困境，催生了重整化群理论的诞生——威尔逊的工作不仅拯救了临界现象理论，更深刻重塑了我们对标度律与普适类的理解。

更根本的局限来自拓扑序的发现。分数量子霍尔效应、自旋液体等新奇物态揭示了一个朗道范式无法触及的隐秘世界：这些体系没有局域序参量，没有对称性自发破缺，其宏观简并与非局域激发由拓扑量子场论刻画，基态的爱因斯坦-玻色统计被任意子统计取代。朗道的对称性破缺范式在此失效，取而代之的是更抽象的拓扑量子场论与张量范畴理论。

正是这些边界与失效，非但没有减损朗道范式的光辉，反而彰显了科学范式的演进之道：伟大的理论不仅因其解释世界的广度而被铭记，更因其留下的问题为后人开辟了新的疆域。朗道的遗产，既是一座辉煌的丰碑，也是一扇通向未知的门——在那里，对称性的语言依然回响，但更深刻的秩序正等待被诉说。

### 3. 重整化群理论

回溯重整化群理论的源头，必须首先致敬两项更为早期却影响深远的奠基性成就。其中首要的便是列夫·D·朗道的开创性工作——他实为系统有效场论的先驱，尽管这一称谓在当时或许尚未定型。公允而言，朗道“发明”了序参量这一概念，其重要性无论怎样强调都不为过。

所谓序参量，其核心洞见在于：即便对微观机制一无所知（正如早期对于超流氦的理解），只要承认微观世界的存在，并坚信其遵循某些普遍的整体性法则——特别是关于局域性与对称性的规律——这些性质便足以决定原子尺度之上最具标志性的宏观行为。诚然，朗道与金兹堡（这一理念的主要合作者与发展者）后来在涨落效应和临界奇异性等关键问题上出现了偏差，但这丝毫未减损这种审视复杂凝聚态系统方式的深刻价值。一旦把握了序参量的本质——譬如，当其恰如波函数般呈现为复数形式——整个物理系统的宏观奥秘便豁然开朗。

朗道引入序参量的划时代意义，在于揭示了一个介于微观与宏观之间、令人意想不到的全新认知层面。传统观念将统计力学视作架设在两个极端世界之间的桥梁：一端是原子核与原子构成的微观领域（尺度在  $10^{-13}$  至  $10^{-8}$  厘米之间），另一端则是毫米乃至米量级的宏观世界。然而，序参量作为动态涨落的实体，在诸多情形下恰恰嵌入了一个居间的介观层次，其特征尺度横跨几十至数百埃，直至微米范围（约  $10^{-6.5}$  至  $10^{-3.5}$  厘米）。随着威尔逊重整化群思想的兴起，源自朗道-金兹堡传统的有效（“粗粒化”）哈密顿量——亦即自由能按序参量展开的数学形式——终于获得了更为精确的理论诠释。如今，朗道-金兹堡-威尔逊（LGW）哈密顿量被理解为经过深刻重整化的真实哈密顿量，那些更为精细的

微观自由度已然被系统地“积掉”。（具体显式表达详见后文。）在现代凝聚态理论的实践中，研究者往往直接从这个介观层次出发，以物理上合理的 LGW 哈密顿量替代那些过于复杂或真实的微观哈密顿量，进而运用统计力学的方法解读宏观现象。至于如何从原子层次严格推导出各类初始 LGW 哈密顿量，并论证其适用边界，则已发展成为一门独立而重要的研究课题。

朗道的序参量框架为相变理论注入了前所未有的明晰性与形式化力量，最终实现了对多临界点的精确刻画，并深化了对有序态诸多特性的理解。然而，1944 年的一声惊雷震撼了学界：拉斯·昂萨格以出神入化的数学技艺（连朗道本人都为之叹服）严格求解了最为简单的铁磁体模型——伊辛模型——的配分函数与全部热力学性质。该模型确实展现出尖锐的相变临界点，但其临界行为的细节，特别是那些奇异性特征，竟与朗道理论的几乎每一项精细预言都大相径庭——其中的缘由将在后文详述。面对这一理论危机，以及日益明确的实验证据，物理学界相继孕育出普适而深刻的临界指数概念、指数间隐藏的特殊关系式，以及针对临界区域的标度理论描述。这些突破性洞见极大地激发了肯尼斯·威尔逊对量子场论的探索。事实上，一旦洞见到以有效哈密顿量处理统计力学问题与在费曼路径积分框架下开展量子场论计算之间深刻的数学同构性，两个领域之间的桥梁便赫然显现。不过，这种类比的全面把握绝非朝夕之功：威尔逊率先在最本质的层面上透彻地理解了这一联系，而这一思想遗产至今仍在持续滋养着两个学科的发展。1971 年，历经四五年艰苦卓绝的探索，肯·威尔逊终于将重整化群思想锻造成一个既概念自治又具备实际计算能力的理论框架。

经由这一系列发展，学界对所谓“反常”（实则标准）临界行为的理解获得了质的飞跃。需要特别指出的是，理论物理学的终极追求始终在于获得深刻的洞察力与真正的理解。至于“理解”一词的确切内涵——“什么构成了理解？”——这本身就是一个智能时代的根本性难题。但，让我们立即指出，完整的重整化群理论不应被视为（尽管这种说法很常见）基于量子场论微扰展开，正如宏伟的吉布斯统计力学结构不能被视为建立于理想经典气体、玻尔兹曼动理论和稀薄流体的集团展开之上一样，因为这将严重歪曲统计力学的力量和范围。

### 3.1 临界指数与普适性

#### 3.1.1 反常量纲

如果要用一个特征来体现重整化群理论的非凡力量与辉煌成就，那么我们不得不说，最值得强调的，或许是重整化群揭示了非经典临界指数的可能性——比如那个不为零的“反常”量纲  $\eta$ 。这很可能是该理论最重大的贡献。

让我们先聚焦于一个“局域”的微观变量，记作  $\psi(\mathbf{r})$ 。在铁磁体中，它或许是某一点  $\mathbf{r}$  处的局域磁化强度  $\mathbf{m}(\mathbf{r})$ ；在流体系统中，它可能是该点密度涨落相对于平均值的偏离  $\delta\rho(\mathbf{r})$ 。而在量子场论中， $\psi(\mathbf{r})$  则化身为基本的量子场，成为“算符值”的实体。对于那些涨落效应不可忽视的磁性系统， $\mathbf{m}(\mathbf{r})$  同样具有算符的本质。不过，这种量子与经典的区分在此相对次要——为简便起见，我们不妨将  $\psi(\mathbf{r})$  视作一个普通的经典变量。当  $\psi$  与我们关注的相变及临界现象中的“序参量”紧密相连时，它便展现出最迷人的一面。

借助散射实验（无论是光、X 射线、中子还是电子），科学家们能够观测到一个关键物理量——对关联函数（或称“两点函数”）：

$$G(\mathbf{r}) = \langle \psi(\mathbf{0})\psi(\mathbf{r}) \rangle \quad (3.1)$$

这里的尖括号  $\langle \cdot \rangle$  代表着统计物理学平均值，也是对有限温度下平衡态系统中所有热涨落的统计平均。

$G(\mathbf{r})$  的物理意义至关重要：它直接度量了原点处的微观涨落如何影响距离  $r = |\mathbf{r}|$  外的远端行为。然而，正是在临界点附近——比如铁磁体的居里点（此时  $\psi \equiv \mathbf{m}$ ）或气-液临界点（此时  $\psi = \delta\rho$ ）——一种强烈的“有序”效应或关联会延伸至宏观尺度。结果，在精确的临界点，我们通常会惊异地发现：

$$G_c(\mathbf{r}) \approx D/r^{d-2+\eta} \quad \text{当 } r \rightarrow \infty \quad (3.2)$$

这是一个幂律衰减，其特征由临界指数  $d-2+\eta$  所主宰。

然而，历史上所有“经典”理论——无论是朗道-金兹堡理论还是范德瓦尔斯理论——都斩钉截铁地预言  $\eta$  必须为零。在量子场论语境下，这对应于自由无质量粒子的行为。其数学根源在于，这些理论假设基本函数具有光滑、解析、无奇点的性质，因此可以像在经典力学中那样，在临界点处自由地做泰勒展开。在场论中，经典指数值  $d-2$ （即  $\eta=0$ ）可通过简单的量纲分析得到，这被称为“正则量纲”；而如果  $\eta$  不为零，则代表着一种“反常量纲”。

物理上， $\eta=0$  的预言往往源于对涨落的忽视，或者更准确地说，源于威尔逊所揭示的那个关键假设：只有更小尺度上的涨落才重要。在这种图景下，涨落可以被安全地“吸收”进有效参数（如质量、耦合常数等），而不会改变理论的根基。

但幂律衰减意味着系统中不存在确定的长度尺度——这正是“标度不变性”的体现。为了看清这一点，让我们将距离按因子  $b$  重新标度：

$$\mathbf{r} \Rightarrow \mathbf{r}' = b\mathbf{r}. \quad (3.3)$$

同时，用某个“协变”因子  $b^\Delta$  重新标度序参量  $\psi$ ，其中  $\Delta$  是表征  $\psi$  的临界指数，称作  $\psi$  的标度量纲。于是：

$$G_c(\mathbf{r}) = \langle \psi(\mathbf{0})\psi(\mathbf{r}) \rangle_c \Rightarrow G'_c(b\mathbf{r}) = b^{2\Delta} \langle \psi(\mathbf{0})\psi(b\mathbf{r}) \rangle_c \quad (3.4)$$

$$\approx b^{2\Delta} D / b^{d-2+\eta} r^{d-2+\eta} \quad (3.5)$$

注意，若取  $\Delta = \frac{1}{2}(d-2+\eta)$ ，则  $b$  的因子恰好抵消，(3.2) 式的形式得以保持。换句话说， $G_c(\mathbf{r})$  是标度不变的：无论放大或缩小多少倍，其数学形式始终如一，不泄露任何特征长度的信息！

这恰恰暴露了经典理论在临界点处的致命弱点——既然幂律意味着所有尺度同等重要，那种假设“只有小尺度重要”的经典图景自然站不住脚。事实上，实验与理论都表明，“反常量纲”  $\eta$  通常确实不为零（至少在凝聚态物理关心的  $d < 4$  维度如此）。早在 1949 年，考夫曼和昂萨格就严格证明：对于二维伊辛模型， $\eta = \frac{1}{4}$ 。这一发现宣判了经典

理论中解析性和泰勒展开假设的“死刑”，也向理论物理学家提出了严峻挑战：如何构造一个能自然导出  $\eta \neq 0$  的理论？

重整化群的力量正在于，它提供了一个概念清晰且计算可行的框架，让  $\eta$ （以及其他指数，如  $\Delta$  及其对能量密度等局域量的类似物）的反常值得以自然涌现，而非强行植入。

在凝聚态物理中，式 (3.2) 的幂律显然只在远大于原子间距  $a$  的距离上成立。因此，关联函数的标度不变性仅是“渐近”的——这正是式中使用  $\approx$  和条件  $r \rightarrow \infty$  的原因。更精细的理论还需考虑非零  $a$  的修正。相比之下，在量子场论中，微观距离  $a$  扮演着“紫外截断”的角色。由于这一截断通常是未知的，物理学家希望将其从理论中移除。若移除不当——这正是场论中“重整化”程序所要规避的——当  $a \rightarrow 0$  时理论将遭受灾难性的无穷大发散。而在统计物理中，短距离截断总是物理存在的，它甚至决定了临界温度  $T_c$  等参数；无穷大项既不会出现，更不会像在场论中那样“驱动”理论。

在场论的现行表述中，“参数的标度依赖性”是常用概念，其物理图景是：在特定长度（或时间）尺度上测量的系统性质，会随观测尺度的变化而缓慢演化。其实，这种表述往往只是重整化群流的一种简化说法——即只追踪单一变量或轨迹，因为粒子物理学家通常只关心描写“现实世界”的唯一理论。对某些凝聚态问题，这种简化或许够用；但更多时候，我们需要更复杂的全景视角。

不过，我们可以通过式 (3.2) 的幂律给“标度依赖性”一个更生动的诠释。若  $\eta = 0$ ，即  $\psi$  具有正则量纲  $\Delta = \Delta_{\text{can}} = \frac{1}{2}(d-2)$ ，则振幅  $D$  是一个固定可测的物理常数，具有明确的物理意义。但当  $\eta$  虽小却不为零时（对许多三维系统， $\eta \approx 0.035$ ），我们可以引入一个“跑动”的或“尺度依赖”的参数：

$$\tilde{D}(R) \approx D/R^\eta \quad \text{当 } R \rightarrow \infty. \quad (3.6)$$

于是原关联函数可改写为：

$$G_c(\mathbf{r}) = \tilde{D}(r)/r^{d-2}. \quad (3.7)$$

由于  $\eta$  很小， $\tilde{D}(R)$  随测量尺度  $R$  的变化相当缓慢。在场论的许多情形下，场  $\psi$  的量纲仅受所谓的边际微扰影响，这转化为  $\tilde{D}(R)$  对  $\log R$  的依赖，其随尺度的变化比  $\eta \neq 0$  时更为微弱。

### 3.1.2 临界指数与普适性

当理论与哲学交汇时，最好的锚点永远是实验事实。因此，让我们先回到实验室，看看那些重塑了我们对物质世界认知的经典观测——正是这些现象，让重整化群理论在应对挑战时展现出了惊人的解释力。

#### 临界点的曙光

1869 年，托马斯·安德鲁斯向皇家学会展示了一个令人着迷的实验：在坚固的玻璃管中密封着二氧化碳。在常温下，容器内的流体清晰地分成两层——底部是密度较高的液态  $\rho_{\text{liq}}(T)$ ，顶部是轻得多的气态  $\rho_{\text{gas}}(T)$ ，两者之间界面分明。然而，当温度缓缓攀升至

$T_c \simeq 31.04^\circ\text{C}$  这个神奇的数值时，液相与气相的界限开始模糊。两者密度逐渐趋近，最终在  $\rho_{liq} = \rho_{gas} = \rho_c$  处合而为一，界面在“临界乳光”的朦胧雾霭中悄然消失。超过这个临界温度后，液态与气态的界限彻底消融，呈现出完美的状态连续性。

这种临界现象如同一首重复的乐章，在所有单质和简单分子流体中不断上演——只是每个“演奏者”的调性不同：氦-4 的临界温度低至 5.20K，而汞则高达约 1764K。临界密度同样千差万别。这些参数像是物质的“指纹”，深深烙印着原子与分子的个性，反映着微观截断尺度  $a$  上的物理细节。在这里，最大密度  $\rho_{max}$ （大致等于低温下晶体的密度）的量级为  $1/a^3$ ，而  $k_B T_c$  则由微观吸引势能设定。虽然这些数值对化学家和工程师至关重要，但对我们理解普遍规律而言，真正引人入胜的不是这些具体数字，而是当系统逼近临界点时，共存曲线  $\rho_{liq}(T)$  与  $\rho_{gas}(T)$  所展现出的那种普适的形态。

### 对称性与临界指数

乍一看，共存的液体与气体毫无对称性可言——一个是拥挤的原子集合，一个是稀疏的分子云团。然而，若定义一个比值

$$R(T) = \frac{\rho_c - \rho_{gas}(T)}{\rho_{liq}(T) - \rho_c} \quad (3.8)$$

来比较共存曲线的两翼，一个精妙的对称性便浮出水面。当温度从下方逼近临界点，即  $t \equiv (T - T_c)/T_c \rightarrow 0^-$  时， $R(T)$  精确地趋向于 1。这意味着，物理流体在临界点附近自发地“学会”了完美的镜像对称——无论何种物质，这一规律普适地成立。

比对称性更令人惊叹的是临界点附近曲线的平坦程度。我们通常用幂律来刻画这一临界区域的形貌：

$$\Delta\rho \equiv \frac{1}{2}[\rho_{liq}(T) - \rho_{gas}(T)] \approx B|t|^\beta \quad \text{as } t \rightarrow 0^-. \quad (3.9)$$

其中  $B$  是依赖于具体物质的非普适振幅，而临界指数  $\beta$  则取一个对所有流体物质都一样的神奇普适值：

$$\beta = 0.325. \quad (3.10)$$

强调一下：这不是简单的分数，而是一个非平庸的数值，它对所有流体的临界点都一模一样！而且这与经典平均场理论预测的  $\beta = 1/2$  形成鲜明对照（经典平均场论的这个结果我们稍后会介绍）。对于二维系统（ $d = 2$ ），昂萨格 1949 年对伊辛模型的精确解给出了  $\beta = 1/8$ ，这一预言已经被实验证实——人们研究了吸附在石墨表面上的甲烷单分子层，这种“二维流体”完美印证了这一普适规律。

### 跨越领域的普适性

这个  $\beta = 0.325$  不仅适用于各类流体，还主宰着三维各向异性磁性材料的世界，特别是具有单一“易轴”的伊辛型磁体。铁磁体的相图如图 (3.1) 所示。这里为了简单起见，我们同样假设铁磁体只沿着竖直轴（所谓“易轴”）磁化，外磁场也只沿着竖直轴。

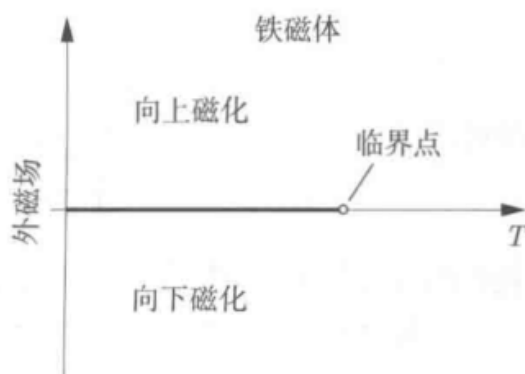


Figure 3.1: 铁磁体的相图

微观上看，铁磁体是由宏观数量相互作用的自旋规则地排布而成的，这些自旋排布成规则的晶格，每个格点上有一个自旋，相邻格点上的自旋会有相互作用，这种作用趋向于使得相邻自旋的取向相同，具体细节与我们关系不大。这些自旋只能取向上和向下两种朝向。如果没有外磁场，那么向上和向下对于系统来说是等价的，我们说系统的微观相互作用有  $\mathbf{Z}_2$  对称性， $\mathbf{Z}_2$  就是  $\{+1, -1\}$  两种可能性（按照通常的乘法进行运算），可以分别指代自旋向上和自旋向下。另外，这些自旋磁矩在宏观局部上的平均值，就是磁化强度，记作  $m(x)$ 。

假设外磁场为零，只调节系统的温度，也就是只关心相图 (3.1) 中水平轴上的情况。在高温时，热运动的能量远远超过相邻自旋之间相互作用的能量，因此，这些自旋各自为政，每个自旋的取向都在向上和向下之间疯狂而随机地改变，自旋向上和自旋向下完全对称，整个系统是向上和向下对称而且无序的。这时候平均来说，每个自旋的自旋磁矩都是零，因此宏观上的磁化强度  $m(x)$  也是零。我们说这时候系统处于具有  $\mathbf{Z}_2$  对称性的相，也就是系统的宏观表现有  $\mathbf{Z}_2$  对称性，它以之前微观相互作用的  $\mathbf{Z}_2$  对称性为基础，但不是一回事。高温相的对称性是微观相互作用所能允许在宏观上表现出来的最大对称性。

当温度降到临界点以下时，自旋之间的相互作用效应开始超过无规则热运动的效应，这时候会发生一件神奇的事情：这些自旋开始自发且协调一致地行动，形成一种要么都向上（总体上）、要么都向下（总体上）的宏观模式。换言之，系统从无序的热运动中自发产生了有序。这种“有序的程度”，正好可以用磁化强度  $m$  来刻画。由于各自旋协调一致地要么都向上、要么都向下，这时候磁化强度当然就不是零，而是一个非零常数，系统出现了自发磁化，这就是临界点处铁磁相变的物理本质。由于磁化强度为零对应无序的高温相，而磁化强度非零对应有序的铁磁相，所以朗道就把这里的磁化强度称之为**序参量**。

由于自发磁化选取了向上或者选取了向下，因此在铁磁相，原来高温相的  $\mathbf{Z}_2$  对称性就没有了，通常称这为  **$\mathbf{Z}_2$  对称性破缺**了。但是，由于向上磁化还是向下磁化原则上是机会均等的，系统只是随机地选取了其中一个方向，所以也称这种对称性破缺为**对称性自发破缺**。

在零外场下，当温度降至居里温度  $T_c$  以下，磁体展现出自发磁化  $m = \pm m_0(T)$ ，符号取

决于外场从正或负方向趋近于零。由于系统具有严格的  $H \rightarrow -H$  ( $H$  代表外磁场)、 $m \rightarrow -m$  的  $\mathbf{Z}_2$  对称性 (这与流体有所不同, 流体的对称性只在临界点附近出现),  $m_c$  严格为零, 对应于式 (3.8) 的对称性在所有温度下都精确成立。然而, 磁化曲线的整体轮廓与流体共存曲线看似迥异, 但在临界点附近, 它们却遵循相同的渐近规律:

$$m_0(T) \approx |t|^\beta \quad \text{as } t \rightarrow 0^- \quad (3.11)$$

其中  $\beta = 0.325$  再次适用于  $d = 3$ , 而经典理论仍固执地预测  $\beta = 1/2$ 。对于  $d = 2$ , 昂萨格的结果  $\beta = 1/8$  依然成立!

流体与单轴铁磁体只是“伊辛普适类”的冰山一角。反铁磁体、亚铁磁体、二元合金的有序-无序转变、某些铁电体……这些看似毫不相干的系统, 在临界点附近都跳动着相同的数学韵律。每个系统都有其特定的序参量, 通过关联函数  $G(R; T)$  定义的特征衰减指数  $\eta$  同样具有普适性。实际上, 任何可测量的物理性质都展现出普适的临界奇异性。

### 更多的临界指数

特别重要的是比热临界指数  $\alpha \approx 0.11$  (三维伊辛类), 它描述了比热在临界点的发散行为:

$$C(T) \approx A_\pm / |t|^\alpha \quad \text{as } t \rightarrow 0^\pm \quad (3.12)$$

(对流体是定容比热, 对磁体是零场比热)。振幅  $A_+$  与  $A_-$  虽非普适, 但其无量纲比值  $A_+/A_- \approx 0.52$  却是普适的。在二维世界中, 昂萨格 1944 年的精确解给出了  $A_+/A_- = 1$ , 且  $|t|^{-\alpha}$  被  $\log|t|$  所取代——而经典理论则错误地预言所有维度下都只有一个有限的比热跃变!

另外两个核心量是发散的等温压缩率 (流体) 或磁化率 (磁体)  $\chi(T) \propto (\partial m / \partial B)_T$ , 以及刻画“影响范围”的发散关联长度  $\xi(T)$ , 在临界点  $\xi \rightarrow \infty$ 。在临界点附近,  $\xi(T)$  的定义是 (注意, 现在是偏离临界点处的关联函数),

$$G(\mathbf{r}) \approx D / r^{d-2+\eta} e^{-r/\xi(T)} \quad \text{当 } r \rightarrow \infty. \quad (3.13)$$

值得一提的是, 在量子场论中, 关联长度的倒数  $\xi^{-1}$  基本上等价于场  $\psi$  的重整化质量: 临界性则等同于无质量性, 因为  $\xi^{-1} \rightarrow 0$ 。对于  $\chi(T)$ 、 $\xi(T)$  这两者的发散, 我们写作: 当  $t \rightarrow 0^\pm$  时,

$$\chi(T) \approx C_\pm / |t|^\gamma \quad \text{and} \quad \xi(T) \approx \xi_0^\pm / |t|^\nu. \quad (3.14)$$

对于三维伊辛类系统:

$$\gamma = 1.24 \quad \text{and} \quad \nu = 0.63. \quad (3.15)$$

而对于  $d = 2$ ,  $\gamma = 7/4$  且  $\nu = 1$ 。

### 普适类的版图

理论上还存在着其他的普适类，尽管实验上观察到的相对较少。重整化群理论的早期辉煌成就之一，便是描绘并深化了我们对这些普适类别的理解。很大程度上，系统的命运由其序参量的矢量或张量特性（标量、复数即二维矢量、三维矢量等）所决定。但究竟是何种深层机制导致了这种分类，以及普适性本身为何存在，这就是临界现象理论所要回答的核心问题。

到 1960-62 年间，关于**普适临界指数**的存在——那些与经典理论预测背道而驰的神秘数字——已在理论与实验的双重检验下尘埃落定，确凿无疑。理论探索的下一步随之浮出水面：科学家们发现这些指数之间并非孤立，而是遵循着简洁优雅的代数方程——即所谓的指数关系——且无论属于哪一类普适类，这些关系都普遍成立。其中最早被揭示的指数关系包括

$$\gamma = (2 - \eta)\nu \quad \text{以及} \quad \alpha + 2\beta + \gamma = 2. \quad (3.16)$$

正如读者可用前文引用的数值自行验证：这些关系在二维伊辛模型中精确吻合，在三维世界中则在实验精度与（理论估算的）数值精度内毫厘不爽；即便以经典平均场所给出的指数值代入，它们也被精确满足——如今我们深知，这些数值恰是  $d > 4$  时的有效结果。但是，为什么临界指数之间一定满足这些关系式？

### 3.2 统计力学回顾

在凝聚态物理的研究中，一个至关重要的深层要求是：重整化群理论应当与吉布斯所完善的统计力学体系建立起紧密而自然的联系。当然，我们既无必要也不应当试图用重整化群理论去取代经典统计力学——后者作为描述均匀系统平衡态现象的基石，其地位不可动摇。因此，我们需要以一种契合重整化群变换语言的方式，重新梳理统计力学的基本框架，这正是本节要做的事。

让我们从一组微观且持续涨落的力学变量出发：在量子场论中，这些变量是定义于欧几里得（或闵可夫斯基）空间各点的各类量子场  $\psi(\mathbf{r})$ ；而在统计物理中，我们设想一个体积为  $V$  的物理系统，其中包含  $N$  个离散的“自由度”。对于经典流体，人们通常采用组成粒子的坐标  $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N$  来描述；然而，从数学处理的简便性出发——同时也为了与量子场论建立更清晰的类比——我们在此采用与离散晶格点（位于等间距位置  $\mathbf{x}$  上）相关联的“自旋”变量  $s_{\mathbf{x}}$ （可以是矢量、张量或算符等）。若晶格常数为  $a$ ，则系统体积可表示为  $V = Na^d$ ，在  $d$  维空间中自由度的密度即为  $N/V = a^{-d}$ 。

基于这些基本变量  $s_{\mathbf{x}}$ ，我们可以构造出各种“局域算符”（或称“可观测变量”），例如局域磁化强度和能量密度：

$$m_{\mathbf{x}} = \mu_B s_{\mathbf{x}}, \quad E_{\mathbf{x}} = -\frac{1}{2} J \sum_{\delta} s_{\mathbf{x}} s_{\mathbf{x}+\delta}, \dots \quad (3.17)$$

其中  $\mu_B$  和  $J$  为固定参数，而  $\delta$  遍历所有最近邻晶格矢量。

一个物理系统由其哈密顿量  $\mathcal{H}[\{s_{\mathbf{x}}\}]$ （即力学中的能量函数）完全刻画，它通常是上述局域算符的空间均匀叠加。为了方便统计物理的应用，不妨假定我们的哈密顿量  $\mathcal{H}$  式

中吸收了逆温度系数  $1/k_B T$ , 即我们的  $\mathcal{H}$  实际上是  $\mathcal{H}/k_B T$ , 有时候也称之为约化哈密顿量。引入约化量

$$t \equiv (T - T_c)/T_c, \quad h = \mu_B H/k_B T, \quad \dots \quad (3.18)$$

$t, h, \dots, h_j, \dots$  是各类“热力学场”（在量子场论中对应耦合常数）。则现在的哈密顿量可以写成

$$\mathcal{H}[s; t, h, \dots, h_j, \dots]. \quad (3.19)$$

这里  $s$  代表所有微观自旋  $s_x$  的集合。我们可以假设其中某些热力学场（特别是温度）可由实验者直接调控；但另一些可能是“内禀给定”的，因为它们体现了系统的微观细节，是“自然固定”的。

在凝聚态物理的传统研究中，人们通常关注那些具有特定简单形式的  $\mathcal{H}$ ，其可调参数往往只有两三个——伊辛模型便是一个典型特例，仅含约化温度  $t$  和约化外场  $h$  两个变量。然而，威尔逊方法的一个革命性洞见在于：任何一个这样的“物理哈密顿量”都应被看作是在一个无比庞大的可能（约化）哈密顿量空间  $\mathbb{H}$  中所确定的一个低维子空间（例如由“坐标”  $t$  和  $h$  张成的平面）。这种视角的转换对于重整化群理论的正确表述至关重要。

一旦给定了微观哈密顿量，统计力学的任务便是揭示相应宏观系统的热力学性质。首要步骤是计算配分函数：

$$\mathcal{Z}[\mathcal{H}] = \text{Tr}_s \left\{ e^{-\mathcal{H}[s]} \right\} \quad (3.20)$$

这里的求迹运算  $\text{Tr}_s \{ \cdot \}$  表示对全部  $N$  个自旋变量  $s_x$  在其所有可能取值上的求和或积分。玻尔兹曼因子  $\exp(-\mathcal{H}[s])$  自然度量了在温度  $T$  的平衡系综中，出现由特定组态  $\{s_x\}$  所描述的微观状态的概率。系统的热力学性质则由总自由能密度给出：

$$f[\mathcal{H}] \equiv f(t, h, \dots, h_j, \dots) = - \lim_{N, V \rightarrow \infty} V^{-1} \log \mathcal{Z}[\mathcal{H}]. \quad (3.21)$$

其中特别包含我们关注的在临界点附近的奇异部分  $f_s[\mathcal{H}]$ 。各类关联函数亦可依标准方式类似定义。

倘若我们能够对某个具体模型切实执行式 (3.20) 中的求迹运算，并完成式 (3.21) 中的“热力学极限”取法，便可精确地得到临界指数、标度函数等关键结果。这正是昂萨格 (Onsager, 1944) 在零磁场下求解二维伊辛模型时所开创的路径。初看起来，这似乎意味着重整化群理论已无用武之地。然而，这种印象与事实相去甚远。问题关键在于**理解**！（即便有朝一日我们真的能够在计算机上以极高精度模拟临界系统上述根本问题依然如故。）简而言之，尽管我们确切知晓二维伊辛模型的  $\alpha = 0$ （对数发散）、 $\beta = 1/8$ 、 $\gamma = 7/4$ 、 $\nu = 1$ 、 $\eta = 1/4$ ；但我们并不知道这些指数为何取这些特定数值，也不理解它们为何满足指数关系式 (3.16)。事实上，即便对于这种在物理上高度简化的模型，其精确解所不可避免的数学复杂性，也几乎掩盖了所有可能“解释”这些结果的普适性底层机制与原理的线索。因此，即便采用实空间重整化群方法对二维伊辛模型进行相当粗糙的近似求解，也可能真正具有深刻的启发意义。

### 与序参量的联系

我们将配分函数表示为：

$$\mathcal{Z} = \sum_n e^{-E_n}. \quad (3.22)$$

式中  $n$  表示系统的第  $n$  个微观状态，记作  $|n\rangle$ ， $E_n$  是这个微观状态所对应的约化能量。

系统的每一个微观态  $|n\rangle$  都可以与某个特定的、空间变化的序参量函数  $m(\mathbf{x})$  相联系。为了说明这一点，我们将使用伊辛模型的语言进行讨论，尽管该讨论可以推广到任何系统。在伊辛模型中，我们可以考虑通过对某点附近一个物理小（所谓物理小就是宏观上看足够小，但是微观上看依然包含了许多的格点在内）的区域内的所有自旋取平均，来为这个物理小的区域赋予一个磁化强度  $m(\mathbf{x})$ 。显然，这样定义的磁化强度在宏观上可以近似看作是平滑的连续函数，这些函数在小于晶格间距的距离尺度上基本上是常数。通过这种方式，我们就得到了从微观态空间到磁化强度函数的一个映射： $|n\rangle \rightarrow m(\mathbf{x})$ 。但这个映射并不是一一对应的。例如，由于平均过程是在足够多的格点上进行的，那么仅仅翻转某一个格点的自旋，不太可能对平均值产生显著影响。因此，许多微观态会映射到同一个磁化强度函数上。仅对这些微观态求和，就可以从第一性原理构造出一个关于序参量的泛函  $F[m(\mathbf{x})]$ ，其定义为：

$$e^{-F[m(\mathbf{x})]} \equiv \sum_{n|m(\mathbf{x})} e^{-E_n}. \quad (3.23)$$

因为在公式 (3.23) 中，我们只对那些对应于特定  $m(\mathbf{x})$  的微观态进行了求和。而要计算完整的配分函数，我们需要对所有微观态求和。但我们可以通过对所有可能的  $m(\mathbf{x})$  值求和来实现这一点。换句话说：

$$\mathcal{Z} = \int \mathcal{D}m(\mathbf{x}) e^{-F[m(\mathbf{x})]}. \quad (3.24)$$

等式右边是一个棘手的对象：它是一个泛函积分。我们是在对所有可能的函数  $m(\mathbf{x})$  进行积分，这相当于进行无限多重积分。当然，实际上，由于序参量  $m(\mathbf{x})$  源自底层的晶格结构，并且在短距离尺度上是足够平滑的，所以这个问题有所缓解。

(3.24) 式在物理上非常优美。它意味着我们应该将自由能泛函  $F[m(\mathbf{x})]$  视为一个关于连续变量  $m(\mathbf{x})$  的新有效哈密顿量。它是通过对配分函数的大部分微观信息进行求和而得到的，但最终仍然留下了一个对这些微观信息的平均量（即序参量）的涨落进行求和或积分的过程。后文会具体讨论如何进行这样的积分。

### 3.3 卡丹诺夫的物理图像

1966 年，卡丹诺夫 (Kadanoff) 以一项大胆的创新颠覆了临界现象的研究——他提出可以通过“粗粒化”手段，将系统在临界点附近“自我映射”，从而在减少有效自由度的同时保持物理本质。这一方案不仅优雅地导出了所有标度律，更为计算临界指数指明了一条看似可行的捷径。然而，当人们仔细审视这一理论框架时，却发现其中潜藏着难以逾越的

障碍。不过回望历史，卡丹诺夫这直觉式的标度图像，恰恰孕育了后来威尔逊完整重整化群理论的核心精髓。因此，梳理这位先驱的开创性思想，对我们理解临界现象的演进至关重要。

让我们跟随卡丹诺夫 (1966) 的思路，设想一个简单模型：一个晶格常数为  $a$  的  $d$  维 ( $d > 1$ ) 伊辛系统，每个格点上坐落着自旋  $s_x = \pm 1$ ，只能取向上 (+1) 或向下 (-1) 两种状态 (见图 (3.2))。最近邻自旋间通过耦合常数  $J > 0$  相互作用，这倾向于让它们保持

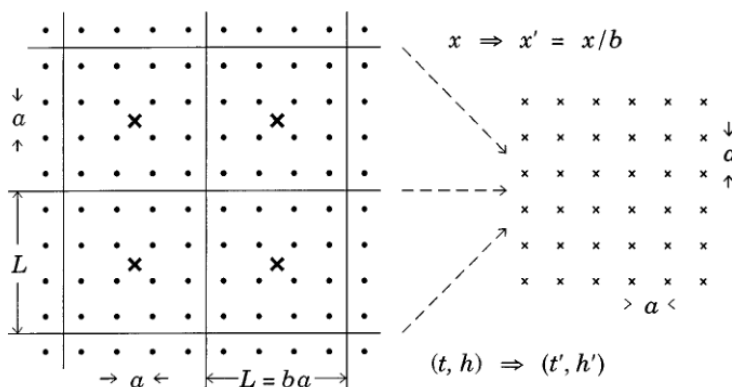


Figure 3.2: 图中展示了一个晶格常数为  $a$  的伊辛自旋晶格 (在  $d = 2$  维空间中)，自旋  $s_x$  取值为  $\pm 1$ ，以实心圆点标示。该晶格被划分为若干卡丹诺夫区块 (或称单元)，每个区块的尺寸为  $(L = ba) \times (L = ba)$ ，内含一个取值为  $\pm 1$  的区块自旋  $s'_x$ ，以十字标示。经过坐标重标度  $\mathbf{x} \Rightarrow \mathbf{x}' = \mathbf{x}/b$  后，区块自旋构成的新晶格在外观上与原始晶格完全相同。然而，人们假设原始晶格的温度  $t$  和磁场  $h$  可以被重整化，从而为重标度后的区块自旋晶格给出恰当的数值  $t'$  和  $h'$ ：详见正文。在此示意图中，空间重标度因子取为  $b = 4$ 。

平行排列 [参见式 (3.17)]。于是，在低温下，绝大多数自旋会自发选择同一个方向，要么集体“向上”，要么集体“向下”——这就是自发磁化强度  $m_0(T)$ ，它随温度升高而衰减，直至在临界温度  $T_c > 0$  处彻底消失，正如式 (3.11) 所描述的那样。

现在，施展一个“显微镜缩放”的魔术：将整个晶格切割成一个个互不相交的区块，每个区块是  $L \times L \times \dots \times L$  的超立方体，其中  $L = ba$ ，包含  $b^d$  个原始自旋 (见图 (3.2))。对于中心位于  $\mathbf{x}'$  的区块  $\mathfrak{B}_{\mathbf{x}'}$ ，我们赋予它一个新的“有效自旋”  $s'_{\mathbf{x}'}$ 。接着，对所有空间坐标进行重标度：

$$\mathbf{x} \Rightarrow \mathbf{x}' = \mathbf{x}/b, \quad (3.25)$$

奇妙的是，这样一来，新的块自旋晶格在几何形态上竟与原始晶格别无二致——自由度的密度丝毫未变 (见图 (3.2))。

但表象可能具有欺骗性。要让这种几何相似真正成为物理上的自洽，我们必须建立有效自旋间的新耦合常数  $J'$  与原始耦合  $J$  之间的桥梁，或者说，将描写有效自旋的重整化后温度偏差  $t'$  与原始  $t$  联系起来。同样，重整化磁场  $h'$  也要与原始场  $h$  建立对应关系。

