



利用构型空间的距离统计表征相变

刘煜靖¹, 苏文瑜¹, 马女森², 程晨^{1*}

1. 兰州大学物理科学与技术学院, 兰州理论物理中心, 量子理论及应用基础教育部重点实验室, 甘肃省理论物理重点实验室, 甘肃省量子物理基础学科研究中心, 兰州 730000

2. 北京航空航天大学物理学院, 北京 102206

*联系人, E-mail: chengchen@lzu.edu.cn

收稿日期: 2025-11-10; 接受日期: 2026-01-13; 网络出版日期: 2026-04-09

国家自然科学基金(编号: 12174167, 12474139, 12247101)、中央高校基本科研业务费(编号: lzujbky-2024-jdxx06)和甘肃省自然科学基金(编号: 22JR5RA389)资助

摘要 本文系统性回顾了我们的课题组在“构型空间统计几何”这一相变研究新范式上的一系列工作. 该方法通过分析蒙特卡罗采样的微观构型间的距离分布, 从原始数据中无监督地识别相变并提取普适临界行为. 我们展示了该范式如何从一个针对复杂模型的数值探索, 发展为一个普适的经验性标度律, 并最终建立起连接构型空间几何量与实空间关联函数的解析理论框架. 该工作揭示了宏观临界性在微观构型几何中的编码方式, 为研究序参量未知的复杂相变问题提供了一个全新的、“白箱”式的可解析数据驱动视角.

关键词 相变, 晶格模型, 构型空间, 临界现象, 蒙特卡罗模拟

PACS: 05.50.+q, 05.70.Fh, 05.10.Ln

1 引言

相变作为统计物理与凝聚态物理学的核心概念, 深刻揭示了构成物质的微观粒子在宏观尺度上集体行为的普适规律^[1]. 朗道-金兹堡-威尔逊(Landau-Ginzburg-Wilson, LGW)理论框架^[2-4]为理解连续相变提供了坚实的理论基础, 其关键在于通过分析定义系统对称性破缺所对应的局域序参量, 来描述临界现象中的普适行为. 然而, 随着相关研究的深入与各类新奇物态的发现, 物理学逐渐遇到越来越多难以完全纳入传统LGW范式的复杂系统. 在几何阻挫磁体^[5]、拓扑物态^[6,7]以及非平衡物态如多体局域化(Many-body

Localization, MBL)^[8,9]等前沿领域中, 寻找一个简单的局域序参量往往异常困难甚至不可行. 这一挑战促使人们迫切需要发展超越传统序参量的新研究范式.

一条重要的研究路径源于对系统微观自由度进行直接统计分析. 例如, 在MBL等非平衡量子问题中, 利用波函数本身刻画相与(可能发生的)相变是一种常见的分析方法. 进一步, 定义在希尔伯特空间中的汉明距离(r_H)已成为一个核心可观测量, 它能够有效刻画量子态的延展性或局域性, 并在量子模拟实验中发挥关键作用^[10-14]. 在这一类非平衡问题中, 希尔伯特空间中的构型分布, 以及构型之间的距离分布蕴含了相与相变的全部信息, 这自然引出一个更深层次的问题: 我们

引用格式: 刘煜靖, 苏文瑜, 马女森, 等. 利用构型空间的距离统计表征相变. 中国科学: 物理学 力学 天文学, 2026, 56: 250006
Liu Y J, Su W Y, Ma N S, et al. Characterizing phase transitions from statistics of configuration-space distances (in Chinese). Sci Sin-Phys Mech Astron, 2026, 56: 250006, doi: [10.1360/SSPMA-2025-0549](https://doi.org/10.1360/SSPMA-2025-0549)

能否从希尔伯特空间出发、用一种较为普适的方法去描述更一般的平衡态相与相变? 这正是本系列研究最初的动机.

在数据科学迅速发展的背景下, 利用无监督机器学习方法从海量无标签数据中自动识别相变, 已成为一个新兴且活跃的研究方向^[15]. 其中一类工作从服从系统物理规则的抽样微观构型出发, 通过分析这些微观构型组成的高维数据结构得到系统的相变信息. 值得注意的是, 诸多先进流形学习算法, 如主成分分析^[16–18]、扩散图嵌入^[19,20]、内禀维度分析^[21–23]等, 其底层输入信息亦为采样构型间的两两距离矩阵——这些方法成功的物理内涵应来源于我们所关注的希尔伯特空间构型距离的分布($P(r_H)$). 然而, 这类机器学习方法通常依赖复杂且物理意义不明确的“黑箱”算法流程, 并引入如核函数选择、超参数调节等人为因素, 这在一定程度上削弱了其“无监督”的纯粹性, 也为深入理解相关物理机制带来困难.

面对这一现状, 我们提出了一条算法更直接、内涵更物理的研究路径: 与其将希尔伯特空间中的距离分布作为复杂机器学习算法的输入, 不如直接利用统计物理的工具来分析 $P(r_H)$ 这一信息源头的统计几何性质. 这一思路旨在建立一个“白箱”式的、可解析的无监督数据驱动方法. 它既继承了传统物理学的严谨性——其核心观测量最终可以被解析地与基础物理量(如关联函数)联系起来; 又具备了现代数据科学的强大能力——能够在没有先验知识的情况下, 从原始数据中提取出普适的临界行为.

本文通过系统回顾我们课题组在此方向上的一系列工作, 揭示希尔伯特空间距离的统计几何分析与传统物理量的解析联系及其深刻物理内涵. 我们将展示, 这一研究范式如何从针对复杂阻挫模型的初步数值探索, 逐步发展为一条经广泛验证的普适经验标度律, 并最终在为应对数值挑战所构建的理论框架中获得完整而统一的解析诠释. 我们将阐明, 宏观临界普适性如何被深刻编码于微观构型空间的全局统计几何特征中, 从而为理解相变与临界现象提供新的理论视角. 文章的结构安排如下: 第2节以笼目自旋冰模型为例, 介绍了利用构型距离分布探测Berezinskii-Kosterlitz-Thouless(BKT)相变的初步唯象探索; 第3节将该方法推广至一般模型, 并总结出关于距离方差的经验性标度律; 第4节建立了连接构型空间几何量与实

空间关联函数的解析理论框架, 并解释了量子相变中的基矢依赖性问题; 第5节进一步探讨了如费雪信息和宇称指数等全局或非局域探针在拓扑相变中的应用; 最后, 第6节对全文进行总结并展望了未来的研究方向.

2 方法的初步探索: 构型距离分布与BKT相变

如引言所述, 我们工作的核心思想是, 相变的普适信息蕴含在系统所有可能构型构成的希尔伯特空间的几何结构中. 本节将回顾我们最初的探索, 即如何通过直接分析这一空间的几何特征来探测相变. 这项初步研究选择了一个具有高度挑战性的模型作为试金石——笼目晶格上的自旋冰模型, 它不仅具有几何阻挫, 还经历两次连续的、由拓扑缺陷主导的BKT相变^[24].

2.1 核心思想与距离的定义

在探索超越传统序参量的方法时, 我们首先注意到, 对于BKT相变这类特殊的临界现象, 一些传统的热力学量在探测相变时会遇到困难. 例如, 比热 C_v 峰值位置往往偏离真实的临界点, 且峰高不随系统尺寸表现出显著的标度行为, 使得精确确定相变点变得棘手^[25–28]. 这一挑战促使我们去寻找一个对临界行为更敏感、更普适的新探针.