为此，卡丹诺夫设定  $b$  足够大，却又小于关联长度  $\xi(t, h)$  与晶格常数  $a$  的比值  $\xi/a$ ；由于  $\xi$  在临界点处发散，这为我们渐近地任意选择  $b$  留下了空间。很显然，原晶格的自由能密度  $f(t, h)$  和重整化后块自旋晶格的自由能密度  $f(t', h')$  之间满足

$$f(t, h) = b^{-d} f(t', h'). \quad (3.26)$$

卡丹诺夫敏锐地指出：外磁场  $h$  与包含  $b^d$  个自旋的区块总耦合，等价于与区块平均自旋的耦合：

$$\bar{s}_{\mathbf{x}'} \equiv b^{-d} \sum_{\mathbf{x} \in \mathfrak{B}_{\mathbf{x}'}} s_{\mathbf{x}} \equiv \zeta(b) s'_{\mathbf{x}'}, \quad (3.27)$$

这里求和遍历区块  $\mathfrak{B}_{\mathbf{x}'}$  内的所有格点  $\mathbf{x}$ ，而与新伊辛块自旋  $s'_{\mathbf{x}'}$  的“渐近等价性”则由某个“自旋重标度因子”  $\zeta(b)$  来刻画。根据自由能间的关系

$$h \sum_{\mathbf{x} \in \mathfrak{B}_{\mathbf{x}'}} s_{\mathbf{x}} = h' s'_{\mathbf{x}'} \Rightarrow h b^d \bar{s}_{\mathbf{x}'} = h' s'_{\mathbf{x}'}. \quad (3.28)$$

从而即有

$$h' = b^d \zeta(b) h. \quad (3.29)$$

类似地，引入温度重标度因子  $\vartheta(b)$ ，我们得到递推关系：

$$t' \approx \vartheta(b) t \quad \text{和} \quad h' \approx b^d \zeta(b) h. \quad (3.30)$$

相应地，基本关联函数应该按如下方式重整化：

$$G(\mathbf{x}; t, h) \equiv \langle s_0 s_{\mathbf{x}} \rangle \approx \langle \bar{s}_0 \bar{s}_{\mathbf{x}'} \rangle \approx \zeta^2(b) G(\mathbf{x}'; t', h'). \quad (3.31)$$

换言之，当我们在空间上“zoom out”（拉远视角）并积掉  $b^{-d}$  比例的微观自由度后，系统在渐近意义上竟映射回了自身——尽管是在重整化后的温度和磁场下！这种映射是完全的，所有统计性质都通过相似性相互关联。

然而，一个关键问题悬而未决：如何确定重标度因子  $\zeta$  和  $\vartheta$ ？让我们聚焦临界点  $t = h = 0$  的情形，此时根据式 (3.30) 也有  $t' = h' = 0$ 。若采纳实验与理论预期的幂律衰减式，即  $G_c(\mathbf{x}) \sim 1/|\mathbf{x}|^{d-2+\eta}$ ，与 (3.5) 式比较，我们很快发现，为了确保两点函数的标度不变性  $\zeta(b)$  必须是  $b^{-\Delta}$ 。追随卡丹诺夫 (1966)，自然假设：

$$\zeta(b) = b^{-\Delta} \quad \text{和} \quad \vartheta(b) = b^{\lambda}, \quad (3.32)$$

其中两个临界指数  $\Delta = (d-2+\eta)/2$  和  $\lambda$  刻画了所研究的相变点，而  $b$  是一个本质上不受限制的标度参数。从而由 (3.30) 式，知

$$t' \approx b^{\lambda} t \quad \text{和} \quad h' \approx b^{d-\Delta} h. \quad (3.33)$$

利用在  $t, h \rightarrow 0$  时自由选择  $b$  的特权, 或者等价地, 迭代递推关系式 (3.30) 和 (3.31), 我们可以导出所有的标度律。当然, 所有临界指数现在都由  $\Delta$  和  $\lambda$  决定: 例如, 将 (3.33) 式代入自由能密度的 (3.26) 式, 即有

$$f(t, h) = b^{-d} f(b^\lambda t, b^{d-\Delta} h). \quad (3.34)$$

取  $b = t^{-1/\lambda}$ , 则有如下标度关系

$$f(t, h) = t^{d/\lambda} \mathcal{F}(h/t^{(d-\Delta)/\lambda}). \quad (3.35)$$

式中  $\mathcal{F}(h/t^{(d-\Delta)/\lambda}) \equiv f(1, h/t^{(d-\Delta)/\lambda})$ 。

例如, 磁化强度

$$m(t, h) = -\frac{\partial f}{\partial h} = -b^{-d} \frac{\partial h'}{\partial h} \frac{\partial f}{\partial h'} = b^{-\Delta} m(t', h'). \quad (3.36)$$

若取  $h = 0, t' = 1 \Rightarrow b = t^{-1/\lambda}$ , 则有

$$m(t) = t^{\Delta/\lambda} m(1, 0) \Rightarrow \beta = \Delta/\lambda. \quad (3.37)$$

若取  $t = 0, h' = 1 \Rightarrow b = h^{1/(\Delta-d)}$ , 则有

$$m(h) = h^{\Delta/(d-\Delta)} m(0, 1) \Rightarrow \delta = (d - \Delta)/\Delta. \quad (3.38)$$

磁化率  $\chi = \frac{\partial m}{\partial h}$ , 可以算得

$$\chi(t, h) = b^{d-2\Delta} \chi(t', h') \Rightarrow \chi(t) = t^{(2\Delta-d)/\lambda} \chi(1, 0) \Rightarrow \gamma = (d - 2\Delta)/\lambda. \quad (3.39)$$

再比如在  $h = 0$  时算比热  $C = -T \frac{\partial^2 f}{\partial T^2}$ , 得

$$C(t) = b^{-d+2\lambda} C(t', 0) = t^{-(2\lambda-d)/\lambda} C(1, 0) \Rightarrow \alpha = 2 - d/\lambda. \quad (3.40)$$

再以上临界指数中消去  $\Delta$  和  $\lambda$ , 则可以得到指数关系

$$\alpha + 2\beta + \gamma = 2, \quad \alpha + \beta(\delta + 1) = 2. \quad (3.41)$$

另外, 很显然, 关联长度满足

$$\xi(t) = b\xi(t') = t^{-1/\lambda} \xi(1) \Rightarrow \nu = 1/\lambda. \quad (3.42)$$

通过消去  $\lambda$ , 不难得到

$$\alpha = 2 - d\nu, \quad \gamma = (2 - \eta)\nu. \quad (3.43)$$

正因为能自然解释所有这些指数关系, 卡丹诺夫的图像因此获得强力支撑: 这一关系在  $d = 2$  伊辛模型中精确成立! 而且对于  $d < 4$  的所有其他可精确求解模型也都成立。更进一步, 这一分析还预言了新的指数关系, 即所谓的超标度律  $\alpha = 2 - d\nu$ , 它显式地依赖于空间维度。

然而，与此前所有与维度  $d$  无关的指数关系不同，超标度在经典理论中失效，除非  $d = 4$ 。既然我们知道（对某些模型甚至是严格证明的）经典指数值在  $d > 4$  时仍然有效，超标度就不可能普遍成立。显然，卡丹诺夫的图像中缺失了某块关键拼图。

不过，超标度问题并非阻碍卡丹诺夫理论发展的主要拦路虎。真正的困难在于如何论证重标度因子式 (3.30)、(3.31)、(3.32) 的幂律形式的合理性，特别是如何证明相邻块自旋（如  $s'_{\mathbf{x}}$  和  $s'_{\mathbf{x}+\sigma}$ ）之间真的可以用单个有效重整化耦合  $J'$  来描述。想象一下，两个相邻的  $L \times L \times L$  区块（以  $d = 3$  为例）的界面，每个界面包含  $b^2$  个强耦合的原始晶格自旋  $s_{\mathbf{x}}$ 。在远低于  $T_c$  时，这些自旋统统“冻结”在向上或向下状态，此时用单一耦合描述或许尚可；但在  $T_c$  及以上，这些自旋在多个尺度上剧烈涨落，单个有效自旋耦合似乎远不足以刻画这种内在的复杂性。

此外，卡丹诺夫的图像并未真正揭示普适性的起源：式 (3.32) 中的重标度指数  $\Delta$  和  $\lambda$  原则上可能因系统而异。威尔逊的重整化群理论不仅回答了“微观细节如何在粗粒化过程中丢失”这一难题，更为普适性提供了自然而然的解释。

### 3.4 威尔逊重整化群理论

#### 3.4.1 重整化群变换的构建

当试图以具体可计算的方式实现卡丹诺夫的标度图像时，必须额外加上一个关键洞察：任何具体物理系统——或其约化哈密顿量  $\mathcal{H}(t, h, \dots)$ ——本质上只是浩瀚哈密顿量空间  $\mathbb{H}$  中一个低维的子流形。这一视角的重要性远非仅具形式意义。

在卡丹诺夫的图像中，通常假设经“重标度”或“重整化”后，新的重整化哈密顿量应保持原来的形式不变，仅少数参数（如温度  $t$  与场  $h$ ）发生重整化。然而，这一假设如同危险的紧身衣，除非特例眷顾，否则往往失效。威尔逊的突破正是从这一束缚中“解放”出来——正是这种自由，为系统设计的重整化群变换敞开了大门。

为阐明此点，不妨回顾吉布斯计算配分函数的经典方案（见式 (3.20)）：对所有  $N$  个自旋变量  $s_{\mathbf{x}}$  的允许值求和或积分。这一直截了当的计算却极其困难。取而代之，“分而治之”的策略是将  $\{s_{\mathbf{x}}\}$  划分为两组：第一组  $\{s_{\mathbf{x}}^{\leq}\}$  包含  $N' = N/b^d$  个自旋，作为保留的涨落变量；第二组  $\{s_{\mathbf{x}}^{\geq}\}$  包含剩余的  $N - N'$  个自旋，将被积分或求和掉，从而从问题中消去。借鉴卡丹诺夫的块自旋图像，一种自然的选择是对每个  $b^d$  自旋块中除中心自旋外的所有自旋进行积分。这一被称为“抽点”的过程，保持了平移不变性，构成了“粗粒化”的具体实现。

完成部分求迹后，必然得到一个仅涉及保留自旋的有效哈密顿量  $\mathcal{H}_{\text{eff}}[s^{\leq}]$ 。为忠实于原始物理，该有效哈密顿量须通过其玻尔兹曼因子定义。根据统计力学基础，可以直接导出显式公式：

$$e^{-\mathcal{H}_{\text{eff}}[s^{\leq}]} = \text{Tr}_{s^{\geq}} \left[ e^{-\mathcal{H}[s^{\leq} \cup s^{\geq}]} \right], \quad (3.44)$$

其中并集  $s^{\leq} \cup s^{\geq}$  即原始自旋的完整集合  $s \equiv \{s_{\mathbf{x}}\}$ 。继而通过重标度与重新标记（如式 (3.27) 所示）， $s_{\mathbf{x}}^{\leq} \Rightarrow s'_{\mathbf{x}'}$ ，得到“重整化哈密顿量”  $\mathcal{H}'[s'] \equiv \mathcal{H}_{\text{eff}}[s^{\leq}]$ 。形式上，从原来的  $\mathcal{H}[s]$  到

$\mathcal{H}'[s']$  的重整化变换可记为:

$$\mathcal{H}'[s'] = \mathbb{R}_b\{\mathcal{H}[s]\}, \quad (3.45)$$

其中下标  $b$  用于追踪空间重标度因子。

值得注意的是, 若接着对重整化自旋求迹以完成吉布斯方案, 即得到完整的配分函数  $\mathcal{Z}[\mathcal{H}]$ 。因此并无信息损失: 重整化哈密顿量完整保留了所有热力学信息。然而经验表明, 与其直接从  $\mathcal{H}'$  计算  $\mathcal{Z}$ , 不如迭代重整化变换, 获得一系列重整化哈密顿量  $\mathcal{H}^{(l)}$ :

$$\mathcal{H}^{(l)} = \mathbb{R}_b[\mathcal{H}^{(l-1)}] = \mathbb{R}_b^l[\mathcal{H}], \quad (3.46)$$

其中  $\mathcal{H}^{(0)} \equiv \mathcal{H}$ ,  $\mathcal{H}^{(1)} = \mathcal{H}'$ 。正是这些迭代赋予了重整化群变换以半群特性。

如果取  $b = 1 + \varepsilon$ , 我们也可以把上面的迭代方程转化成单参微分方程

$$\frac{d}{dl}\mathcal{H}_l = \mathbb{B}[\mathcal{H}_l]. \quad (3.47)$$

式中  $\mathbb{B}$  表示某个变换。

然而关键在于: 经重标度与重新标记后, 微观变量  $\{s'_x\}$  虽与原始自旋  $\{s_x\}$  完全等价, 但当着手通过式 (3.44) 确定  $\mathcal{H}_{\text{eff}}$  及随后的  $\mathcal{H}'$  时, 立即显现出一个基本事实——不能期望  $\mathcal{H}'$  重现  $\mathcal{H}$  的原始形式。

具体考察零磁场下正方晶格上的伊辛自旋 ( $s_x = \pm 1$ ), 其初始哈密顿量仅含最近邻相互作用, 耦合强度为  $K_1 = J_1/k_B T$ 。在最保守的卡丹诺夫图像中, 重整化耦合本应遵循确定的递推关系  $K'_1 = \mathcal{T}_1(K_1)$ , 体现在某个特定函数  $\mathcal{T}(\cdot)$  中。但实际上,  $\mathcal{H}_{\text{eff}}$  必然包含更多非零自旋耦合: 次近邻间的  $K_2$ 、第三近邻间的  $K_3$ , 直至无限高阶。更甚之, 四自旋耦合项如  $K_{\square_1 s_{x_1} s_{x_2} s_{x_3} s_{x_4}}$  也会出现, 且涵盖所有可能的四自旋排列; 六自旋、八自旋耦合亦然。事实上, 晶格上任意  $2m$  个伊辛自旋的集合  $Q$  (及其平移等价类) 都会生成非零耦合常数  $K'_Q$  并出现在  $\mathcal{H}'$  中。

唯一的慰藉在于, 进一步迭代式 (3.44) 的抽点变换 (零场下) 不会导致更糟的情况。换言之, 零场伊辛自旋哈密顿量空间  $\mathbb{H}_{\text{Is}}$  可由所有可能自旋耦合的无限集合  $\{K_Q\}$  刻画, 且在抽点变换下封闭。形式上,  $\mathbb{R}_b$  因此由全套递推关系描述:

$$K'_P = \mathcal{T}_P(\{K_Q\}) \quad (\text{所有 } P). \quad (3.48)$$

这显然回答了卡丹诺夫原始图像中那些复杂的“跨块面”相互作用的归宿: 它们将重整化哈密顿量推出了 (过小的) 最近邻伊辛模型流形, 引入了 (无穷多) 进一步的耦合。图 (3.3) 示意性地描绘了由此产生的局面:  $\mathcal{H}'(t', h')$  的重整化流形通常与原始流形并无重叠。假如我们考虑的不是离散的重整化群迭代, 而是它的连续版本, 那么连续的“迭代”过程就生成了图 (3.3) 中的流线或“轨迹”。注意, (3.48) 式实际上是一个无穷维动力系统, 它的持续迭代会在哈密顿量空间  $\mathbb{H}$  生成出很多流线。但是, 动力系统当然可能出现固定点, 比如图 (3.3) 中的  $\mathcal{H}^*$  就是一个临界固定点, 因为既有流线流入这个固定点, 也有流线从这个固定点流出, 而这往往对应相变。更清楚地显示这样的相变图像的是图 (3.4)。

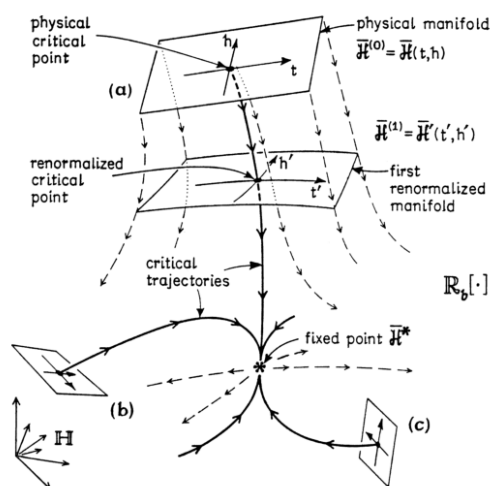


Figure 3.3: 这是哈密顿量空间  $\mathbb{H}$  的示意图——其中展示了初始的或物理的流形 [标记为 (a)、(b)、 $\dots$  等], 以及由离散 RG 变换  $\mathbb{R}_b$  (具有空间重标度因子  $b$ ) 的重复作用、或由相应的连续重整化群所诱导的流动。临界轨迹以粗线显示: 在图示的  $\mathbb{H}$  区域内, 它们全都终止于一个固定点  $\mathcal{H}^*$ 。一般而言, 完整空间还包含其他非平庸的临界固定点, 描述多临界点和不同的临界点普适类; 此外, 通常还会出现平庸的固定点, 典型的如对应高温相的“汇” (sink), 其没有流出的轨迹。[改绘自 Fisher (1983)。]

实践中, 式 (3.44) 指定的朴素抽点变换通常难以成为有效计算的基础。设计实用的重整化群变换与其说是一门科学, 不如说是一门艺术: 并无标准配方可循。然而, 威尔逊阐述的一般哲学提供了指导原则: 首先积去对所研究宏观现象“最不直接重要”的微观变量或自由度, 同时保留最重要的那些。对于铁磁或气-液临界点, 最重要的现象发生在长的长度尺度上——关联长度  $\xi$  发散; 临界关联  $G_c(r)$  在长距离上缓慢衰减;  $T_c$  以下出现长程有序。

### 3.4.2 流、不动点、普适性与标度

为完善这一理论叙事并填补先前略过的逻辑缺环, 此处需阐明威尔逊于哈密顿量空间  $\mathbb{H}$  中构建的重整化群变换, 如何使该理论得以解释普适性, 以及原则上如何实现对临界指数的计算。如图 (3.3) 所示, 重整化群变换  $\mathbb{R}_b$  的递归应用在哈密顿量空间  $\mathbb{H}$  中诱导出一道动力学流。细致观察可见, 那些“明智的”、“合理的”, 或更恰当地说, “设计精良的”重整化群变换具有光滑性, 因此原始物理流形  $\mathcal{H}^{(0)} = \mathcal{H}(t, h)$  中彼此靠近的点——例如在温度参数上相近者——经重整化后在  $\mathcal{H}^{(1)} \equiv \mathcal{H}'$  中依然邻近, 且随着流参数  $l$  的递增, 在  $\mathcal{H}^{(l)}$  中始终保持这一邻近性。值得注意的是, 由于空间尺度依  $\mathbf{x} \Rightarrow \mathbf{x}' = \mathbf{x}/b^l$  重整化, 可将  $l = -\log_b(|\mathbf{x}'|/|\mathbf{x}|)$  对数式地理解为描述系统的尺度; 然而需注意, 随着“尺度”变更或  $l$  增大, 哈密顿量的形式通常亦随之演化。故而, 部分重整化的哈密顿量预期将呈现某种通用的介观形态: 这使其成为赋予朗道-金兹堡 (现称 LGW) 有效哈密顿量确切含义的恰当候选者。

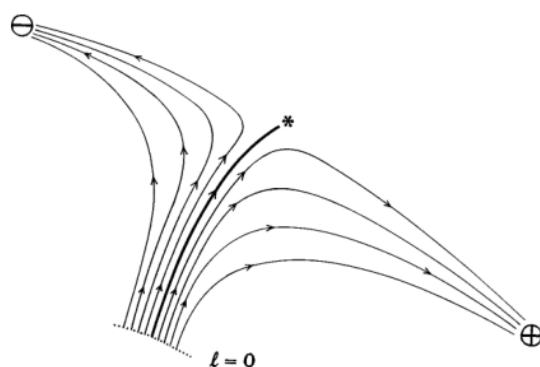


Figure 3.4: 关于某个大空间中“流”的构想图景。此图传达的观念是：在流的起点——即标记为  $l=0$  的轨迹处——原本相互接近且平滑相连的点，最终会彼此分离，流向代表截然不同的“终态”的遥远区域：这正是相变的本质所在。用现代术语来说，这一流发生在哈密顿量的空间  $\mathbb{H}$  中；图中以加粗线条显示的分界线（separatrix）与初始轨迹（ $l=0$ ）的交点即代表物理临界点； $*$  表示控制性固定点，而  $\oplus$  和  $\ominus$  则分别代表渐近的高温无序态和低温有序态。

鉴于重整化群变换的光滑特质，若已获知第  $l$  次重整化阶段的自由能密度  $f_l \equiv f[\mathcal{H}^{(l)}]$ ，则原始自由能密度  $f[\mathcal{H}]$  及其临界行为便随之确定：明确而言，有

$$f(t, h, \dots) \equiv f[\mathcal{H}] = b^{-dl} f[\mathcal{H}^{(l)}] \equiv b^{-dl} f_l(t^{(l)}, h^{(l)}, \dots). \quad (3.49)$$

此外，光滑性保证了所有普适临界性质在重整化过程中保持守恒。同理，可发现  $\mathcal{H}^{(0)} \equiv \mathcal{H}$  的临界点映射至  $\mathcal{H}^{(1)} \equiv \mathcal{H}'$  的临界点，并依此类推，如图 (3.3) 中的粗流线所示。因此，追踪  $\mathbb{H}$  空间中的临界轨迹——即从物理临界点出发的重整化群流线——极具启发意义。原则上，这些轨迹的拓扑结构可能极为复杂，甚至呈现混沌特征；然而实践中，对于设计精良或“贴切”的重整化群变换，最常见的情形是临界流终止——或更精确地说，渐近地停滞——于重整化群的一个不动点（固定点） $\mathcal{H}^*$ ：见图 (3.3)。此类不动点由方程 (3.46) 或 (3.47) 简定义为

$$\mathbb{R}_b[\mathcal{H}^*] = \mathcal{H}^* \quad \text{或} \quad \mathbb{B}[\mathcal{H}^*] = 0. \quad (3.50)$$

不动点何以至关重要？其中某些实则平淡无奇，仅对应于无相互作用或全部自旋冻结等平庸情形。然而非平庸不动点代表着临界态；更有甚者，其临界本质及其邻域内的自由能，必须与所有那些临界轨迹收敛至同一不动点的迥异哈密顿量完全一致！换言之，特定不动点定义了一个普适类，它“统治”或“吸引”所有那些临界点终将映射于其上的系统：见图 (3.3)。

至此，终于获得关于普适性的自然诠释：物理性质迥异的系统，可能仍归属于  $\mathbb{H}$  中同一不动点  $\mathcal{H}^*$  的吸引域。不同的流入轨迹反映了它们相关与无关变量的各异物理内涵。

自每个临界不动点至少流出两条“不稳定”或“向外”轨迹。这些轨迹对应于一个或多个**相关变量**，具体而言，对于图 (3.3) 和图 (3.4) 所示情形，对应于温度或热场  $t$ ，以及

磁场或有序场  $h$ 。若存在更多相关轨迹，可预期发生向不同临界行为的交叉。在  $\mathbb{R}$  空间中，此类轨迹通常导向描述（一般而言）全新普适类的不同不动点。

设计精良的重整化群变换之光滑性意味着，它总可至少在局部展开为泰勒级数。值得强调的是，正是这一性质在临界区的自由能中失效：为重获此能力，庞大的哈密顿量空间至关重要。在满足方程 (3.50) 的不动点附近，可相当普遍地期望实现线性化，写作 ( $g$  很小)

$$\mathbb{R}_b[\mathcal{H}^* + g\mathcal{Q}] = \mathcal{H}^* + g\mathbb{L}_b\mathcal{Q} + o(g), \quad (3.51)$$

或微分形式

$$\frac{d}{dl}(\mathcal{H}^* + g(l)\mathcal{Q}) = \frac{d}{dl}(g(l)\mathcal{Q}) = g(l)\mathbb{B}_1\mathcal{Q} + o(g). \quad (3.52)$$

这里  $\mathbb{L}_b$  与  $\mathbb{B}_1$  为线性算符（尽管作用于大空间  $\mathbb{R}$  上）。因此，可寻求本征值及相应的“本征算符”，例如  $\mathcal{Q}_k$ （这将是“部分哈密顿量”）。于是，有

$$\mathbb{L}_b\mathcal{Q}_k = \Lambda_k(b)\mathcal{Q}_k \quad \text{或} \quad \mathbb{B}_1\mathcal{Q}_k = \lambda_k\mathcal{Q}_k. \quad (3.53)$$

实际上（由半群性质）本征值必须满足  $\Lambda_k(b) = b^{\lambda_k}$ 。如同任何此类线性问题，了解本征值谱与本征算符，或至少其主导部分，将揭示大量关键信息。合理推测， $\mathcal{Q}_k$  应构成一般算符的展开之基

$$\mathcal{H} \equiv \mathcal{H}^* + \sum_{k \geq 1} g_k \mathcal{Q}_k. \quad (3.54)$$

物理上，展开系数  $g_k (= g_k^{(0)})$  则代表与“临界算符”  $\mathcal{Q}_k$  共轭的热力学场，而  $\mathcal{Q}_k$  本身通常接近于某些局域算符的组合。将本征值方程代入 (3.52) 即有

$$\frac{d}{dl}g_k(l) = \lambda_k g_k(l). \quad (3.55)$$

很显然，对于  $\lambda_k > 0$  的算符，相应扰动将导致流线从临界不动点向外，因此是**相关算符**。反之，对于  $\lambda_k < 0$  的算符，相应扰动总是被吸向临界不动点，因此这类算符是**无关算符**。对于  $\lambda_k = 0$  的算符，则称之为**边际算符**。在重整化下，每个  $g_k$  简单地演化为  $g_k^{(l)} \approx b^{\lambda_k l} g_k^{(0)}$ 。

实际上，在典型临界点问题中，通常可发现两个相关算符，例如  $\mathcal{Q}_1$  与  $\mathcal{Q}_2$ ，具有  $\lambda_1 > 0, \lambda_2 > 0$ 。通常，这些算符之一，例如通过其对称性，可被识别为局域能量密度  $\mathcal{Q}_1 \cong \mathcal{E}$ ，因而  $g_1 \equiv t$  为热场；第二个则表征序参量  $\mathcal{Q}_2 \cong \psi$ ，其场为  $g_2 \equiv h$ 。

最后，检视自由能的流方程 (3.49)。关键点在于，重整化程度  $b^l$  可任意选择得足够大。当  $t \rightarrow 0$  时，即处于待理解的临界区，明智之选是令  $b^l = 1/|t|^{1/\lambda_1}$ ，其显然发散至  $\infty$ 。于是发现方程 (3.49) 导出如下基本标度关系

$$f_s(t, h, \dots, g_j, \dots) \approx |t|^{\frac{d}{\lambda_1}} \mathcal{F} \left( \frac{h}{|t|^{\frac{\lambda_2}{\lambda_1}}}, \dots, \frac{g_j}{|t|^{\phi_j}}, \dots \right). \quad (3.56)$$

其中  $\phi_j = \frac{\lambda_j}{\lambda_1}$ 。

继而，由于  $\lambda_1 > 0$ ，所以给定  $\phi_j$  的符号，从而相应  $\lambda_j$  的符号，就决定了相应临界算符  $\mathcal{Q}_j$  及其共轭场  $g_j$  的相关性。此场可能（但对大多数  $j$  值并非如此）处于实验的直接控制之下。当不存在边际变量且负得最少的  $\phi_j$  之模大于 1 时，简单的标度描述通常效果良好，卡丹诺夫图像几乎适用，因为这时候，随着  $|t| \rightarrow 0$ ，标度函数  $\mathcal{F}$  中所有  $g_j$  变量的贡献都趋于零，从而可以忽略。

### 3.4.3 算符乘积展开

重整化群变换相当于不断在空间上“zoom out”（拉远视角）系统，同时平均掉（积掉）小尺度上的微观自由度。所以重整化群变换告诉我们的是如何把关注重心逐步从小的空间尺度转移向大的空间尺度。特别的，在重整化群的不动点位置，由于可以无限地往大尺度走，所以是看不见系统的微观晶格结构的，看到的只能是连续的场，在固定点  $\mathcal{H}^*$  处求配分函数就相当于以玻尔兹曼因子  $e^{-\mathcal{H}^*}$  为权重对连续的场进行泛函积分，这对应的是一个连续的欧几里德空间量子场论，它以  $\mathcal{H}^*$  为欧空间作用量，有时候也记作  $S_*$ 。而且，由于固定点在重标度下不变，所以固定点处的这个量子场论还得有标度不变性，或者说尺度不变性，即在空间尺度的放缩下保持不变。

因此，在固定点附近，可以按照标度量纲来将局域场算符分类。根据 (3.5) 式，对于标度量纲为  $\Delta_i$  的局域场算符  $\mathcal{O}_i$ （它可能是个复合算符），我们有关联函数

$$\langle \mathcal{O}_i(\mathbf{x}) \mathcal{O}_i(0) \rangle = \frac{D}{|\mathbf{x}|^{2\Delta_i}}. \quad (3.57)$$

现在，考虑两个不同空间点上算符的乘积， $\mathcal{O}_i(\mathbf{x}_1) \mathcal{O}_j(\mathbf{x}_2)$ ，当  $\mathbf{x}_1 \rightarrow \mathbf{x}_2$  时  $\mathcal{O}_i(\mathbf{x}_1)$  对  $\mathcal{O}_j(\mathbf{x}_2)$  作用的结果就相当于对后者进行了一个算符空间的变换，变换的结果当然依然是一个  $\mathbf{x}_2$  位置处的算符，因此可以把变换的最终结果按照  $\mathcal{O}_k(\mathbf{x}_2)$  进行展开，即有

$$\mathcal{O}_i(\mathbf{x}_1) \mathcal{O}_j(\mathbf{x}_2) = c_{ij}^k(\mathbf{x}_1 - \mathbf{x}_2) \mathcal{O}_k(\mathbf{x}_2), \quad (3.58)$$

式中重复的指标要默认求和（下文同）。展开系数  $c_{ij}^k(\mathbf{x}_1 - \mathbf{x}_2)$  包含所有短距离的奇异行为。

为了确定这些系数，我们考虑标度变换  $\mathcal{O}(\mathbf{x}) \rightarrow b^\Delta \mathcal{O}(b\mathbf{x})$ 。将这样的标度变换作用在 (3.58) 式的两边，即有

$$b^{\Delta_i + \Delta_j} \mathcal{O}_i(b\mathbf{x}_1) \mathcal{O}_j(b\mathbf{x}_2) = b^{\Delta_k} c_{ij}^k(\mathbf{x}_1 - \mathbf{x}_2) \mathcal{O}_k(b\mathbf{x}_2). \quad (3.59)$$

另一方面

$$\mathcal{O}_i(b\mathbf{x}_1) \mathcal{O}_j(b\mathbf{x}_2) = c_{ij}^k(b(\mathbf{x}_1 - \mathbf{x}_2)) \mathcal{O}_k(b\mathbf{x}_2). \quad (3.60)$$

两个式子比较，即有

$$c_{ij}^k(b(\mathbf{x}_1 - \mathbf{x}_2)) = b^{\Delta_k - \Delta_i - \Delta_j} c_{ij}^k(\mathbf{x}_1 - \mathbf{x}_2). \quad (3.61)$$

这一定是下面这样的齐次函数

$$c_{ij}^k(\mathbf{x}_1 - \mathbf{x}_2) = \frac{c_{ij}^k}{|\mathbf{x}_1 - \mathbf{x}_2|^{\Delta_i + \Delta_j - \Delta_k}}. \quad (3.62)$$

代入前面的 (3.58) 式, 立即有

$$\mathcal{O}_i(\mathbf{x}_1)\mathcal{O}_j(\mathbf{x}_2) = \frac{c_{ij}^k}{|\mathbf{x}_1 - \mathbf{x}_2|^{\Delta_i + \Delta_j - \Delta_k}} \mathcal{O}_k(\mathbf{x}_2), \quad \text{当 } \mathbf{x}_1 \rightarrow \mathbf{x}_2. \quad (3.63)$$

这就是具有标度不变性的量子场论所必须满足的算符乘积展开关系。它描述的是标度不变量子场论的短距离结构。

#### 3.4.4 固定点附近的重整化群流

假定我们有一个欧空间量子场论, 它的作用量是  $S(a)$ , 由于欧空间量子场论就相当于对连续场变量的吉布斯统计力学, 所以  $S(a)$  也相当于这个统计力学中的有效哈密顿量  $\mathcal{H}(a)$ , 其中  $a$  是定义这个理论所需的一个尺度, 比方说如果这个理论是定义在格点上, 那么  $a$  就是格距。现在我们改变  $a$ , 将之变为  $\tilde{a}$ , 并重标度理论, 同时调节理论的耦合常数, 使得作用量变为  $S(\tilde{a})$ 。重整化群变换的要求是, 用新尺度  $\tilde{a}$  和新作用量  $S(\tilde{a})$  算出来的关联函数依然和原来用  $S(a)$  算出来的一样。特别的, 在重整化群变换下, 配分函数  $\mathcal{Z}$  应该保持不变。重整化群的固定点就是这样一种特别的作用量  $S_*$  (相当于前面的  $\mathcal{H}^*$ ), 它满足

$$S_*(\tilde{a}) = S_*(a). \quad (3.64)$$

我们这一节要研究的就是固定点附近的重整化群流。考察对固定点理论  $S_*(a)$  的一个扰动,

$$S_* \rightarrow S = S_* + \int (d^d x) g_i a^{\Delta_i - d} \mathcal{O}_i, \quad (3.65)$$

式中  $\mathcal{O}_i$  是一个标度量纲为  $\Delta_i$  的算子,  $a$  是定义理论的尺度, 把因子  $a^{\Delta_i - d}$  单独写出来, 是为了确保只需要重标度  $a$  就能实现对理论的重标度。  $g_i$  是无量纲的耦合常数, 是一个小扰动, 因子  $a^{\Delta_i - d}$  也确保了整个扰动项无量纲。当然, 我们默认对  $i$  求和。那么理论的配分函数近似就是,

$$\begin{aligned} \mathcal{Z} = \text{Tr}(e^{-S}) &= \mathcal{Z}_* \left( - \int (d^d x) g_i a^{\Delta_i - d} \langle \mathcal{O}_i \rangle \right) + \\ &\mathcal{Z}_* \left( \frac{1}{2} \int \int_{|x_1 - x_2| > a} (d^d x_1)(d^d x_2) g_i g_j a^{\Delta_i + \Delta_j - 2d} \langle \mathcal{O}_i(x_1) \mathcal{O}_j(x_2) \rangle \right), \end{aligned} \quad (3.66)$$

式中  $\mathcal{Z}_* = \text{Tr}(e^{-S_*})$  是固定点理论的配分函数。由于我们定义理论的尺度是  $a$ , 因此在积分时我们要求两个不同点之间的距离大于  $a$ , 即  $|x_1 - x_2| > a$ 。

现在我们改变定义理论的尺子至  $\tilde{a}$ , 为了保持配分函数不变, 我们同时调节耦合常数  $g_i \rightarrow \tilde{g}_i$ , 则

$$\begin{aligned} \tilde{\mathcal{Z}} &= \mathcal{Z}_* \left( - \int (d^d x) \tilde{g}_i \tilde{a}^{\Delta_i - d} \langle \mathcal{O}_i \rangle \right) + \\ &\mathcal{Z}_* \left( \frac{1}{2} \int \int_{|x_1 - x_2| > \tilde{a}} (d^d x_1)(d^d x_2) \tilde{g}_i \tilde{g}_j \tilde{a}^{\Delta_i + \Delta_j - 2d} \langle \mathcal{O}_i(x_1) \mathcal{O}_j(x_2) \rangle \right). \end{aligned} \quad (3.67)$$

下面我们考虑一个无穷小的尺度变换  $\tilde{a} = (1 + \varepsilon)a$ , 即取重标度的尺度  $b = 1 + \varepsilon$ ,  $\varepsilon$  是一个无穷小量。利用算子乘积展开

$$\mathcal{O}_i(x_1)\mathcal{O}_j(x_2) \sim \frac{1}{|x_1 - x_2|^{\Delta_i + \Delta_j - \Delta_k}} c_{ij}^k \mathcal{O}_k(x_2), \quad (3.68)$$

我们有

$$\begin{aligned} \tilde{\mathcal{Z}} &= \mathcal{Z}_* \left( - \int (d^d x) \tilde{g}_i \tilde{a}^{\Delta_i - d} \langle \mathcal{O}_i \rangle \right) + \\ &\mathcal{Z}_* \left( \frac{1}{2} \int \int_{|x_1 - x_2| > \tilde{a}} (d^d x_1)(d^d x_2) \tilde{g}_i \tilde{g}_j \tilde{a}^{\Delta_i + \Delta_j - 2d} \frac{c_{ij}^k \langle \mathcal{O}_k(x_2) \rangle}{|x_1 - x_2|^{\Delta_i + \Delta_j - \Delta_k}} \right). \end{aligned} \quad (3.69)$$

上式后一项括号中的部分显然可以写成

$$\begin{aligned} &\frac{1}{2} \int \int_{|x_1 - x_2| > a} (d^d x_1)(d^d x_2) \tilde{g}_i \tilde{g}_j \tilde{a}^{\Delta_i + \Delta_j - 2d} \langle \mathcal{O}_i(x_1)\mathcal{O}_j(x_2) \rangle \\ &- \frac{1}{2} \int \int_{(1+\varepsilon)a > |x_1 - x_2| > a} (d^d x_1)(d^d x_2) \tilde{g}_i \tilde{g}_j \tilde{a}^{\Delta_i + \Delta_j - 2d} \frac{c_{ijk} \langle \mathcal{O}_k(x_2) \rangle}{|x_1 - x_2|^{\Delta_i + \Delta_j - \Delta_k}}. \end{aligned} \quad (3.70)$$

保留到一阶无穷小, 上式的第二项就是

$$-\frac{1}{2} S_d \varepsilon \int (d^d x) g_i g_j a^{\Delta_k - d} c_{ijk} \langle \mathcal{O}_k \rangle, \quad (3.71)$$

式中  $S_d$  表示  $d$  维空间中  $d-1$  维单位球面的面积。因此, 我们就有

$$\begin{aligned} \tilde{\mathcal{Z}} &= \mathcal{Z}_* \left[ - \int (d^d x) \left( \tilde{g}_k (1 + \varepsilon)^{\Delta_k - d} + \frac{1}{2} S_d \varepsilon g_i g_j c_{ijk} \right) a^{\Delta_k - d} \langle \mathcal{O}_k \rangle \right] + \\ &\mathcal{Z}_* \left( \frac{1}{2} \int \int_{|x_1 - x_2| > a} d^d x_1 d^d x_2 \tilde{g}_i \tilde{g}_j (1 + \varepsilon)^{\Delta_i + \Delta_j - 2d} a^{\Delta_i + \Delta_j - 2d} \langle \mathcal{O}_i(x_1)\mathcal{O}_j(x_2) \rangle \right). \end{aligned}$$

显然, 为了保持配分函数不变, 即为了保证  $\tilde{\mathcal{Z}} = \mathcal{Z}$ , 我们应该将  $\tilde{g}_i$  取成

$$\tilde{g}_k = g_k (1 + \varepsilon)^{d - \Delta_k} - \varepsilon \frac{1}{2} S_d g^i g^j c_{ijk}, \quad (3.72)$$

这里我们精确到一阶无穷小。也即是说, 我们有重整化群流

$$\begin{aligned} \frac{dg_k}{d\varepsilon} &= (d - \Delta_k) g_k - \frac{1}{2} S_d g^i g^j c_{ijk} \\ &= - \frac{\partial}{\partial g_k} \left( \frac{1}{2} (\Delta_i - d) g_i^2 + \frac{1}{3!} S_d c^{ijk} g_i g_j g_k \right). \end{aligned} \quad (3.73)$$

从上面的重整化群流方程我们很容易看出, 对于  $\Delta_k > d$  的扰动, 重整化群固定点是稳定的, 也就是说在这个扰动方向上重整化群流指向固定点, 这样的扰动我们称之为无关扰动。然而, 对于  $\Delta_k < d$  的那些扰动方向, 重整化群流的流向是背离固定点的, 这是一些不稳定的扰动方向, 这样的扰动称之为相关扰动。用前面小节的话来说即是, 线性化之后的本征值  $\lambda_k = d - \Delta_k$ 。

### 3.4.5 重整化群方程

前面我们讲过，重整化变换就是积去小尺度短程自由度。有时候我们也将这样的操作说成是从物理上忽视短程自由度，它有点像我们给一副图片打码赛克，使得图片的某些细节被忽视。物理上系统地进行这种操作的办法是将“图片”进行缩放。我们总可以通过将“图片”不断缩小，从而把打码的像素缩小成基本像素单位，进而看到整幅图片越来越宏观的结构。在重整化变换中即是，我们总可以通过不断重标度（即缩小），从而把浮动的截断尺度  $a$  化约为 1 个基本单位，从而得到大尺度的有效理论，当然，这么做的代价是会引入放大倍率  $e^{-l}$ （即缩小）（即取  $b = e^l$ ）。这一节我们就来研究，系统的关联函数如何随着这个缩放倍率的变化而变化，这就叫做重整化群方程。

假设我们考察的量子场论系统有一组局域的可观测量算符  $\mathcal{O}_i(x)$ ，它们的标度量纲分别为  $\Delta_i$ ，当我们将这个量子场论系统放大  $e^{-l}$  倍时（即缩小）， $\mathcal{O}_i(x)$  将变为  $\mathcal{O}'_i(x)$ ，由于缩小系统会使得能量密度更大，所以我们应该有如下算符混合关系，

$$\mathcal{O}_i(x) \simeq e^{-l\Delta_i} \mathcal{O}'_i(e^{-l}x). \quad (3.74)$$

等价地，我们也可以将这个方程写成

$$\mathcal{O}_i(e^l x) \simeq e^{-l\Delta_i} \mathcal{O}'_i(x). \quad (3.75)$$

这里我们没有写等于号，而是用的  $\simeq$  号，这是因为定义加撇算符的理论和定义不加撇算符的理论其实不是一回事，两者之间相差一个重整化，下面我们会进一步澄清这一点。

对于一个量子场论而言，假设基本场变量为  $\phi$ ，欧氏作用量为  $S$ ，那么量子场论的关键在于计算泛函积分  $\text{Tr}(e^{-S}) \equiv \int [\mathcal{D}\phi] e^{-S}$ ，比方说，在计算关联函数时，我们就是要计算这样的泛函积分。因此，一个关键的问题是，当我们将量子场论系统放大  $e^{-l}$  倍时，这个泛函积分会怎么变？换言之，我们得搞清楚理论本身如何重整化。

这里出现一个关键的问题，即实际上，当我们说把量子场论系统放大  $e^{-l}$  倍（即缩小）时，我们实际在物理上并不能真正做到这一点，我们实际上做的只是离这个系统远一点，在更远的地方来观测这个系统从而使得自己能够忽略一些小尺度细节而已。换句话说，我们实际上是在用一把更大的尺子来衡量这个系统。或者说在重整化过程中，系统中各点的物理距离是不变的，将坐标距离放大  $e^{-l}$  倍（即缩小）必然同时要尺子放大  $e^l$  倍。物理系统的尺子就反映为度规张量，所以，将量子场论系统的坐标距离放大  $e^{-l}$  倍必然同时要度规张量  $h$  放大  $e^{2l}$  倍，即同时进行如下变换

$$h \rightarrow e^{2l}h. \quad (3.76)$$

式中  $h$  为度规张量，其用指标写出来的分量形式为  $h_{ab}$ 。

根据上面的分析我们就能写出关联函数的基本缩放（重整化）方程

$$\begin{aligned} & \langle \mathcal{O}_1(e^l x_1) \mathcal{O}_2(e^l x_2) \dots \mathcal{O}_n(e^l x_n) \rangle_h \\ &= e^{-(\Delta_1 + \Delta_2 + \dots + \Delta_n)l} \langle \mathcal{O}'_1(x_1) \mathcal{O}'_2(x_2) \dots \mathcal{O}'_n(x_n) \rangle_{e^{2l}h}. \end{aligned} \quad (3.77)$$

$\langle \dots \rangle_{\mathbf{h}}$  表示计算相应的泛函积分时, 作用量中的时空度规取作  $\mathbf{h}$ , 也即是说,  $\langle \dots \rangle_{\mathbf{h}}$  意味着理论是定义在度规为  $\mathbf{h}$  的黎曼流形上。注意, 与度规  $e^{2l}\mathbf{h}$  对应的算子都是加撇的, 因为根据以上所说, 算子加撇和度规放大为  $e^{2l}\mathbf{h}$  是同步的, 两者一起就是所谓的重整化变换。下面我们就是要从这个方程 (3.77) 出发, 得到关联函数在缩放下满足的关于  $l$  的微分方程。

首先我们注意到, 在经典物理的层次上, 对于欧氏作用量  $S$ , 它在度规变动  $\delta h_{ab}$  下满足  $\delta S = -\int_M \frac{1}{2} T^{ab} \delta h_{ab}$ ,  $T^{ab}$  就是经典的能动量张量<sup>1</sup>,  $\int_M$  表示在时空流形上积分。我们关心的度规无穷小变动是度规  $e^{2l}h_{ab}$  在  $l$  作无穷小改变时引起的, 这时候在经典物理的层次上必有  $\delta_l S = \frac{\partial S}{\partial l} = -\int_M T_a^a(l)$ ,  $T_a^a(l)$  表示定义在度规  $e^{2l}h_{ab}$  上的能动量张量的缩并。为此, 我们可以假设在量子的层次上, 在  $l$  做无穷小变动时, 作用量和泛函积分测度  $[\mathcal{D}\phi]e^{-S}$  的变动为

$$\int_M \mathbf{T}'(l), \quad (3.78)$$

式中  $\mathbf{T}'(l)$  为前面  $T_a^a(l)$  的量子版本, 是一个算符。

根据上一段的分析我们容易知道, 当我们将方程 (3.77) 中的  $l$  作无穷小变动时, 必有

$$\begin{aligned} & \frac{\partial}{\partial l} \langle \mathcal{O}_1(e^l x_1) \mathcal{O}_2(e^l x_2) \dots \mathcal{O}_n(e^l x_n) \rangle_{\mathbf{h}} \\ &= -(\Delta_1 + \Delta_2 + \dots + \Delta_n) \langle \mathcal{O}_1(e^l x_1) \mathcal{O}_2(e^l x_2) \dots \mathcal{O}_n(e^l x_n) \rangle_{\mathbf{h}} \\ & \quad + e^{-(\Delta_1 + \Delta_2 + \dots + \Delta_n)l} \langle \mathcal{O}'_1(x_1) \mathcal{O}'_2(x_2) \dots \mathcal{O}'_n(x_n) \int_M \mathbf{T}'(l) \rangle_{e^{2l}\mathbf{h}}. \end{aligned} \quad (3.79)$$

下面假设  $\int_M \mathbf{T}'(l)$  可以用局域算子  $\mathcal{O}'_i$  展开成

$$\int_M \mathbf{T}'(l) = \sum'_i \int_M \beta^i(g) \mathcal{O}'_i(x), \quad (3.80)$$

求和号中的 ' 号表示我们仅仅对标度量纲小于等于时空维数  $d$  的算子求和, 因为能动量张量的标度量纲不能超过  $d$ 。另一方面, 我们也可以将度规  $e^{2l}\mathbf{h}$  上的有效作用量用局域算子展开为

$$S = \sum'_i g_i \int_M \mathcal{O}'_i(x). \quad (3.81)$$

同样, 由于我们关心的是长程物理, 所以我们仅仅只需要关心相关算子, 或者说可重整项。上式中的  $g_i$  是一个一般性的记号, 它包括相互作用耦合常数以及质量等等参数。由此我们可以知道, 方程 (3.79) 可以重写成

$$\begin{aligned} & \frac{\partial}{\partial l} \langle \mathcal{O}_1(e^l x_1) \mathcal{O}_2(e^l x_2) \dots \mathcal{O}_n(e^l x_n) \rangle_{\mathbf{h}} \\ &= -(\Delta_1 + \Delta_2 + \dots + \Delta_n) \langle \mathcal{O}_1(e^l x_1) \mathcal{O}_2(e^l x_2) \dots \mathcal{O}_n(e^l x_n) \rangle_{\mathbf{h}} \\ & \quad - e^{-(\Delta_1 + \Delta_2 + \dots + \Delta_n)l} \sum'_i \beta^i(g) \frac{\partial}{\partial \lambda^i} \langle \mathcal{O}'_1(x_1) \mathcal{O}'_2(x_2) \dots \mathcal{O}'_n(x_n) \rangle_{e^{2l}\mathbf{h}}. \end{aligned} \quad (3.82)$$

<sup>1</sup>这个结果的正负号不用死记, 取一个标量场论算一下就知道了。

将加撇算符变回不加撇算符，简单整理一下即有

$$\begin{aligned} & \frac{\partial}{\partial l} \langle \mathcal{O}_1(e^l x_1) \mathcal{O}_2(e^l x_2) \dots \mathcal{O}_n(e^l x_n) \rangle_h \\ &= -(\Delta_1 + \Delta_2 + \dots + \Delta_n) \langle \mathcal{O}_1(e^l x_1) \mathcal{O}_2(e^l x_2) \dots \mathcal{O}_n(e^l x_n) \rangle_h \\ & \quad - \sum_i' \beta^i(g) \frac{\partial}{\partial g_i} \langle \mathcal{O}_1(e^l x_1) \mathcal{O}_2(e^l x_2) \dots \mathcal{O}_n(e^l x_n) \rangle_h. \end{aligned} \quad (3.83)$$

这就是我们要推导的**重整化群方程**。

如果不用度规的办法进行重整化，而像前面的小节一样认为重整化就是调节耦合常数  $g_i$ ，因此它是一个依赖于  $l$  的  $g_i(l)$ ，则我们还有

$$\frac{dg_i}{dl} = -\beta^i(g). \quad (3.84)$$

可见， $\beta^i(g)$  就是通常所说的 Beta 函数。只不过我们这里的  $l$  是对长度尺度的放大倍率，而不是对能量尺度的放大倍率，所以 Beta 函数的方程 (3.84) 多了一个负号。

## 3.5 临界指数的计算

### 3.5.1 经典平均场理论

能给出具体临界指数的最简单办法就是朗道的平均场理论，虽然它给出的临界指数是错的，但不失为一个好的起点。本节以伊辛铁磁体为例讨论朗道平均场论。

首先，自由能  $F$  是序参量场  $m(\mathbf{x})$  的泛函  $F[m(\mathbf{x})]$  (或者也可以称作有效哈密顿量)。朗道-金兹堡假设这个泛函可以解析地进行泰勒展开，不过，由于  $m(\mathbf{x})$  可以随空间变化，所以我们要在展开中包含  $m(\mathbf{x})$  的空间梯度项，忽略无关的常数项，这个展开即是

$$F[m(\mathbf{r})] = \int d^d \mathbf{x} \left[ \frac{1}{2} b(t) m^2(\mathbf{x}) + \frac{1}{4} c(t) m^4(\mathbf{x}) + \frac{1}{2} (\nabla m(\mathbf{x}))^2 + \dots \right] \quad (3.85)$$

式中，通过合适地将一个常数系数吸收到  $m(\mathbf{x})$  的定义中，我们已经把  $\frac{1}{2} (\nabla m(\mathbf{x}))^2$  项的系数归一化为 1 了。由于微观相互作用有  $\mathbf{Z}_2$  对称性，自由能作为  $m$  的函数必定会保持这种对称性，因此我们已经要求  $F[m(\mathbf{x})]$  在  $m \rightarrow -m$  的  $\mathbf{Z}_2$  变换下保持不变了，这也就相当于要求  $F[m(\mathbf{x})]$  是  $m$  的偶函数，所以上述展开式中没有  $m$  的奇次项。上述表达式忽略了高于四次的泰勒展开项，因为在临界点附近，自发磁化的  $m$  其实是小量，四次以上的项太小而可以忽略，而且，正如我们马上会看到的，保留到四次项已经足以解释连续相变的临界行为了。

自由能极小的条件现在变成了泛函  $F[m(\mathbf{x})]$  对  $m(\mathbf{x})$  的变分求极小。不难求出

$$\delta F = \int d^d \mathbf{x} \left[ b(t) m(\mathbf{x}) + c(t) m^3(\mathbf{x}) - \nabla^2 m(\mathbf{x}) + \dots \right] \delta m(\mathbf{x}). \quad (3.86)$$

式中我们进行了分部积分，并丢掉了边界项 (假设边界上的变分  $\delta m(\mathbf{x})$  为零)。由  $\delta F = 0$ ，即可以得

$$b(t) m(\mathbf{x}) + c(t) m^3(\mathbf{x}) - \nabla^2 m(\mathbf{x}) = 0. \quad (3.87)$$

上面这个方程当然不止一个解，其中最简单的解是  $m(\mathbf{x})$  取常数  $m$ ，这时候方程变成

$$[b(t) + c(t)m^2]m = 0. \quad (3.88)$$

注意到临界温度以上，即  $t > 0$  时，没有自发磁化，物理上序参量只能取  $m = 0$ ，且这个  $m = 0$  要使  $F[m]$  取极小值，因此朗道假设  $t > 0$  时， $b(t) > 0, c(t) > 0$ ，否则要么  $m = 0$  不是极小，要么上述方程还存在  $m \neq 0$  的解。同样， $t < 0$  时，物理上要求  $m \neq 0$ ，因此上述  $m = 0$  的解必定不是  $F[m]$  的极小值，这就要求  $t < 0$  时， $b(t) < 0$ ，进一步要求上述方程存在  $m \neq 0$  的解，因此必定还有  $t < 0$  时， $c(t) > 0$ 。此时这个非零的  $m$  解为

$$m = \pm m_0 \quad \text{其中 } m_0 = \sqrt{-\frac{b(t)}{c(t)}}, \quad (3.89)$$

$\pm$  号表示这时候向上磁化和向下磁化均有可能，虽然实际的物理系统只会随机取其中一种磁化。 $b(t) > 0$  时，自由能密度  $f(m) = F[m]/V$  的函数图像如图 (3.5) 所示，而  $b(t) < 0$  时，

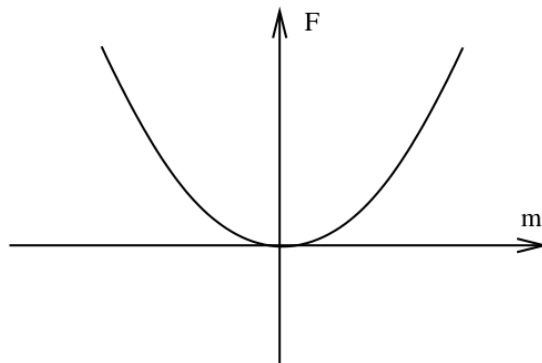


Figure 3.5:  $b(t) > 0$  时，自由能密度  $f(m)$  的函数图像

自由能密度  $f(m)$  的函数图像如图 (3.6) 所示.

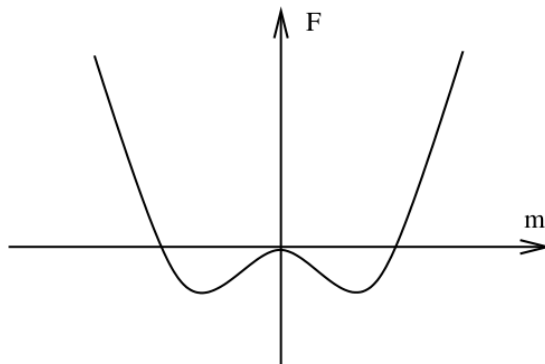


Figure 3.6:  $b(t) < 0$  时，自由能密度  $f(m)$  的函数图像

朗道理论清楚地把序参量由零变成非零的过程与对称性自发破缺联系起来。首先，理论本身具有  $\mathbf{Z}_2$  对称性，因为自由能函数在  $m \rightarrow -m$  的  $\mathbf{Z}_2$  变换下保持不变。其次， $t > 0$  时，自由能函数只有  $m = 0$  这一个极小值，它本身就是  $\mathbf{Z}_2$  不变的。但是， $t < 0$  时， $m = 0$  不再是极小值，而是局部极大值，相反，这时候极小值分裂成了两个，分别在  $\pm m_0$ 。虽然这两个极小值是  $\mathbf{Z}_2$  对称的，但是，只看其中一个的话，每一个都没有  $\mathbf{Z}_2$  对称性，而真实的系统只能选取两个极小值中的一个，但只要选取了一个，那么  $\mathbf{Z}_2$  对称性就自发破缺了。

综合以上分析 (特别要注意  $b(t)$  在  $t > 0$  和  $t < 0$  时要改变正负号而  $c(t)$  则不变号)，朗道假设在临界点附近，参量  $a(t), b(t), c(t)$  可以泰勒展开成

$$\begin{aligned} a(t) &= a_0 + a_1 t + \cdots \\ b(t) &= b_0 t + \cdots \quad \text{其中 } b_0 > 0 \\ c(t) &= c_0 + c_1 t + \cdots \quad \text{其中 } c_0 > 0 \end{aligned} \quad (3.90)$$

由此立即可以得到沿着相界线趋于临界点时磁化强度  $m(t)$  的临界行为

$$m(t) = m_0 \sim (-t)^{1/2}. \quad (3.91)$$

这就是前面所说的，经典平均场论预言  $\beta = 1/2$  的来源，当然这个结果与实验是不吻合的，但是，稍微想想你就明白了，它是具有普适性的。

为了进一步研究经典平均场论的临界指数。我们引入外磁场  $h$ ，由于外磁场和自旋磁矩的耦合，此时自由能密度将成为

$$f(m) = -hm + \frac{1}{2}b(t)m^2 + \frac{1}{4}c(t)m^4 + \cdots \quad (3.92)$$

自由能极小的条件将告诉我们

$$h = [b(t) + c(t)m^2]m. \quad (3.93)$$

当我们从高温趋于临界点时，前面已经看到，磁化强度一直是零，这一点当外磁场很弱时只有很弱的修正，因此可以忽略 (3.93) 式中的  $m^3$  次方项，进而可以计算磁化率  $\chi(t)$  如下

$$h \approx b(t)m \Rightarrow \chi(t) \equiv \frac{\partial m}{\partial h} \sim 1/b(t) \sim t^{-1}. \quad (3.94)$$

也即是说，对于伊辛类，从经典平均场论可以得到  $\gamma = 1$ ，这当然也是不对的。

回到方程 (3.87) 的一般情况。朗道-金兹堡理论最重要的应用之一，是理解系统在空间不同点上涨落之间的关联。具体来说，假设我们在空间某点——比如说  $\mathbf{x} = 0$  的原点处——加上一个无限尖锐的外磁场  $h(\mathbf{x}) = \varepsilon \delta^d(\mathbf{x})$  (式中  $\varepsilon$  为小量， $\delta^d(\mathbf{x})$  为  $d$  维 delta 函数) 来扰动它，以激发起当地的磁化强度，我们想知道这对别处的磁化强度会产生什么影响？加了这样的一个外磁场扰动以后，原来的自由能泛函就要多加上一项，即  $\int d^d \mathbf{x} [-m(\mathbf{x})h(\mathbf{x})] = \int d^d \mathbf{x} [-m(\mathbf{x})\varepsilon \delta^d(\mathbf{x})]$ 。进而变分求自由能极小时对应的方程就是

$$b(t)m(\mathbf{x}) + c(t)m^3(\mathbf{x}) - \nabla^2 m(\mathbf{x}) = \varepsilon \delta^d(\mathbf{x}). \quad (3.95)$$

假设温度低于临界值，即  $t < 0$ ，且扰动前系统原本处在  $m(\mathbf{x}) = m_0 = \sqrt{-\frac{b(t)}{c(t)}}$  的极小值处，外磁场把系统扰动到了

$$m(\mathbf{x}) = m_0 + \delta m(\mathbf{x}), \quad (3.96)$$

式中扰动  $\delta m(\mathbf{x})$  是小量。代入 (3.95) 式，并保留到扰动  $\delta m(\mathbf{x})$  的线性阶，即有

$$\begin{aligned} (b(t) + 3c(t)m_0^2)\delta m(\mathbf{x}) - \nabla^2 \delta m(\mathbf{x}) &= \varepsilon \delta^d(\mathbf{x}) \\ \Rightarrow -2b(t)\delta m(\mathbf{x}) - \nabla^2 \delta m(\mathbf{x}) &= \varepsilon \delta^d(\mathbf{x}). \end{aligned} \quad (3.97)$$

这个方程的解是

$$\delta m(\mathbf{x}) \sim \frac{e^{-|\mathbf{x}|/\xi(t)}}{|\mathbf{x}|^{d-2}} \quad (3.98)$$

这个解其实就是前面的两点函数  $G(\mathbf{x})$ ，从这里很容易读出经典理论的结果  $\eta = 0$ 。式中  $\xi(t)$  为

$$\xi(t) = \sqrt{\frac{1}{-2b(t)}}. \quad (3.99)$$

$\xi(t)$  的含义是，原点处外磁场扰动的效应在离原点远大于  $\xi(t)$  的距离上会迅速衰减为零，因此通常称  $\xi(t)$  为关联长度。

不难看出，沿着相界线趋于临界点时，关联长度的临界行为是

$$\xi(t) \sim (-t)^{-1/2}, \quad (3.100)$$

特别的，这意味着，在临界点附近，关联长度会发散。换言之，在临界点附近，序参量的涨落会在长程上保持关联。正是因为这种长程关联，使得气液相变临界点时对光的散射会增强，从而使得原来透明的气体或者液体，在接近临界点时变得浑浊起来，呈现一片乳白色，这就是气液临界点附近的临界乳光现象。

朗道理论给出来的这些临界指数与实验不吻合。问题出在，我们忽略了在不同的空间局部上，序参量可以有涨落，在算配分函数时，要对所有这些序参量涨落进行泛函积分，而不是仅仅求一个泛函极小值就可以了。尤其在临界点附近，序参量的涨落将变得非常强，而且长程关联，平均场近似此时是失效的。不过，如何系统地考虑序参量的涨落并不是一件简单的事情。这个困难最终是威尔逊通过他的重整化群计算解决的。

### 3.5.2 积去小尺度自由度

重整化群实际上是一个关于如何处理多尺度多自由度系统的纲领，重整化变换的物理实质就是积去小尺度的短程自由度，这样的操作之所以起作用是基于两个基本的观察：第一，小尺度短程自由度对于更大尺度的长程物理是有影响的，第二，但是只要合适地选取表达理论的变量，那么这个影响就可以用几个相关参数来概括，几个参数就可以概括小尺度的所有相关信息，至于短程结构的更多细节信息则与大尺度物理无关。因此，通过积

去短程自由度你就能忽略小尺度结构的无关细节，得到一个与你所关心的尺度的物理密切相关的有效理论。的确，有效场论的有效二字可以理解成无效的反义词，它暗示说如果你的理论与你所要考察的物理不在同一尺度上，那么你的理论就是一种徒劳。下面我们就一个具体的算例来看看通过积去小尺度短程自由度进而得到有效理论这个操作可以如何进行。

考虑到对序参量涨落的泛函积分以后，朗道平均场理论中的自由能泛函就成为一个欧氏量子场论的有效作用量。为此，不妨考察一个定义在  $d$  维欧空间的标量场论，场变量为  $\phi(x)$ ，你可以将它理解为上一小节的磁化强度  $m(\mathbf{x})$ 。根据朗道平均场理论的自由能展开可知，在截断尺度  $a$ ，它的有效作用量（自由能泛函）可以写成（写法和上一小节看起来不同，实际上一样的）

$$S(a) = \int d^d x \left( \frac{1}{2} (\partial_i \phi \partial_i \phi + t a^{-2} \phi^2) + \frac{1}{4!} g a^{d-4} \phi^4 \right), \quad (3.101)$$

式子明显地写出尺度  $a$  相关的因子  $a^{-2}$  和  $a^{d-4}$ ，是为了保证我们只需重标度  $a$  就能实现对理论的重标度，同时这两个因子也确保了相应的项无量纲。式中的  $t$  和  $g$  是无量纲的耦合常数（ $t$  可以理解为约化温度或者说热场），我们就是要研究它们如何依赖于重整化尺度  $a$ 。为此，我们考虑一个更长一点的临近尺度  $\tilde{a} = a(1 + \varepsilon)$ ，即取重标度尺度  $b = 1 + \varepsilon$ ， $\varepsilon$  是一个无穷小量。根据定义，在截断尺度  $\tilde{a}$  上的理论当然应该是

$$S(\tilde{a}) = \int d^d x \left( \frac{1}{2} (\partial_i \phi \partial_i \phi + \tilde{t} \tilde{a}^{-2} \phi^2) + \frac{1}{4!} \tilde{g} \tilde{a}^{d-4} \phi^4 \right). \quad (3.102)$$

重整化的物理实质相当于说， $S(\tilde{a})$  可以通过在  $S(a)$  的基础上积去波长在  $\lambda \in [a, a(1 + \varepsilon)]$  区间之内的自由度而得到。我们不妨以  $\phi$  来表示波长比  $a(1 + \varepsilon)$  更长的自由度，而以  $\eta(x)$  来表示波长在  $\lambda \in [a, a(1 + \varepsilon)]$  之内的自由度，因此  $S(a)$  中的场就应该替换成  $\phi + \eta$ ，这样一来，重整化群的物理实质就相当于说

$$e^{-S(\tilde{a})} = \int [d\eta] e^{-S(a)}. \quad (3.103)$$

根据自由度按波长的这种分解，我们可以将  $S(a)$  重写成

$$\begin{aligned} S(a) = & \int d^d x \left( \frac{1}{2} (\partial_i \phi \partial_i \phi + t a^{-2} \phi^2) + \frac{1}{4!} g a^{d-4} \phi^4 \right) \\ & + \int d^d x \left( \partial_i \eta \partial_i \phi + t a^{-2} \eta \phi + \frac{1}{3!} g a^{d-4} \phi^3 \eta \right) \\ & + \frac{1}{2} \int d^d x \left( \partial_i \eta \partial_i \eta + t a^{-2} \eta^2 + \frac{1}{2} g a^{d-4} \phi^2 \eta^2 \right). \end{aligned} \quad (3.104)$$

这个式子中  $\eta$  的线性项取  $\int d^d x A(x) \eta$  的形式， $A(x)$  是一个关于  $\phi$  的函数， $\phi$  的波长大于  $a(1 + \varepsilon)$ ，而  $\eta(x)$  的波长小于  $a(1 + \varepsilon)$ ，因此  $A(x)$  相对于  $\eta(x)$  来说是变化缓慢的，由于  $\eta(x)$  的相对较快的振荡线性项  $\int d^d x A(x) \eta$  的积分实际上等于零。因此，在进行 (3.103) 的泛函积分计算时，我们实际上要算的就是  $\eta(x)$  的二次项的高斯积分，由熟知的高斯积分公式可知这个积分算出来是

$$\det[-\partial^2 + t a^{-2} + \frac{1}{2} g a^{d-4} \phi^2]^{-\frac{1}{2}}. \quad (3.105)$$

利用  $\det(A) = e^{\text{Tr} \ln(A)}$ , 我们可以得到

$$\begin{aligned}
& \exp \left\{ -\frac{1}{2} \int \frac{d^d x d^d p}{(2\pi)^d} \Big|_{(1-\varepsilon)1/a < p < 1/a} \ln(p^2 + ta^{-2} + \frac{1}{2}ga^{d-4}\phi^2) \right\} \\
& \simeq \exp \left\{ -\varepsilon \frac{1}{2} \frac{S_d}{(2\pi a)^d} \int d^d x \ln(a^{-2} + ta^{-2} + \frac{1}{2}ga^{d-4}\phi^2) \right\} \\
& = \exp \left\{ \varepsilon C - \varepsilon \frac{1}{2} \frac{S_d}{(2\pi a)^d} \int d^d x \ln(1+t + \frac{1}{2}ga^{d-2}\phi^2) \right\} \\
& = \exp \left\{ \varepsilon C' - \varepsilon \frac{1}{2} \frac{S_d}{(2\pi a)^d} \int d^d x \left( \frac{1}{2}g(1-t)a^{d-2}\phi^2 - \frac{1}{8}g^2a^{2d-4}\phi^4 \right) \right\}, \quad (3.106)
\end{aligned}$$

式中  $S_d$  表示  $d$  维欧空间中单位球面的面积,  $C$  和  $C'$  均是与场无关的常数, 可以通过在原来的标量场作用量中加上一个常数来重整化, 因此我们将忽略它。

现在, 方程 (3.103) 告诉我们

$$\begin{aligned}
e^{-S(\bar{a})} & = \exp \left\{ -\int d^d x \left( \frac{1}{2}(\partial_i \phi \partial_i \phi + ta^{-2}\phi^2) + \frac{1}{4!}ga^{d-4}\phi^4 \right) \right. \\
& \quad \left. - \varepsilon \frac{1}{2} \frac{S_d}{(2\pi a)^d} \int d^d x \left( \frac{1}{2}g(1-t)a^{d-2}\phi^2 - \frac{1}{8}g^2a^{2d-4}\phi^4 \right) \right\} \\
& = \exp \left\{ -\int \left( \frac{1}{2}(\partial_i \phi \partial_i \phi + \tilde{t}(a(1+\varepsilon))^{-2}\phi^2) + \frac{1}{4!}\tilde{g}(a(1+\varepsilon))^{d-4}\phi^4 \right) \right\},
\end{aligned}$$

这个式子的最后一行用到了 (3.102)。比较上面这个方程的相应项, 我们可以得到

$$\begin{aligned}
\tilde{t} & = t + \varepsilon \left( 2t + \frac{S_d}{2(2\pi)^d}g(1-t) \right) \\
\tilde{g} & = g + \varepsilon \left( (4-d)g - \frac{3S_d}{2(2\pi)^d}g^2 \right). \quad (3.107)
\end{aligned}$$

也即是说, 我们有重整化群方程

$$\begin{aligned}
\frac{dt}{d\varepsilon} & = \left( 2t + \frac{S_d}{2(2\pi)^d}g(1-t) \right) \\
\frac{dg}{d\varepsilon} & = \left( (4-d)g - \frac{3S_d}{2(2\pi)^d}g^2 \right). \quad (3.108)
\end{aligned}$$

如果  $d = 4$ , 即假如我们研究的是一个四维量子场论, 那么重整化群方程 (3.108) 就是

$$\begin{aligned}
\frac{dt}{d\varepsilon} & = \left( 2 - \frac{1}{16\pi^2}g \right) t + \frac{1}{16\pi^2}g \\
\frac{dg}{d\varepsilon} & = -\frac{3}{16\pi^2}g^2. \quad (3.109)
\end{aligned}$$

这个方程只有平凡的固定点  $t = 0, g = 0$ , 在固定点附近  $\phi^2$  项扰动是相关扰动,  $\phi^4$  扰动则是无关扰动。

如果  $d < 4$ , 当然重整化群方程 (3.108) 依然有平凡的固定点  $t = 0, g = 0$ , 在这个固定点附近  $\phi^2$  扰动和  $\phi^4$  扰动均是相关扰动。但是除了这个平凡的固定点之外, 这时候方程 (3.108) 还有一个非平凡的固定点, 即  $g = \frac{2(2\pi)^d}{3S_d}(4-d)$ ,  $t = -\frac{4-d}{2+d}$ , 在这个固定点附近重整

化群可以线性化为

$$\begin{aligned}\frac{d\delta t}{d\varepsilon} &= \frac{2+d}{3}\delta t + \frac{3S_d}{(2+d)(2\pi)^d}\delta g \\ \frac{d\delta g}{d\varepsilon} &= -(4-d)\delta g,\end{aligned}\quad (3.110)$$

很显然，这时候  $\phi^2$  扰动依然是相关的，但是  $\phi^4$  扰动已经变成无关扰动了。如果一个物理系统的重整化群流从这个非平凡固定点附近经过，那么由线性化的重整化群方程 (3.110) 人们容易知道，这个系统的关联长度  $\xi$  的临界指数  $\nu$  为 ( $\lambda_1 = \frac{2+d}{3}$ )

$$\nu = \frac{3}{2+d}, \quad \alpha = 2 - \frac{3d}{2+d} \quad (3.111)$$

代入  $d=3$ ，得  $\nu=0.6$ ，与测量值 0.63 吻合得还可以，值少比平均场理论的值  $1/2=0.5$  好多了，同样，由  $d=3$  得  $\alpha=0.2$ ，与测量值 0.11 比虽然有所偏离，但不算离谱，要知道平均场理论甚至只能算出一个比热跳变。关于临界指数更系统的计算，请参阅下面这篇写得非常漂亮的经典论文，Kenneth G. Wilson, John B. Kogut, The renormalization group and the  $\varepsilon$  expansion, Physics Reports, DOI: 10.1016/0370-1573(74)90023-4

### 3.5.3 从重整化群的物理实质到理解生命现象

首先，让我们重复一下前面的这段话：重整化群实际上是一个关于如何处理多尺度多自由度系统的纲领，重整化变换的物理实质就是积去小尺度的短程自由度，这样的操作之所以起作用是基于两个基本的观察：第一，小尺度短程自由度对于更大尺度的长程物理是有影响的，第二，但是只要合适地选取表达理论的变量，那么这个影响就可以用几个相关参数来概括，几个参数就可以概括小尺度的所有相关信息，至于短程结构的更多细节信息则与大尺度物理无关。因此，通过积去短程自由度你就能忽略小尺度结构的无细节，得到一个与你所关心的尺度的物理密切相关的有效理论。

重整化群这种处理多尺度多自由度系统的纲领已经被证明反映了凝聚态物理系统和量子场论系统的本质。然而，生命现象也是从无生命的原子分子中呈展出来的，任何有生命的系统都是一个多尺度多自由度系统。但是这种多尺度多自由度系统和传统凝聚态物理或量子场论中研究的系统有所不同，其中一个关键的不同是，凝聚态系统和量子场论系统的大尺度物理都是由小尺度结构决定的，重整化群流只能从小尺度流向大尺度，然而生命系统的典型特征之一是，大尺度可以反过来影响和调节小尺度的结构，也即是说，在生命系统这样的系统中，多尺度之间的相互沟通更为复杂和丰富，因此重整化群的纲领很可能不能完全适用，这里很可能还需要全新的想法！

与此相关的另一个重要不同是，一般来说在重整化群流之下，系统的熵一定是增加的，因为重整化操作本身就意味着不断舍弃小尺度的无关信息。然而，生命系统与此相当不同，当然，在生命系统中从小尺度到大尺度也一定有大量信息是无关而应该被舍弃的，不同的是，在生命系统中许多真正有意思的信息并不是原来就存在于小尺度结构之中，反而正是来源于宏观大尺度的输入，因此，生命系统的从小尺度到大尺度不完全是一个提取相关信息舍弃无关信息的过程（正是这一点和重整化群有本质区别），它同时还需要相关信息（负