我们的出发点是这样一个假设: 系统的热力学状态, 无论是通过其配分函数还是波函数描述, 其全部信息都编码在构型 s 的概率分布中. 因此, 构型与构型之间的“差异”或“相似度”理应是一个核心的描述量. 对于由离散自由度(如伊辛自旋)构成的系统, 最直接的差异度量是汉明距离 r_H (本文采用下标 H 标记定义在希尔伯特空间中的距离, 用以区别实空间距离 r). 其物理定义是两个构型中自旋指向不同的格点数目. 对于一个拥有 N 个格点的伊辛自旋系统, 其归一化的数学形式可以巧妙地写为

$$r_H(s, s') = \frac{1}{2N} \sum_{i=1}^N (1 - s_i s'_i), \quad (1)$$

其中 $s_i, s'_i = \pm 1$. 该距离的取值范围为 $[0, 1]$, 0代表构型完全相同, 1代表完全相反. 值得强调的是, 汉明距离并

非一个纯粹的数学工具,它在MBL等领域已被广泛用作一个核心的物理可观测量,用以探测量子态在希尔伯特空间中的结构^[10-12].这为我们将其应用于平衡态相变研究提供了坚实的物理基础.

2.2 可行性验证: 笼目自旋冰模型

为了在一个严苛的环境下验证上述思想,即 $P(r_H)$ 包含了描述相与相变的完整信息,我们考察了笼目晶格上的伊辛模型,其哈密顿量包含最近邻的反铁磁相互作用($J_1 > 0$)和次近邻的铁磁相互作用($J_2 < 0$)^[29,30]:

$$H = J_1 \sum_{\langle i,j \rangle} s_i s_j + J_2 \sum_{\langle\langle i,j \rangle\rangle} s_i s_j, \quad (2)$$

这里 $s_i = \pm 1$ 表示晶格 i 上的经典自旋.在 $J_2/J_1 = -1/3$ 的参数下,该模型因其基态的六重简并(形成所谓的“六态时钟自旋冰”)以及从低温有序相到高温无序相过程中经历一个中间临界相(准长程有序),而成为一个理想的测试平台^[24].

我们通过大规模经典蒙特卡罗模拟,系统地研究了该模型中 $P(r_H)$ 随温度变化的演化规律^[31],其结果如图1所示.在三个不同的物相中, $P(r_H)$ 呈现出截然不同的、信息量丰富的形态特征.

(1) 低温有序相($T < T_{c1}$): $P(r_H)$ 呈现出位于0、1/3、2/3、1等处的多个尖锐峰,这直接反映了系统在六个简并基态附近涨落的 $\mathbb{Z}_6$ 离散对称性.一个重要的细节是,随着温度升高,代表单个基态内部涨落的0峰和1峰会向中心0.5移动;而代表不同基态间相对距离的1/3峰和2/3峰的位置则几乎保持不变,这体现了低温有序态 $\mathbb{Z}_6$ 对称性结构的鲁棒性.

(2) 中间临界相($T_{c1} < T < T_{c2}$): 进入BKT相后,离散的峰结构消失.此时的分布呈现出一种复杂的、非单峰的宽阔形态,在数值上可以被很好地拟合为两个宽缓高斯分布的叠加.这种形态是系统从 $\mathbb{Z}_6$ 离散对称性破缺“融化”为具有幂律关联的 $\mathbb{U}(1)$ 连续对称性临界相的直接体现.

(3) 高温无序相($T > T_{c2}$): 在高温极限下,所有构型等权出现, $P(r_H)$ 最终演变为一个以0.5为中心的高斯分布,表现出无关联的统计特征.

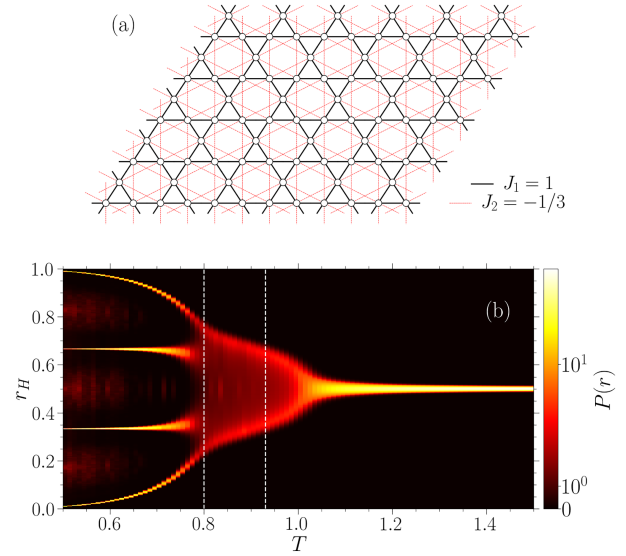


图 1 笼目自旋冰模型的晶格示意图(a)及构型距离分布 $P(r_H)$ 随温度 T 的演化(b). (b)中热图的颜色代表概率密度,其清晰地展示了低温有序相的多峰结构、中间临界相的宽缓分布以及高温无序相的单一高斯峰结构.图片改编自文献^[31]

Figure 1 Schematic diagram of the lattice (a) and the evolution of the configuration distance distribution $P(r_H)$ with temperature T (b) for the kagome spin ice model. In (b), the colors of the heat map represent the probability density, which clearly shows the multi-peak structure of the low-temperature ordered phase, the broad distribution of the intermediate critical phase, and the single Gaussian peak structure of the high-temperature disordered phase. The figure is adapted from ref. ^[31].

这种 $P(r_H)$ 分布形态的突变,不仅清晰地区分了不同的物相,更促使我们对其开展定量分析.通过观察数据,可尝试对不同温度 T 下的 $P(r_H)$ 多峰结构进行正态分布拟合,并进一步将首个峰的有效高度 $1/\Delta \sqrt{2\pi}$ 作为序参量的候选量,进行有限尺寸标度分析.假设该量满足BKT相变所特有的尺度行为:

$$\frac{1}{\Delta \sqrt{2\pi}} L^b = \mathcal{F}_{\text{BKT}} \left(\frac{L}{e^{\frac{c}{\sqrt{t}}}} \right), \quad (3)$$

其中 b 和 c 为待定的临界指数, $t = |T - T_c|/T_c$ 表示归一化的温度偏离.采用这一具有BKT特征的标度形式,对上述量进行有限尺寸缩放分析,如图2所示,我们成功地以较高精度确定了两个相变点 T_{c1} 和 T_{c2} ^[31].

此外,为了验证该方法的优越性,我们还将其与当时流行的另一种无监督机器学习方法——内禀维度(I_d)^[21-23]——进行了对比.结果表明,虽然 I_d 也能感知到相变的存在,但其难以较为准确地确定相变点.更

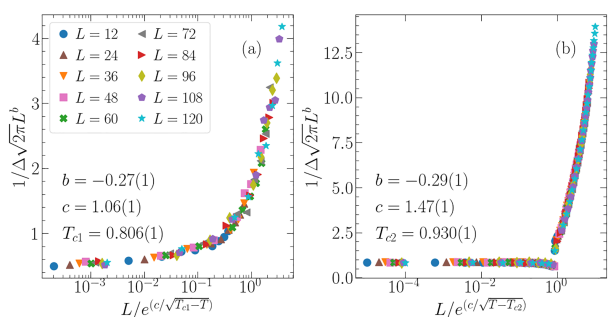


图 2 基于构型空间测量、服从式(3)的笼目自旋冰模型的有限尺寸标度分析。(a) T_{c1} ; (b) T_{c2} 。两图共享图例。图片改编自文献[31]

Figure 2 Finite-size scaling analysis of the kagome spin ice model based on configuration space measurements and obeying eq. (3). (a) T_{c1} ; (b) T_{c2} . Two panels share the same legend. The figure is adapted from ref. [31].