熵) 的输入 (也即是, 生命系统总体上要保持向环境排出更多熵而吸收更少熵的状态)。也许, 正是这个相关信息 (负熵) 的输入使得生命系统的大尺度结构可以反过来调节小尺度结构。归纳来说, 在生命系统中存在两个不同的相关信息流向, 一个是小尺度的相关信息随着重整化群流流向大尺度的结构, 然而, 在生命系统中同时还存在一个负熵流, 或者说一个反向的相关信息流, 它是从大尺度流向小尺度的。生命现象正是这两种相关信息流向共同作用的结果, 因此要真正理解生命现象我们也许就必须发展一个处理多尺度多自由度系统的新框架, 重整化群和负熵流也许必须同时被包括进这样的框架之中。

通过相关信息在多尺度之间的振荡, 生命系统将体现出一种全息性质, 也就是生命系统每一个小尺度局部都将在一定程度上编码系统整体的信息。这是因为, 假设原来系统有两个相距遥远的小尺度局部, 它们的信息是不相关联的, 然而, 随着小尺度信息跟着重整化群流流向大尺度, 这两个局部的信息就可以在大尺度上关联起来, 然后它们再被负熵流压缩回小尺度, 因此, 最后每一个小尺度局部都将含有最初相距遥远的两个局部的信息, 从而体现出一种全息的性质, 或者也可以说是某种非局域性。

在最后这一节, 我们花了一点篇幅来讨论重整化群的物理实质和它的可能推广。目的是为了说明, 生命系统同样是一个多尺度多自由度物理系统, 然而这个系统和我们以前用重整化群来处理的那些系统都有所不同, 人们也许需要一个新的框架才能处理生命系统这样不可思议的多尺度多自由度系统。

## 参考文献

1. 开创性双论文: 建立完整的动量空间重整化群框架, 引入粗粒化与重整化的系统方法, 解决临界现象中的多尺度耦合问题。

Kenneth G. Wilson, Renormalization Group and Critical Phenomena. I. Renormalization Group and the Kadanoff Scaling Picture, Phys. Rev. B 4, 3174, DOI: <https://doi.org/10.1103/PhysRevB.4.3174>

Kenneth G. Wilson, Renormalization Group and Critical Phenomena. II. Phase-Space Cell Analysis of Critical Behavior, Phys. Rev. B 4, 3184-3205, DOI: <https://doi.org/10.1103/PhysRevB.4.3184>

2. 引入  $\epsilon$  展开 ( $\epsilon = 4 - d$ ) 技术, 开创了用微扰方法计算临界指数的先河, 是 RG 从概念走向定量计算的关键。

Kenneth G. Wilson, Michael E. Fisher, Critical Exponents in 3.99 Dimensions, Phys. Rev. Lett. , DOI: <https://doi.org/10.1103/PhysRevLett.28.240>

3. 这篇长篇综述 (125 页) 是 RG 理论的圣经级文献, 系统总结了动量空间 RG、 $\epsilon$  展开、固定点理论, 是深入理解 Wilson RG 的必读原文。

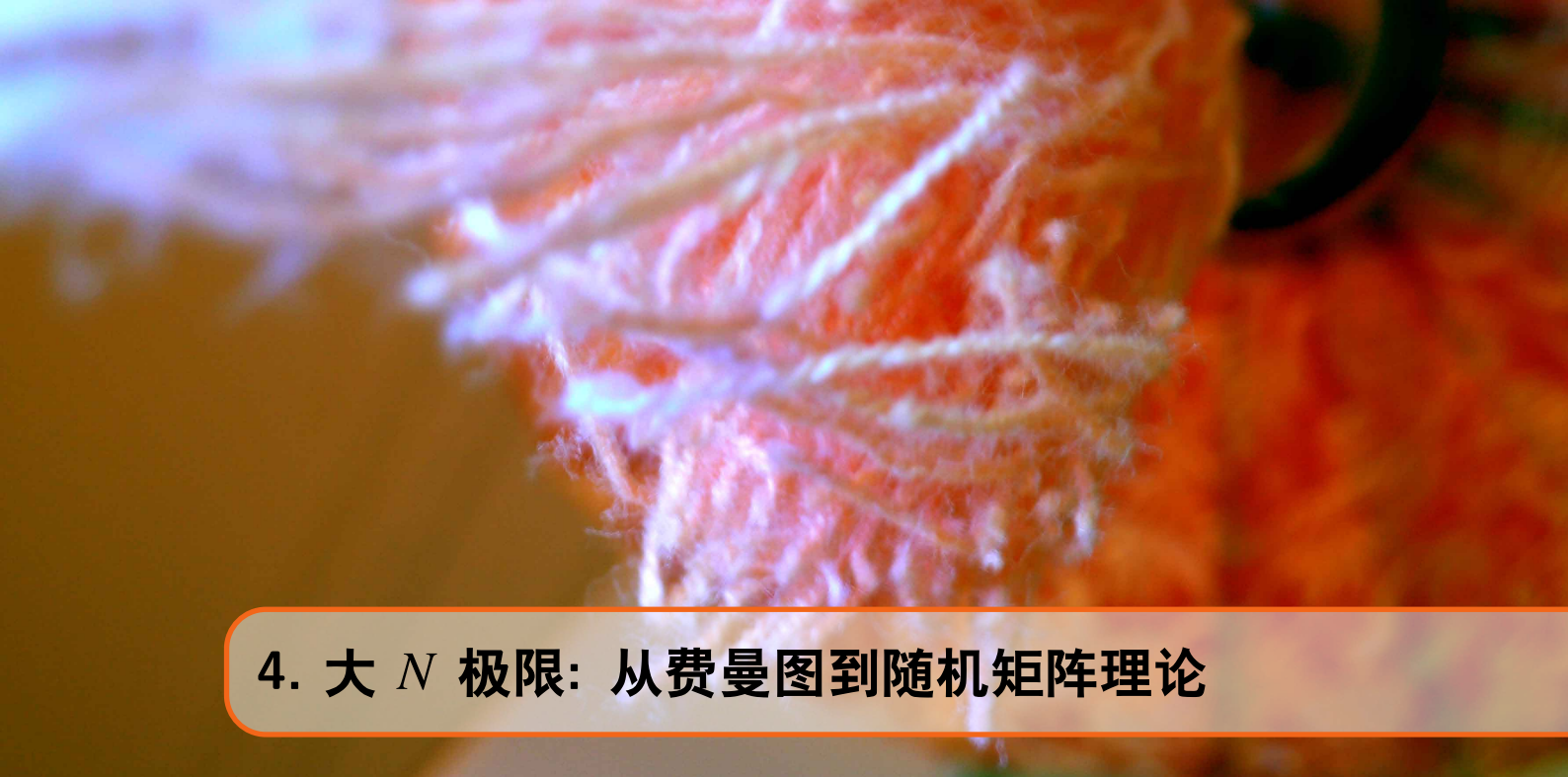
Kenneth G. Wilson, John B. Kogut, The renormalization group and the  $\epsilon$  expansion, Physics Reports, DOI: [10.1016/0370-1573\(74\)90023-4](https://doi.org/10.1016/0370-1573(74)90023-4)

4. Wilson 的诺贝尔演讲综述, 回顾 RG 思想的发展历程。

Wilson (1983), The Renormalization Group and Critical Phenomena, Rev. Mod. Phys. 55, 583-600

5. 深刻影响了本文写作的一篇回顾

Michael E. Fisher, Renormalization group theory: Its basis and formulation in statistical physics, Rev. Mod. Phys. 70, 653 –Published 1 April, 1998



## 4. 大 $N$ 极限：从费曼图到随机矩阵理论

### 4.1 引言

在研究复杂的物理或数学系统时，我们经常会遇到一个令人惊叹的现象：无论微观层面的个体行为多么千差万别，当系统的自由度（比如粒子数  $N$ ）变得极大时，宏观层面往往会涌现出极其简洁且普适的规律。本章正是围绕着这一主题展开，试图为大家搭建一座连接概率论、量子场论图解技术与随机矩阵理论的桥梁。

本章的内容安排遵循由浅入深、从单一到多维的逻辑脉络：首先，我们将从概率论中最核心的中心极限定理出发。为了获得直观的物理图像，我们会通过“高尔顿钉板”和“醉汉随机游走”这两个经典模型，手把手地推导出独立随机变量之和为何必然趋向于高斯分布，并借此引入“特征函数”与“累积量（集团阶矩）”这两个贯穿全讲义的基础数学工具。其次，我们将视线转向对高斯分布的扰动。当概率分布不再是纯粹的高斯型时，传统的直接积分往往会变得异常困难。此时，我们将借用物理学中极为强大的武器——费曼图。大家将会看到，如何把抽象的微扰展开转化为直观的“画图游戏”，以及如何通过判断图的“连通性”来直接提取累积量，并进一步理解“圈展开”与统计力学中温度参数的对应关系。最后，我们将把前两部分发展出的图解思想推广到高维系统，进入大  $N$  展开与随机矩阵理论的殿堂。我们会发现，对于  $N \times N$  随机厄米矩阵，大  $N$  极限在图解语言中惊人地等价于“平面图极限”。借助这一洞察，我们将极其优雅地推导出著名的魏格纳半圆定律，并通过“戴森气体”这一绝妙的物理类比，掌握求解一般势场下本征值分布的普适方法。希望本章能够帮助大家在不需要翻阅厚重的量子场论教材的前提下，直击这些优美数学物理方法的核心精髓。

## 4.2 中心极限定理与随机游走

中心极限定理可以说是概率论中最为重要的定理之一。它解决的是  $N$  个独立随机变量的和在  $N$  很大时的概率分布问题。中心极限定理说, 不管什么随机变量, 在满足一些合理的条件以后这样的概率分布一定会趋向于一个高斯分布。这样一个普适性的结果是非常令人吃惊的, 它首次揭示了系统在大  $N$  极限下可能会展现出普遍性的规律。这在整个数学和物理学上都影响深远。我们这里主要讨论这个重要结果的数学证明。为简单起见, 这里我们将仅仅处理  $N$  个独立同分布随机变量的情形。

### 4.2.1 高尔顿钉板实验

英国生物统计学家高尔顿是达尔文的表弟, 高尔顿钉板实验是他设计来直观演示中心极限定理的。

高尔顿在木板上钉了很多钉子, 这些钉子分成了  $N$  层, 每一层的相邻两颗钉子之间间距相等, 上一层钉子的水平位置都恰好落在下一层的相邻两颗钉子正中间。在实验时, 从上方放入一个小圆球, 小圆球的直径小于任意两颗钉子之间的几何距离, 但是大于相邻两层钉子之间的水平距离。因此, 小圆球在下落的过程中总会碰到钉子, 它碰到钉子以后可以往左边滚下来, 也可以往右边滚下来, 概率均为  $1/2$ 。然后再碰到下一层的钉子, 又以  $1/2$  的概率往左滚或往右滚, 如此不断进行下去, 直到经过  $N$  层钉子以后, 小圆球最终会落到下方的格子中。

在第  $i$  层, 小圆球往左滚还是往右滚是随机的, 往右滚则它的水平位移增加 1, 往左滚则减少 1, 我们可以定义小圆球在第  $i$  层的水平位移为随机变量  $X_i$ , 小圆球往右滚  $X_i = 1$ , 往左滚则  $X_i = -1$ 。很显然, 经过  $N$  层钉子以后, 小圆球最终会落在哪个方格子中由总水平位移  $X = X_1 + X_2 + \dots + X_N$  决定。  $X$  当然也是一个随机变量, 它的分布情况可以通过将实验重复很多很多次, 然后统计下方各格子中小圆球的数目, 由这个数目的分布情况反映出来。

网上有这个实验的视频, 从视频中可以清楚地看到, 最终小圆球的数目分布趋向于一个正态分布 (当  $N$  很大时), 因此也就是说, 随机变量  $X$  最终趋向于一个正态分布。这就是中心极限定理的一个最直观例子。下面我们开始讨论中心极限定理的证明。

### 4.2.2 随机变量的特征

在证明中心极限定理之前, 我们先简单地讨论一个有用的数学概念, 即一个随机变量的特征。假设有一个随机变量  $X$ , 其概率分布为  $p(x)$ , 则  $X$  的特征  $\hat{p}(k)$  定义为

$$\hat{p}(k) = \langle e^{ikX} \rangle = \int dx e^{ikx} p(x). \quad (4.1)$$

很显然,  $\hat{p}(0) = 1$ 。更一般的,  $\hat{p}(k)$  在 0 处关于  $ik$  泰勒展开的  $n$  阶项系数就是随机变量  $X$  的  $n$  阶矩  $M_n$ ,  $M_n = \langle X^n \rangle = \int dx x^n p(x)$ , 即

$$\hat{p}(k) = 1 + \sum_{n=1}^{+\infty} \frac{1}{n!} (ik)^n M_n. \quad (4.2)$$

一个更有用的概念是所谓的累积量 (cumulants), 我们这里也称之为集团  $n$  阶矩  $K_n$ , 它由下式给出

$$\hat{p}(k) = \exp\left(\sum_{n=1}^{+\infty} \frac{1}{n!} (ik)^n K_n\right), \quad (4.3)$$

人们也常常将  $K_n$  记为  $K_n = \langle X^n \rangle_c$ .

由  $n$  阶矩和集团  $n$  阶矩的两个定义式子可以得出

$$\begin{aligned} M_1 &= K_1, \\ M_2 &= K_2 + K_1^2, \\ M_3 &= K_3 + 3K_1K_2 + K_1^3, \\ M_4 &= K_4 + 4K_3K_1 + 3K_2^2 + 6K_2K_1^2 + K_1^4 \dots \end{aligned} \quad (4.4)$$

实际上, 如果我们将  $K_m$  表示成一个  $m$  个点的集团, 然后将任意  $n$  个点完全分解成一些这样的集团, 那么  $M_n$  就可以看成是  $n$  个点所有可能集团分解的和, 如图 (4.1) 所示。很

$$\begin{aligned} \langle x \rangle &= \bullet \\ \langle x^2 \rangle &= \begin{array}{c} \bullet \\ \bullet \end{array} + \bullet \bullet \\ \langle x^3 \rangle &= \begin{array}{c} \bullet \\ \bullet \\ \bullet \end{array} + 3 \begin{array}{c} \bullet \\ \bullet \end{array} \bullet + \bullet \bullet \bullet \\ \langle x^4 \rangle &= \begin{array}{c} \bullet \\ \bullet \\ \bullet \\ \bullet \end{array} + 4 \begin{array}{c} \bullet \\ \bullet \end{array} \bullet \bullet + 3 \begin{array}{c} \bullet \\ \bullet \end{array} \begin{array}{c} \bullet \\ \bullet \end{array} + 6 \begin{array}{c} \bullet \\ \bullet \end{array} \bullet \bullet + \bullet \bullet \bullet \bullet \end{aligned}$$

Figure 4.1: 图中每一个未被圈起来的单独黑点自身是一个单点集团, 每一个被虚线圈起来的多点是一个多点集团。

明显, (4.1) 中的图形正好对应等式 (4.4)。其实,  $K_1 = M_1 = \langle X \rangle$  就是通常所说的平均值, 我们也记为  $\mu$ ,  $K_2 = M_2 - M_1^2 = \langle X^2 \rangle - \langle X \rangle^2$  就是我们通常所说的方差  $\sigma^2$ 。因此, 根据累积量, 或者说集团  $n$  阶矩的定义, 我们有

$$\hat{p}(k) = \exp\left(ik\mu - \frac{1}{2}k^2\sigma^2 + \dots\right). \quad (4.5)$$

### 4.2.3 中心极限定理的证明与随机游走

假设有  $N$  个独立的随机变量  $X_i, i = 1, 2, \dots, N$ , 每一个的概率分布都是同一个函数  $p(x_i)$ , 其平均值为  $\mu$ , 方差为  $\sigma^2$ 。这里我们以小写的  $x_i$  来表示随机变量  $X_i$  的取值。则随机变量  $X = \sum_i X_i$  的概率分布  $q(x)$  可以表示成,

$$q(x) = \int \left[ \prod_i dx_i \right] \delta\left(x - \sum_i x_i\right) p(x_1)p(x_2)\dots p(x_N). \quad (4.6)$$

利用  $\delta(x) = \int \frac{dk}{2\pi} e^{-ikx}$ , 我们就有

$$q(x) = \int \frac{dk}{2\pi} e^{-ikx} \int \left[ \prod_i dx_i \right] e^{ik \sum_i x_i} p(x_1) p(x_2) \dots p(x_N) \quad (4.7)$$

$$= \int \frac{dk}{2\pi} e^{-ikx} \left[ \int dx' e^{ikx'} p(x') \right]^N \quad (4.8)$$

$$= \int \frac{dk}{2\pi} e^{-ikx} \widehat{p}^N(k), \quad (4.9)$$

式中  $\widehat{p}(k) = \int dx' e^{ikx'} p(x')$ 。利用 (4.5) 式我们容易将上面的结果重写成,

$$q(x) = \int \frac{dk}{2\pi} e^{-ik(x-N\mu) - N\frac{1}{2}\sigma^2 k^2 + NO(k^3)}. \quad (4.10)$$

假设我们将  $Nk$  看成一个新变量, 则容易看出在  $N \rightarrow \infty$  的时候式中的  $O(k^3)$  可以忽略。由此就可以得到

$$q(x) \rightarrow \frac{1}{\sqrt{2\pi N\sigma}} \exp\left(-\frac{1}{2} \frac{(x-N\mu)^2}{N\sigma^2}\right). \quad (4.11)$$

这就证明了中心极限定理。从最终结果我们可以看到随机变量  $X$  的平均值为  $N\mu$ , 而方差为  $N\sigma^2$ 。

下面我们给出中心极限定理的一个简单解释。假设我们考虑一个醉汉在  $x$  轴上随机游走, 我们将上文的随机变量  $X_i$  看成是醉汉第  $i$  步这一步所产生的位移, 由于这是一个醉汉, 因此这个位移是随机变量。 $\mu$  就是醉汉平均每一步前进的距离。那么随机变量  $X$  就是醉汉在  $N$  步中的总位移, 中心极限定理告诉我们, 在  $N$  很大时, 这个总位移趋向于一个高斯分布, 这个高斯分布的方差是  $N\sigma^2$ 。

关于方差的这个结果其实可以用下面的方法容易地得到, 下面为了简单起见我们不妨假设醉汉醉得很厉害, 以致于  $\mu = 0$ 。我们记醉汉在  $i$  步中的总位移为  $X(i)$ , 显然

$$X^2(i+1) = (X(i) + X_{i+1})^2 = X^2(i) + X_{i+1}^2 + 2X(i)X_{i+1}. \quad (4.12)$$

然后将上面这个方程两边取期望值, 由于醉汉是随机游走, 因此  $X_{i+1}$  与  $X(i)$  统计无关, 因此  $\langle X(i)X_{i+1} \rangle = \langle X(i) \rangle \langle X_{i+1} \rangle = 0$ , 因此

$$\langle X^2(i+1) \rangle = \langle X^2(i) \rangle + \langle X_{i+1}^2 \rangle = \langle X^2(i) \rangle + \sigma^2. \quad (4.13)$$

将这个式子进行递推就可以得到我们需要的结论。当然, 从这个推导可以看出, 关于方差的这个结果不需要大  $N$  极限, 这一点也可以从我们在上文中那个更加数学化的证明方法中看出来。实际上, 只有最终的分布是一个高斯分布这一点才是依赖于大  $N$  极限的。

### 4.3 扰动高斯分布与费曼图

我们将考察两种很有用的随机变量, 第一种是满足高斯分布的随机变量, 第二种随机变量的分布函数是对高斯分布的一个扰动。我们关心的要点是, 在这两种情况下我们如何计算  $n$  阶矩以及累积量 (cumulants), 也就是我们所谓的集团  $n$  阶矩。我们将从高斯分布

开始，但是我们更关心的其实是扰动高斯分布，这时候我们将发展一种画图的办法来计算  $n$  阶矩以及集团  $n$  阶矩的微扰展开。这种图就是微扰统计场论中著名的费曼图。当然，我们并不会讨论微扰统计场论，但值得说明的是，微扰统计场论其实就是我们考虑的扰动高斯分布的推广。只不过在统计场论中，我们需要考虑无穷多个随机变量，它们形成所谓的场算符，而我们这里仅仅只考虑一个随机变量。总之，微扰统计场论其实是我们将要考察的扰动高斯分布的无穷维推广。因此，这里发展的费曼图方法同样适用于微扰统计场论是毫不令人意外的。

### 4.3.1 高斯分布

首先考虑高斯分布。假设随机变量  $\Phi$  满足如下分布

$$P(\phi) = \frac{1}{\mathcal{Z}} \exp\left(-\frac{1}{2}m^2\phi^2\right), \quad (4.14)$$

式中  $\mathcal{Z}$  为归一化系数。这就是高斯分布。对于这种分布我们很容易计算随机变量的特征，现在记为  $\mathcal{Z}[J]$

$$\mathcal{Z}[J] = \frac{1}{\mathcal{Z}} \int d\phi \exp\left(-\frac{1}{2}m^2\phi^2 + iJ\phi\right) \quad (4.15)$$

$$= \exp\left(-\frac{1}{2}\frac{J^2}{m^2}\right). \quad (4.16)$$

从这个表达式可以读出在所有的集团  $n$  阶矩中，只有集团 2 阶矩  $K_2$  非零，很显然  $K_2 = \frac{1}{m^2}$ 。

计算  $n$  阶矩的办法有两种，一种是将特征函数  $\mathcal{Z}[J]$  在  $J=0$  处泰勒展开到  $n$  阶，然后读出其展开系数即为  $n$  阶矩。很显然  $n$  为奇数时结果为零。而对于偶数情形，比方说，对于  $2n$  阶矩  $M_{2n} = \langle \Phi^{2n} \rangle$ ，我们用这种办法可以得到

$$M_{2n} = \frac{(2n)!}{n!2^n} \frac{1}{m^{2n}}. \quad (4.17)$$

第二种计算  $n$  阶矩的办法是利用  $n$  阶矩的集团展开，由于非零的集团只有 2 集团，因此这就是将  $n$  阶矩按照 2 集团进行展开。具体的办法可以进行如下，画出  $n$  个完全等同的点，每个点代表一个  $\Phi$ ，然后将这些点两两配对组合成 2 集团，每个 2 集团中的两个点用一根线连起来，这样就构成一个不完全连通的  $n$  点图。每一个 2 集团的贡献是  $K_2 = 1/m^2$ ，将一幅集团展开图中所有集团的贡献乘起来就是这幅图对  $n$  阶矩的贡献。而将所有可能的集团展开图的贡献加起来就得到总的  $n$  阶矩。

很显然，按照这个办法，当  $n$  为奇数时，任何集团展开中必然会涉及到 1 集团，而由于  $K_1 = 0$ ，所以 1 集团的贡献为零，因此最后总的  $n$  阶矩必然为零。对于  $2n$  阶矩，集团展开图有很多，每个图都包含  $n$  个 2 集团，但是由于这  $2n$  个点完全等同，所有这些图的贡献都是一样的，都为  $1/m^{2n}$ 。那么一共有多少个集团展开图呢？这有多种不同的计算方法，最简单的一种是这样的：首先，随意标记这  $2n$  个点中的任何一个，与之配对形成 2 集团的点有  $(2n-1)$  个不同选择，配对完成以后，再次在剩下的  $2n-2$  个点中随意标记一个，与之配对的点就有  $(2n-3)$  种可能，如此不断进行下去，直到所有点都完成了配对。很显然，可能的集团展开图数目为

$$(2n-1)!! \quad (4.18)$$

很容易看出来  $(2n-1)!! = (2n)!/(n!2^n)$ , 因此这种办法计算出来的最终结果与之前的第一种办法计算的结果完全吻合。

### 4.3.2 受扰动的高斯分布

下面我们假定随机变量  $\Phi$  的分布为

$$P(\phi) = \frac{1}{\mathcal{Z}[g]} \exp\left(-\frac{1}{2}m^2\phi^2 - \frac{1}{4!}g\phi^4\right), \quad (4.19)$$

$\mathcal{Z}[g]$  来自于这个概率分布的归一化, 很显然, 这种概率分布的归一化计算起来很困难。不过, 我们只是在形式上使用这一表达式, 实际上  $g$  应该理解为一个很小的参数,  $\exp(-\frac{1}{4!}g\phi^4)$  其实是对高斯分布的一个小扰动。

这一概率分布的特征函数, 现在记为  $\mathcal{Z}[g, J]$ , 由下式给出

$$\mathcal{Z}[g, J] = \frac{1}{\mathcal{Z}[g]} \int d\phi \exp\left(-\frac{1}{2}m^2\phi^2 - \frac{1}{4!}g\phi^4 + iJ\phi\right). \quad (4.20)$$

如果不把  $g$  处理成微扰,  $\mathcal{Z}[g, J]$  是很难计算的, 但它关于  $J$  的泰勒展开同样是  $n$  阶矩, 并且  $\log(\mathcal{Z}[g, J])$  关于  $J$  的泰勒展开同样是集团  $n$  阶矩。

我们不妨来考虑  $2n$  阶矩  $M_{2n} = \langle \Phi^{2n} \rangle$  (根据被积函数的奇偶性,  $2n+1$  阶矩显然为零),

$$\begin{aligned} M_{2n} &= \frac{1}{\mathcal{Z}[g]} \int d\phi \phi^{2n} \exp\left(-\frac{1}{2}m^2\phi^2 - \frac{1}{4!}g\phi^4\right) \\ &= \frac{1}{\mathcal{Z}[g]} \int d\phi \phi^{2n} \left[1 - \frac{1}{4!}g\phi^4 + \frac{1}{2}\left(-\frac{1}{4!}g\phi^4\right)^2 + \dots\right] \exp\left(-\frac{1}{2}m^2\phi^2\right). \end{aligned} \quad (4.21)$$

显然计算的结果必然是一个关于  $g$  的泰勒级数, 级数中每一项的计算都化简成了标准高斯分布情形的计算。因此前文所描述的 2 集团展开的办法依然完全适用, 但这样虽然能够计算出结果, 但相应的集团展开依然是高斯分布的集团展开, 而不是 (4.19) 的集团展开, 因此我们无法直接从中得到扰动高斯分布的集团多阶矩。

注意到在我们现在考察的扰动高斯分布 (4.19) 中,  $-\frac{1}{4!}g\phi^4$  是指数上的同一项, 它的四个  $\phi$  肯定不应该被拆开, 尤其是不能拆开到不同的集团中去。因此, 现在我们引入一种新的办法来画集团展开图。在现在的画法中, 除了  $\phi^{2n}$  所对应的  $2n$  个点 (这些点我们称之为外点) 以外, 对于泰勒展开 (4.21) 中的每一个  $-\frac{1}{4!}g\phi^4$  因子, 我们都在图中引入一个 (而不是四个) 不同的点, 我们称之为顶角, 因此每个顶角不是代表一个  $\Phi$ , 而是代表 4 个  $\Phi$  乘在一起, 从上一节的研究中我们知道, 每个  $\Phi$  都要和其它的  $\Phi$  配对, 所以我们在每个顶角上画出四条腿, 每条腿代表一个待配对的  $\Phi$ 。对于 (4.21) 中  $g$  的泰勒级数的每一阶, 当我们将所有的配对 (同样, 每个配对画一条连线) 都完成以后就生成了一幅不完全连通的图, 一组配对方式就对应一幅图, 这样的图称之为费曼图。和前文一样, 每一幅费曼图中的每一根代表配对的连线贡献都是  $1/m^2$ , 而每一个顶角额外还有一个  $-g$ 。将一幅图中所有连线和顶角的贡献都乘起来就是这幅费曼图对  $2n$  阶矩的贡献。将所有不同费曼图的贡献都加起来就得到  $2n$  阶矩的最终值。

读者可能注意到我没有提到顶角上的那个  $\frac{1}{4!}$  因子, 这是因为, 当我们将顶角上的四个  $\Phi$  和其它  $\Phi$  进行配对时, 一般来说这四个  $\Phi$  的排列组合会产生  $4!$  种可能性, 刚好将

$\frac{1}{4!}$  抵消了。除非我们画出来的费曼图对于某个顶角上的  $\Phi$  的配对有对称性，这时候我们就要额外除以对称操作的数目（恒等操作也算一个对称操作），因为对称的配对方式结果是完全一样的。同样的，上面我们也没有谈泰勒展开  $g^2$  阶项前面的  $1/2!$  因子，这是因为  $g^2$  阶其实是两个顶角的乘积，一般来说，当将其它  $\Phi$  与  $g^2$  阶上的  $\Phi$  进行配对时，两个顶角就产生了两种可能性，正好将  $1/2!$  消去。除非我们画出来的费曼图这两个顶角是完全对称的，这时候就要额外除以对称操作的数目 2。总之，每一幅费曼图的贡献还要除以一个对称因子，它就是使得费曼图保持不变的对称操作的数目。

另外，很显然，不同拓扑结构的费曼图有不同的配对方式，而由于  $\Phi$  的等同性，相同拓扑结构的费曼图对  $2n$  阶矩的贡献必定是相同的（有相同的连线数和顶角数），因此对于每一拓扑类，我们常常只画一幅图为代表，然后在最终贡献中乘上一个拓扑权因子，它就是具有相同拓扑结构的不同费曼图的个数。

下面的两幅图 (4.2)、(4.3) 是将  $\langle \Phi^4 \rangle$  计算到  $g^2$  阶所涉及到的所有费曼图，每幅图前面的数字是它的拓扑权因子除以对称因子。比如图 (4.2) 中 (c) 图的系数  $3 = C_4^2 \times \frac{1}{2}$ ，其中  $C_4^2$  是从四个外点中选出两个进行直接配对的组合数，而  $1/2$  来自于图中那个圈的左右对称性。比如图 (4.2) 中 (f) 图的系数  $3/2$  是因为将四个外点分成全同的两组（两组外点分别和一个顶角配对）有 3 种可能性，另外图中圈的对称性导致还要除以 2。比如图 (4.2) 中 (i) 图的系数  $3/8 = 3 \times 1/8$ ，其中的  $1/8$  是因为图中的双圈有 8 个对称操作。再比如图 (4.3) 中的 (n) 图的系数  $1 = C_4^2 \times (1/3!)$ ，其中的  $C_4^2$  是四个外点中选出两个直接配对的组合数， $1/3!$  是因为图中那个圈中的 3 根线具有重排对称性。类似的可以得到其它图前面的系数。

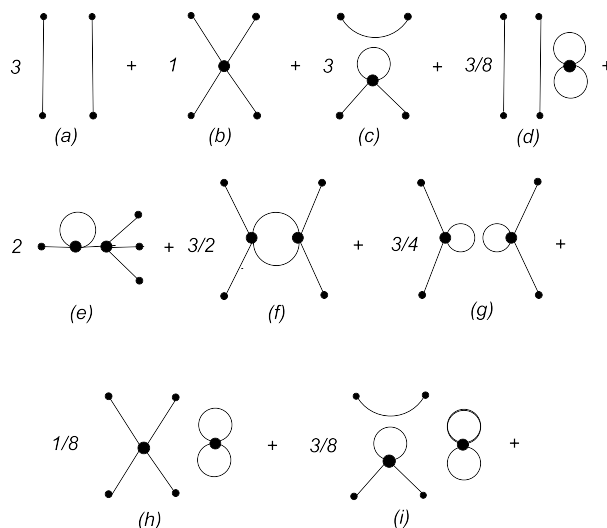
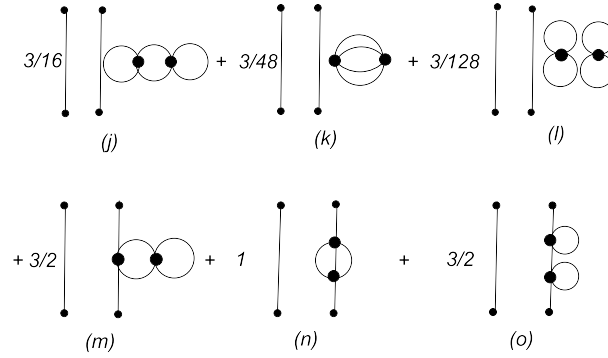


Figure 4.2: The feynman diagrams concern  $\langle \Phi^4 \rangle$ .

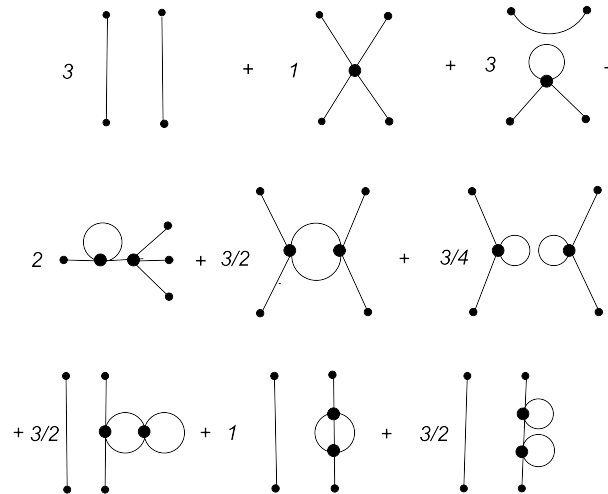
很容易看出来，上面这些费曼图的和可以因式分解成如下两部分，图 (4.4) 和图 (4.5)，的乘积。人们可能注意到，这个乘积中还多出来一些  $g^3$  阶甚至  $g^4$  阶的部分，这些部分是我们暂时不关心的。图 (4.5) 中的第一项那个  $I$  表示单位 1。另外，图 (4.5) 有一个特征，那就是它完全不涉及外点，我们称之为真空图。这些真空图实际上对最终结果是没有贡献

Figure 4.3: The feynman diagrams concern  $\langle \Phi^4 \rangle$ .

的原因就在于，我们回顾计算公式 (4.21)，就会发现，要最后算出  $\langle \Phi^4 \rangle$ ，我们还需要除以  $\mathcal{Z}[g]$ ，而  $\mathcal{Z}[g]$  来自于归一化，其计算也可以按照微扰展开的形式来进行，只不过这时候没有外点，稍微想一下我们就能明白，图 (4.5) 的这些真空图，就是计算  $\mathcal{Z}[g]$  时所涉及到的那些费曼图。所以在最终计算  $\langle \Phi^4 \rangle$  时，这些真空图的贡献就从分子和分母中消去了。因此，为了计算  $\langle \Phi^4 \rangle$  (到  $g^2$  阶)，我们只需要将图 (4.4) 中的那些费曼图的贡献加起来。最后的结果容易计算得

$$\langle \Phi^4 \rangle = 3 \frac{1}{m^4} - 4g \frac{1}{m^8} + \frac{33}{4} g^2 \frac{1}{m^{12}}. \quad (4.22)$$

这个结果也可以不用费曼图方法，而是用直接的计算来验证 (要记得除以  $\mathcal{Z}[g]$ )。

Figure 4.4: The feynman diagrams for calculating  $\langle \Phi^4 \rangle$ .

费曼图的方法在直接计算如  $\langle \Phi^4 \rangle$  这样的多阶矩时显示出来的方便之处还不够大。但如果我们想计算的是集团多阶矩，比如  $K_4$ ，费曼图方法就会显得特别方便。为了说清楚这一点，让我们观察图 (4.4) 中的那些费曼图，我们发现，其中有些费曼图是连通图，因此四个外点属于同一个集团，而有些图是不连通的，其四个外点被分割成了不同的集团。也就

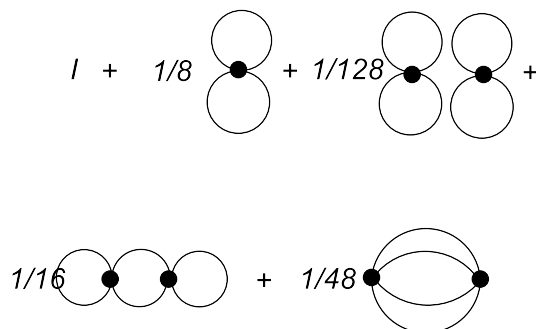


Figure 4.5: The vacuum diagrams.