重要的是, 计算内禀维度 I_d 时需引入人为调节的参数, 这减弱了其无监督属性和数据的准确性, 同时加大了数据处理的计算量(详见文献[31]第IV节的论述)。这初步证明了我们基于 $P(r_H)$ 的分析方法不仅可行, 而且相对于一般的机器学习算法更加稳健和精确。

2.3 初步结论与待解决的问题

这项在笼目自旋冰模型上的初步研究取得了成功。它首次证明, 通过直接分析蒙特卡罗采样构型的距离分布 $P(r_H)$, 我们可以在无需预先知道系统序参量的情况下, 有效地区分物相并精确定位复杂的BKT相变点。这表明, $P(r_H)$ 作为一个新颖的数值分析工具具有巨大的潜力。

然而, 这一探索本质上是唯象的, 并留下了一系列亟待解决的根本性问题, 这些问题是我们后续研究的核心驱动力:

(1) 方法的普适性, 首要任务是验证该方法是否仅适用于笼目晶格的BKT相变, 还是对其他类型的相变(如连续相变、一级相变)和更广泛的模型(如具有连续自由度的XY模型)也同样有效;

(2) 核心观测量, 需要探究是否存在一个比“峰高”更基础、更普适的统计矩, 以统一地描述各类临界现象;

(3) 物理内涵, 必须阐明构型空间的几何统计量探测相变的微观物理机制, 并建立其与实空间物理量(如关联函数和临界指数)的解析联系。

解答这些问题, 将是把这一新颖的数值观测推广至一个自洽的物理理论框架的关键。

3 方法的广泛适用性与一个经验性标度律的发现

上一节的论述验证了利用构型距离分布 $P(r_H)$ 探测复杂相变的可行性, 但该方法仍具有较强的唯象特征。本节将系统阐述如何将这一原本针对特定模型的数值方法, 推广为一个具有广泛适用性的通用分析框架, 并在此过程中揭示一个深刻的经验性标度律[32]。

3.1 从特例到普适: 框架的推广

要回答该方法是否具有普适性, 我们必须首先克服其在方法论上的局限。文献[31]所讨论的汉明距离仅适用于伊辛自旋等二能级系统。因此, 构建一种更为广义的距离度量, 并从中提炼出更具普遍性的统计观测量, 是实现该方法普适化的重要前提之一。

3.1.1 从汉明距离到广义度量: 方法的拓展

为了将分析框架推广至具有更丰富局域自由度的系统, 我们以 q -态时钟模型作为一个典型范例。该模型的哈密顿量通常写为

$$H = -J \sum_{\langle i,j \rangle} \cos(\theta_i - \theta_j), \quad (4)$$

其中, 格点 i 上的自旋自由度由角度变量 θ_i 描述, 其取值被限制在 q 个离散方向上, 即 $\theta_i \in \{2\pi k/q \mid k = 0, 1, \dots, q-1\}$ 。显然, 定义在二能级格点中的汉明距离仅能区分“相同”与“不同”两种状态, 已不足以刻画这类多值角度变量之间更为精细的“差异程度”。

为此, 可以引入一个更为广义的距离定义——基于角度变量的1-范数距离。对于两个构型 s_α 和 s_β , 设它们在格点 i 上的局域角度分别为 θ_i^α 和 θ_i^β , 则二者之间的归一化局域距离定义为这两个角度在圆周上的最短弧长:

$$d_i(\theta_i^\alpha, \theta_i^\beta) = \frac{1}{\pi} \min \{ |\theta_i^\alpha - \theta_i^\beta|, 2\pi - |\theta_i^\alpha - \theta_i^\beta| \}. \quad (5)$$

该距离将圆环上的角差映射到区间 $[0, 1]$, 表示两点间相对位置的最小归一化距离。整个构型空间中的

归一化距离 r_H 定义为所有格点上局域距离的平均值:

$$r_H = \frac{1}{N} \sum_{i=1}^N d_i. \quad (6)$$

这一距离度量具有两个重要优势: 首先, 在物理意义上直观明确——它刻画了构型在连续状态空间中的几何差异, 可视为传统汉明距离中“差异度”概念在连续角度模型中的自然推广. 其次, 在数学上具备良好的兼容性: 当系统退化为伊辛模型(即 $q = 2$, 角度仅取 0 或 π)时, 上述距离取值非 0 即 1 , 恰好还原为归一化的汉明距离. 因此, 该定义不仅适用于更广泛的序参量结构, 还保持了与离散情形的一致性.

3.1.2 核心观测量的提炼: 从“峰高”到统计矩

文献[31]采用“峰高”作为特征量, 依赖于对分布曲线的拟合过程, 不可避免引入了一定的唯象性与主观选择偏差. 因此, 文献[32]构建了一个更为基础、普适且无偏的分析框架, 转向直接考察构型距离分布 $P(r_H)$ 本身的无偏统计性质, 具体重点关注以下两个核心观测量.

(1) 距离标准差: 定义为 $\sqrt{\text{Var}(r_H)} = \sqrt{\langle r_H^2 \rangle - \langle r_H \rangle^2}$, 它直接刻画了构型距离分布的整体展宽程度. 从物理图像上看, 该量反映了在热力学或量子涨落作用下, 系统在希尔伯特空间中采样到的构型集合的“弥散”范围. 我们预计在相变临界区域, 随着长程关联的形成和系统对称性的演化, 构型空间的几何结构将发生显著重构, 这种变化将在分布的宽度上清晰体现, 表现为标准差在临界点附近出现异常增强或标度行为.

(2) 参与熵: 定义为香农熵 $S_{r_H} = -\sum_{r_H} P(r_H) \times \ln P(r_H)$, 用于量化距离分布的信息含量. 该熵值表征了构型距离取值的多样性与不确定性. 通常情况下, 在临界点附近系统涨落最为复杂, 多尺度结构共存, 导致构型采样更加均匀或呈现特定的非均匀模式, 从而使参与熵表现出极值、拐点或奇异的标度特性, 成为识别相变点的敏感探针.

通过将分析视角从模型依赖的“峰高”提升至更具普适性的统计矩——即距离标准差 $\sqrt{\text{Var}(r_H)}$ 与参与熵 S_{r_H} , 并结合广义构型距离的定义, 我们建立起一套不依赖具体模型细节、适用于多种量子或多体系统的通用分析框架. 这一框架不仅增强了方法的客观性与

可迁移性, 也为探索未知相变和复杂序提供了新的工具. 接下来的关键步骤是对该框架的普适性与有效性进行系统性的数值验证, 涵盖不同维度、对称性和相互作用类型的模型, 以检验其在实际应用中的鲁棒性与分辨能力.

3.2 系统性验证与一个经验性标度律的发现

在构建了包含广义距离定义与普适统计观测量的分析框架后, 需对其有效性与普适性进行系统的数值验证. 为全面评估该框架在不同物理情境下的适用性, 文献[32]选取了一系列凝聚态物理中具有代表性的二维经典晶格模型, 涵盖多种相变类型、对称性结构以及局域自由度形式. 具体包括: 具有 $\mathbb{Z}_2$ 对称性的伊辛模型(经历连续相变)、具有 $\mathbb{Z}_q$ 对称性的 q -态时钟模型(当 $q \geq 5$ 时呈现两次BKT相变), 以及具有 $\mathbb{U}(1)$ 连续对称性的XY模型(经历BKT相变). 通过这些典型体系中的测试, 系统地检验了新框架在识别相变和刻画临界行为方面的鲁棒性与广泛适用性.