是说，从图 (4.4) 中我们可以发现，通过顶角和连线的作用可以将四个外点自然地分割成集团展开结构，其中一个集团就对应费曼图中的一个连通部分。又考虑到按照  $g$  的微扰展开，我们就可以说，集团多阶矩对应的就是连通图的贡献之和。比如说  $K_4$ ，计算到  $g^2$  阶，它就是图 (4.2) 中 (b), (e), (f) 三个费曼图的贡献之和，很容易算得

$$K_4 = -g \frac{1}{m^8} + \frac{7}{2} g^2 \frac{1}{m^{12}}. \quad (4.23)$$

总之，对于任何扰动高斯分布，其集团  $n$  阶矩  $K_n$  其实就是  $n$  个外点的连通图之和，因此我们也可以称之为连通  $n$  阶矩，并记作  $\langle \Phi^n \rangle_c$ ，即

$$K_n = \langle \Phi^n \rangle_c. \quad (4.24)$$

### 4.3.3 圈展开

读者可能发现了，上面画出的所有费曼图中，有些费曼图包含了闭合回路，我们称之为圈，而有些则没有。为了看清楚圈的含义我们再次修改概率分布，假设我们记  $S(\phi) = \frac{1}{2} m^2 \phi^2 + \frac{1}{4!} g \phi^4$ ，那么我们将要考虑的概率分布为

$$P(\phi) = \frac{1}{\mathcal{Z}[g, h]} \exp\left(-\frac{1}{h} S(\phi)\right). \quad (4.25)$$

我们同样可以按照对参数  $g$  的微扰展开来计算  $n$  阶矩和连通  $n$  阶矩，稍微想一下人们就能知道，现在所有的费曼图和上一节中的费曼图都是一样的，与我们在上一节中考察的情况的唯一区别在于，现在费曼图中每一根连线的贡献是  $h/m^2$ ，而每一个顶角的贡献是  $-g/h$ 。为了让  $h$  的贡献平等地进入外点，现在假设我们要计算的是  $\langle \Phi^n \rangle_c / h^n$ ，也就是说每一个外点要额外贡献一个  $1/h$ 。假设我们正在考察的费曼图共有  $v$  个点 (包括外点和顶角)， $e$  根连线， $l$  个独立的圈，则很显然这幅费曼图会贡献一个  $h^{e-v}$ ，由连通图的欧拉公式可知，它就是

$$h^{l-1}. \quad (4.26)$$

也即是说，每一幅费曼图的圈的个数其实是它按照  $h$  展开的阶数加 1。

因此, 如果我们假定  $h$  是一个小量, 那么在计算  $\langle \Phi^n \rangle_c / h^n$  的时候, 高圈图的贡献相比于低圈图的贡献就处在  $h$  微扰展开的更高阶上, 概率分布 (4.25) 按照  $h$  进行微扰展开其实就是相应费曼图的圈展开。这个结论对于任何扰动高斯分布都是成立的。如果我们将  $\phi$  看成是某个粒子的坐标,  $S(\phi)$  看成是这个粒子的势能,  $h$  看成是温度, 那么 (4.25) 其实就是这个粒子在热平衡之后的概率分布, 也就是所谓的玻尔兹曼分布。而我们考虑的圈展开其实就是一种低温展开。

#### 4.3.4 复变量

现在, 假设我们考察的是一个复随机变量  $\Phi$ , 其取值为  $\phi = \phi_1 + i\phi_2$ , 其中  $\phi_1, \phi_2$  是两个实数, 相应的高斯分布为

$$P(\phi, \bar{\phi}) = \frac{1}{\mathcal{Z}} \exp(-m^2 \bar{\phi} \phi). \quad (4.27)$$

很显然, 这个概率分布的特征函数可以计算得

$$\langle \exp(iJ\bar{\phi} + i\bar{J}\phi) \rangle = \exp\left(-\frac{\bar{J}J}{m^2}\right). \quad (4.28)$$

由此容易得到  $\langle \Phi\Phi \rangle = \langle \bar{\Phi}\bar{\Phi} \rangle = 0$ ,  $\langle \Phi\bar{\Phi} \rangle = 1/m^2$ 。也就是说,  $\Phi$  仅仅在与其复共轭进行配对的时候才会有非零的贡献。这一点在我们考虑扰动高斯分布画费曼图的时候尤其重要, 这时候为了强调  $\Phi$  只能与其复共轭进行连线我们常常在这个连线上加上一个箭头, 这个箭头从  $\Phi$  指向  $\bar{\Phi}$ 。另外, 概率分布必然是实的, 因此我们不妨假定任何扰动高斯分布的顶角都是  $\bar{\phi}\phi$  的函数, 而不仅仅是  $\phi$  或者  $\bar{\phi}$  的函数。这样一来在画微扰展开的费曼图的时候, 任何一个进入顶角的箭头都必然对应一个离开这个顶角的箭头, 因此, 从一个  $\Phi$  外点出发顺着箭头的方向走必然会终止于某个  $\bar{\Phi}$  外点。如此一来, 我们必然有, 在这样的扰动高斯分布中  $\langle \Phi\Phi \rangle = \langle \bar{\Phi}\bar{\Phi} \rangle = 0$  (画不出任何相应的费曼图)。这个结果也可以直接从对称性来证明, 这是因为, 如果我们的概率分布是  $\bar{\phi}\phi$  的函数, 那么它在  $\phi \rightarrow e^{i\alpha}\phi, \bar{\phi} \rightarrow e^{-i\alpha}\bar{\phi}$  的变换下就是不变的, 因此我们必然有  $e^{-2i\alpha} \langle \Phi\Phi \rangle = \langle \Phi\Phi \rangle$ , 由于  $\alpha$  是任意的, 这就意味着  $\langle \Phi\Phi \rangle = 0$ , 类似的也可以证明  $\langle \bar{\Phi}\bar{\Phi} \rangle = 0$ 。

#### 4.3.5 微扰统计场论

微扰统计场论可以看成是上面我们所考察的这一类问题的推广。只不过这时候我们有无穷多个随机变量, 它们形成某个欧氏空间 (比方说  $\mathbb{R}^1$ ) 上的场分布, 我们可以用  $\Phi(x)$  来表示这些随机变量,  $x$  表示欧氏空间上的坐标。  $\Phi(x)$  的取值我们可以记为  $\phi(x)$ , 它是一个场, 为了简单起见我们不妨假定这些随机变量的取值都为实数。统计场论的关键是给每一个场位形  $\phi(x)$  都指定一个概率  $P[\phi]$ , 它由下面的式子给出,

$$P[\phi] = \frac{1}{\mathcal{Z}} \exp(-S[\phi]), \quad (4.29)$$

式中  $S[\phi]$  的一个典型例子如下

$$S[\phi] = \int dx \left[ \frac{1}{2} ((\partial_x \phi)^2 + m^2 \phi^2) + \frac{1}{4!} g \phi^4 \right]. \quad (4.30)$$

归一化常数  $\mathcal{Z}$  通常被称作配分函数，它的计算方法是

$$\mathcal{Z} = \int \mathcal{D}\phi \exp(-S[\phi]), \quad (4.31)$$

这里的积分表示对所有的场位形进行求和，称作泛函积分。在当前，泛函积分还不是一个有精确定义的数学概念，如何严格地定义泛函积分是最为重要的待解决数学难题之一。不过在形式上，我们可以将泛函积分理解成对每个空间点  $x$  上的场变量都进行积分，是一个无穷重积分，也即是说，在形式上我们可以将积分测度  $\mathcal{D}\phi$  理解成， $\mathcal{D}\phi = \prod_x d\phi(x)$ ，这里的  $d\phi(x)$  不表示对  $x$  求微分，而是表示将  $x$  点的场变量  $\phi(x)$  本身看成是一个积分变量。

由于  $\Phi(x)$  依赖于  $x$ ，所以原来的多阶矩的概念就可以推广成多点关联函数的概念。比方说两点关联函数的定义如下

$$\langle \Phi(x)\Phi(y) \rangle = \frac{1}{\mathcal{Z}} \int \mathcal{D}\phi \phi(x)\phi(y) \exp(-S[\phi]). \quad (4.32)$$

我们可以用类似于前面发展的微扰展开的办法来计算多点关联函数。比方说，对于  $S[\phi]$  由 (4.30) 式给出来的统计场论，我们可以将顶角项  $\int dx \frac{1}{4!} g \phi^4$  看成是一个微扰项。很显然，(4.30) 式去掉这一顶角项剩下的部分同样是某种高斯积分，通过计算这一高斯积分我们就能把两个不同点上的场进行配对。整个费曼图的画法和我们前文的讨论是类似的，不同的是，在计算每一费曼图对多点关联函数的贡献时采用的计算规则略微有些不同，比方说，在现在的情形中每一根代表配对的连线的贡献不再是一个简单的数了，再比方说，每一个顶角的贡献不再是一个简单的  $-g$  了，而是还要包含一个对  $x$  的积分。

实际上，我们前面发展的所有概念都可以平行地照搬到微扰统计场论中去。当然，微扰统计场论也有其自身的复杂性，在微扰统计场论中，费曼图的每一个独立的圈都会额外贡献一个积分，而这个积分往往是发散的，物理学家用重整化的办法来处理这一发散，以从中得出有意义的有限的结果。不过在数学上，这一问题还没有得到严格的处理，实际上，它和我们之前提到的泛函积分的严格数学定义问题密切相关，是有待解决的重大数学难题。当然，我们并不打算详细地讨论微扰统计场论，所以让我们就在这里停下来吧。

#### 4.4 大 N 展开与随机矩阵理论

在中心极限定理的证明过程中我们看到，对于一个有  $N$  个随机变量的系统，它的某些问题在  $N$  很大的极限下会得到极大的简化。这一观察促使人们去研究各种  $N$  个随机变量的系统在大  $N$  极限下的行为，以及这些系统按照  $1/N$  的微扰展开，这就是所谓的大  $N$  展开思想。在物理学中，大  $N$  展开思想已经有好几十年的历史发展脉络，对于今天的人来说，大  $N$  展开最著名的例子也许是 AdS/CFT 对应。不过，在这里我们并不想系统性地回顾大  $N$  展开的这些发展。相反，我们想介绍大  $N$  展开的一个重要例子，即随机矩阵理论。除了因为这一理论本身有广泛的应用以外，我们介绍它是因为，就好像中心极限定理一样，这一理论的大  $N$  极限是已经被了解得很清楚的，是一个漂亮的数学物理定理。

下面我们来介绍随机矩阵理论。这时候，我们要考察的随机变量是一个  $N \times N$  的厄米矩阵  $\Phi$ ，其取值我们依然记为  $\phi$ ， $\phi$  的第  $i$  行第  $j$  列我们记为  $\phi_j^i$ 。我们假定这个随机变量

的概率分布如下

$$P(\phi) = \frac{1}{\mathcal{Z}} \exp(-N\text{Tr}V(\phi)), \quad (4.33)$$

式中  $V(\phi)$  是关于  $\phi$  的一个实多项式, 比如说  $V(\phi) = \frac{1}{2}m^2\phi^2 + g\phi^4$  等等。

我们主要关心  $\phi$  的本征值密度的期望值  $\rho(E)$ , 其定义如下

$$\rho(E) = \frac{1}{N} \langle \text{Tr} \delta(E - \Phi) \rangle = \frac{1}{N} \langle \sum_i \delta(E - E_i) \rangle, \quad (4.34)$$

式中  $E_i$  为某个按照概率 (4.33) 随机产生的  $\phi$  的本征值。  $\rho(E)$  也就是随机产生许许多多  $\phi$  以后, 它们的本征值的分布函数, 它显然满足  $\int dE \rho(E) = 1$ 。当  $N$  有限时,  $\rho(E)$  很复杂很难精确地求解出来, 但是, 正如我们将要看到的, 在  $N \rightarrow +\infty$  的大  $N$  极限下,  $\rho(E)$  将趋向于一个简单的光滑函数。我们下面要做的事情就是找到这个光滑函数, 它就类似于我们在中心极限定理中的高斯分布。

注意到  $\lim_{\epsilon \rightarrow 0} \text{Im} \left( \frac{1}{x+i\epsilon} \right) = -\pi \delta(x)$ , 我们可以引入一个更容易计算的函数  $G(z)$ , 其定义为

$$G(z) = \frac{1}{N} \langle \text{Tr} \frac{1}{z - \Phi} \rangle. \quad (4.35)$$

很显然,  $\rho(E) = -(1/\pi) \lim_{\epsilon \rightarrow 0} \text{Im} G(E + i\epsilon)$ 。另外, 注意到我们的概率分布 (4.33) 在么正变换下是不变的, 即  $P(\phi) = P(U^\dagger \phi U)$ , 这里  $U$  为一个任意的  $N \times N$  的么正矩阵。因此我们必然有

$$G_j^i(z) = \left\langle \left( \frac{1}{z - \Phi} \right)_j^i \right\rangle = \delta_j^i G(z). \quad (4.36)$$

#### 4.4.1 魏格纳半圆定律

下面我们考虑一种简单的情况, 即  $V(\phi) = \frac{1}{2}m^2\phi^2$  的情况。这时候由高斯积分我们容易有

$$\langle \Phi_j^i \Phi_l^k \rangle = \frac{1}{\mathcal{Z}} \int [d\phi] \phi_j^i \phi_l^k \exp \left( -N \frac{1}{2} m^2 \sum_{p,q} \phi_q^p \phi_p^q \right) = \delta_l^i \delta_j^k \frac{1}{Nm^2}. \quad (4.37)$$

这就是高斯分布情况下随机矩阵理论中两个外点进行配对所需满足的规则。由于  $\Phi_j^i$  带有上下两个不同的指标, 因此我们常常用如图 (4.6) 所示的双线来表示这种配对。注意图中箭头的方向总是从下指标指向上指标的, 并注意图中两根线的箭头方向相反。

现在, 我们可以计算  $G_j^i(z)$  了, 注意到奇数个  $\Phi$  的期望值一定为零, 我们可以将  $G_j^i(z)$  按照  $1/z$  泰勒展开为

$$G_j^i(z) = \sum_{n=0}^{\infty} \frac{1}{z^{2n+1}} \langle (\Phi^{2n})_j^i \rangle \quad (4.38)$$

$$= \delta_j^i \frac{1}{z} + \frac{1}{z^3} \langle (\Phi^2)_j^i \rangle + \frac{1}{z^5} \langle (\Phi^4)_j^i \rangle + \dots \quad (4.39)$$

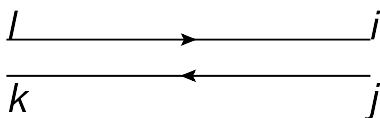


Figure 4.6: The pairing between  $\Phi_j^i$  and  $\Phi_l^k$  with contribution  $\delta_l^i \delta_j^k \frac{1}{Nm^2}$ . The  $i, j$  end of the double lines stands for  $\Phi_j^i$ , the  $k, l$  end stands for  $\Phi_l^k$ .

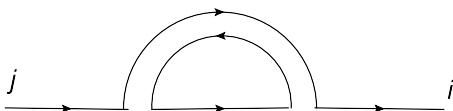


Figure 4.7: The diagram for  $\frac{1}{z^3} \langle \Phi_k^i \Phi_j^k \rangle$ .

首先让我们来看这个展开中的第二项，即  $\frac{1}{z^3} \langle \Phi_k^i \Phi_j^k \rangle$  (当一个表达式中的一个上指标和一个下指标用同样的字母表示时，我们默认对这个指标进行求和，也称之为将这两个指标进行收缩)，稍微想一下就能发现，这一项可以用如下的图形 (4.7) 来表示。

图中每一根单线的贡献是  $1/z$ ，当用一根单线将双线的两端连起来的时候表示要将这两端的两个指标进行收缩。因此，这幅图的总贡献为  $\delta_j^i N \frac{1}{Nm^2}$ ，总的来说是在  $N$  的零次方阶，我们记为  $O(N^0)$ 。注意这里额外的那个  $N$ ，它显然来自于上下指标的收缩，也就是对应于图 (4.7) 中那个闭合的回路。稍微想一下就能发现，任何闭合回路总意味着一个这样的收缩，因此会额外贡献一个因子  $N$ 。

下面我们来考察展开式 (4.39) 中的第三项，也就是  $\frac{1}{z^5} \langle \Phi_k^i \Phi_l^k \Phi_h^l \Phi_j^h \rangle$ ，这涉及到四个矩阵，根据高斯积分两两配对的规则，它有多种配对的可能性。暂时不考虑这四个  $\Phi$  如何配对，仅仅注意到这四个  $\Phi$  的上下指标收缩关系，我们容易知道相应的图形必然具有如下结构，如图 (4.8)。图中的阴影部分表示双线间的可能配对。

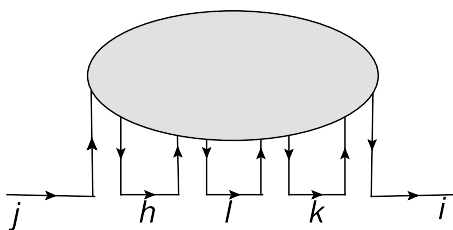


Figure 4.8: The diagram for  $\frac{1}{z^5} \langle \Phi_k^i \Phi_l^k \Phi_h^l \Phi_j^h \rangle$ .

现在，让我们来考虑双线间的可能配对形式。它给出如图 (4.9) 所示的 3 种可能性。

图 (4.9) 中的每一幅图都有两根双线，每根双线贡献一个  $1/N$ 。但是图 (4.9) 中的 (a)、(c) 两图和 (b) 图有根本性的不同，(a)、(c) 两图包含了两个闭合回路，每个闭合回路会额外贡献一个  $N$ ，所以总的来说 (a)、(c) 两图是  $O(N^0)$  阶的。但是 (b) 图完全可以一笔画出，它不包含任何闭合回路，所以总的来说，(b) 图是  $O(1/N^2)$  阶的。也就是说，在大  $N$  极限

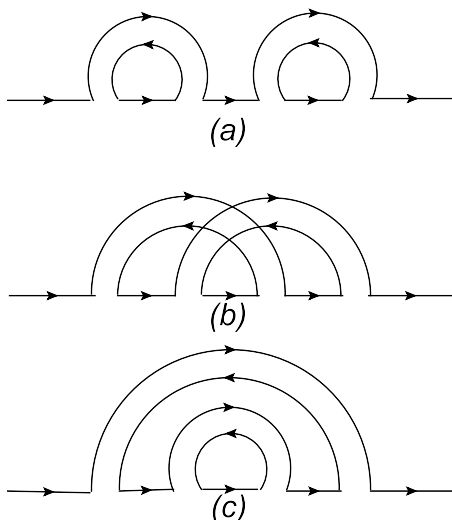


Figure 4.9: The diagrams for  $\frac{1}{z^3} \langle \Phi_k^i \Phi_l^k \Phi_h^l \Phi_j^h \rangle$ .

下, (b) 图的贡献完全可以忽略不计。实际上, (b) 图通常被称为非平面图, 也就是说当它在两维平面上画出来的时候必然要涉及一些交叉线。相反, (a)、(c) 这样的图被称之为平面图, 它们可以直接在两维平面上画出来而不涉及到任何交叉线。图 (4.9) 其实示例了一条一般规则, 即任何非平面图相对于平面图来说总有更少的回路, 由于每一个闭合回路都会额外贡献一个  $N$ , 因此非平面图在大  $N$  展开的时候就代表  $1/N$  的更高阶贡献。特别的, 在  $N \rightarrow \infty$  的大  $N$  极限下, 我们可以忽略掉所有的非平面图。正因为如此, 大  $N$  极限有时候也被称之为平面图极限。

下面我们考虑完整的  $G_j^i(z)$  在平面图极限下的结构, 这就要考虑泰勒展开 (4.39) 的所有阶。为此让我们首先观察一下图 (4.9) 的 (a)、(c) 两图的特征, 很容易发现, (a) 图是将同一个组成单元重复两次而成, 而 (c) 图是在图 (4.7) 的基础上嵌套一根双线而成。稍微尝试一下就能发现, 在平面图极限下, 这两个特征是泰勒展开所有阶共有的, 也即是说, 在平面图极限下, 任何对  $G_j^i(z)$  有贡献的图都可以通过重复和嵌套这两种方式来生成。假定通过嵌套生成的所有图的贡献之和为  $S(z)$ , 并将两种操作生成的所有图的贡献之和记为  $G(z)$  (因为  $G_j^i(z) = \delta_j^i G(z)$ ), 如下图 (4.10) 所示。

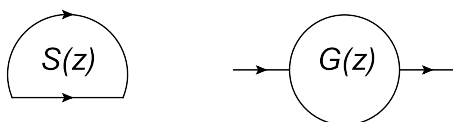


Figure 4.10:  $S(z)$  and  $G(z)$ .

很清楚,  $G(z)$  是将  $S(z)$  不断重复而成, 反过来,  $S(z)$  是在  $G(z)$  的基础上嵌套一根双

线而成，用图形来表示就是 (4.11)。从这幅图中我们很容易读出来

$$\begin{aligned} G(z) &= \frac{1}{z} + \frac{1}{z} S(z) \frac{1}{z} + \frac{1}{z} S(z) \frac{1}{z} S(z) \frac{1}{z} + \dots \\ &= \frac{1}{z - S(z)}. \end{aligned} \quad (4.40)$$

另外，从图中也可以看出来，从  $G(z)$  通过嵌套得到  $S(z)$  的过程我们加上了一根双线，其贡献为  $1/(Nm^2)$ ，但同时也增加了一个回路，其贡献为  $N$ ，因此我们有

$$S(z) = \frac{1}{m^2} G(z). \quad (4.41)$$

通过 (4.40) 和 (4.41) 这两个方程，我们容易解得

$$G(z) = \frac{m^2}{2} \left( z - \sqrt{z^2 - \frac{4}{m^2}} \right). \quad (4.42)$$

式中在二次方程两个根的  $\pm$  号中，我们选择了  $-$  号，这是因为只有这样相应的  $G(z)$  才能满足在  $z \rightarrow \infty$  时  $G(z) \rightarrow 1/z$ 。

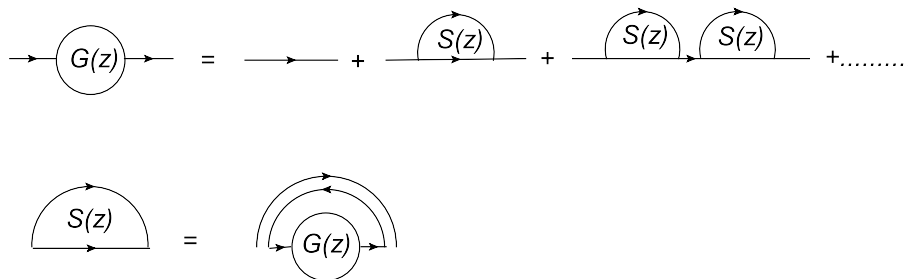


Figure 4.11: The diagrams for  $G(z)$  and  $S(z)$ .

由  $G(z)$  的表达式 (4.42)，我们很容易得到本征值分布  $\rho(E)$ ，

$$\rho(E) = \frac{m^2}{2\pi} \sqrt{\frac{4}{m^2} - E^2}, \quad (4.43)$$

式中  $E$  的取值在区间  $[-2/m, +2/m]$  之内。这就是魏格纳半圆定律。

#### 4.4.2 戴森气体

魏格纳半圆定律是大  $N$  极限下随机矩阵本征值分布的一个简单结果。但是这个结果只适用于  $V(\phi) = \frac{1}{2}m^2\phi^2$  的情形。这一节我们就是要把类似的结论推广到  $V(\phi)$  是一个一般性的多项式的情形，比方说  $V(\phi) = \frac{1}{2}m^2\phi^2 + g\phi^4$  这样的情形。如果用按照  $g$  进行微扰展开的方式来处理问题，我们就能发现，这时候，顶角  $g\phi^4$  的引入使得我们需要考虑的典型平面图为如 (4.12) 所示的样子。很清楚，我们很难通过直接将所有类似这样的图的贡献都加起来以得到一个  $G(z)$  的表达式，所以我们可能需要用图形之外的方法来处理现在的一般性情形。

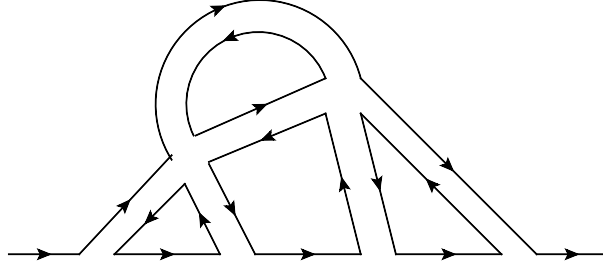


Figure 4.12: A diagram with two vertexes.

为此让我们来考察一个观测量  $\text{Tr}(\mathcal{O}(\phi))$  的期望值, 这里  $\mathcal{O}(\phi)$  表示  $\phi$  的一个函数。很清楚,  $\text{Tr}(\mathcal{O}(\phi))$  在么正变换  $\phi \rightarrow U^\dagger \phi U$  下是不变的。这个期望值由下式计算

$$\langle \text{Tr}(\mathcal{O}(\phi)) \rangle = \frac{1}{\mathcal{Z}} \int [d\phi] \text{Tr}(\mathcal{O}(\phi)) e^{-N\text{Tr}V(\phi)}. \quad (4.44)$$

前面我们已经说过了概率分布  $P(\phi)$  在么正变换下是不变的,  $\text{Tr}(\mathcal{O}(\phi))$  也是么正不变的, 值得讨论的是积分测度  $[d\phi]$ , 不过可以证明如果将它自然地定义成  $[d\phi] = \prod_{i,j} d\phi_j^i$ , 则它在么正变换下也是不变的。

另一方面, 对于任何厄米矩阵  $\phi$ , 如果将其本征值所构成的对角矩阵记为  $\Lambda = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_N\}$ , 则  $\phi$  必定可以分解成  $\phi = U^\dagger \Lambda U$  的形式, 式中  $U$  为某个么正矩阵。那么积分测度  $[d\phi]$  必定可以重写成  $[d\phi] = J \prod_i d\lambda_i [dU]$ , 式中  $J$  为雅可比行列式,  $[dU]$  表示对么正矩阵  $U$  的积分测度。由于积分表达式 (4.44) 的被积函数是么正变换不变的, 因此对  $[dU]$  的积分是平凡, 不妨将之归一化为 1, 如此一来, 积分 (4.44) 就可以重写成,

$$\langle \text{Tr}(\mathcal{O}(\phi)) \rangle = \frac{1}{\mathcal{Z}} \int J \prod_i d\lambda_i \sum_j (\mathcal{O}(\lambda_j)) e^{-N \sum_k V(\lambda_k)}. \quad (4.45)$$

这个表达式中唯一还不清楚的就是那个雅可比矩阵  $J$ 。

为了求出雅可比矩阵  $J$ , 我们不妨考虑  $\phi$  在对角矩阵邻域内的情形, 即  $\phi = e^{-i\varepsilon} \Lambda e^{i\varepsilon}$ , 其中  $\varepsilon$  为一个无穷小厄米矩阵, 因此有  $\phi = \Lambda - i[\varepsilon, \Lambda]$ , 由这个式子可以看出, 对  $\Lambda$  的积分 (即  $[d\Lambda] = \prod_i d\lambda_i$ ) 带来的雅可比行列式为 1, 同时由于矩阵元  $[\varepsilon, \Lambda]_{ij} = (\lambda_i - \lambda_j)\varepsilon_{ij}$ , 它带来的积分测度为  $\prod_{i,j} [(\lambda_i - \lambda_j)d\varepsilon_{ij}] = \prod_{i < j} (\lambda_i - \lambda_j)^2 [d\varepsilon]$ , 其中积分测度  $[d\varepsilon] = \prod_{i,j} d\varepsilon_{ij}$  就是测度  $[dU]$  在恒等矩阵邻域内的具体形式。因此, 我们有, 雅可比行列式  $J = \prod_{i < j} (\lambda_i - \lambda_j)^2$ 。将之代入积分式 (4.45) 我们就有

$$\langle \text{Tr}(\mathcal{O}(\phi)) \rangle = \frac{1}{\mathcal{Z}} \int \prod_i d\lambda_i \sum_j (\mathcal{O}(\lambda_j)) \exp[-NS(\lambda_1, \lambda_2, \dots, \lambda_N)]. \quad (4.46)$$

式中  $S(\lambda_1, \lambda_2, \dots, \lambda_N) = \sum_i V(\lambda_i) - \frac{1}{N} \frac{1}{2} \sum_{i \neq j} \log(\lambda_i - \lambda_j)^2$ 。可以看到, 现在问题变成了一个  $N$  粒子系统的问题, 粒子的坐标就是  $\lambda_i, i = 1, 2, \dots, N$ 。这  $N$  个粒子处在一个势阱  $V(\lambda)$  中, 并且任意两个粒子之间有排斥势  $-\frac{1}{N} \log(\lambda_i - \lambda_j)^2$  的相互作用。这  $N$  个粒子就构成某种有相互作用的气体, 称之为戴森气体。

从表达式 (4.46) 很容易看出, 在大  $N$  极限下, 只有函数  $S(\lambda_1, \lambda_2, \dots, \lambda_N)$  的极小值对积分有贡献, 因此有  $\frac{\partial}{\partial \lambda_i} S(\lambda_1, \lambda_2, \dots, \lambda_N) = 0, i = 1, 2, \dots, N$ , 由此就得到如下方程

$$V'(\lambda_i) = \frac{2}{N} \sum_{j \neq i} \frac{1}{\lambda_i - \lambda_j}. \quad (4.47)$$

当  $N \rightarrow \infty$  时, 这些本征值就会趋向于一个连续谱, 引入谱密度  $\rho(\lambda)$  (它就是我们前面引入的本征值分布函数  $\rho(E)$  的大  $N$  极限, 因此满足  $\int d\lambda \rho(\lambda) = 1$ ), 我们就可以把上面这个方程重写为

$$V'(\lambda) = 2\mathcal{P} \int d\mu \frac{\rho(\mu)}{\lambda - \mu}. \quad (4.48)$$

式中  $\mathcal{P}$  表示取积分主值。

从方程 (4.48) 容易看出, 大  $N$  极限下的本征值谱必然分布在一个有限的区间之内, 因为在允许  $|\lambda| \rightarrow \infty$  的情况下, 方程 (4.48) 右边趋于  $\sim 1/\lambda$ , 因此是无法和左边相等的。下面我们将把这个本征值的分布区间记为  $A$ 。

下面我们引入解析函数  $G(z) = \int_A d\mu \frac{\rho(\mu)}{z - \mu}$ , 其中变量  $z$  定义在整个复平面上。很容易看出来  $\rho(\lambda) = -\frac{1}{\pi} \text{Im}G(\lambda + i\epsilon)$ , 因此这里的  $G(z)$  实际上就是我们前面引入的  $G(z)$  函数的大  $N$  极限,  $G(z)$  现在的这种表达形式称之为谱表示, 从这种表达形式可以看出来,  $G(z)$  除了在区间  $A$  上有一条割线以外, 它在整个复平面都是解析的。很显然方程 (4.48) 告诉我们的不过是  $\text{Re}G(\lambda + i\epsilon) = \frac{1}{2}V'(\lambda)$  (在区间  $A$  上)。由这个条件, 加上  $G(z)$  的解析性, 再加上  $z \rightarrow \infty$  时  $G(z) \rightarrow 1/z$  的条件, 我们就能唯一性地确定解析函数  $G(z)$ , 由此就能求出  $\rho(\lambda)$ 。

比方说, 假如  $V(\lambda) = \frac{1}{2}m^2\lambda^2 + g\lambda^4$ , 这时候由于  $V(\lambda)$  是一个偶函数, 我们可以知道, 谱分布  $\rho(\lambda)$  也必然是一个偶函数, 因此区间  $A$  必定可以取成  $[-a, a]$  的形式, 其中  $a$  是待定的。魏格纳半圆定律启发我们可以试探性地将  $G(z)$  取成如下形式

$$G(z) = \frac{1}{2}V'(z) - P(z)\sqrt{z^2 - a^2}, \quad (4.49)$$

其中  $P(z)$  是一个待确定的多项式。很显然, 这时候

$$\rho(\lambda) = \frac{1}{\pi}P(\lambda)\sqrt{a^2 - \lambda^2}, \quad (4.50)$$

由于  $\rho(\lambda)$  是一个偶函数, 所以  $P(z)$  必定是一个偶多项式。又由于  $z \rightarrow \infty$  时  $G(z) \rightarrow 1/z$ , 可知  $P(z)$  是一个 2 次偶多项式, 因此包含两个待定系数。再加上前面待定的  $a$ , 我们就有 3 个待定参数。它们可以这样来确定, 由于  $z \rightarrow \infty$  时  $G(z) \rightarrow 1/z$ , 所以在  $z \rightarrow \infty$  时  $G(z)$  渐进式的  $z^3, z^1$  项系数都要消去, 而  $1/z$  项的系数必定要等于 1, 这就给出了三个方程, 刚好可以确定 3 个待定参数。这就完成了我们想要的求解, 相应的 (4.50) 式就是魏格纳半圆定律的推广。

得到  $\rho(\lambda)$  后, 观察量  $\text{Tr}\mathcal{O}(\phi)$  的期望值在大  $N$  极限下就可以写成

$$\frac{1}{N} \langle \text{Tr}(\mathcal{O}(\phi)) \rangle = \frac{1}{N} \sum_i \mathcal{O}(\lambda_i) = \int d\lambda \rho(\lambda) \mathcal{O}(\lambda). \quad (4.51)$$

## 4.5 总结

至此，我们走完了从简单的标量随机变量到高维随机矩阵的完整旅程。回顾本章，核心的思想可以凝练为以下几个层面的“统一”：第一，是“大数目极限”下的统一。无论是中心极限定理中  $N$  个随机变量之和趋向高斯分布，还是随机矩阵理论中  $N \rightarrow \infty$  时本征值趋向确定的光滑连续谱（如魏格纳半圆），都揭示了“大  $N$  极限”是压制涨落、催生普适性规律的温床。第二，是“图形化思维”的统一。从最初用“点的集团分解”来理解普通矩与累积量的关系，到用带有顶角和连线的费曼图计算微扰高斯分布，再到用双线费曼图计算矩阵的格林函数。我们贯穿始终地演示了如何将复杂的积分与配对问题，降维打击为直观的拓扑计数问题。第三，是“连通性与物理可观测量”的统一。讲义中反复强调了一个核心结论：普通矩对应所有图的求和，而真正具有物理意义的累积量（或连通格林函数），仅仅对应于连通图的贡献。这个规则在标量场和矩阵场中都是成立的。第四，是“数学结构与物理图像”的统一。我们看到，矩阵积分中极其复杂的雅可比行列式  $\prod_{i < j} (\lambda_i - \lambda_j)^2$ ，在物理图像上完美等价于处于势阱中且带有对数排斥势的“戴森气体”；而图论中的“圈数”，在统计力学中直接对应着“温度微扰的阶数”。掌握这些思想，不仅是为了计算几个特定的积分，更是为了获得一种跨学科的直觉。当未来大家在量子场论、多体物理甚至机器学习的高维统计中再次遇到类似的结构时，希望本章所建立的这套“语言体系”，能成为你们迅速看透问题本质的利器。

## 5. KT 相变

### 5.1 引言

KT 相变是凝聚态物理中的一场革命，它打破了二维世界无法发生相变的传统认知，并揭示了一种由拓扑缺陷主导的全新物理秩序。其核心洞察在于，这种相变并非源于传统的对称性破缺，而是由一类被称为涡旋（Vortex）的拓扑缺陷的“解绑（Unbinding）”所驱动的。

这一突破性的理论由物理学家 David J. Thouless 和 J. Michael Kosterlitz 在 20 世纪 70 年代初提出。值得注意的是，苏联物理学家 Vadim Berezinskii 在更早的时候也独立提出了类似理论，因此该相变也被称为 Berezinskii-Kosterlitz-Thouless (BKT) [B, KT, K] 相变。因在这一领域的工作，Thouless 和 Kosterlitz 与另一位科学家共同获得了 2016 年诺贝尔物理学奖。

20 世纪中叶，物理学家对二维世界是否能够发生相变普遍持否定态度，这源于两个关键事件：首先是理论的禁令，1966 年，Mermin-Wagner 定理 [MerminWagner, Hohenberg, Coleman] 从数学上证明，在二维系统中，热涨落会强大到足以破坏任何连续对称性自发破缺下的长程有序，换言之，两维系统不会发生连续对称性的自发破缺，因此根据朗道相变理论，似乎也就不会发生相应的常规相变。然而，同期实验却在二维超流氦-4 薄膜中观测到了类似相变的行为，理论与实验之间产生了尖锐的矛盾。

正是在这个背景下，Kosterlitz 与 Thouless 决定挑战这一难题。他们意识到，要解释实验现象，必须跳出传统思维，考虑更高能量的激发——即涡旋，而不是只关注低能涨落（如自旋波）。

理论的核心是一个简单而深刻的模型——二维 XY 模型。它描述了一组在平面上可以自由旋转的“自旋”（可类比为小磁针）的相互作用，是理解超流薄膜、二维磁体等系统的理想框架。Mermin-Wagner 定理虽然禁止了真正的长程有序，但二维系统在低温下仍可以维持一种微妙的状态——准长程有序。在这种状态下，自旋-自旋关联函数随距离以幂律形式衰减，而不是高温无序状态下的指数衰减。当系统的温度升高，热涨落不仅会引起自旋波（平滑的扰动），更会激发拓扑缺陷——涡旋（Vortex）和反涡旋（Anti-vortex）。这些缺陷具有一个重要的拓扑不变量——“绕数”，表征了绕着平面上的一条闭合回路走一圈自旋旋转的总圈数，通常为 +1（涡旋）或 -1（反涡旋）。由于绕数是一个整数，无法通过连续的变形将其变为 0，这赋予了它们“拓扑保护”的特性，使其极其稳定，不像普通的局部扰动那样容易消失。在低温下，涡旋与反涡旋总是成对出现，并相互束缚在一起。从宏观上看，由于自旋波的存在，这涡旋束缚态仍然维持着系统的准长程有序。随着温度升高，熵（描述系统混乱度）的作用越来越大。当达到临界温度  $T_c$  时，熵的增加足以克服涡旋-反涡旋对之间的对数型吸引势能。这对束缚被“熔断”，涡旋与反涡旋解绑，开始在系统中自由移动。这种自由运动的涡旋会像湍流一样，彻底扰乱系统的准长程有序，使其变为无序态。这种涡旋与反涡旋的相互作用与二维库仑气体中的正负电荷行为有着深刻的数学对应 [KT, K][cmft][nagaosa]。低温下的束缚涡旋对，就好比库仑气体中的绝缘态（正负电荷被束缚在一起）；高温下的自由涡旋，则对应导体的导电态（正负电荷自由移动）。

BKT 相变超越了连续相变的朗道范式（该范式将相变与对称破缺联系起来）。因此，人们传统上研究 BKT 相变，不是通过连续场论，而是作为格点理论（2D XY 自旋模型）或作为等效的库仑等离子体系统。但 BKT 相变也可以用有效场论来描述，即二维正弦-戈登模型 [cmft, WXG]。这里，我们将讲述关于这一有效场论的直接连续场论推导。

## 5.2 二维没有连续对称性自发破缺

引言中提到 Mermin-Wagner 定理，即二维系统没有连续对称性自发破缺，这里给出一个场论证明，这源自于 Coleman 的工作，虽然这个证明针对的是场论系统，但其本质适用于一般性的二维系统。

为了简单起见，假设我们的场论系统有一个  $U(1)$  连续全局对称性（推广到其它连续对称性是容易的），系统的序参量为局域复标量场  $\phi(x)$ ，在  $U(1)$  对称性的作用下，它按照如下规则变换

$$\phi(x) \rightarrow e^{i\alpha} \phi(x). \quad (5.1)$$

下面用反证法，假设低温时，这个  $U(1)$  对称性自发破缺了，从而  $\phi(x)$  获得一个非零的真空期望值：

$$\langle \phi(x) \rangle = v \neq 0.$$

如果这个假设成立，那么即有，当  $|x| \rightarrow \infty$  时，两点关联函数

$$\langle \phi(x) \bar{\phi}(0) \rangle = \langle \phi(x) \rangle \langle \bar{\phi}(0) \rangle = v^2.$$

下面我们就是要证明，如果考虑到序参量的涨落，这个式子将不可能成立，从而导出矛盾，进而说明  $U(1)$  对称性不可能自发破缺。

另一方面，在相变温度之下，考虑到序参量的涨落，序参量将由两维空间到陪集空间  $U(1)/1 = U(1)$  的连续映射所描述，因此，在相变温度之下，可将  $\phi(x)$  参数化为

$$\phi(x) = v e^{i\pi(x)/v},$$

通常称式中的  $\pi(x)$  为 Goldstone 场。描写序参量涨落的低能有效理论由 Goldstone 场  $\pi(x)$  的如下作用量主导：

$$S_{\text{eff}} = \int d^2x \frac{1}{2} (\partial\pi)^2 + \text{高阶导数项}.$$

式中  $(\partial\pi)^2 \equiv \nabla\pi \cdot \nabla\pi$ 。因此在长距离上（从而  $\pi$  的小尺度剧烈变化可以忽略），高阶导数项可以忽略，于是， $\pi(x)$  的两点函数为

$$\langle \pi(x)\pi(0) \rangle = \int \frac{d^2p}{(2\pi)^2} \frac{e^{ip \cdot x}}{p^2} = -\frac{1}{2\pi} \log(\Lambda^2 x^2),$$

其中  $\Lambda$  是紫外截断，因为否则的话式中的积分是发散的。

进而即可以计算

$$\langle \phi(x)\bar{\phi}(0) \rangle = v^2 \left\langle e^{i(\pi(x) - \pi(0))/v} \right\rangle.$$

对于高斯分布，有  $\langle e^{iA} \rangle = e^{-\frac{1}{2}\langle A^2 \rangle}$ ，其中  $A = (\pi(x) - \pi(0))/v$ 。于是

$$\langle \phi(x)\bar{\phi}(0) \rangle = v^2 \exp\left(-\frac{1}{2v^2} \langle (\pi(x) - \pi(0))^2 \rangle\right).$$

利用  $\langle (\pi(x) - \pi(0))^2 \rangle = 2\langle \pi(0)^2 \rangle - 2\langle \pi(x)\pi(0) \rangle$ ，代入  $\langle \pi(x)\pi(0) \rangle$  得

$$\langle (\pi(x) - \pi(0))^2 \rangle = K + \frac{1}{\pi} \log(\Lambda^2 x^2),$$

其中  $K = 2\langle \pi(0)^2 \rangle$  是与  $x$  无关的常数（可能发散但可通过重整化吸收）。因此，

$$\langle \phi(x)\bar{\phi}(0) \rangle = v^2 e^{-K/(2v^2)} (\Lambda^2 x^2)^{-1/(2\pi v^2)}.$$

当  $|x| \rightarrow \infty$  时，上式趋于零。这与对称性破缺假设导出的 (5.2) 式（通常称之为聚类性质）矛盾，因此  $U(1)$  对称性不能自发破缺。

对于更一般的连续对称群  $G$ ，假设其破缺到子群  $H$ ，则为了参数化陪集空间  $G/H$ ，会存在多个 Goldstone 场，每个都贡献类似的对数因子。序参量可以是  $G/H$  上的某个矩阵或向量，其两点函数在长距离下表现为若干指数因子的乘积，每个因子对应一个 Goldstone 模式。由于每个因子均幂律衰减到零，整体仍趋于零，与聚类性质矛盾。因此结论对任意连续对称性成立。

### 5.3 涡旋算符与 BKT 相变的正弦-戈登描述

#### 5.3.1 BKT 相变的连续场论描述

在本小节中，我们将通过路径积分发展 BKT 相变的连续场论描述。我们将从二维 XY 模型开始设定我们的约定，但我们的描述本质上是基于连续场论的，并不依赖于 XY 模型的格点理论。

二维 XY 模型是一个自旋被约束在格点平面（格距为  $a$ ）内旋转的系统。该模型由配分函数描述：

$$\begin{aligned} \mathcal{Z} &= \int_{-\pi}^{\pi} \prod_x \frac{dA_x}{2\pi} \exp \left[ \frac{\beta}{2\pi} \sum_{x,\delta} (\mathbf{S}_x \cdot \mathbf{S}_{x+\delta} - 1) \right] \\ &= \int_{-\pi}^{\pi} \prod_x \frac{dA_x}{2\pi} \exp \left[ \frac{\beta}{2\pi} \sum_{x,\delta} (\cos(A_x - A_{x+\delta}) - 1) \right]. \end{aligned} \quad (5.2)$$

这里  $x$  表示二维格点上的一个格点， $\delta$  是连接该格点与其最近邻之一的单位矢量， $\beta = 1/T$  是逆温度。其中  $A_x \sim A_x + 2\pi$  是第  $x$  个自旋与某个任意轴所成的角度，并且我们将  $|\mathbf{S}_x|$  归一化为 1。

由于我们只关心该系统的长距离行为，我们可以为其发展一个连续场论描述。在这个描述中，只有缓慢变化的构型才会对配分函数有显著贡献，因此我们可以将作用量展开到角度  $A_x$  的二次项：

$$S = \frac{1}{4\pi T} \sum_{x,\delta} (A_x - A_{x+\delta})^2 \approx \frac{1}{4\pi T} \int_{\mathbb{R}^2} \|dA\|^2 \quad (5.3)$$

其中  $A(x) \sim A(x) + 2\pi$  是一个连续标量场， $dA = \frac{\partial A}{\partial x^i} dx^i$  是  $A$  的外微分， $\|dA\|^2 = (dA) \wedge *(dA)$ ，这里  $*$  是二维的霍奇星算符， $\wedge$  表示微分形式的外积。如果我们首先忽略角度变量  $A$  的周期性（因此忽略涡旋构型），那么系统的配分函数可以表示为  $A$  场的适当路径积分，权重为  $W[A]$ ，

$$W[A] = \exp \left[ -\frac{1}{4\pi T} \int_{\mathbb{R}^2} \|dA\|^2 \right]. \quad (5.4)$$

为了考虑  $A$  的周期性，我们必须对某些携带非平凡拓扑数的奇异构型（在连续场论描述中）求和。也就是说，我们应该允许场  $A$  是一个多值函数，这样它可以在某些分支割线处有  $2\pi$  的跃变。这些场构型代表涡旋激发。

现在我们描述如何在  $A$  场路径积分中处理涡旋构型。当在点  $p$  处存在一个涡旋奇点时，我们在  $p$  周围画一个无穷小的逆时针圆  $C_p$ ，切掉  $C_p$  的内部——以消除涡旋奇点——并将该内部区域上的场构型路径积分替换为一个小常数  $g$ ，它代表涡旋内部的贡献。之后，我们将  $A$  沿  $C_p$  的构型（该构型由位于  $p$  处的涡旋决定）作为  $A$  场在  $C_p$  外部区域的边界条件，并且我们将对  $C_p$  外部区域进行带有此涡旋边界条件的路径积分。因此，一个涡旋构型对路径积分的总贡献是这个外部区域路径积分与内部区域贡献  $g$  的乘积。

事实上，对于位于  $p$  处、携带拓扑量子数（绕数） $n \in \mathbf{Z}$  的  $n$  涡旋，我们有

$$\int_{C_p} \frac{F_A}{2\pi} = n, \quad (5.5)$$

其中  $F_A = dA$ 。因此，对于这个涡旋，沿  $C_p$  的  $A$  场边界条件可以简单地描述为：在  $C_p$  上逆时针绕行一圈后， $A$  的值将增加  $n2\pi$ 。如果  $n$  为负，我们将称这个  $n$  涡旋为反涡旋。

对于拓扑绕数为 2 的涡旋，即 2-涡旋，我们可以想象它由两个紧密束缚的 1-涡旋组成。然而，由于涡旋-涡旋相互作用总是相互排斥（正如我们将在第 2.3 小节看到的），这两个复合的 1-涡旋总是倾向于彼此分离，因此非束缚构型具有更低的能量和更高的路径积分权重。同样的论证也适用于具有更高绕数的涡旋和反涡旋。因此，我们可以忽略所有拓扑量子数大于 1 的涡旋（或绕数小于 -1 的反涡旋）对配分函数的贡献，而只包括那些良好分离的 1-涡旋和 1-反涡旋（-1-涡旋）的贡献，它们形成一种稀薄气体。

对于每个具有  $m$  个 1-涡旋和  $n$  个 -1-涡旋的稀薄气体，我们可以进行相应的路径积分，这是我们描述过的单涡旋路径积分的推广。最后，我们将对所有具有不同数量涡旋和反涡旋的稀薄气体的贡献求和，包括没有涡旋的贡献，最终得到一个 1-涡旋近似的配分函数  $\mathcal{Z}_{\text{vortexes}}$ ，

$$\mathcal{Z}_{\text{vortexes}} = \sum_{m,n} \frac{1}{m!} \frac{1}{n!} g^{m+n} \int_{m,n} [\mathcal{D}A] \exp \left[ -\frac{1}{4\pi T} \int_{\mathbb{R}^2} \|dA\|^2 \right], \quad (5.6)$$

其中  $\int_{m,n} [\mathcal{D}A]$  表示在每一个 1-涡旋和每一个 -1-涡旋的外部区域上，对  $A$  场构型进行路径积分，并带有相应的涡旋边界条件（在无穷小圆上）。注意，每个涡旋和每个反涡旋的位置坐标也必须被积分，因为位于不同位置的涡旋代表不同的构型。(5.6) 中的因子  $\frac{1}{m!}$  和  $\frac{1}{n!}$  分别来自全同涡旋和全同反涡旋的排列，因子  $g^{m+n}$  来自这  $m$  个涡旋和  $n$  个反涡旋的内部构型。我们使用符号  $\mathcal{Z}_{\text{vortexes}}$  是为了提醒我们，我们做了 1-涡旋（和 -1-涡旋）以及稀薄气体近似。

在接下来的两个小节中，我们将通过发展所谓的阿贝尔对偶性，以及针对涡旋的涡旋算符描述，来计算 (5.6) 中的路径积分。

### 5.3.2 阿贝尔对偶性

在本小节中，我们将利用所谓的二维阿贝尔对偶性，为我们在第 (5.3.1) 节中描述的 BKT 相变，发展一个用某个  $B$  场表示的对偶描述，以取代  $A$  场描述。简而言之，其主要思想是将  $A$  场描述中的涡旋变换为对偶描述中的某些涡旋算符。

我们从没有涡旋的构型开始，注意到路径积分  $\int [\mathcal{D}A] W[A]$  等价于

$$\mathcal{Z} = \int [\mathcal{D}A \mathcal{D}B] \exp \left[ \frac{i}{2\pi} \int_{\mathbb{R}^2} (dB \wedge F_A) \right] W[A], \quad (5.7)$$

其中  $F_A = dA$ ，并且我们将  $B$  场归一化为具有周期性  $B \sim B + 2\pi$ 。这个等价性可以通过注意到 (5.7) 中的指数因子在分部积分后是平凡的，因此对  $B$  的路径积分只贡献一个平凡的发散常数，而这个常数很容易被正规化。

现在我们交换  $A$  和  $B$  泛函积分的顺序。注意对  $A$  的泛函积分基本上等同于对  $F_A = dA$  的泛函积分，因为  $A$  场构型是拓扑平凡的。因此，我们首先进行 (5.7) 中对  $dA$  的积分，结果（忽略来自  $dA$  高斯积分的常数因子）是

$$\mathcal{Z} = \int [\mathcal{D}B] \exp \left[ -\frac{T}{4\pi} \int \|dB\|^2 \right]. \quad (5.8)$$

换句话说，配分函数  $\int [\mathcal{D}A] W[A]$  (无涡旋) 完全等价于 (5.8)，即对偶场  $B$  的泛函积分。

为了熟悉这个对偶变换，我们想再算一个例子，计算带有插入项  $dA$  的路径积分  $\int [\mathcal{D}A](dA)W[A]$ 。通过使用上述对偶变换，我们可以看到  $\int [\mathcal{D}A](dA)W[A]$  等价于

$$\int [\mathcal{D}B](iT * dB) \exp \left[ -\frac{T}{4\pi} \int \|dB\|^2 \right]. \quad (5.9)$$

有了这些准备，我们现在可以将带有 1-涡旋边界条件的  $A$  场路径积分重新表达为对偶  $B$  场路径积分加上适当的局域算符插入，这些局域算符就是能够产生涡旋的涡旋算符。这是下一小节的内容。

### 5.3.3 涡旋算符与正弦-戈登描述

我们现在可以为  $\mathcal{Z}_{1\text{vortexes}}$  (5.6) 推导出一个对偶的  $B$  场描述，结果证明它就是 BKT 相变的正弦-戈登描述。为此，我们首先计算单个 1-涡旋的贡献。

回忆一下，在  $A$  场描述中，位于  $p$  处的单个 1-涡旋是一个带有相应边界条件的涡旋边界，该边界条件定义在无穷小圆  $C_p$  上 ( $C_p$  的内部已被切掉并替换为因子  $g$ )。沿着  $C_p$  逆时针走一圈后， $A$  的值将增加  $2\pi$ 。我们将带有此涡旋边界条件的  $A$  场路径积分记为  $\mathcal{Z}[C_p]$ 。

为继续推导，我们将  $C_p$  外部的空间记为  $C \times \mathbb{R}^+$ ，并在其上选取右手定向。显然，作为  $C \times \mathbb{R}^+$  的边界，所考虑的 1-涡旋边界具有相关的顺时针定向，我们将此边界记为  $C_p^-$ ，上标  $-$  用来表示其顺时针定向。然后我们考虑指数因子

$$\exp \left[ \frac{i}{2\pi} \int_{C \times \mathbb{R}^+} dB \wedge F_A \right]. \quad (5.10)$$

这个因子可以重写为  $\exp \left[ \frac{i}{2\pi} \int_{C \times \mathbb{R}^+} d(BF_A) \right]$ ，它可以在边界  $C_p^-$  上积分为

$$\exp \left[ iB(p) \int_{C_p^-} \frac{F_A}{2\pi} \right], \quad (5.11)$$

因为  $C_p^-$  是一个无穷小圆。注意到  $C_p^-$  是所考虑的 1-涡旋的涡旋边界，但具有顺时针定向，我们有  $\int_{C_p^-} F_A/2\pi = -\int_{C_p} F_A/2\pi = -1$ ，因此

$$\exp \left[ iB(p) \int_{C_p^-} \frac{F_A}{2\pi} \right] = \exp[-iB(p)]. \quad (5.12)$$

显然，可以通过将指数因子 (5.10) 乘以一个额外的因子  $\exp[iB(p)]$  来抵消这个最终结果，因此

$$\exp[iB(p)] \exp \left[ \frac{i}{2\pi} \int_{C \times \mathbb{R}^+} dB \wedge F_A \right] = 1. \quad (5.13)$$

从这些讨论可以看出， $A$  场路径积分  $\mathcal{Z}[C_p]$  等价于下面的路径积分

$$\mathcal{Z}[C_p] = \int [\mathcal{D}A \mathcal{D}B] \exp[iB(p)] \exp \left[ \frac{i}{2\pi} \int_{C \times \mathbb{R}^+} (dB \wedge F_A) \right] W[A], \quad (5.14)$$

因为对  $B$  的路径积分平凡地给出一个无穷常数 (由于 (5.13))，该常数很容易被正规化。在交换  $A$  和  $B$  泛函积分的顺序并进行  $A$  场积分之后，就像我们在第 (5.3.2) 小节中所做的那样，我们可以得到 (忽略来自  $dA$  高斯积分的一个无关常数因子)

$$\mathcal{Z}[C_p] = \int [\mathcal{D}B] \exp[iB(p)] \exp \left[ -\frac{T}{4\pi} \int_{\mathbb{R}^2} \|dB\|^2 \right]. \quad (5.15)$$

上述推导告诉我们，在  $B$  场描述中，位于  $p$  处的单个 1-涡旋可以通过在  $B$  场路径积分中插入一个局域算符  $\exp[iB(p)]$  来产生。类似地，要在  $p$  处产生一个单个 -1-涡旋，应该插入局域算符  $\exp[-iB(p)]$ 。一般来说，位于  $p$  处的  $m$  涡旋由指数算符  $\exp[imB(p)]$  产生。这些指数算符就是涡旋算符。

作为练习，我们可以计算位于  $x$  处的  $m$  涡旋和位于原点 0 处的  $n$  涡旋之间的势能  $U_{eff}(x)$  和力  $F(x)$ 。显然， $U_{eff}(x)$  由我们在  $B$  场描述中刚刚发展的涡旋算符的两点关联函数决定：

$$\exp[-\beta(U_{eff}(x) + U_L)] = \langle \exp[imB(x)] \exp[inB(0)] \rangle, \quad (5.16)$$

其中引入可调常数  $U_L$  是为了设定  $U_{eff}(L) = 0$ ， $L$  表示所考虑系统的空间尺度。通过类似于 (5.2) 节中的计算可以发现  $U_{eff}(x) = \frac{mn}{2} \ln(L^2/x^2)$ ，这是电荷  $m$  和电荷  $n$  之间的二维库仑势。力  $F(x)$  为  $F(x) = mn/|x|$ ，当  $m, n$  同号时（两个涡旋或两个反涡旋）为排斥力，否则为吸引力（一个涡旋和一个反涡旋）。这将导致平面 XY 模型的库仑气体描述 [KT]。

我们现在可以利用 1-涡旋和 -1-涡旋的涡旋算符描述，将  $\mathcal{Z}_{\text{vortexes}}$  (5.6) 重新表述为

$$\begin{aligned} \mathcal{Z}_{\text{vortexes}} = & \sum_{m,n} \frac{1}{m!} \frac{1}{n!} g^{m+n} \int [\mathcal{D}B] \int d^2x_1 d^2x_2 \dots d^2x_m \exp[i \sum_{i=1}^m B(x_i)] \\ & \cdot \int d^2x'_1 \dots d^2x'_n \exp[-i \sum_{j=1}^n B(x'_j)] \exp \left[ -\frac{T}{4\pi} \int_{\mathbb{R}^2} \|dB\|^2 \right], \end{aligned} \quad (5.17)$$

其中  $x_i, x'_j$  分别表示第  $i$  个涡旋和第  $j$  个反涡旋的位置坐标。并且我们通过对涡旋的位置坐标进行积分来计入所有可能构型的贡献。

看起来似乎任意  $m$  个 1-涡旋和任意  $n$  个 -1-涡旋的构型都有贡献，但实际上只有 1-涡旋和 -1-涡旋数量相等的中性构型才有非零贡献。这可以通过首先在 (5.17) 的路径积分中积分掉  $B$  场的零模式（即  $B$  等于常数的模式）看出。因此，我们只需要考虑涡旋和反涡旋的中性库仑气体。

但是，如果我们首先对 (5.17) 中的非负整数  $m, n$  进行求和，我们可以将  $\mathcal{Z}_{\text{vortexes}}$  重新表达为

$$\mathcal{Z}_{\text{vortexes}} = \int [\mathcal{D}B] \exp \left[ -\frac{T}{4\pi} \int_{\mathbb{R}^2} \|dB\|^2 + g \int_{\mathbb{R}^2} \cos(B) \right]. \quad (5.18)$$

这是一个二维正弦-戈登模型，即众所周知的 BKT 相变的有效场论描述 [cmft, WXG]。至此，我们已经达到了我们的主要目的，即为 BKT 相变的正弦-戈登描述提供了一个优雅的推导，我们的推导直接基于连续场论，而不是基于格点理论或库仑气体模型。

## 5.4 关于 BKT 相变的讨论

现在我们利用正弦-戈登描述 (5.18) 来讨论 BKT 相变。本节的所有结果都可以与文献 [KT, K, cmft, WXG] 进行比较。从表达式 (5.18) 可以看出，在极低温度下， $B$  场的涨落非常大，这些涨落会抹去涡旋的贡献。因此，在极低温度下我们可以完全忽略涡旋和

反涡旋。那么自旋-自旋关联可以在  $A$  场描述 (5.4) 中容易地计算出来:

$$\langle \exp[iA(x)] \exp[-iA(0)] \rangle \sim \left( \frac{1}{x^2} \right)^{T/2}. \quad (5.19)$$

而在高温下, (5.18) 的相位刚度很大, 余弦项变得重要, 因此场  $B$  被经典地固定在 0 附近。在这种情况下, 可以在 (5.18) 中将场  $B$  在 0 附近展开, 以看到系统是有能隙的, 并且系统的序是短程的。

为了确定  $g=0$  时的临界温度  $T_{c,g=0}$ , 我们将有效理论 (5.18) 视为对自由场论 (作用量为  $S_0 = \frac{1}{4\pi\beta} \int_{\mathbb{R}^2} \|dB\|^2$ ) 的  $g \cos(B)$  微扰。根据重整化群理论, 临界温度出现在使  $\cos(B)$  扰动成为边际扰动的数值处, 这时  $\cos(B)$  的标度量纲  $\Delta = \beta/2$  (可以通过两点关联函数  $\langle e^{iB(x)} e^{-iB(0)} \rangle$  确定  $e^{iB}$  的标度量纲, 这也就是  $\cos(B)$  的标度量纲) 变为 2。这给出

$$1/(2T_{c,g=0}) = 2, \quad T_{c,g=0} = 1/4. \quad (5.20)$$

此外, 在  $g=0$  附近, 可以计算无量纲耦合常数  $y = ga^2$  的重整化群 (RG) 流 (到  $y$  的一阶) (这里  $a$  为有效的格距)。该流应与  $x = \Delta - 2 = \beta/2 - 2$  成正比, 因为  $\Delta = 2$  使得  $y \cos(B)$  变成边际扰动。一个详细的微扰计算 (比如用背景场方法) 告诉我们

$$dy = -xy d \ln a. \quad (5.21)$$

可以看到, 当  $T < 1/4$  时,  $x > 0$ ,  $y \cos(B)$  是不相关的, 在长距离下 RG 将流向一个自由玻色子的共形场论 (CFT), 并且两个  $SO(2)$  自旋之间存在代数长程关联 (5.19)。事实上, 在这种情况下, 涡旋和反涡旋激发将配对并组合成中性偶极子 (由于它们之间的库仑吸引力), 系统处于介电相。这样的偶极子对系统的热力学行为影响可以忽略。而当  $T > 1/4$  时,  $x < 0$ , 涡旋微扰  $y \cos(B)$  是相关的, 系统具有能隙, RG 将在长距离下流向平庸理论。在这种情况下, 偶极子将解离, 形成涡旋和反涡旋的等离子体, 并动态产生一个屏蔽长度。当  $T = 1/4$  时,  $x = 0$ ,  $y \cos(B)$  是边际扰动, 如果  $y = 0$ , 系统将流向一个  $Z_2$  轨道折线 CFT [pol]。

在不动点  $x = \beta/2 - 2 = 0$ ,  $y = 0$  (即  $T = 1/4$ ,  $g = 0$ ) 附近, 我们也可以计算  $x$  的 RG 流到  $y$  和  $x$  的二阶, 结果为

$$dx = -Cy^2 d \ln a, \quad (5.22)$$

其中  $C$  是一个正常数, 可以通过改变长度尺度  $a$  的标度设为任意正数 (由于  $y = ga^2$  依赖于  $a$ ), 比如不妨设  $C = 1$ , 从而

$$dx = -y^2 d \ln a. \quad (5.23)$$

利用微分方程组 (5.21) 和 (5.23) (关于这两个方程的推导, 参见 [cmft, WXG]), 我们可以更好地理解系统的低温 ( $T < T_c$ ) 行为和高温 ( $T > T_c$ ) 行为。比如, 在临界温度  $T_c$ , 系统将流向不动点  $x = y = 0$ , 此时理论由  $Z_2$  轨道折线 CFT 描述。在  $A$  场描述中, 这个 CFT 可以描述为对  $W[A]$  (5.4) 的路径积分, 其中  $T = 1/4$ 。因此, 在  $T_c$  处自旋-自旋关联的临界指数  $\eta$  将是  $\eta = 1/4$ 。

另外，将方程 (5.23) 除以方程 (5.21)，可以得到  $\frac{dx}{dy} = \frac{y}{x}$ ，积分即得

$$x^2 - y^2 = -\delta^2, \quad (5.24)$$

这就是重整化群流在  $(x, y)$  平面上的轨迹方程，式中  $-\delta^2$  为积分常数。

为了考察高温相（从而  $x < 0$ ）的关联长度，假设我们从  $(x, y) = (x_0, y_0)$  的临界点附近（从而  $x_0, y_0$  很小）开始演化重整化群，演化到  $(x, y) \sim (-\infty, \infty)$  的强耦合区域（高温相  $x < 0$ ），此时有效的  $a$  就是关联长度  $\xi$ ，即  $\xi \sim a \sim e^{\ln a}$ 。因此我们有  $x^2 - y^2 = -\delta^2$ ，或者  $y^2 = x^2 + \delta^2$ ，代入方程 (5.23) 即有

$$d \ln a = -dx/y^2 = -\frac{dx}{\delta^2 + x^2} \Rightarrow \ln a = -\int_{x_0}^{-\infty} \frac{dx}{\delta^2 + x^2} \approx \int_0^{\infty} \frac{dx}{\delta^2 + x^2}. \quad (5.25)$$

这里需要注意到，在临界点附近  $|x_0| \approx |y_0|$  均很小，从而上述积分约为

$$\ln a \approx \frac{\pi}{2\delta}. \quad (5.26)$$

所以，关联长度  $\xi \sim e^{\ln a}$  为

$$\xi \sim e^{\pi/(2\delta)}. \quad (5.27)$$

另一方面，在临界点附近  $|x_0| \approx |y_0|$ ，从而  $\delta = \sqrt{y_0^2 - x_0^2} \sim \sqrt{2|x_0|(y_0 - x_0)}$ 。而根据  $x$  的定义， $x_0 = \beta/2 - 2 \sim -2\frac{T-T_c}{T_c} \equiv -2t$ ，注意，由于定义在紫外截断处，所以这里的  $\beta$  和  $T$  都是物理的，而不是像原本  $x$  的定义式中那样是有效的，从而这里的  $t$  就是物理的约化温度。通过这个分析我们知道了，

$$\delta \sim \sqrt{t}, \quad (5.28)$$

代入 (5.27) 式，即有

$$\xi \sim e^{b/\sqrt{t}}, \quad (5.29)$$

式中  $b > 0$  为某个常数。注意，在  $t \rightarrow 0^+$  时，关联长度是发散的，但是，这种发散不是像通常临界现象那样的幂律发散，而是快于任何幂律。

最后，正如我们已经强调的，在推导有效理论 (5.18) 时，我们做了 1-涡旋近似。可以自然地预期，双涡旋会贡献一个类似的项  $g_2 \cos(2B)$ ，但系数  $g_2 \ll g$  小得多。当温度升高到 1 时， $\cos(2B)$  也将变得相关。因此，我们的 1-涡旋近似仅在  $T$  小于 1 时有效。

## 5.5 总结

KT 相变理论的普适性使其应用远超最初的超流氦薄膜，深刻影响了多个物理学领域：比如，在二维超导体薄膜中，库珀对（电子对）可以形成类似的涡旋-反涡旋束缚态，其解绑导致超导性消失。比如，这个理论被成功用于解释二维晶体的熔化过程，并预言了在真实二维磁性材料（如  $\text{TmMgGaO}_4$ ）中存在的 KT 相变。再比如，在约瑟夫森结阵列中（也就是由超导体-绝缘体-超导体结构构成的阵列），其相变行为也属于 KT 普适类。

总而言之，KT 相变超越了相变的朗道范式，它不仅成功解决了二维世界能否有序的理论悖论，更开创性地将数学中的拓扑学引入物理学，催生了“拓扑物态”这一全新的研究范式。它所揭示的由缺陷主导的相变机制，深刻地改变了我们对物质状态和相变本质的理解。

## Bibliography

- [1] V. B. Berezinskii, Sov. Phys. JETP 32, 493 (1971).
- [2] J. M. Kosterlitz, D. J. Thouless, Journal of Physics C: Solid State Physics, Vol. 6 pages 1181-1203 (1973).
- [3] J. M. Kosterlitz, J. Phys. C 7, 1046 (1974).
- [4] N. D. Mermin, H. Wagner, Phys. Rev. Lett. 17, 1133-1136 (1966).
- [5] P. C. Hohenberg, Phys. Rev. 158, 383 (1967).
- [6] Sidney Coleman, Commun. Math. Phys. 31, 259 (1973).
- [7] A. Altland, B. Simons, Condensed Matter Field Theory, Cambridge University Press 2010.
- [8] Xiao-Gang Wen, Quantum Field Theory of Many-body Systems: From the Origin of Sound to an Origin of Light and Electrons, Oxford University Press 2004.
- [9] N. Nagaosa, Quantum Field Theory in Condensed Matter Physics, Springer-Verlag Berlin Heidelberg 1999.
- [10] J. Polchinski, String Theory, Cambridge University Press, Cambridge, UK (1998).



## 6. 两维伊辛模型的精确解

本文是一个关于两维伊辛模型的讲义，核心内容是关于两维伊辛模型的精确解。我们介绍了两种不同的求解方法：一种是 Schultz-Mattis-Lieb 1964 年在 RMP 的综述文章中给出的方法，当然这种方法其实是 Kaufman 1949 年方法的修订版；另一种方法是，1980 年 Stuart Samuel 给出的利用 Grassmann 变量积分来求解的办法。当然，为了给这两种方法做准备，我们还粗略介绍了伊辛模型的高（低）温展开，以及 Kramers-Wannier 对偶。

### 6.1 模型的提出和伊辛的“错误证明”

本文要处理的是两维伊辛模型。但我们想首先回到一维情形，因为伊辛模型首先是在一维情形下被提出的。这是一个关于科学史上著名“失败”的精彩故事，其中蕴含着深刻的物理教训。让我们回顾一下这段历史。

Wilhelm Lenz (1888-1957) 是德国物理学家，曾在汉堡大学任教，是量子力学发展中的重要人物（著名的“兰斯-艾森伯交换作用”中的“兰斯”就是他）。1920 年，他在研究铁磁性的微观机制时，提出了一个极简的模型：其核心思想是，铁磁性可能源于相邻原子磁矩之间的相互作用，而不需要外场。Lenz 将这个想法告诉了他的研究生 Ernst Ising (1900-1998)，建议他用这个模型来研究铁磁性的起源。这就是伊辛 (Ising) 模型的来源。

#### 模型的定义与 Ising 的“结论”

Lenz 提出的模型如下：一维链上排列着  $N$  个格点，每个格点  $i$  上有一个可以取两个值的“自旋”变量： $s_i = \pm 1$ （+1 表示自旋向上，-1 表示自旋向下）。最近邻自旋之间有相

互作用，哈密顿量为：

$$H = -J \sum_{i=1}^{N-1} s_i s_{i+1},$$

其中  $J > 0$  为铁磁耦合（也即是相邻自旋同向时能量更低）。加上外磁场  $h$  后：

$$H = -J \sum_{i=1}^{N-1} s_i s_{i+1} - h \sum_{i=1}^N s_i.$$

这就是著名的 **Ising 模型**——但注意，其实是 Lenz 提出了它，Ising 只是做了计算。

Ising 在 1924 年完成了他的博士论文，研究了一维 Ising 模型的统计力学性质。他的核心结论是：“**一维 Ising 模型在任何有限温度下都不存在铁磁相变。**”换句话说，一维情况下没有自发磁化。

Ising 的原始论证大体如下（简化版）：Ising 考虑了长程序（long-range order）的存在性。他的思路大致如下：

1. 基态分析 ( $T = 0$ ): 在温度  $T = 0$  时，所有自旋同向排列时能量最低： $E_0 = -J(N-1)$ ，此时存在完全的铁磁长程序，形如 ( $\uparrow\uparrow\uparrow \cdots \uparrow\uparrow$ )。

2. 激发态分析 ( $T > 0$ ): 热激发会产生畴壁 (domain walls)，在畴壁两边的自旋指向刚好相反，形如 ( $\uparrow \cdots \uparrow\downarrow \cdots \downarrow$ )。因此，产生每个畴壁的能量代价为  $\Delta E = 2J$ 。

3. Ising 的关键论证：在一维链中，畴壁是一个点（零维）。在有限温度  $T$  之下，熵的贡献倾向于产生畴壁。Ising 估计：畴壁的熵约为  $S \sim k_B \ln N$ （畴壁可以在  $N$  个位置中的任意一个）。因此每个畴壁对应的自由能贡献为： $\Delta F = 2J - k_B T \ln N$

4. Ising 的结论：对于任何有限温度，在  $N \rightarrow \infty$  的热力学极限之下， $k_B T \ln N \rightarrow \infty$ 。因此  $\Delta F < 0$ ，系统总是趋向于产生很多畴壁。从而使得所有自旋均同向的长程序被破坏，因此平均来说，没有自发磁化，从而没有相变。

Ising 的论证大体是正确的（一维确实没有相变），虽然推理过程有一些缺陷，尤其是这种论证未严格计算配分函数，从而逻辑不够严密。Ising 幸运地得出了正确结论，但他的论证不够严谨。这在科学史上并不罕见——正确的结论往往比正确的证明更早出现。但问题是 Ising 错误地将这个一维情形下得到的结论推广到了更高维。Ising 在论文中**断言**：“**这个模型在二维或三维也不会有相变。**”

这是完全错误的。Onsager 在 1944 年证明了二维 Ising 模型存在严格的相变。Onsager 的结果打破了 Ising 的悲观预言，证明了统计物理模型的丰富性。更重要的是，证明了统计物理可以用来理解相变现象。

Ising 为何做出这个错误推断？他可能认为“维度增加不会改变”畴壁总是破坏有序”的图像。他没有意识到高维中畴壁的拓扑性质改变（一维是点，二维是线，但不需要是直线，三维是面，但不需要是平面），从而导致高维中畴壁的能量-熵竞争与一维有本质不同，可以存在有序相。

Ising 的错误断言使该模型被忽视了约 20 年。Onsager 1944 年的精确解因此更加震撼，开创了统计物理的新纪元。有意思的是：Onsager 在 1944 年的论文中只给出了结果，没有展示推导过程。他后来在 1947 年的一个会议黑板上写下了关键的旋量方法，但从未

发表完整推导。Kaufman 在 1949 年给出了更清晰的方法。讨论两维 Ising 模型的精确解就是本文后面的核心内容。

### 一维没有相变的严格证明 (转移矩阵方法)

一维 Ising 模型的严格解实际上非常简单, 可以用所谓的**转移矩阵方法**。为了说清楚这种方法, 我们不妨设想  $N$  个自旋围成了一个首尾相连的圆周, 满足  $s_{N+1} \equiv s_1$  的周期性边界条件, 从而可以将一维 Ising 模型的哈密顿量重写为

$$H = -J \sum_{i=1}^N s_i s_{i+1} - h \frac{1}{2} \sum_{i=1}^N (s_i + s_{i+1}). \quad (6.1)$$

于是系统的配分函数可以写成

$$\mathcal{Z} = \sum_{\{s_i\}} \prod_{i=1}^N \exp \left[ \beta J s_i s_{i+1} + \beta h \frac{1}{2} (s_i + s_{i+1}) \right]. \quad (6.2)$$

现在我们引入一个  $2 \times 2$  的转移矩阵  $T$ , 矩阵元为  $T_{s_i s_{i+1}}$ , 其定义为:

$$T = \begin{pmatrix} e^{K+h'} & e^{-K} \\ e^{-K} & e^{K-h'} \end{pmatrix}.$$

其中  $K = \beta J$ ,  $h' = \beta h$ 。于是配分函数  $\mathcal{Z}$  刚好可以写成  $T_{s_i s_{i+1}}$  相乘的形式

$$\mathcal{Z} = \sum_{\{s_i\}} T_{s_1 s_2} T_{s_2 s_3} \cdots T_{s_N s_1} = \text{Tr}(T^N) = \lambda_+^N + \lambda_-^N.$$

式中  $\lambda_{\pm}$  为转移矩阵  $T$  的本征值:

$$\lambda_{\pm} = e^K \cosh h' \pm \sqrt{e^{2K} \sinh^2 h' + e^{-2K}}.$$

**关键观察:** 对于任何  $T > 0$  (即  $K < \infty$ ),  $\lambda_+ > \lambda_- > 0$ , 因此在  $N \rightarrow +\infty$  的热力学极限之下,  $\lambda_-$  对配分函数的贡献可以忽略, 即有  $\mathcal{Z} \approx \lambda_+^N$ 。从而不难发现自由能密度  $f = -k_B T \log \mathcal{Z} / N \approx -k_B T \ln \lambda_+$  是  $T$  的解析函数, 从而系统不存在相变。而且不难算出, 磁化强度  $m = -\frac{\partial f}{\partial h}$  在  $h \rightarrow 0^+$  时趋于零, 从而没有自发磁化。**严格结论:** 一维 Ising 模型的自由能在  $T > 0$  时解析, 不存在相变。

## 6.2 两维伊辛模型, Kramers-Wannier 对偶

下面我们把伊辛模型从一维推广到两维。我们考虑一个二维正方晶格上的零场伊辛模型, 所谓零场指的是外磁场  $h = 0$ 。其中所有自旋变量  $s_i$  都定义在正方晶格的格点上, 下标  $i$  表示第  $i$  个格点。这个模型的哈密顿量为

$$H = -J \sum_{\langle i, j \rangle} s_i s_j, \quad (6.3)$$

其中  $\langle i, j \rangle$  表示  $i, j$  两个格点是最近邻格点, 我们要对这样的最近邻格点对进行求和。从而系统的配分函数为 ( $K = \beta J$ )

$$\mathcal{Z} = \sum_{\{s_i\}} \prod_{\langle i, j \rangle} e^{K s_i s_j}. \quad (6.4)$$