3.2.1 框架的广泛有效性

一系列经典自旋模型的数值模拟结果表明, $\sqrt{\text{Var}(r_H)}$ 与 S_{r_H} 这两个核心观测量在所有研究的模型中均能有效捕捉相变信号. 以距离标准差 $\sqrt{\text{Var}(r_H)}$ 为例, 其在各类模型的相变点附近均展现出显著的临界标度行为. 如图3(a)所示, 二维经典Ising模型的构型距离分布 $P(r_H)$ 同样清晰表征了两个相. 进一步, 假设距离标准差 $\sqrt{\text{Var}(r_H)}$ 满足连续相变的有限尺寸标度行为:

$$L^{\frac{b}{\nu}} \sqrt{\text{Var}(r_H)} = \mathcal{F}_{\text{con}} \left(t L^{\frac{1}{\nu}} \right), \quad (7)$$

这里 ν 是关联长度指数, b 是未知的拟合指数. 如图3(c)所示, 不同尺寸的数据塌缩至一条光滑的曲线, 所得临界温度 T_c 和临界指数 ν 和理论值一致.

类似地, 对于六态clock模型中的两个BKT相变, 假设其标度率满足

$$L^b \sqrt{\text{Var}(r_H)} = \mathcal{F}_{\text{BKT}} \left(\frac{L}{e^{\frac{c}{\nu}}} \right), \quad (8)$$

这里 b 和 c 均为未知的拟合指数. 图4所示, 相应的数据塌缩得到了与文献中一致的相变温度. 尤为值得注意

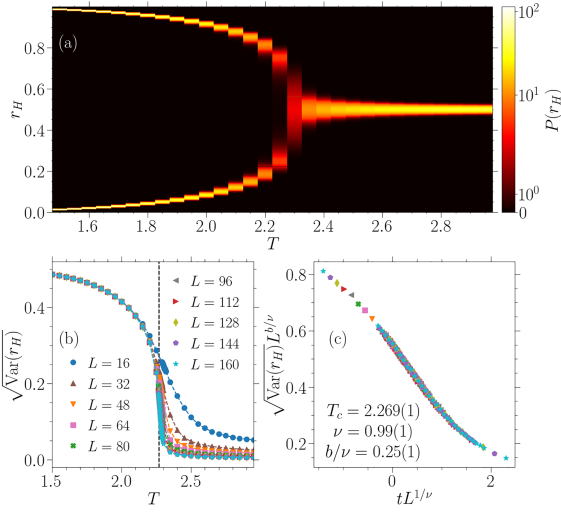


图 3 (a) 二维经典伊辛模型构型空间距离分布 $P(r_H)$ 随温度 T 的变化; (b) 构型距离分布标准差 $\sqrt{\text{Var}(r_H)}$; (c) 其有限尺寸数据塌缩结果. 图片数据来自文献[32]

Figure 3 (a) Evolution of the configuration-space distance distribution $P(r_H)$ with temperature T for the 2D classical Ising model; (b) the standard deviation of the configuration distance distribution $\sqrt{\text{Var}(r_H)}$; (c) its finite-size data collapse result. The data is adapted from ref. [32].

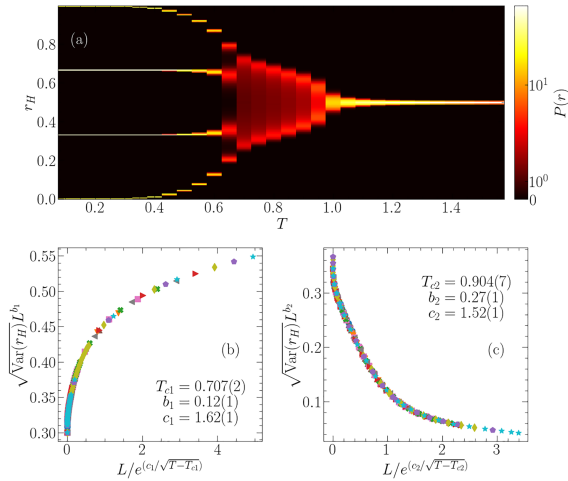


图 4 (a) 二维 6-state clock 模型构型空间距离分布 $P(r_H)$ 随温度 T 的变化; (b) 构型距离分布标准差 $\sqrt{\text{Var}(r_H)}$; (c) 其有限尺寸数据塌缩结果. (a) 对应系统边长为 $L = 80$; (b) 和 (c) 数据点对应尺寸为 $L = 12$ –174, 间隔为 18. 图片数据来自文献[32]

Figure 4 (a) Evolution of the configuration-space distance distribution $P(r_H)$ with temperature T for the 2D 6-state clock model; (b) the standard deviation of the configuration distance distribution $\sqrt{\text{Var}(r_H)}$; (c) its finite-size data collapse result. (a) corresponds to a system size of $L = 80$; data points in (b) and (c) correspond to sizes from $L = 12$ to 174 with a step of 18. The data is adapted from ref. [32].

的是, 通过数值拟合对比不同相变类别下的数据塌缩函数形式, 距离标准差 $\sqrt{\text{Var}(r_H)}$ 还可用于判别相变的类型, 这为研究未知系统的相变特性提供了有力工具 [32].

此外, 图 4 中 6 态时钟模型的 $P(r_H)$ 与图 1 中笼目自旋冰模型的结果极为相似——尽管这两个模型的局域自由度不同, 但其相变属于同样的相变普适类. 该结果预示了构型距离分布可能包含了相变的普适类信息.

与此同时, 参与熵 S_{r_H} 为临界点的确定提供了有力的交叉验证. 如图 5 所示, 二维伊辛模型的归一化参与熵 $S_{r_H}/S_{\max}$ 在相变点表现出锐利的尖峰; 通过对峰的位置做有限尺寸外推, 所得的相变点在误差范围内和对距离标准差 $\sqrt{\text{Var}(r_H)}$ 数据塌缩的结果是吻合的. 文献[32]展示了一系统经典自旋模型的结果, 面对各类模型的不同局域自由度和相变类型, 参与熵均能较为准确地确定相变点.

3.2.2 普适经验性标度律的发现

在系统性数值实验取得成功的基础上, 进一步的分析揭示出一条更为深刻的数值规律. 值得注意的是, 从 $\sqrt{\text{Var}(r_H)}$ 的有限尺寸标度行为中直接提取的标度指数——即 b/ν (适用于连续相变) 或 b (适用于 BKT 相变)——在文献[32]研究的所有经典自旋模型中, 均与相应相变普适类的反常维度 η 高度一致. 这一结果表明, 可能存在一个跨越多种相变类型的普适性经验标

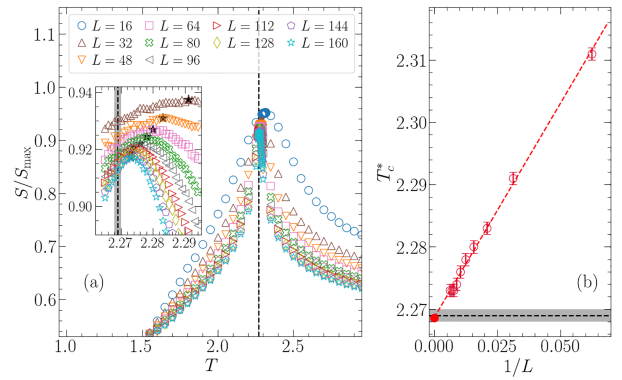


图 5 (a) 二维经典伊辛模型中, 由 $S_{r_H}/S_{\max}$ 峰值位置定义的有限尺寸临界点 $T_c^*(L)$; (b) 其有限尺寸外推外插. 图片改编自文献[32]

Figure 5 (a) The critical points $T_c^*(L)$ defined by the peak positions of $S_{r_H}/S_{\max}$ for the Ising model; (b) its finite-size extrapolation. The figure is adapted from ref. [32].