### 高温展开

注意到  $s_i = \pm 1$ , 从而  $(s_i)^2 = 1$ , 因此有

$$e^{K s_i s_j} = \cosh(K) + s_i s_j \sinh(K). \quad (6.5)$$

进而可以将上述配分函数写成

$$\mathcal{Z} = [\cosh(K)]^{2N} \sum_{\{s_i\}} \prod_{\langle i, j \rangle} [1 + s_i s_j \tanh(K)]. \quad (6.6)$$

注意对于  $N$  个格点的正方格子, 每个格点有 4 条边, 每条边被 2 个格点共享, 从而总边数为  $4N/2 = 2N$ 。在温度很高时,  $K$  是小量, 从而  $\tanh(K) \ll 1$  可以看作小量, 因此可以将上式对  $\tanh(K)$  进行级数展开。这个展开的领头阶就是每个乘积因子都取 1。接下来的贡献是取少量  $\tanh(K)$  因子, 而多取因子 1。但需要注意的是, 每一个  $\tanh(K)$  因子必然连带着一对自旋因子  $s_i s_j$  一起出现。由于最后需要对每一个格点上的自旋  $s_i$  的两个不同值  $\pm 1$  求和, 所以如果某个特定的  $s_i$  在级数展开中仅出现奇数次, 那由于  $(+1) + (-1) = 0$ , 这一项最终的贡献就必定为零。因此, 所有的非零贡献中, 必定每个自旋变量都出现偶数次 (包括零次), 从而所有的自旋变量都会抵消 (由于  $(s_i)^2 = 1$ )。

我们可以将每一个  $s_i s_j \tanh(K)$  因子的近邻对  $\langle i, j \rangle$  画一条粗线连接起来。那么级数展开的每一非零贡献中, 这些近邻对的连线必定会形成一条闭合回路, 因为只有闭合回路才能保证其连接的每一个格点上的  $s_i$  都出现偶数次。如此一来, 就可以将 (6.6) 式展开的每一项与一个如图 (6.1) 所示的回路图形  $\Gamma$  联系起来。每一个特定回路图形中的每一个近邻对链接对配分函数的贡献为  $\tanh(K)$ 。所以, 最终可以将配分函数 (6.6) 写成

$$\mathcal{Z}(K) = [\cosh(K)]^{2N} 2^N \sum_{\Gamma} [\tanh(K)]^{L(\Gamma)}, \quad (6.7)$$

式中需要对所有可能的回路图形求和,  $L(\Gamma)$  表示图形  $\Gamma$  的总周长 (以晶格格距为单位), 式中前面的因子  $2^N$  是来自于每个格点上的自旋都有 2 种取值, 而算配分函数时要对所有自旋的不同取值进行求和。值得注意的是, 这里的多边形回路  $\Gamma$  是可以交叉重叠的回路。

### 低温展开

而当温度很低时,  $K$  很大。这时候可以将配分函数  $\mathcal{Z}$  写成  $\mathcal{Z} = \sum_n \Omega_n e^{-\beta E_n}$ , 式中  $E_n$  是伊辛模型第  $n$  个可能的总能量取值,  $\Omega_n$  表示具有这种总能量值的自旋位形数目。当温度很低时,  $\beta$  很大, 我们可以将不同的能量取值按从低到高排列,  $E_0 < E_1 < E_2 < \dots$ , 由于  $\beta$  很大, 所以显然, 能量越低的位形对配分函数的贡献越大。

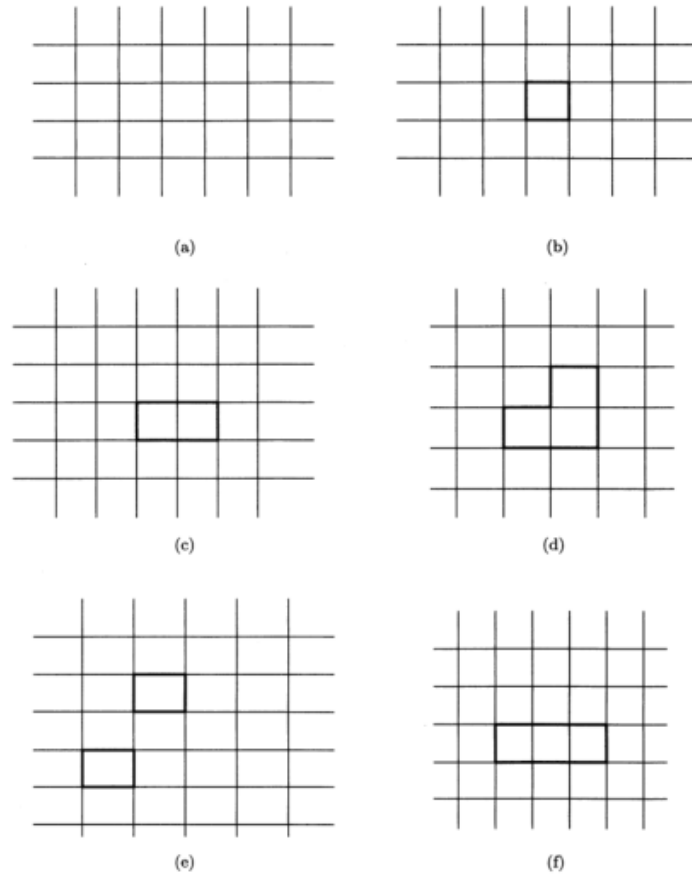


Figure 6.1: 高温展开图示例。图 (a)、(b)、(c) 分别为长度等于 0, 4, 6 的回路；图 (d)、(e)、(f) 是长度等于 8 的回路，它有三种不同的拓扑结构。

首先，二维方格子 Ising 模型有两个基态，即两个最低能的自旋位形，分别对应自旋全向上或者全向下，能量为  $E_0 = -2NJ$  (注意总边数为  $2N$ )，或者说  $\beta E_0 = -2NK$ 。

激发态 (高能量的自旋位形) 可以通过在基态的基础上翻转自旋而成，所以关键是考虑两个不同自旋取向区域的分界线 (称作畴壁, domain walls)，因为自旋的相对翻转就是发生在这样的分界线上。我们可以把这样的分界线画在对偶的格子上 (所谓对偶格子，如图 (6.2) 所示，就是以原格子每个正方形的中心为新格点所形成的新格子，很显然，正方格子的对偶格子依然是正方格子)，不难明白，原格子上的畴壁正好是对偶格子上的闭合回路  $\Gamma$ 。记  $L(\Gamma)$  为畴壁总长度，也就是  $\Gamma$  在对偶格子上的边数。则很显然，每一个  $\Gamma$  贡献的能量为  $E = E_0 + 2JL(\Gamma)$ ，或者说  $\beta E = -2NK + 2KL(\Gamma)$ 。很显然，配分函数的低温展开就是对畴壁进行展开，最后即有

$$\mathcal{Z}(K) = 2e^{2NK} \sum_{\Gamma} [e^{-2K}]^{L(\Gamma)}. \quad (6.8)$$

其中前面的因子 2 来自于基态的二重简并。另外，由于现在的回路  $\Gamma$  是原格子上两个区域的分界线，所以它是不能交叉重叠的 (与高温展开的情形不同)。不难看出，在这种低温

的情况下，大的畴壁是被压制的。但是，当温度很高时，这种低温展开就失效了，这意味着高温时，畴壁回路可以自由涨落。

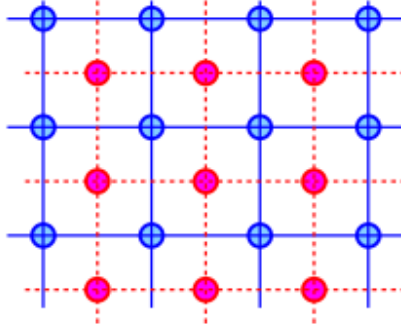


Figure 6.2: 用两种不同颜色格点标记的原格子与其对偶格子

对比 (6.7) 式的高温展开和 (6.8) 式的低温展开，可以看出，两者都是对某种闭合回路进行展开，只不过高温展开是原格子上的闭合回路，而低温展开是对偶格子上的闭合回路。另外，高温展开是以  $\tanh(K)$  为每条边的权重，而低温展开则是以  $e^{-2K}$  为每条边的权重。因此，两个展开具有相同的组合结构！这暗示存在变量替换使得它们等价。

### Kramers-Wannier 对偶

Kramers-Wannier 指出，我们可以在对偶格子上定义一个对偶的伊辛模型，它不是以  $K$  而是以对偶的  $K^*$  为耦合常数，式中  $K^*$  的定义为

$$e^{-2K^*} \equiv \tanh(K). \quad (6.9)$$

我们也可以利用双曲函数的恒等式将上式改成更对称的

$$\sinh(2K) \sinh(2K^*) = 1. \quad (6.10)$$

则原格子配分函数的高温展开为

$$\begin{aligned} \frac{\mathcal{Z}(K)}{[2 \sinh(2K)]^{N/2}} &= \frac{2^N [\cosh(K)]^{2N}}{[2 \sinh(2K)]^{N/2}} \sum_{\Gamma} [\tanh(K)]^{L(\Gamma)} \\ &= \frac{2^N [\cosh^2(K)]^N}{[2 \sinh(2K)]^{N/2}} \sum_{\Gamma} [e^{-2K^*}]^{L(\Gamma)}. \end{aligned} \quad (6.11)$$

利用  $\sinh(2K) \sinh(2K^*) = 1$  以及  $\cosh^2(K) = \frac{e^{4K^*}}{e^{4K^*} - 1}$  (请证明)，不难得到

$$\begin{aligned} \frac{2^N [\cosh(K)]^{2N}}{[2 \sinh(2K)]^{N/2}} &= \frac{2^N e^{4NK^*}}{(e^{4K^*} - 1)^N \left(\frac{2}{\sinh(2K^*)}\right)^{N/2}} = \frac{2^{N/2} e^{4NK^*} \sinh^{N/2}(2K^*)}{(e^{4K^*} - 1)^N} \\ &= \frac{2^{N/2} e^{4NK^*} \sinh^{N/2}(2K^*)}{e^{2NK^*} 2^N \sinh^N(2K^*)} = \frac{e^{2NK^*}}{[2 \sinh(2K^*)]^{N/2}}. \end{aligned} \quad (6.12)$$

代入上面的关于原格子的高温展开，即有

$$\frac{\mathcal{Z}(K)}{[2\sinh(2K)]^{N/2}} = 2 \frac{e^{2NK^*}}{[2\sinh(2K^*)]^{N/2}} \sum_{\Gamma} [e^{-2K^*}]^{L(\Gamma)}. \quad (6.13)$$

很显然，结果正好是对偶格子配分函数的低温展开除以  $[2\sinh(2K^*)]^{N/2}$ ，式中已经考虑到高温展开的  $\Gamma$  可以交叉重叠，而低温展开的不能，这使得式子前面多出一个因子 2 (严格的讨论要稍微复杂一些，这里从略)。换言之，我们有

$$\frac{\mathcal{Z}(K)}{[2\sinh(2K)]^{N/2}} = \frac{\mathcal{Z}(K^*)}{[2\sinh(2K^*)]^{N/2}}. \quad (6.14)$$

Kramers-Wannier 对偶变换将  $K \leftrightarrow K^*$ 。对于自对偶点， $K = K^* = K_c$ ，由 (6.10) 式，很显然有

$$\sinh(2K_c) = 1. \quad (6.15)$$

不难解得

$$\begin{aligned} K_c &= \frac{1}{2} \operatorname{arsinh}(1) = \frac{1}{2} \ln(1 + \sqrt{2}) \approx 0.440686\dots \\ \Leftrightarrow \frac{J}{k_B T_c} &\approx 0.440686\dots \Rightarrow T_c \approx 2.269J/k_B. \end{aligned} \quad (6.16)$$

人们可以合理假设 Ising 模型只有一个相变点，则这个点必定只能是自对偶点  $K_c$ ，相应的相变温度必定是  $T_c$ ，否则相变点就至少为两个，因为  $K$  和与之对偶的  $K^*$  在物理上是等价的。另外，由 (6.10) 式不难看出， $K$  小 (原格子高温相) 则  $K^*$  大 (对偶格子低温相)，反之亦然。所以 Kramers-Wannier 对偶变换交换了原格子和对偶格子的高低相。对于正方格子，由于其对偶格子依然为一样的正方格子，所以相变点只能是自对偶点。如此一来，通过 Kramers-Wannier 对偶性，我们就可以在不用精确求解完整配分函数的情况下，确定了相变点的温度。这是 Kramers 和 Wannier 在 1941 年的工作，早于 Onsager 1944 年的精确解。

### 6.3 费米子方法

本节将讨论本质上由 Kaufman 在 1949 年给出的求解两维 Ising 模型的费米子方法，当然我们采用的是 Schultz-Mattis-Lieb 的修改版本。

为了精确求解两维伊辛模型 (假定  $h=0$ )，不妨考虑一个  $N$  行  $M$  列的正方格子，在格子的四周加上周期性边界条件。相应的伊辛模型哈密顿量可以写成

$$H = -J_1 \sum_n \sum_m s_{n,m} s_{n+1,m} - J_2 \sum_n \sum_m s_{n,m} s_{n,m+1}, \quad (6.17)$$

式中我们已经把第  $n$  行第  $m$  列的格点上的自旋写成了  $s_{n,m}$ ，并且我们记相邻行之间的耦合常数为  $J_1$ ，而相邻列之间的耦合常数为  $J_2$ 。类似的，我们记  $K_1 = \beta J_1$ ， $K_2 = \beta J_2$ 。

#### 转移矩阵

不妨记  $\mathbf{s}_n = (s_{n,1}, s_{n,2}, \dots, s_{n,M})$ , 由于每一个自旋有两个不同取值, 所以这是一个  $2^M$  个分量的矢量。则不难看出, 系统的配分函数可以写成

$$\mathcal{Z} = \sum_{\{\mathbf{s}_n\}} (e^{K_1 \mathbf{s}_1 \cdot \mathbf{s}_2} e^{K_2 \sum_m^M s_{1,m} s_{1,m+1}}) \dots (e^{K_1 \mathbf{s}_N \cdot \mathbf{s}_1} e^{K_2 \sum_m^M s_{N,m} s_{N,m+1}}). \quad (6.18)$$

其中小括号里的项仅仅包含同一行之内的相互作用以及相邻行之间的相互作用。类似的, 记  $\mathbf{s} = (s_1, s_2, \dots, s_M)$ 。不妨定义如下  $2^M \times 2^M$  的矩阵  $V_1, V_2$  (分别描述行间和行内相互作用)

$$V_{1, \mathbf{s} \mathbf{s}'} \equiv e^{K_1 \mathbf{s} \cdot \mathbf{s}'} \quad V_{2, \mathbf{s} \mathbf{s}'} \equiv e^{K_2 \sum_m^M s_m s_{m+1}} \delta_{\mathbf{s}, \mathbf{s}'}. \quad (6.19)$$

则很显然, 上述配分函数可以重写成

$$\mathcal{Z} = \sum_{\{\mathbf{s}_n\}} [(V_1 V_2)_{\mathbf{s}_1 \mathbf{s}_2} (V_1 V_2)_{\mathbf{s}_2 \mathbf{s}_3} \dots (V_1 V_2)_{\mathbf{s}_N \mathbf{s}_1}] = \text{Tr}(V_1 V_2)^N. \quad (6.20)$$

利用矩阵求迹的性质, 也可以写成更对称的

$$\mathcal{Z} = \text{Tr}(V_2^{1/2} V_1 V_2^{1/2})^N. \quad (6.21)$$

当然,  $V \equiv V_2^{1/2} V_1 V_2^{1/2}$  就是所谓的转移矩阵。

引入作用在第  $m$  列上的泡利矩阵  $\vec{\sigma}_m = (\sigma_m^x, \sigma_m^y, \sigma_m^z)$ , 其行列指标分别用自旋变量来标记, 即记作  $s_m, s'_m$ , 从而  $(\sigma_m^z)_{s_m s'_m} = s_m \delta_{s_m s'_m}$ 。由此不难看出

$$V_2 = e^{K_2 \sum_m^M \sigma_m^z \sigma_{m+1}^z}. \quad (6.22)$$

注意, Kramers-Wannier 对偶变换是完全对称的, 所以必定也有

$$e^{-2K} \equiv \tanh(K^*). \quad (6.23)$$

据此不妨定义

$$e^{K_1} \equiv A(K_1) \cosh(K_1^*), \quad e^{-K_1} \equiv A(K_1) \sinh(K_1^*). \quad (6.24)$$

从而  $1 = A^2 \sinh(K_1^*) \cosh(K_1^*) = A^2 \frac{1}{2} \sinh(2K_1^*) = A^2 / (2 \sinh(2K_1))$ , 从而

$$A(K_1) = [2 \sinh(2K_1)]^{1/2}. \quad (6.25)$$

为了处理非对角矩阵  $V_1$ , 我们注意到  $e^{K_1 s_m s'_m}$  可以用  $2 \times 2$  矩阵形式写成

$$e^{K_1 s_m s'_m} = \begin{pmatrix} e^{K_1} & e^{-K_1} \\ e^{-K_1} & e^{K_1} \end{pmatrix} = A(K_1) \begin{pmatrix} \cosh(K_1^*) & \sinh(K_1^*) \\ \sinh(K_1^*) & \cosh(K_1^*) \end{pmatrix} \quad (6.26)$$

另一方面, 注意到  $e^{K_1^* \sigma_m^x}$  (利用  $(\sigma^x)^2 = 1$ ) 正好可以写成

$$e^{K_1^* \sigma_m^x} = \cosh(K_1^*) 1_{2 \times 2} + \sigma_m^x \sinh(K_1^*) = \begin{pmatrix} \cosh(K_1^*) & \sinh(K_1^*) \\ \sinh(K_1^*) & \cosh(K_1^*) \end{pmatrix}. \quad (6.27)$$

所以我们有

$$e^{K_1 s_m s'_m} = [2 \sinh(2K_1)]^{1/2} (e^{K_1^* \sigma_m^x})_{s_m s'_m} \quad (6.28)$$

从而很显然有

$$V_1 = [2 \sinh(2K_1)]^{M/2} e^{K_1^* \sum_m^M \sigma_m^x}. \quad (6.29)$$

为了把矩阵  $V_1, V_2$  写成更方便的形式, 我们做一个相似变换 (不改变矩阵求迹), 也就是一个绕  $y$  轴的  $90$  度旋转, 使得

$$\sigma^x \rightarrow -\sigma^z, \quad \sigma^z \rightarrow \sigma^x. \quad (6.30)$$

从而可以将  $V_1, V_2$  写成

$$V_1 = [2 \sinh(2K_1)]^{M/2} e^{-K_1^* \sum_m^M \sigma_m^z}, \quad V_2 = e^{K_2 \sum_m^M \sigma_m^x \sigma_{m+1}^x}. \quad (6.31)$$

注意, 在这个新的形式下,  $V_1$  才是对角的。

### Jordan-Wigner 变换

类似于前面的一维情形, 为了算出配分函数, 我们至少需要求出转移矩阵最大的本征值。这可以通过利用所谓的 Jordan-Wigner 变换的方法来完成。

我们定义  $\sigma^+ \equiv \frac{1}{2}(\sigma^x + i\sigma^y)$ ,  $\sigma^- \equiv \frac{1}{2}(\sigma^x - i\sigma^y)$ , 不难得到  $(\sigma^\pm)^2 = 0$ ,  $\sigma^+ \sigma^- = \frac{1}{2}(1 + \sigma^z)$ ,  $\sigma^- \sigma^+ = \frac{1}{2}(1 - \sigma^z)$ , 从而  $[\sigma^+, \sigma^-] = \sigma^z$ , 同时反对易子  $\{\sigma^+, \sigma^-\} = 1$ 。

Jordan-Wigner 变换是定义如下算符

$$\begin{aligned} c_m &\equiv \exp\left(i\pi \sum_{j=1}^{m-1} \sigma_j^+ \sigma_j^-\right) \sigma_m^- \\ c_m^\dagger &\equiv \exp\left(i\pi \sum_{j=1}^{m-1} \sigma_j^+ \sigma_j^-\right) \sigma_m^+. \end{aligned} \quad (6.32)$$

我们的目标是要验证这两类算符构成了费米子的产生湮灭算符。为了验证一些代数关系, 下面式子是有用的 (请证明)

$$(\sigma^+ \sigma^-)^2 = \sigma^+ \sigma^- \Rightarrow e^{\theta \sigma^+ \sigma^-} = 1 + \sigma^+ \sigma^- (e^\theta - 1) \Rightarrow e^{2\pi i \sigma^+ \sigma^-} = 1. \quad (6.33)$$

由此不难验证,  $c_m$  与  $c_m^\dagger$  的确互为厄米共轭。另外, 由此不难算得

$$c_m^\dagger c_m = \sigma_m^+ \sigma_m^-, \quad c_m c_m^\dagger = \sigma_m^- \sigma_m^+. \quad (6.34)$$

所以的确有

$$\{c_m^\dagger, c_m\} = 1. \quad (6.35)$$

但是, 为了真正描述费米子, 我们还需要验证, 不同指标  $m$  的这两类算符是反对易的。验证如下: 对于  $n > m$ , 我们有

$$\begin{aligned}
c_n c_m^\dagger &= \sigma_n^- \exp\left(i\pi \sum_{j=1}^{n-1} \sigma_j^+ \sigma_j^-\right) \exp\left(i\pi \sum_{j=1}^{m-1} \sigma_j^+ \sigma_j^-\right) \sigma_m^+ \\
&= \sigma_n^- e^{2\pi i \sum_{j=1}^{m-1} \sigma_j^+ \sigma_j^-} e^{i\pi \sigma_m^+ \sigma_m^-} e^{i\pi \sum_{j=m+1}^{n-1} \sigma_j^+ \sigma_j^-} \sigma_m^+ \\
&= e^{i\pi \sum_{j=m+1}^{n-1} \sigma_j^+ \sigma_j^-} \sigma_n^- e^{i\pi \sigma_m^+ \sigma_m^-} \sigma_m^+ \\
&= e^{i\pi \sum_{j=m+1}^{n-1} \sigma_j^+ \sigma_j^-} \sigma_n^- [1_m + \sigma_m^+ \sigma_m^- (e^{i\pi} - 1)] \sigma_m^+ \\
&= e^{i\pi \sum_{j=m+1}^{n-1} \sigma_j^+ \sigma_j^-} \sigma_n^- [\sigma_m^+ - 2\sigma_m^+ \sigma_m^- \sigma_m^+] \\
&= e^{i\pi \sum_{j=m+1}^{n-1} \sigma_j^+ \sigma_j^-} \sigma_n^- [\sigma_m^+ - 2\sigma_m^+] \\
&= -e^{i\pi \sum_{j=m+1}^{n-1} \sigma_j^+ \sigma_j^-} \sigma_n^- \sigma_m^+.
\end{aligned} \tag{6.36}$$

类似的, 可以算得

$$c_m^\dagger c_n = e^{i\pi \sum_{j=m+1}^{n-1} \sigma_j^+ \sigma_j^-} \sigma_m^+ \sigma_n^-. \tag{6.37}$$

所以, 的确有  $c_m^\dagger, c_n$  反对易。其它情形的反对易均可以类似地验证。

由于  $\sigma_m^z = 2\sigma_m^+ \sigma_m^- - 1_m = 2c_m^\dagger c_m - 1_m$ , 所以

$$V_1 = [2 \sinh(2K_1)]^{M/2} e^{-K_1^* \sum_m \sigma_m^z} = [2 \sinh(2K_1)]^{M/2} e^{-2K_1^* \sum_m (c_m^\dagger c_m - \frac{1}{2})}. \tag{6.38}$$

为了得到算符  $V_2$  的费米子表达式, 我们注意到

$$\sigma_m^x \sigma_{m+1}^x = (\sigma_m^+ + \sigma_m^-)(\sigma_{m+1}^+ + \sigma_{m+1}^-) = \sigma_m^+ \sigma_{m+1}^+ + \sigma_m^- \sigma_{m+1}^- + \sigma_m^+ \sigma_{m+1}^- + \sigma_m^- \sigma_{m+1}^+.$$

另一方面, 对于  $m < M$ , 利用和前面验证费米子反对易关系时类似的计算, 可以算得

$$c_m^\dagger c_{m+1} = \sigma_m^+ \sigma_{m+1}^-, \quad c_m c_{m+1} = -\sigma_m^- \sigma_{m+1}^-, \quad c_m^\dagger c_{m+1}^\dagger = \sigma_m^+ \sigma_{m+1}^+. \tag{6.39}$$

代入前面的结果即有

$$\begin{aligned}
\sigma_m^x \sigma_{m+1}^x &= c_m^\dagger c_{m+1}^\dagger - c_m c_{m+1} + c_m^\dagger c_{m+1} + c_{m+1}^\dagger c_m \\
&= (c_m^\dagger - c_m)(c_{m+1}^\dagger + c_{m+1}).
\end{aligned} \tag{6.40}$$

而对于  $m = M$ , 我们需要考虑到周期性边界条件, 这时候可以算得

$$\begin{aligned}
\sigma_M^+ \sigma_1^- &= -(-)^F c_M^\dagger c_1 \neq c_M^\dagger c_1 \\
\sigma_M^+ \sigma_1^+ &= -(-)^F c_M^\dagger c_1^\dagger \neq c_M^\dagger c_1^\dagger \\
\sigma_M^- \sigma_1^- &= (-)^F c_M c_1 \neq -c_M c_1.
\end{aligned} \tag{6.41}$$

式中  $F$  是总费米子数目算符

$$F \equiv \sum_m^M \sigma_m^+ \sigma_m^- = \sum_m^M c_m^\dagger c_m. \tag{6.42}$$

因此, 综合上面所有结果, 即知算符  $V_2$  可以用费米子表达成

$$V_2 = e^{K_2 \sum_m^M \sigma_m^x \sigma_{m+1}^x} = e^{K_2 \left[ \sum_{m=1}^{M-1} (c_m^\dagger - c_m)(c_{m+1}^\dagger + c_{m+1}) - (-)^F (c_M^\dagger - c_M)(c_1^\dagger + c_1) \right]}. \quad (6.43)$$

很显然, 算符  $(-)^F$  数的是有偶数个费米子还是奇数个费米子, 所以称作费米宇称算符。由于算符  $V_1$  和  $V_2$  的指数上涉及的都是费米子算符的二次型 (偶数个费米子), 所以很显然

$$[(-)^F, V_1] = [(-)^F, V_2] = 0. \quad (6.44)$$

因此我们可以分别考虑转移矩阵  $V$  在费米奇宇称态 ( $(-)^F$  的本征值为负) 和费米偶宇称态 ( $(-)^F$  的本征值为正) 上的作用。很明显, 在费米偶宇称态上  $V_2$  等价于

$$V_2^+ = e^{K_2 \left[ \sum_{m=1}^M (c_m^\dagger - c_m)(c_{m+1}^\dagger + c_{m+1}) \right]}, \text{ 其中, } c_{M+1} = -c_1, c_{M+1}^\dagger = -c_1^\dagger. \quad (6.45)$$

在费米偶奇称态上  $V_2$  等价于

$$V_2^- = e^{K_2 \left[ \sum_{m=1}^M (c_m^\dagger - c_m)(c_{m+1}^\dagger + c_{m+1}) \right]}, \text{ 其中, } c_{M+1} = c_1, c_{M+1}^\dagger = c_1^\dagger. \quad (6.46)$$

两种情况相应的转移矩阵  $V = V_2^{1/2} V_1 V_2^{1/2}$  我们分别记作  $V^+, V^-$ 。

因此求转移矩阵本征值的问题就变成一个对角化三个算符的乘积算符的问题, 这三个算符中的每一个都是费米子二次型的指数化。这个问题虽然不简单, 但并不是不可解决的。

### 傅里叶变换

因为整个系统是平移不变的, 所以不妨先进行如下傅里叶变换

$$c_m = \frac{1}{\sqrt{M}} e^{-i\pi/4} \sum_q e^{iqm} \eta_q. \quad (6.47)$$

其中引入额外的因子  $e^{-i\pi/4}$  是为了确保当将  $V^\pm$  用  $\eta_q$  来表达时, 所有项的系数均为实数。进而不难看出, 偶费米宇称态的周期性边界条件 (6.45) 要求

$$q \equiv l = \pm\pi/M, \pm 3\pi/M, \dots, \pm(M-1)\pi/M. \quad (6.48)$$

而奇费米宇称态的周期性边界条件 (6.46) 要求

$$q \equiv k = 0, \pm 2\pi/M, \pm 4\pi/M, \dots, \pm(M-2)\pi/M, \pi. \quad (6.49)$$

这里为了方便起见, 我们已经假定  $M$  为偶数。

对于偶费米宇称态, 显然, 傅里叶变换的性质确保

$$\sum_m (c_m^\dagger c_m - 1/2) = \sum_{q=l} (\eta_q^\dagger \eta_q - 1/2) = \sum_{q=l>0} (\eta_q^\dagger \eta_q + \eta_{-q}^\dagger \eta_{-q} - 1). \quad (6.50)$$

对于奇费米宇称态,  $q = k = 0, \pi$  这两种情况要单独处理, 其它都和偶费米宇称态一样。由此即可以知道  $V_1$  如何用  $\eta_q$  来表达。

为了得到算符  $V_2$  用  $\eta_q$  表达的式子, 我们依然先考虑偶费米宇称态, 经过一些繁琐的计算, 并利用  $\sum_{q=l>0} \cos(q) = 0$ , 可以得到此时

$$\begin{aligned} & \sum_{m=1}^M (c_m^\dagger - c_m)(c_{m+1}^\dagger + c_{m+1}) \\ &= 2 \sum_{q=l>0} [(\sin q)(\eta_{-q}^\dagger \eta_q^\dagger + \eta_q \eta_{-q}) + (\cos q)(\eta_q^\dagger \eta_q + \eta_{-q}^\dagger \eta_{-q})]. \end{aligned} \quad (6.51)$$

对奇费米宇称态, 讨论也是类似的, 只是需要单独处理  $q = k = 0, \pi$  的贡献。

综合所有这些结果, 可以得到

$$V^\pm = [2 \sinh(2K_1)]^{M/2} \prod_{0 \leq q \leq \pi} V_q, \quad (6.52)$$

式中,  $q$  要么取  $l$  类可能性, 要么取  $k$  类可能性。对于  $q \neq 0, \pi$ ,

$$V_q = (V_{2q})^{1/2} V_{1q} (V_{2q})^{1/2}, \quad (6.53)$$

其中

$$\begin{aligned} V_{1q} &= \exp[-2K_1^*(\eta_q^\dagger \eta_q + \eta_{-q}^\dagger \eta_{-q} - 1)] \\ V_{2q} &= \exp\{2K_2[(\sin q)(\eta_{-q}^\dagger \eta_q^\dagger + \eta_q \eta_{-q}) + (\cos q)(\eta_q^\dagger \eta_q + \eta_{-q}^\dagger \eta_{-q})]\}. \end{aligned} \quad (6.54)$$

而对于  $q = 0$ ,

$$V_0 = \exp\left[-2(K_1^* - K_2)(\eta_0^\dagger \eta_0 - \frac{1}{2})\right], \quad (6.55)$$

而对于  $q = \pi$ ,

$$V_\pi = \exp\left[-2(K_1^* + K_2)(\eta_\pi^\dagger \eta_\pi - \frac{1}{2})\right]. \quad (6.56)$$

### 对角化

由于  $V_q$  和  $V_{q'}$  这样的算符涉及到费米子的二次型, 费米子是反对易的, 其二次型对易, 因此对于两个不同的  $q$  和  $q'$ ,  $V_q$  和  $V_{q'}$  必定相互对易, 从而可以被同时对角化。对于  $V_0$  和  $V_\pi$ , 很显然, 在占有数表象中, 它们已经是对角的了。而很显然, 对于更一般的  $q$ , 我们只需要在  $\Phi_0, \Phi_q, \Phi_{-q}, \Phi_{-qq}$  这四个矢量中对角化  $V_q$ , 其中  $\Phi_0$  是费米子真空态,  $\Phi_{-qq} = \eta_{-q}^\dagger \eta_q^\dagger \Phi_0$ ,  $\Phi_q$  和  $\Phi_{-q}$  的含义类似。

不难看出,  $\Phi_q$  和  $\Phi_{-q}$  已经是  $V_q$  的本征矢量了。这是因为, 它们同时是  $\eta_q \eta_{-q}$  和  $\eta_{-q}^\dagger \eta_q^\dagger$  的本征值为零的本征矢量, 同时还是  $\eta_q^\dagger \eta_q + \eta_{-q}^\dagger \eta_{-q}$  的本征值为 1 的本征矢量。因此, 对于  $q \neq 0, \pi$ , 我们有

$$V_q \Phi_q = \exp(2K_2 \cos q) \Phi_q, \quad (6.57)$$

$\Phi_{-q}$  的情况类似。

下一步就是在矢量  $\Phi_0$  和  $\Phi_{-qq}$  上对角化  $V_q$ 。很容易计算  $V_{1q}$  和  $V_{2q}$  在这两个矢量上的矩阵元。其中  $V_{1q}$  已经是对角的，为

$$V_{1q} = \begin{pmatrix} \exp(-2K_1^*) & 0 \\ 0 & \exp(2K_1^*) \end{pmatrix}. \quad (6.58)$$

为了计算  $V_{2q}^{1/2}$  的矩阵元，我们引入如下对湮灭和对产生算符

$$\sigma_q^- = \eta_q \eta_{-q}, \quad \sigma_q^+ = \eta_{-q}^\dagger \eta_q^\dagger, \quad (6.59)$$

不难看出，在  $\Phi_0$  和  $\Phi_{-qq}$  张成的两维向量空间中，这两个算符所对应的其实就是泡利矩阵的  $\sigma^-$  和  $\sigma^+$ 。因此（还是在这个两维空间中）

$$\eta_q^\dagger \eta_q + \eta_{-q}^\dagger \eta_{-q} = \sigma_q^z + 1 \equiv 2\sigma_q^+ \sigma_q^-, \quad \eta_q \eta_{-q} + \eta_{-q}^\dagger \eta_q^\dagger = \sigma_q^x. \quad (6.60)$$

因此，不难有

$$\begin{aligned} (V_{2q})^{1/2} &= \exp\{K_2[(\sigma_q^z + 1)\cos q + \sigma_q^x \sin q]\} = \exp(K_2 \cos q) \exp(K_2 \sigma_q') \\ &= \exp(K_2 \cos q) (\cosh K_2 + \sigma_q' \sinh K_2) \\ &= \exp(K_2 \cos q) \begin{pmatrix} \cosh K_2 + \sinh K_2 \cos q & \sinh K_2 \sin q \\ \sinh K_2 \sin q & \cosh K_2 - \sinh K_2 \cos q \end{pmatrix}. \end{aligned}$$

其中  $\sigma_q' \equiv \sigma_q^z \cos q + \sigma_q^x \sin q$ ，它具有  $(\sigma_q')^2 = 1$  的重要性质。

综合上述结果，即可以得到  $V_q$  在  $\Phi_0$  和  $\Phi_{-qq}$  张成的两维空间中的矩阵元

$$V_q = \exp(2K_2 \cos q) \begin{pmatrix} A_q & C_q \\ C_q & B_q \end{pmatrix}, \quad (6.61)$$

其中

$$\begin{aligned} A_q &= \exp(-2K_1^*) (\cosh K_2 + \sinh K_2 \cos q)^2 + \exp(2K_1^*) (\sinh K_2 \sin q)^2 \\ B_q &= \exp(-2K_1^*) (\sinh K_2 \sin q)^2 + \exp(2K_1^*) (\cosh K_2 - \sinh K_2 \cos q)^2 \\ C_q &= (2 \sinh K_2 \sin q) (\cosh 2K_1^* \cosh K_2 - \sinh 2K_1^* \sinh K_2 \cos q). \end{aligned} \quad (6.62)$$

通过一点代数运算可以得到两个本征值为

$$\exp(2K_2 \cos q) \left\{ \frac{1}{2}(A_q + B_q) \pm \left[ \left( \frac{1}{2}(A_q - B_q) \right)^2 + C_q^2 \right]^{1/2} \right\} = \exp(2K_2 \cos q) e^{\pm \varepsilon_q},$$

式中  $\varepsilon_q > 0$  满足

$$\cosh \varepsilon_q = \cosh 2K_2 \cosh 2K_1^* - \sinh 2K_2 \sinh 2K_1^* \cos q. \quad (6.63)$$

与  $+$ ,  $-$  两个符号代表的本征值对应的本征矢量是

$$\begin{aligned} \Psi_0 &= \cos \phi_q \Phi_0 + \sin \phi_q \Phi_{-qq} \\ \Psi_{-qq} &= -\sin \phi_q \Phi_0 + \cos \phi_q \Phi_{-qq}. \end{aligned} \quad (6.64)$$

其中  $\phi_q$  (模  $\pi$ ) 由下式定义

$$\tan \phi_q = C_q / (e^{\varepsilon_q} - A_q). \quad (6.65)$$

通过一些代数运算, 也可以将这个定义换成

$$\tan(2\phi_q) = 2C_q / (B_q - A_q), \quad (6.66)$$

不过, 现在就变成模  $\pi/2$ , 为了保持和上面模  $\pi$  的情形一致, 我们额外要求

$$\text{sgn}(2\phi_q) = \text{sgn}(q). \quad (6.67)$$

我们可以用波哥留波夫简化超导 BCS 理论同样方法来简化  $V_q$  的表达式, 为此, 引入如下变换

$$\begin{aligned} \xi_q &= \cos \phi_q \eta_q + \sin \phi_q \eta_{-q}^\dagger \\ \xi_{-q} &= \cos \phi_q \eta_{-q} - \sin \phi_q \eta_q^\dagger. \end{aligned} \quad (6.68)$$