度律:

$$\frac{b}{\nu} = \eta \text{ (对于连续相变)}, \quad (9)$$

$$b = \eta \text{ (对于BKT相变)}, \quad (10)$$

其中, 反常维度 η 是临界现象理论中的核心临界指数之一, 它刻画了系统在临界点处实空间两点关联函数 $C(r)$ 的幂律衰减行为: $C(r) \sim r^{2-d-\eta}$. 该经验关系将构型涨落统计的标度特性与关联函数的代数衰减直接联系起来, 暗示其背后可能存在更深层次的理论结构, 值得在未来工作中深入探究.

至此, 文献[32]在两个层面上取得了重要进展: 首先, 成功将构型距离分析法从针对特定模型的方法推广为一个普适性理论框架, 并确立了 $\sqrt{\text{Var}(r_H)}$ 与 S_{r_H} 作为其中稳健的核心观测量. 其次, 更为重要的是, 我们发现了一个深刻且普适的经验性标度律. 然而, 这一经验性标度律的出现本身构成了一大理论谜题: 为何一个定义在抽象高维构型空间中的几何统计量——即距离分布的宽度——其标度行为竟会精确对应于描述实空间物理现象的核心临界指数——反常维度 η , 即关联函数的长程衰减特性? 这两个分别属于物理描述不同层次的量之间, 似乎存在着某种尚未被揭示的、深刻的理论联系. 建立连接二者之间的理论桥梁, 并从第一性原理出发为这一数值规律提供根本性解释, 已成为我们下一阶段研究的核心目标与主要驱动力.

4 理论的统一: 从量子相变的挑战到解析框架的建立

前两节的工作成功地建立并验证了一套通过构型距离分布分析经典相变的普适框架, 并最终凝练出一个深刻的经验性标度律. 然而, 一个自然且关键的问题随之而来: 这套在热涨落驱动的经典相变中被证明行之有效的框架, 是否同样适用于由量子涨落驱动的量子相变, 是一个亟待验证的关键问题. 本节将以文献[33]的内容为基础回答这一问题. 在将研究拓展至量子领域时所遭遇的一个关键“反常”现象, 反过来激发了对方法本身的深刻反思, 并最终催生了一个能够统一解释整个系列工作的解析理论框架.

4.1 量子相变与关键反常现象: 横场伊辛模型

为了探究我们方法在量子相变中的有效性, 我们选择横场伊辛模型(Transverse-field Ising Model, TFIM)作为研究范例. 该模型是理论与实验上研究量子相变最具代表性的“范例系统”之一, 其哈密顿量为

$$H_{\text{TFIM}} = -J \sum_{\langle i,j \rangle} \hat{\sigma}_i^z \hat{\sigma}_j^z - h \sum_i \hat{\sigma}_i^x, \quad (11)$$

其中, $J > 0$ 表示铁磁相互作用强度, h 为横向磁场的强度. 在一维情况下, 该模型在临界磁场 h_c 处经历一个从铁磁相(当 $h < h_c$)到顺磁相(当 $h > h_c$)的二阶量子相变, 其普适类与二维经典伊辛模型相同.

我们首先在对应于传统序参量 $\langle \hat{\sigma}^z \rangle$ 的 $\hat{\sigma}^z$ 表象基矢下, 采用随机级数展开(Stochastic Series Expansion, SSE)量子蒙特卡罗方法[34]对系统基态进行采样, 并计算不同构型之间的汉明距离分布 $P(r_H)$. 数值结果表明, 该方法在量子体系中初步取得了预期效果. 如图6所示, 汉明距离的标准差 $\sqrt{\text{Var}(r_H)}$ 在临界点 $h_c = 1$ 附近展现出清晰的临界标度行为, 并呈现出良好的数据塌缩特性. 这一结果初步验证了构型空间几何分析方法在研究量子相变问题中的可行性.

为进一步检验该方法的鲁棒性, 我们设计并开展了一项关键的控制实验: 在与序参量正交的 $\hat{\sigma}^x$ 基底下重复相同的测量与分析流程. 令人意外的是, 出现了一个显著的“反常现象(Anomaly)”: 如图7所示, $\sqrt{\text{Var}(r_H)}$

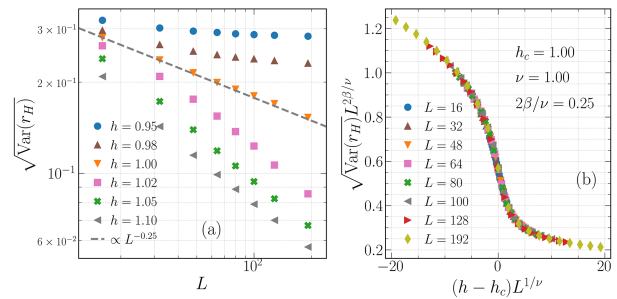


图6 一维横场伊辛模型在 $\hat{\sigma}^z$ 基矢下 $\sqrt{\text{Var}(r_H)}$ 的有限尺寸行为. (a) $\sqrt{\text{Var}(r_H)}$ 在临界点附近随系统尺寸 L 的幂律衰减行为; (b) 采用二维伊辛普适类临界指数的数据塌缩. 图片改编自文献[33]

Figure 6 Finite-size behavior of $\sqrt{\text{Var}(r_H)}$ on the $\hat{\sigma}^z$ basis for the 1D TFIM. (a) The power-law decay behavior of $\sqrt{\text{Var}(r_H)}$ with the system size L near the critical point; (b) data collapse using the critical exponents of the 2D Ising universality class. The figure is adapted from ref. [33].

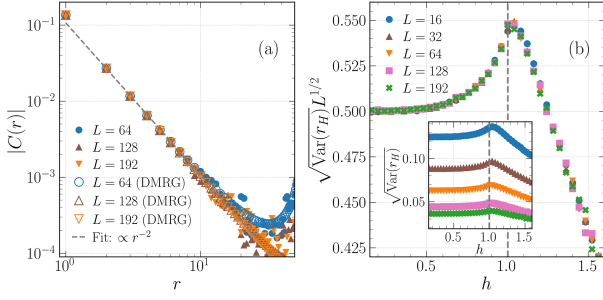


图 7 一维横场伊辛模型在 $\hat{\sigma}^x$ 基矢下 $\sqrt{\text{Var}(r_H)}$ 的有限尺寸行为. (a) $\sqrt{\text{Var}(r_H)}$ 在临界点附近随系统尺寸 L 的快速幂律衰减行为; (b) $\sqrt{\text{Var}(r_H)}$ 经过非临界重标度后的数据塌缩图, 子图为重标度前原始数据. 图片改编自文献[33]

Figure 7 Finite-size behavior of $\sqrt{\text{Var}(r_H)}$ in the $\hat{\sigma}^x$ basis for the 1D TFIM. (a) The rapid power-law decay behavior of $\sqrt{\text{Var}(r_H)}$ with the system size L near the critical point; (b) the data collapse plot of $\sqrt{\text{Var}(r_H)}$ after non-critical rescaling by $L^{-d/2}$, the inset shows the original data before rescaling. The figure is adapted from ref. [33].

的临界行为完全消失. 尽管其数值随横场 h 平滑变化, 但在临界点附近, 不同系统尺寸的曲线既未表现出发散趋势, 也未出现标度交叉, 因而无法进行有效的有限尺寸标度分析[33].

这一反常现象表明, $\sqrt{\text{Var}(r_H)}$ 所呈现的临界行为并非一种普适的、与测量方式无关的数学特性, 而是深刻依赖于系统的物理对称性及观测所采用的基矢选择. 这个看似“失效”的控制实验, 恰恰揭示了我们对于这一方法背后物理机制理解的局限, 进而促使我们重新审视 $\text{Var}(r_H)$ 的物理意义, 特别是其对应的可观测量本质, 以及其基底敏感性的微观起源. 为此, 我们必须回归第一性原理, 构建构型空间中的几何量与实空间物理量之间的解析联系, 从而揭示 r_H 涨落背后的深层物理图像.

4.2 解析的桥梁: 连接构型空间方差与实空间关联函数

让我们回归到最基本的伊辛自旋模型, 对汉明距离的统计矩进行严格而系统的理论推导. 考虑汉明距离的均值 $\langle r_H \rangle$, 根据定义式(1)并利用期望的线性性质, 可得

$$\langle r_H \rangle = \left\langle \frac{1}{2N} \sum_{i=1}^N (1 - s_i s'_i) \right\rangle = \frac{1}{2} \left(1 - \frac{1}{N} \sum_{i=1}^N \langle s_i s'_i \rangle \right). \quad (12)$$

由于 $|s\rangle$ 与 $|s'\rangle$ 是从同一系综中独立采样得到的, 两

点关联可分解为

$$\langle s_i s'_i \rangle = \langle s_i \rangle \langle s'_i \rangle = \langle \hat{\sigma}_i^\gamma \rangle^2,$$

其中 γ 表示测量所采用的自旋基矢方向(如 z 或 x 方向). 若系统具有平移不变性, 则所有格点上的局域磁化强度相等, 记为 $\langle \hat{\sigma}_i^\gamma \rangle = m_\gamma$. 因此, 上述表达式简化为

$$\langle r_H \rangle = \frac{1}{2} (1 - m_\gamma^2), \quad (13)$$

其中 m_γ 具有明确的物理含义: 它代表在 γ 基矢下单个自旋的系综平均磁化强度.

需要特别指出的是, 在热力学极限下可能发生自发对称性破缺的系统(例如横场伊辛模型中的铁磁相), 尽管单次测量或有限尺寸模拟通常会落入某个特定的对称破缺态(如多数自旋向上, 导致 $m_z \neq 0$), 但一个完整且符合哈密顿量对称性的统计系综必须包含所有简并真空态的等权叠加. 对于具有 $\mathbb{Z}_2$ 对称性的系统, 正负磁化态在系综中权重相同, 因而总磁化强度的系综平均严格为零, 即 $m_z = 0$. 此时, 在 $\hat{\sigma}^x$ 基下测得的平均汉明距离恒为

$$\langle r_H \rangle = \frac{1}{2}.$$

这一结果凸显了系综平均与实际观测之间的微妙差异, 也揭示了基矢选择在刻画多体系统几何结构和相变特性中的关键作用.

4.2.1 与实空间关联函数的联系

进一步, 我们转向核心物理量——构型距离的方差 $\text{Var}(r_H)$. 与距离平均值 $\langle r_H \rangle$ 的推导类似, 考虑同一系综中的独立抽样, 将同一格点不同构型的观测分解, $\text{Var}(r_H)$ 可被精确地推导为[33]

$$\text{Var}(r_H) = \frac{1 - m_\gamma^4}{4N} + \frac{1}{4N^2} \sum_{i \neq j} (C_{ij}^2 + 2m_\gamma^2 C_{ij}), \quad (14)$$

其中, $C_{ij} = \langle \hat{\sigma}_i^\gamma \hat{\sigma}_j^\gamma \rangle - \langle \hat{\sigma}_i^\gamma \rangle \langle \hat{\sigma}_j^\gamma \rangle$ 是实空间中的两点连通关联函数.

这个公式构成了我们整个理论框架的基石, 它在抽象的构型空间几何量 $\text{Var}(r_H)$ 与具体的实空间物理量 m_γ 和 C_{ij} 之间, 建立起了一座精确的解析桥梁. 该公式揭示了 $\text{Var}(r_H)$ 可分解为两个本质不同的贡献:

(1) 一项来源于单点涨落(第一项), 其幅值随系统尺寸 N 成反比衰减, 通常表现出非临界特征;

(2) 另一项则源于两点关联函数(第二项), 它将 $\text{Var}(r_H)$ 明确地与系统中所有格点对之间的关联性联系起来, 尤其包含了关联函数平方和 $\sum C_{ij}^2$ 的信息.

正是第二项的存在赋予了 $\text{Var}(r_H)$ 探测临界现象的能力: 在相变点附近, 随着长程关联的建立, $\sum C_{ij}^2$ 显著增强, 从而驱动 $\text{Var}(r_H)$ 展现出明显的临界行为. 特别地, 当满足 $m_\gamma = 0$ 的条件时——例如在 δ^z 基下测量铁磁相变, 或在任意基下探测顺磁相——式(14)可简化为如下更简洁的形式:

$$\text{Var}(r_H) \approx \frac{1}{4N} + \frac{1}{4N^2} \sum_{i \neq j} C_{ij}^2 (m_\gamma = 0). \quad (15)$$

这一简化形式成为后续所有标度分析的基础. 它清晰地表明, 在零序参量条件下, $\text{Var}(r_H)$ 的临界特性完全由实空间中两点关联函数的平方和所主导, 从而为我们提供了一个从几何视角理解多体系统临界性的有力工具.

4.2.2 与静态结构因子的联系

上述建立的解析关系式(15)在动量空间中存在一个等价且更为直观的对应形式. 对于包含 N 个格点的系统, 其静态结构因子 $S(\mathbf{q})$ 定义为实空间关联函数 $C(\mathbf{r})$ 的傅里叶变换:

$$S(\mathbf{q}) = \sum_{\mathbf{r}} e^{-i\mathbf{q} \cdot \mathbf{r}} C(\mathbf{r}).$$

根据帕塞瓦尔定理(Parseval's Theorem), 实空间中关联函数模平方的总和与动量空间中结构因子模平方的总和严格相等:

$$\sum_{i,j} |C(\mathbf{r}_i - \mathbf{r}_j)|^2 = \sum_{\mathbf{q}} |S(\mathbf{q})|^2.$$

将这一恒等式代入我们先前的核心关系式(15), 可得构型空间中汉明距离方差与动量空间物理量之间的基本联系:

$$\text{Var}(r_H) \propto \frac{1}{N^2} \sum_{\mathbf{q}} |S(\mathbf{q})|^2. \quad (16)$$

当式(21)成立, 即关联项主导系统的标度行为时, 该关系近似成立. 此式揭示了一个重要物理图像: 构型空间中距离分布的涨落强度, 正比于在整个布里渊区

内对静态结构因子强度的积分. 这为从动量空间视角理解构型涨落提供了清晰而有力的工具.

这一联系为我们的方法赋予了清晰的物理图像. 任何磁有序相的形成——无论是铁磁序($\mathbf{Q} = 0$)、反铁磁序($\mathbf{Q}$ 位于布里渊区边界), 还是其他更复杂的非公度磁序——其典型特征都是在对应的磁有序波矢 $\mathbf{Q}$ 处, 静态结构因子 $S(\mathbf{q})$ 中出现尖锐的布拉格峰. 这些峰的存在会导致 $S(\mathbf{q})$ 在布里渊区内积分 $\sum_{\mathbf{q}} |S(\mathbf{q})|^2$ 表现出显著的、随系统尺寸变化的奇异行为, 而这种奇异行为恰好被 $\text{Var}(r_H)$ 所捕捉. 因此, $\text{Var}(r_H)$ 能够灵敏地探测磁有序的出现, 其根本原因在于它等效于对整个动量空间中磁涨落的一种全局测量. 尽管对 $S(\mathbf{q})$ 的完整分析可以提供关于磁序类型更为细致的信息, 但 $\text{Var}(r_H)$ 作为一个简单的标量, 其优势在于无需预先知晓有序波矢 $\mathbf{Q}$ 的具体形式, 便能以“普适”的方式对任意类型磁序的萌发做出响应, 展现出广泛的适用性.