进而不难验证

$$\begin{aligned} \xi_q \Psi_0 &= \xi_{-q} \Psi_0 = 0, \quad \Phi_q = \Psi_q = \xi_q^\dagger \Psi_0, \\ \Phi_{-q} &= \Psi_{-q} = \xi_{-q}^\dagger \Psi_0, \quad \Phi_{-qq} = \xi_{-q}^\dagger \xi_q^\dagger \Psi_0. \end{aligned} \quad (6.69)$$

利用  $\xi$  算符, 即可以把  $V_q$  表达成

$$V_q = \exp(2K_2 \cos q) \exp[-\varepsilon_q (\xi_q^\dagger \xi_q + \xi_{-q}^\dagger \xi_{-q} - 1)]. \quad (6.70)$$

对于  $q=0$  和  $q=\pi$  的特殊情况, 我们可以定义

$$\begin{aligned} \phi_0 &= 0 \quad \text{且} \quad \varepsilon_0 = 2(K_1^* - K_2) \\ \phi_\pi &= 0 \quad \text{且} \quad \varepsilon_\pi = 2(K_1^* + K_2). \end{aligned} \quad (6.71)$$

进而即可以得到  $V^\pm$  的表达式

$$V^\pm = (2 \sinh 2K_1)^{M/2} \exp[-\sum_q \varepsilon_q (\xi_q^\dagger \xi_q - \frac{1}{2})], \quad (6.72)$$

这里我们已经利用了  $2K_2 \sum_q \cos q = 0$  的事实。

注意, 一个仅仅涉及偶数 (奇数) 个  $\xi$  粒子的态同样只涉及偶数 (奇数) 个  $\eta$  粒子。所以  $V^+$  的被允许的本征态仅能涉及偶数个  $\xi$  粒子。类似的,  $V^-$  的被允许的本征态仅能涉及奇数个  $\xi$  粒子。不妨分别记  $V^+$  和  $V^-$  的本征态为  $\Psi^+$  与  $\Psi^-$ 。

定义  $T_c$  为使得下式成立的温度

$$K_1^* = K_2, \quad (6.73)$$

或者根据  $K^*$  的定义, 等价的有

$$\sinh(2J_1/k_B T_c) \sinh(2J_2/k_B T_c) = 1. \quad (6.74)$$

当  $T \neq T_c$  时

$$\varepsilon_q \geq \varepsilon_{\min} > 0, \quad q \neq 0. \quad (6.75)$$

且不难看出

$$\begin{aligned} \text{对于 } T > T_c, \quad \varepsilon_0 &= + \lim_{q \rightarrow 0} \varepsilon_q > 0 \\ \text{对于 } T < T_c, \quad \varepsilon_0 &= - \lim_{q \rightarrow 0} \varepsilon_q < 0. \end{aligned} \quad (6.76)$$

类似于前面一维伊辛模型的讨论, 可知, 在  $N, M \rightarrow \infty$  的热力学极限之下, 系统的配分函数由转移矩阵  $V$  的最大本征值的  $N$  次方给出。所以想算出配分函数的话, 只要求出  $V$  的最大本征值。

很显然,  $V^+$  的最大本征值  $\lambda_0^+$  相应于  $\xi_l$  粒子的真空态,  $\Psi_0^+$ ,

$$\lambda_0^+ = (2 \sinh 2K_1)^{M/2} \exp\left[\frac{1}{2} \sum_l \varepsilon_l\right]. \quad (6.77)$$

所有其它的本征值至少比上述本征值小  $\exp(-\varepsilon_{\min})$  因子。

对于  $V^-$ , 我们要分  $T > T_c$  与  $T < T_c$  两种情况来讨论。对于  $T > T_c$  情形, 由于  $\varepsilon_0 > 0$ , 所以这时候  $V^-$  的最大本征值相应于  $\xi_k$  粒子真空态  $\Psi_0^-$ , 对应的本征值为

$$\lambda_0^- = (2 \sinh 2K_1)^{M/2} \exp\left[\frac{1}{2} \sum_k \varepsilon_k\right]. \quad (6.78)$$

但这个本征态不被费米奇宇称允许。所以这时候整个  $V$  的最大本征值就是  $\lambda_0^+$ 。

对于  $T < T_c$  情形, 由于  $\varepsilon_0 < 0$ , 所以这时候  $V^-$  最大本征值对应的态是  $\Psi_{k=0}^- = \xi_{k=0}^\dagger \Psi_0^-$ , 相应的本征值  $\lambda_{k=0}^-$  由下式给出

$$\lambda_{k=0}^- = (2 \sinh 2K_1)^{M/2} \exp\left[\frac{1}{2} \sum_k |\varepsilon_k|\right]. \quad (6.79)$$

注意, 这个态是被费米奇宇称允许的。所以, 在这种情况下, 为了决定到底哪个是  $V$  的最大本征值, 我们需要比较  $\lambda_0^+$  和  $\lambda_{k=0}^-$ 。不过, 不难看出, 在  $M \rightarrow \infty$  (从而可以允许  $q \rightarrow 0$ ) 的热力学极限之下, 这两个本征值其实是简并的, 都趋于下面的  $\lambda_0$

$$\lambda_0 = (2 \sinh 2K_1)^{M/2} \exp\left[\frac{M}{4\pi} \int_{-\pi}^{\pi} \varepsilon_q dq\right]. \quad (6.80)$$

所以无论高温还是低温, 两维伊辛模型的配分函数  $\mathcal{Z}$  均可以写成

$$\mathcal{Z} / [2 \sinh(2K_1)]^{NM/2} = \exp\left[\frac{NM}{4\pi} \int_{-\pi}^{\pi} \varepsilon_q dq\right]. \quad (6.81)$$

对于低温情形, 前面要多乘以一个因子 2, 它来自于本征值的 2 重简并。但是, 这个因子 2 对于自由能的贡献是非延展的, 所以在热力学极限之下可以忽略。

为了简化上述结果, 我们使用下面数学恒等式 (证明稍微放后面一点)

$$\frac{1}{2\pi} \int_0^{2\pi} \log[2 \cosh x - 2 \cos \omega] d\omega = |x|. \quad (6.82)$$

从而即有

$$\mathcal{Z}/[2\sinh(2K_1)]^{NM/2} = \exp\left[\frac{NM}{4\pi} \int_{-\pi}^{\pi} dq \frac{1}{2\pi} \int_0^{2\pi} \log(2\cosh \varepsilon_q - 2\cos \omega) d\omega\right].$$

代入 (6.63) 式, 即有 (忽略掉非延展的  $\log(2)$ )

$$\begin{aligned} & \log [\mathcal{Z}/[2\sinh(2K_1)]^{NM/2}] \\ &= \frac{NM}{8\pi^2} \int_{-\pi}^{\pi} dq \int_0^{2\pi} d\omega \log[2(\cosh 2K_2 \cosh 2K_1^* - \sinh 2K_2 \sinh 2K_1^* \cos q) - 2\cos \omega]. \end{aligned}$$

利用  $\sinh(2K_1^*) = 1/\sinh(2K_1)$ , 以及

$$\begin{aligned} \cosh(2K_1^*) &= \frac{e^{2K_1^*} + e^{-2K_1}}{2} = \frac{\frac{1}{\tanh(K_1)} + \tanh(K_1)}{2} \\ &= \frac{\cosh^2(K_1) + \sinh^2(K_1)}{2\sinh(K_1)\cosh(K_1)} = \coth(2K_1). \end{aligned} \quad (6.83)$$

进而就可以把刚才的结果重写成

$$\begin{aligned} & \log [\mathcal{Z}/[2\sinh(2K_1)]^{NM/2}] \\ &= \frac{NM}{8\pi^2} \int_{-\pi}^{\pi} dq \int_0^{2\pi} d\omega \log[2(\cosh 2K_2 \coth 2K_1 - \sinh 2K_2 \frac{1}{\sinh 2K_1} \cos q) - 2\cos \omega] \\ &= \frac{NM}{8\pi^2} \int_{-\pi}^{\pi} dq \int_0^{2\pi} d\omega \log[(\cosh 2K_1 \cosh 2K_2 - \sinh 2K_2 \cos q) - \sinh 2K_1 \cos \omega] \\ & \quad + \frac{NM}{2} \log(2/\sinh 2K_1) \\ &= \frac{NM}{2\pi^2} \int_0^{\pi} dq \int_0^{\pi} d\omega \log[(\cosh 2K_1 \cosh 2K_2 - \sinh 2K_2 \cos q) - \sinh 2K_1 \cos \omega] \\ & \quad + \frac{NM}{2} \log(2/\sinh 2K_1). \end{aligned}$$

特别的, 对于  $K_1 = K_2 = K$  的情形,

$$\begin{aligned} \log [\mathcal{Z}/[2\sinh(2K)]^{NM/2}] &= \frac{NM}{2} \log(2/\sinh 2K) \\ & + \frac{NM}{8\pi^2} \int_{-\pi}^{\pi} dq \int_{-\pi}^{\pi} d\omega \log[\cosh^2 2K - \sinh 2K(\cos q + \cos \omega)]. \end{aligned} \quad (6.84)$$

利用和差化积公式, 即有

$$\begin{aligned} \log [\mathcal{Z}/[2\sinh(2K)]^{NM/2}] &= \frac{NM}{2} \log(2/\sinh 2K) \\ & + \frac{NM}{8\pi^2} \int_{-\pi}^{\pi} d\phi_1 \int_{-\pi}^{\pi} d\phi_2 \log[\cosh^2 2K - 2\sinh 2K \cos \phi_1 \cos \phi_2] \end{aligned}$$

定义  $\kappa \equiv 2\sinh 2K/\cosh^2 2K$ , 则有

$$\begin{aligned} \log [\mathcal{Z}/[2\sinh(2K)]^{NM/2}] &= NM \log 2 - \frac{NM}{2} \log \kappa \\ & + \frac{NM}{8\pi^2} \int_{-\pi}^{\pi} d\phi_1 \int_{-\pi}^{\pi} d\phi_2 \log[1 - \kappa \cos \phi_1 \cos \phi_2] \\ &= NM \log 2 - \frac{NM}{2} \log \kappa + \frac{NM}{2\pi^2} \int_0^{\pi} d\phi_1 \int_0^{\pi} d\phi_2 \log[1 - \kappa \cos \phi_1 \cos \phi_2] \end{aligned} \quad (6.85)$$

这就是我们的**核心结果**。

回过头来简单说一下数学恒等式 (6.82) 的证明, 这又是用到如下恒等式 (请自行证明): 对于  $|r| \leq 1$ , 有

$$\frac{1}{2\pi} \int_0^{2\pi} \log(1 + re^{i\theta}) d\theta = 0. \quad (6.86)$$

从而对于  $x > 0$ , 可令  $r = e^{-x}$ , 从而

$$\begin{aligned} & \frac{1}{2\pi} \int_0^{2\pi} \log[2 \cosh x - 2 \cos \omega] d\omega \\ &= \frac{1}{2\pi} \int_0^{2\pi} \log[e^x + e^{-x} - 2 \cos \omega] d\omega \\ &= x + \frac{1}{2\pi} \int_0^{2\pi} \log[1 - 2r \cos \omega + r^2] d\omega \\ &= x + \frac{1}{2\pi} \int_0^{2\pi} \log(1 - re^{i\omega})(1 - re^{-i\omega}) d\omega \\ &= x + \frac{1}{2\pi} \int_0^{2\pi} \log(1 - re^{i\omega}) d\omega + \frac{1}{2\pi} \int_0^{2\pi} \log(1 - re^{-i\omega}) d\omega \\ &= x. \end{aligned} \quad (6.87)$$

得证!

## 6.4 Grassmann 变量积分方法

1980 年, Stuart Samuel 给出了也许是最简单优雅的二维伊辛模型精确解推导。他使用 Grassmann 变量的 Berezin 积分, 将该问题表述为一个自由费米子路径积分模型。相关的 Grassmann 作用量是二次型的, 因此可以通过著名的高斯积分公式来求解。不过, Stuart Samuel 的这个工作本质和 Berezin 1969 年的一篇论文是一样的, 只是 Berezin 的论文由于发表在苏联而长期被人们忽视了。

让我们仿照 Samuel 的原始工作, 从两维 Ising 模型的低温展开开始。根据 (6.8) 式, 这时配分函数可以写成

$$\mathcal{Z}(K) = 2z^{-N} \sum_{\Gamma} z^{L(\Gamma)}. \quad (6.88)$$

式中  $z = e^{-2K}$ 。式中的求和是对对偶格子上的畴壁求和, 它们是对偶格子上不交叉重叠的多边形闭合回路。也即是说, 伊辛模型的每个构型对应于一个特定的畴壁构型。因此, 对所有畴壁构型求和等价于对所有自旋态求和, 除了一个 2 重简并, 因为每个畴壁构型对应 2 个可能的自旋构型。

上面的回路  $\Gamma$  由仅连接对偶格子上最近邻格点的边组成, 称每一条这样的边为布洛赫壁, 显然, 它可以是水平向的, 也可以是竖直向的。对于每个这样的“闭合图” $\Gamma$ , 都有相应的有向图, 其中链接是有方向的。由于任何这样的有向图中每条边 (每个布洛赫壁) 最多出现一次, 但从不出现在两次或更多次, 而且每条边都有两个端点, 因此, 我们可以通过为每条边分配一对 Grassmann 变量 (分别对应两个端点) 来枚举这样的闭合有向图。

Samuel 的核心想法是，将上面这种低温展开转换成对 Grassmann 变量的 Berezin 积分。关于这种 Berezin 积分的详细讨论请读者问人工智能或者参阅相关文献。这里只给一个简单回顾，假设有  $N$  个相互反对易的 Grassmann 变量  $\eta_i, (i = 1, 2, \dots, N)$ 。则由于幂零性  $(\eta_a)^2 = 0$ ，可以将这些变量的任意函数  $f(\eta)$  级数展开成

$$f = a_0 + \sum_i a_i \eta_i + \sum_{i < j} a_{ij} \eta_i \eta_j + \dots + a_{12\dots N} \eta_1 \eta_2 \dots \eta_N. \quad (6.89)$$

则 Berezin 积分的定义是

$$\int d\eta f(\eta) = \int d\eta_1 d\eta_2 \dots d\eta_N f(\eta) \equiv a_{12\dots N}. \quad (6.90)$$

称作， $f$  级数展开的  $a_{12\dots N}$  项使得 Berezin 积分饱和。

我们将利用 Grassmann 变量的两个性质。第一个性质是反对易性，因此 Grassmann 变量的平方为零。这种幂零性 (nilpotency) 可以用来对每个畴壁单元计数不超过一次。我们将利用的第二个性质的 Berezin 积分特有的特征。回想一下，只有当被积函数中的 Grassmann 变量与积分变量完全匹配或对应时，Berezin 多重积分才不为零。如果有一个积分变量不匹配，整个多重积分就消失，因为对于任何没有匹配的 Grassmann 变量  $\theta$  的 Berezin 积分  $\int d\theta = 0$ 。从现在开始，如果所有积分变量都被被积函数变量匹配，我们将说被积函数饱和 (saturates) 了 Berezin 积分。因此，第二个性质的多重 Berezin 积分能够仅选择饱和积分的项的能力。

### 关键部分

让我们定义一个修改的配分函数：

$$\mathcal{Z}' = \sum_{\Gamma} z^{L(\Gamma)}. \quad (6.91)$$

使得  $\mathcal{Z} = 2z^{-N} \mathcal{Z}'$ 。因此，我们今后的目标是计算  $\mathcal{Z}'$ 。为此，我们将使用 Grassmann 变量和 Berezin 积分。

让我们在对偶格子的每个格点  $x$  上定义一对垂直方向的 Grassmann 变量  $\eta_{+1}(x), \eta_{-1}(x)$ ，以及另一对水平方向的 Grassmann 变量  $\eta_{+2}(x), \eta_{-2}(x)$ 。对于一个给定的格点，我们约定

$$\int d\eta_{-1} d\eta_{+1} \eta_{-1} \eta_{+1} = 1. \quad (6.92)$$

我们宣称  $\mathcal{Z}'$  可以由如下 Berezin 积分得到 (稍后会作论证)：

$$\mathcal{Z}' = (-1)^N \int e^S \prod_x d\eta_{-1}(x) d\eta_{+1}(x) d\eta_{-2}(x) d\eta_{+2}(x). \quad (6.93)$$

式中  $S$  是上面引入的 Grassmann 变量的二次型，由于每个不重叠回路  $\Gamma$  由 (i) 连接对偶晶格上相邻格点的边 (布洛赫壁单元)，(ii) 畴壁穿过该格点的边格点，以及 (iii) 水平和竖直两种布洛赫壁相交成一个拐角的拐角格点组成。所以  $S$  可以分解成三部分之和  $S = S_L + S_M + S_C$ 。(i) 布洛赫壁或“线”项  $S_L$ ，(ii) “单体” (monomer) 项  $S_M$ ，以及 (iii) 拐角项  $S_C$ 。

技巧的实质是将适当选择的作用量指数化。有限个 Grassmann 变量函数的泰勒级数是一个多项式（由于幂零性）。这个多项式可以编码各种图以及图的部分的其他奇怪组合，例如孤立拐角或单体。然后，通过 Berezin 积分，我们可以仅选择饱和积分的项。如果作用量选择得当，饱和项就精确对应于非重叠的布洛赫畴回路。

上述布洛赫畴项的定义如下：

$$S_L = z \sum_x [\eta_{+1}(x)\eta_{-1}(x+\hat{1}) + \eta_{+2}(x)\eta_{-2}(x+\hat{2})], \quad (6.94)$$

其中  $\hat{1}$  表示沿着垂直方向移动一个格距， $\hat{2}$  表示沿着水平方向移动一个格距。其中  $\eta_{+1}(x)\eta_{-1}(x+\hat{1})$  是用来计数一个布洛赫垂直壁，而  $\eta_{+2}(x)\eta_{-2}(x+\hat{2})$  是用来计数水平壁。

单体项则非常简单。我们希望单体项“什么都不做”，即仅在需要时贡献因子 1。根据约定 (6.92)，因此我们简单地得到：

$$S_M = \sum_x [\eta_{-1}(x)\eta_{+1}(x) + \eta_{-2}(x)\eta_{+2}(x)] \quad (6.95)$$

所谓的“单体”项由 2 个水平或 2 个垂直变量组成，它们表示相应边的缺失。一个格点处的单个单体项表示在相应方向上穿过该格点的畴壁不存在（请读者细想一下这是为什么），同一格点处的两个单体项表示在水平和竖直两个方向上穿过该格点的畴壁都不存在，从而该格点是一个内点，即不在畴壁上的格点。

为了与上述两个定义保持一致，拐角项必须定义为：

$$S_C = \sum_x [\eta_{+1}(x)\eta_{-2}(x) + \eta_{+2}(x)\eta_{-1}(x) + \eta_{-2}(x)\eta_{+1}(x) + \eta_{-1}(x)\eta_{+2}(x)] \quad (6.96)$$

要理解原因，考虑沿着先水平向前再垂直向前的路径形成的拐角。因此你有 2 个布洛赫壁段在拐角处相遇。为了计数此拐角，我们希望先饱和水平壁，然后饱和垂直壁。水平壁贡献  $\eta_{+2}(x)\eta_{-2}(x)$ ，垂直壁贡献  $\eta_{+1}(x)\eta_{-1}(x+\hat{1})$ 。所以我们想用拐角因子  $\eta_{+2}(x)\eta_{-1}(x)$  来饱和该格点  $x$  处的 Berezin 积分。这是拐角作用量中的第 2 项。其他 3 项可以类似地推导出来。

注意，拐角项和单体项每个格点有偶数个 Grassmann 变量，而线项在两个相邻格点上每个只有一个 Grassmann 变量。因此，为了饱和 Berezin 积分，偶数个线项（因此是 0、2 或 4 个）必须在给定格点处汇合。固定格点  $x$  处的 Berezin 积分只能通过以下方式饱和：

- 两个单体项，一个水平和一个垂直。
- 两条线（均垂直或均水平）和一个单体项（表示另一个方向的线缺失）。
- 两条线（一条水平一条垂直）和一个拐角。
- 四条线。

以下在任何格点处都是被禁止的：1. 奇数条线（因为没有办法饱和缺失的 Grassmann 变量）。2. 一个拐角和一个单体（因为一个 Grassmann 变量必然重复，因此幂零性杀死该项，类似地一个变量缺失，也导致零）。

还有另一种有趣的情况：没有线的 2 个拐角项。在这种情况下，格点处的每个 Grassmann 变量恰好出现一次，因此这样的项确实有贡献。而且，每个这样的项都与具有另外两

种拐角的项匹配。例如, 双拐角项  $[\eta_{+1}(x)\eta_{-2}(x)][\eta_{+2}(x)\eta_{-1}(x)]$  与  $[\eta_{-2}(x)\eta_{-1}(x)][\eta_{+2}(x)\eta_{+1}(x)]$  匹配。但是, 不难看出:

$$[\eta_{-2}(x)\eta_{-1}(x)][\eta_{+2}(x)\eta_{+1}(x)] = +\eta_{+1}(x)\eta_{-2}(x)\eta_{+2}(x)\eta_{-1}(x)$$

因此, 结合 2 个拐角的两种方式导致双倍的贡献。但前一个双拐角实际上是有两个单体的项的负值:

$$[\eta_{+1}(x)\eta_{-2}(x)][\eta_{+2}(x)\eta_{-1}(x)] = -\eta_{-1}(x)\eta_{+1}(x)\eta_{-2}(x)\eta_{+2}(x).$$

因此, 在每一个格点上, 双单体项加上两种双拐角项的总贡献为  $+1-2 = -1$ 。由于回路上的格点总数必定为偶数, 所以假设共有  $N$  个格点, 其中  $N'$  个不在回路上, 则必有  $(-1)^{N'} = (-1)^N$ 。这解释了公式 (6.93) 前面那个  $(-1)^N$  因子的来源。

综合上面的讨论, 并注意到最后, 由于每一个畴壁单元有一个权重  $z$ , 所以整个闭合回路  $\Gamma$  的总权重就是  $z^{L(\Gamma)}$ , 可知关键的 (6.93) 式成立。这样我们就成功地将计算两维伊辛模型配分函数的问题转换成了一个计算 Berezin 积分的问题。

其实, 我们可以把 (6.93) 式前面的因子  $(-1)^N$  吸收掉, 吸收的办法就是在每个格点上多加一个因子  $-1$ 。很显然, 这相当于把  $S_C$  和  $S_M$  等式右边都整体地加个负号。

结果将证明, 对于二维伊辛模型, 作用量  $S$  是一个二次型, 因此我们本质上有一个自由费米子模型。具体而言, 二次型作用量通过傅里叶变换被“对角化”, 立即导出 Onsager 最初发现的精确解。在所有求解伊辛模型的方法中, 这种方法也许是最简单、最美丽且最强大的。而且, 可以证明, 对于三维的立方晶格, 人们得到的将是一个包含四费米子相互作用项的作用量, 正如该领域的专家所熟知的那样。因此, 在这个意义上, 三维伊辛模型并不直接等价于自由费米子模型, 而是等价于相互作用费米子模型。

### Berezin 积分的计算

二次型作用量  $S$  具有平移不变性, 因此可以利用傅里叶变换来将其对角化。为此, 定义

$$\eta_i(x) = \frac{1}{\sqrt{N}} \sum_k e^{ik \cdot x} \hat{\eta}_i(k). \quad (6.97)$$

式中  $k = k_x \hat{1} + k_y \hat{2}$ , 其中

$$k_x = \frac{2\pi n_x}{N_x}, \quad k_y = \frac{2\pi n_y}{N_y}, \quad (6.98)$$

式中  $N_x, N_y$  分别为格子的行数和列数,  $N_x N_y = N$ 。为了方便起见, 通常取  $N_x = 2l_x + 1$ ,  $N_y = 2l_y + 1$ , 从而可以取

$$\begin{aligned} n_x &= -l_x, -l_x + 1, \dots, l_x - 1, l_x \\ n_y &= -l_y, -l_y + 1, \dots, l_y - 1, l_y. \end{aligned}$$

由于傅里叶变换的么正性, 不难验证, 傅里叶变换之后的  $\hat{\eta}_\alpha(k)$  也是反对易的 Grassmann 变量。而且, 同样是由于傅里叶变换的么正性, 也不难验证

$$\prod_x d\eta_{-1}(x)d\eta_{+1}(x)d\eta_{-2}(x)d\eta_{+2}(x) = \prod_k d\hat{\eta}_{-1}(k)d\hat{\eta}_{+1}(k)d\hat{\eta}_{-2}(k)d\hat{\eta}_{+2}(k). \quad (6.99)$$

所以, 为了算出 (6.93) 式的 Berezin 积分, 我们只需要将作用量  $S$  用傅里叶变换之后的  $\hat{\eta}_\alpha(k)$  来表达, 并算出相应的 Berezin 积分即可。

注意到作用量在  $\eta$  中是二次的。具体来说, 作用量是 Grassmann 变量中的二次型, 我们可以写成:

$$S = \sum_{x,y,i,j} \eta_i(x)A_{ij}(x,y)\eta_j(y). \quad (6.100)$$

我们将在下面显式定义  $A_{ij}(x,y)$ , 但暂时我们不必担心它, 只需注意它只是一个矩阵。注意, 对于固定的一对不一定不同的位点  $x$  和  $y$ ,  $A$  是一个真正的  $(4 \times 4)$  矩阵, 只有两个指标。类似地, 对于固定的  $i, j$ ,  $A$  是一个  $N \times N$  矩阵。我们将对位点变量  $x, y$  应用傅里叶变换, 注意到作用量具有内积的形式, 并且通过傅里叶变换的么正性, 可以将作用量重写为:

$$S = \sum_{k,k',i,j} \hat{\eta}_i^*(k)\hat{A}_{ij}(k,k')\hat{\eta}_j(k'). \quad (6.101)$$

我们这里所做的只是应用了 Plancherel 定理, 有时也称为 Parseval 定理。带帽符号表示傅里叶变换后的量。

二维 Ising 模型的平移不变二次型作用量只连接相邻位点。作用量的唯一非对角部分是由于这种到相邻位点的平移。但在傅里叶变换后, 我们消除了平移, 获得了一个在傅里叶变换后变量  $k$  中对角的作用量。换句话说,  $\hat{A}_{ij}(k,k') = \hat{A}_{ij}(k,k)\delta_{k,k'}$ , 其中  $\delta_{k,k'}$  是 Kronecker delta 符号。

让我们引入更紧凑的符号以使计算更容易。让我们定义向量:

$$\eta = \begin{bmatrix} \eta_{-1} \\ \eta_{+1} \\ \eta_{-2} \\ \eta_{+2} \end{bmatrix}. \quad (6.102)$$

对  $\hat{\eta}$  也类似。那么我们可以写成:

$$S = \sum_{x,y} \eta^T(x)A(x,y)\eta(y). \quad (6.103)$$

上标  $\cdot^T$  表示转置。矩阵  $A_C$  和  $A_M$  由 (6.96) 和 (6.95) 给出, 且在位点指标  $x, y$  中是对角

的, 考虑到吸收  $(-1)^N$  因子后:

$$\begin{aligned} A_C(x,y) &= -\delta_{x,y} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix} \\ A_M(x,y) &= -\delta_{x,y} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \end{aligned} \quad (6.104)$$

矩阵  $A_L$  由 (6.94) 给出, 它在位点指标中不是对角的, 因为它连接相邻位点:

$$A_L(x,y) = z \begin{bmatrix} 0 & 0 & 0 & 0 \\ \delta_{x+\hat{1},y} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \delta_{x+\hat{2},y} & 0 \end{bmatrix}. \quad (6.105)$$

用  $\hat{\eta}$  来表示, 我们可以将作用量写成:

$$S = \sum_k \hat{\eta}^\dagger(k) \hat{A}(k,k) \hat{\eta}(k). \quad (6.106)$$

求和现在只对单个指标  $k$  的原因是  $A$  具有平移不变性, 应用 Parseval 定理允许我们将位点平移转换为相位, 因此  $\hat{A}(k_1, k_2) = \delta_{k_1, k_2} \hat{A}(k_1, k_1)$ 。换句话说,  $\hat{A}$  在傅里叶变量 (如  $k$ ) 中是对角的。

作为一个说明性例子, 让我们显式地展示最后这一点。考虑  $\sum_y C(x,y) \eta(y)$  的傅里叶变换, 其中  $C(x,y) = C_0 \delta_{x+\hat{1},y}$  是平移不变的:

$$\begin{aligned} \sum_x e^{-ik \cdot x} \sum_y C(x,y) \eta(y) &= \sum_x e^{-ik \cdot x} \sum_y C_0 \delta_{x+\hat{1},y} \eta(y) \\ &= \sum_x e^{-ik \cdot x} C_0 \eta(x+\hat{1}) \\ &= \sum_{x'} e^{-ik \cdot (x'-\hat{1})} C_0 \eta(x') \\ &= e^{ik \cdot \hat{1}} C_0 \sum_{x'} e^{-ik \cdot x'} \eta(x') \\ &= e^{ik_x} C_0 \hat{\eta}(k). \end{aligned} \quad (6.107)$$

因此  $\hat{A}(k,k)$  在之前解释的意义上是对角的。它很容易计算为:

$$\hat{A}(k,k) = \begin{bmatrix} 0 & -1 & 0 & 0 \\ ze^{ik_x} & 0 & -1 & 0 \\ -1 & 0 & 0 & -1 \\ -1 & -1 & ze^{ik_y} & 0 \end{bmatrix}. \quad (6.108)$$

注意到作用量中出现了共轭转置变量  $\hat{\eta}^\dagger(k)$ 。我们需要处理这个问题。由于  $\eta(x)$  原则上是一个“实” Grassmann 变量, 其傅里叶变换满足  $[\hat{\eta}(k)]^* = \hat{\eta}(-k)$ 。实函数傅里叶变换

的这一“厄米”性质允许我们将作用量写成：

$$S = \sum_k \hat{\eta}^T(-k) \hat{A}(k, k) \hat{\eta}(k). \quad (6.109)$$

我们可以显式地计算所需的 Berezin 积分。但需要一些小心。观察到作用量“混合”了频率  $k$  和  $-k$ 。当我们将上述作用量中和的指数重写为各项的指数的乘积时，这些指数不仅包含构成向量  $\hat{\eta}(k)$  的 4 个 Grassmann 变量，还包含  $\hat{\eta}(-k)$  中的另外 4 个变量。因此完整的 Berezin 积分不是分解为对 4 个变量的 Berezin 积分，而是对 8 个变量的积分。因此我们将 Grassmann 微分重新分组为 8 个一组，而不是 4 个一组，如下所示：

$$\prod_k d\hat{\eta}_{-1}(k) d\hat{\eta}_{+1}(k) d\hat{\eta}_{-2}(k) d\hat{\eta}_{+2}(k) = \prod_{k \geq 0} d\hat{\eta}(k) d\hat{\eta}(-k). \quad (6.110)$$

(这里我们稍微滥用一点符号，写  $k \geq 0$  表示  $k > 0$  和  $-k > 0$  不会同时包含在连乘中) 式中

$$\begin{aligned} d\hat{\eta}(k) d\hat{\eta}(-k) &\equiv \\ d\hat{\eta}_{-1}(k) d\hat{\eta}_{+1}(k) d\hat{\eta}_{-2}(k) d\hat{\eta}_{+2}(k) & d\hat{\eta}_{-1}(-k) d\hat{\eta}_{+1}(-k) d\hat{\eta}_{-2}(-k) d\hat{\eta}_{+2}(-k). \end{aligned}$$

为了正确地分解完整的 Berezin 积分，我们需要收集所有含  $k$  和  $-k$  的项，这可以通过将作用量中的求和重写为  $k$  的一半值来实现。我们用  $k \geq 0$  表示  $k > 0$  和  $-k > 0$  不会同时包含在求和中。使用这个符号，我们有：

$$e^S = \prod_{k \geq 0} \exp \left[ \hat{\eta}^T(-k) \hat{A}(k, k) \hat{\eta}(k) + \hat{\eta}^T(k) \hat{A}(-k, -k) \hat{\eta}(-k) \right]. \quad (6.111)$$

如果我们定义：

$$B(k) = \hat{A}(k, k) - \hat{A}^T(-k, -k). \quad (6.112)$$

那么可以将作用量写成：

$$S = \sum_{k \geq 0} \hat{\eta}^T(-k) B(k) \hat{\eta}(k). \quad (6.113)$$

完整的 Berezin 积分现在可以写成：

$$\mathcal{Z}^l = \prod_{k \geq 0} \int \exp \left[ \hat{\eta}^T(-k) B(k) \hat{\eta}(k) \right] d\hat{\eta}(k) d\hat{\eta}(-k). \quad (6.114)$$

每个 Berezin 积分因子都是一个高斯积分。回忆一下，对于 Grassmann 数  $x_i$  和  $y_i$  以及复矩阵  $A$ ：

$$\int e^{\sum_{ij} x_i A_{ij} y_j} dx dy = \det(A). \quad (6.115)$$

因此：

$$\mathcal{Z}^l = \prod_{k \geq 0} \det B. \quad (6.116)$$

移除对  $k \geq 0$  的限制, 我们有:

$$\mathcal{Z}' = \prod_k \sqrt{\det B}. \quad (6.117)$$

忽略非延展的  $\log(2)$ , 我们有:

$$\log(\mathcal{Z}') = -N\log(z) + \log \mathcal{Z}' = -N\log(z) + \frac{1}{2} \sum_k \log \det B. \quad (6.118)$$

在热力学极限下, 求和变为积分, 因此:

$$\log(\mathcal{Z}') = -N\log(z) + N \frac{1}{2} \int_{-\pi}^{\pi} \frac{dk_x}{2\pi} \int_{-\pi}^{\pi} \frac{dk_y}{2\pi} \log \det B. \quad (6.119)$$

反对称矩阵  $B$  不难求出来, 为:

$$B(k) = \begin{bmatrix} 0 & -(1+ze^{-ik_x}) & 1 & 1 \\ (1+ze^{ik_x}) & 0 & -1 & 1 \\ -1 & 1 & 0 & -(1+ze^{-ik_y}) \\ -1 & -1 & (1+ze^{ik_y}) & 0 \end{bmatrix}. \quad (6.120)$$

其行列式为:

$$\det B = (1+z^2)^2 - 2z(1-z^2)(\cos k_x + \cos k_y). \quad (6.121)$$

代入上面的积分, 即可以得到配分函数的表达式:

$$\begin{aligned} \log(\mathcal{Z}') &= -N\log(z) + \\ & N \frac{1}{2} \int_{-\pi}^{\pi} \frac{dk_x}{2\pi} \int_{-\pi}^{\pi} \frac{dk_y}{2\pi} \log [(1+z^2)^2 - 2z(1-z^2)(\cos k_x + \cos k_y)] \\ &= N \frac{1}{2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{dk_x dk_y}{(2\pi)^2} \log [(z^{-1}+z)^2 - 2(z^{-1}-z)(\cos k_x + \cos k_y)]. \end{aligned} \quad (6.122)$$

在最后的結果中代入  $z = e^{-2K}$ , 不难看出这个結果正与上一节的结果 (6.84) 完全一致。

我们完成了! (6.84) 和 (6.122) 就是 Onsager 著名的結果!

### 三维 Ising 模型

自从 1944 年 Onsager 发表他的开创性论文以来, 多年来人们提出了许多针对三维或立方 Ising 模型的 tentative 精确解猜想。如前所述, 立方 Ising 模型的 Grassmann 作用量是四次的。在量子场论中, 四次 Grassmann 作用量与相互作用费米子模型相关联, 而二次作用量与自由费米子模型相关联。后者可以通过 Pfaffian 和行列式公式轻松求解, 就像我们上面所做的那样, 但目前还没有已知的方法能够给出具有四次 Grassmann 作用量的晶格模型的精确解 (在热力学极限下)。因此, 任何声称对立方 Ising 模型有精确解的人都必须解释他们如何克服处理四次作用量的数学困难, 或者至少解释新方法如何绕过这一数学障碍。

Barry Cipra 在《科学》杂志的一篇文章中, 将三维 Ising 问题称为“统计力学的圣杯”。该文章列出了我们可能永远无法实现找到立方 Ising 模型显式精确解目标的许多其他原因。立方 Ising 模型的精确解也许是一个不可能的问题!

**参考文献**

1. Two-Dimensional Ising Model as a Soluble Problem of Many Fermions, T. D. SCHULTZ, D. C. MATTIS, and E. H. LIEB, *Rev. Mod. Phys.* 36, 856 –Published 1 July, 1964, DOI: <https://doi.org/10.1103/RevModPhys.36.856>
2. Samuel, Stuart (1980). "The Use of Anticommuting Integrals in Statistical Mechanics. 1". *J. Math. Phys.* 21 (12): 2806–2814. Bibcode:1980JMP....21.2806S. doi:10.1063/1.524404.
3. Samuel, Stuart (1980). "The Use of Anticommuting Integrals in Statistical Mechanics. 2". *J. Math. Phys.* 21 (12): 2815. Bibcode:1980JMP....21.2815S. doi:10.1063/1.524405.
4. Samuel, Stuart (1980). "The Use of Anticommuting Integrals in Statistical Mechanics. 3". *J. Math. Phys.* 21 (12): 2820. Bibcode:1980JMP....21.2820S. doi:10.1063/1.524406
5. F. A. Berezin, The Plane Ising Model, *Russ. Math. Surv.* 24, 1 (1969).



## 参考文献

参考文献见各章最后。