4.3 普适标度律及其适用判据

上一节建立的解析关系式(15)为从理论上理解 $\text{Var}(r_H)$ 的临界行为提供了坚实的基础. 该式表明, 在总磁化强度期望为零的系综中, $\text{Var}(r_H)$ 的临界特性完全由实空间两点关联函数的平方和 $\sum_{ij} C_{ij}^2$ 所主导. 本节将以此为核心出发点, 进一步得到 $\text{Var}(r_H)$ 在临界点附近所遵循的普适标度律.

4.3.1 从解析关系到普适标度律

为了推导 $\text{Var}(r_H)$ 的有限尺寸标度行为, 我们分析其在临界点($t = 0$)附近的行为. 根据式(15), $\text{Var}(r_H)$ 由两项构成. 第一项为非临界的背景项, 其标度行为为 $1/4N \sim L^{-d}$. 第二项是包含临界物理的关键项, 我们可以通过将求和近似为积分来分析其标度行为:

$$\begin{aligned} \frac{1}{4N^2} \sum_{i \neq j} C_{ij}^2 &\sim \frac{1}{L^{2d}} N \int^L C(r)^2 r^{d-1} dr \\ &\sim \frac{1}{L^d} \int^L C(r)^2 r^{d-1} dr. \end{aligned} \quad (17)$$

在临界点, 关联函数 $C(r)$ 呈幂律衰减, $C(r) \sim r^{2-d-z-\eta}$ (其中量子系统动力学指数 $z > 0$, 经典系统 $z = 0$). 代入上式积分, 我们得到该项的主要标度

行为:

$$\frac{1}{L^d} \int_0^L r^{2(2-d-z-\eta)} r^{d-1} dr \sim \frac{1}{L^d} L^{2(2-z-\eta)-d} \sim L^{4-d-2z-2\eta}. \quad (18)$$

利用超标度关系式(Hyperscaling Relation) $2\beta/\nu = d + z - 2 + \eta$, 我们可以将上式最终写为

$$\frac{1}{4N^2} \sum_{i \neq j} C_{ij}^2 \sim L^{-\frac{4\beta}{\nu}}. \quad (19)$$

因此, 距离方差 $\text{Var}(r_H)$ 的临界行为由 L^{-d} 项和 $L^{-4\beta/\nu}$ 项竞争决定. 只要临界项的衰减慢于或等于背景项, 它就将主导系统的临界标度. 由此, 我们得到了本系列工作的核心理论成果——距离标准差的普适标度律:

$$\sqrt{\text{Var}(r_H)} \sim L^{-\frac{2\beta}{\nu}}. \quad (20)$$

这一结果为第3.2.2节中发现的经验性标度律提供了坚实的理论基础: 有限尺寸标度指数等于反常维度 η 的现象, 正是该普适标度律的直接体现. 例如, 对于二维经典Ising普适类 ($d = 2, z = 0, \eta = 1/4$), 超标度关系给出 $2\beta/\nu = 2 + 0 - 2 + 1/4 = 1/4 = \eta$; 而对于BKT相变边界, 有 $\sqrt{\text{Var}(r_H)} \sim L^{-\eta}$, 我们的理论亦可自然推广至此类情形^[33]. 这一推导深刻揭示了构型空间中距离分布的展宽行为, 其临界特性本质上由实空间关联函数的长程衰减性质所决定.

4.3.2 理论的边界: 适用性判据

上述推导的成立依赖于临界项 $L^{-4\beta/\nu}$ 的衰减速率慢于或等于非临界背景项 L^{-d} . 这种幂律衰减之间的竞争关系, 自然为所采用的标度律划定了适用范围. 为了确保临界行为能够在有限尺寸系统中被清晰分辨, 必须满足

$$\frac{4\beta}{\nu} \leq d. \quad (21)$$

这个不等式与推导中所使用的 $m_\gamma = 0$ 条件一起, 构成了我们这套分析框架的两个基本适用性判据. 对于常见理论模型的相变, 该判据在大多数情况下都是满足的. 例如, 二维经典Ising模型 ($4\beta/\nu = 0.5 < 2$), 三维经典Ising模型 ($4\beta/\nu \approx 2.07 < 3$), 以及一维横场Ising模型 ($4\beta/\nu = 0.5 < 1$) 都满足条件.

4.4 理论的验证与对数值现象的解释

在构建了连接构型空间几何与实空间物理的解析框架之后, 我们现在能够基于该理论对第4.1节中TFIM模型的数值结果进行系统性的回顾与物理解释. 这一过程不仅为理论本身提供了严格的验证, 也为最初观察到的数值现象赋予了清晰且自洽的物理图像.

4.4.1 理论的直接验证(1D TFIM, $\hat{\sigma}^z$ 基)

我们首先考察一维横场伊辛模型在 $\hat{\sigma}^z$ 基矢下的测量结果. 该量子相变属于二维经典Ising普适类, 其临界指数为 $\beta = 1/8, \nu = 1$ ^[3]. 为验证理论的适用性, 我们检查以下两个判据: (1) 在包含所有简并基态贡献的完备系综平均中, 总磁化强度 m_z 严格为零; (2) 判据 $4\beta/\nu \leq d$ 成立, 即 $4(1/8)/1 = 0.5 \leq 1$. 两项条件均满足, 因此理论预言 $\sqrt{\text{Var}(r_H)}$ 应遵循式(20). 基于已知临界指数, 理论预测的有限尺寸标度指数为 $-2\beta/\nu = -2(1/8)/1 = -0.25$. 这一结果与我们在图6(a)中通过数值拟合得到的幂律行为高度一致, 从而为该理论框架提供了直接而有力的支持.

4.4.2 对 $\hat{\sigma}^x$ 基下反常现象的解释

最后, 该理论框架为第4.1节中观察到的核心现象——即在与序参量正交的 $\hat{\sigma}^x$ 基矢下, $\sqrt{\text{Var}(r_H)}$ 的临界行为完全消失——提供了清晰而深刻的物理解释.

根据式(14), $\text{Var}(r_H)$ 的临界特性源于关联函数 C_{ij} 在相变点附近所展现出的长程幂律行为. 然而, 在 $\hat{\sigma}^x$ 表象中, 相应的两点关联函数 C_{ij}^x 并不对应于序参量的关联. 已有理论分析与数值计算一致表明, 即使在临界点 $h_c = 1$ 处, C_{ij}^x 也仅呈现快速衰减的幂律形式 ($\sim r^{-2}$), 而在远离临界区的区域则表现为指数衰减^[35]. 这种缺乏长程发散特征的关联结构, 无法通过式(14)中的空间求和机制诱导出普适的临界标度行为.

因此, 我们的理论明确预言: 在 $\hat{\sigma}^x$ 测量基矢下, $\text{Var}(r_H)$ 不会呈现出与临界点 h_c 相关的普适临界现象. 这一预测与图7所示的数值结果高度吻合. 该一致性深刻揭示了 $\sqrt{\text{Var}(r_H)}$ 并非一个单纯的数学或统计构造, 而是一个具有明确物理内涵的“类序参量”观测量, 其敏感性直接根植于系统对称性破缺所伴随的临界关联. 换言之, $\sqrt{\text{Var}(r_H)}$ 是否能够有效探测量子相变, 取

决于测量基矢是否能捕获到序参量主导的长程关联信息. 这一发现进一步强调了观测手段与物理本质之间的紧密联系.

4.5 框架的统一: 对广义距离的理论支撑

本节到目前为止建立的解析框架, 其严格推导是基于伊辛自旋的汉明距离. 然而, 第3节所展示的方法的广泛普适性, 却是通过在 q 态时钟模型等上使用更广义的1-范数距离来验证的. 这就提出了一个关键问题: 即基于汉明距离推导出的理论, 为何能够完美解释基于1-范数距离得到的数值结果. 本节旨在回答这一问题, 从而将系列工作统一在一个自洽的理论框架之下.

为了更清晰地揭示其临界行为的本质, 我们首先引入一个在物理和数学上更易于处理的广义距离——余弦距离(Cosine Distance):

$$d'_i = \frac{1}{2}[1 - \cos(\theta_i - \theta'_i)]. \quad (22)$$

这一定义具有多重优势: 它与式(4)中的相互作用项直接关联, 物理意义明确; 作为一个光滑且解析的函数, 它便于进行理论分析与数学推导; 在 $q = 2$ 的极限下, 该距离能自然退化为汉明距离.

利用余弦距离, 其协方差 $\text{Cov}(d'_i, d'_j)$ 可以被精确地计算. 在满足 $\mathbb{Z}_q$ 或 $\mathbb{U}(1)$ 对称性的系综中, 可以严格证明:

$$\text{Cov}(d'_i, d'_j) = \frac{1}{8} \langle \cos(\theta_i - \theta_j) \rangle^2 = \frac{1}{8} C_{ij}^2. \quad (23)$$

这个简洁而严格的结果清晰表明, 余弦距离方差的临界行为完全由实空间基频关联函数的平方和 $\sum C_{ij}^2$ 主导, 因此必然遵循与汉明距离相同的普适标度律.

现在我们回到1-范数距离. 至于其行为与余弦距离一致的原因, 并非因为它们在数值上处处接近. 事实上, 对于小的角度差涨落 $\Delta\theta \rightarrow 0$ ——这对应于有序相中某个特定简并态内部的涨落——两者的行为截然不同:

$$d(\Delta\theta) = \frac{1}{\pi} |\Delta\theta| \propto |\Delta\theta|, \quad (24)$$

$$d'(\Delta\theta) = \frac{1}{2}(1 - \cos \Delta\theta) \approx \frac{1}{4}(\Delta\theta)^2 \propto (\Delta\theta)^2. \quad (25)$$

1-范数距离对小涨落的响应是线性的, 而余弦距离是二次的. 真正的等价性隐藏在它们完整的数学结

构以及临界现象的本质之中: 1-范数距离的协方差可以被精确地展开为一个包含所有谐波模式的级数:

$$\text{Cov}(d_i, d_j) = \frac{1}{2} \sum_{m=1}^{\lfloor \frac{q}{2} \rfloor} a_m^2 [C_{ij}^{(m)}]^2 + \mathcal{R}_q, \quad (26)$$

其中 $C_{ij}^{(m)}$ 是 m 阶谐波关联函数, $\mathcal{R}_q$ 是一个仅在特定离散情况下存在的修正项.

在临界点, 这个级数的行为被其第一项($m = 1$)以极高的精度主导. 这并非一个物理假设, 而是由两个因素共同决定的.

(1) 数学上的快速收敛因子: 级数中傅里叶系数的权重 a_m^2 以 $1/m^4$ 的速度快速衰减, 极大地抑制了高阶项. 例如, $m = 3$ 项的权重仅为 $m = 1$ 项的 $1/3^4 = 1/81$, $m = 5$ 项的权重仅为 $1/5^4 = 1/625$. 这是一个纯粹的、与模型无关的数学事实.

(2) 物理上关联函数的衰减特性: 在临界点, 系统的长程物理由衰减最慢的模式所主导, 这正是基频模式 $C_{ij}^{(m=1)}$. 更高阶的谐波关联函数 $C_{ij}^{(m>1)}$ 对应于能量更高的激发, 其关联长度更短, 衰减速度远快于基频模式.

正是这两个因素的共同作用, 使得1-范数距离的协方差在临界点有效地近似正比于 C_{ij}^2 , 从而表现出与余弦距离完全相同的临界标度行为.

根据本节的分析以及前期工作的经验可知, 在确定测量基矢的情况下, 距离的具体定义主要影响观测量的绝对数值, 而不会改变系统的临界标度行为及其在有限尺寸分析中的统计收敛性——至少对于 $\text{Var}(r_H)$ 而言是如此. 因此在实际应用中, 可依据系统局域自由度的几何特性灵活选用适配的度量形式. 例如, 针对二能级系统, 可采用汉明距离^[31]; 而对于定义在平面内的二维经典自旋系统, 既可在极坐标系下使用角度差或余弦距离作为度量标准^[32], 也可在正交坐标系中采用二维向量的欧几里得距离^[22].

5 超越局域探针: 全局与拓扑信息的提取

通过建立构型空间距离方差 $\text{Var}(r_H)$ 与实空间两点关联函数 C_{ij} 之间的解析关系, 我们深入揭示了 $\sqrt{\text{Var}(r_H)}$ 作为有效相变探针背后的物理机制. 随之而来的问题是这一理论框架的潜力是否止步于此.

概率分布 $P(r_H)$ 作为一个富含信息的对象, 极有可能还蕴藏着超越两点关联函数的、更为深刻的物理内涵。

我们将从 $\sqrt{\text{Var}(r_H)}$ 的物理内涵及其局限性入手, 阐明从分布 $P(r_H)$ 中提取更高阶信息的必要性. 在此基础上, 引入两类不同层次的“全局”探针: 其一是基于信息论、用于刻画分布整体形态的度量工具, 如参与熵与费雪信息; 其二是为捕捉非局域物理特征而专门构建的全局观测量——宇称指数. 通过将这些新型探针应用于传统方法难以有效处理的物理情境, 我们将展示构型空间几何分析法在研究范式上的拓展潜力: 从传统的对称性破缺相变, 延伸至拓扑物相等更为前沿的领域。

5.1 $\sqrt{\text{Var}(r_H)}$ 的物理本质与局限性

如第4节所述, $\sqrt{\text{Var}(r_H)}$ 的临界行为由实空间中两点关联函数的平方和 $\sum C_{ij}^2$ 所主导; 在动量空间中, 这等价于对静态结构因子 $S(q)$ 在整个布里渊区上的积分. 这一对应关系清晰地揭示了 $\sqrt{\text{Var}(r_H)}$ 的物理本质: 它本质上是一种与两点关联函数等效的实空间准局域探针。

这一本质属性解释了为何 $\sqrt{\text{Var}(r_H)}$ 在探测由对称性破缺所描述的传统相变(即LGW范式下的相变)时极为有效. 这是因为此类相变的特征——无论是铁磁、反铁磁, 还是其他更复杂的磁有序结构——最终都会反映在两点关联函数 C_{ij} 的长程行为上. 而 $\sqrt{\text{Var}(r_H)}$ 作为对 C_{ij} 的一种全局积分测度, 能够灵敏地捕捉到长程关联的形成与演化, 从而成为识别这类相变的有力工具。

然而, 正是这一物理本质也决定了 $\sqrt{\text{Var}(r_H)}$ 的局限性. 在面对那些无法通过局域序参量和两点关联函数有效描述的相变时, 该量可能失效. 一个典型的例子是对称性保护拓扑相(Symmetry-protected Topological, SPT)之间的相变. 在这类相变中, 系统在相变前后可能均具有有限的关联长度, 其两点关联函数不表现出任何奇异性, 相变的本质隐藏于波函数的非局域拓扑特性之中. 因此, 依赖于两点关联信息的探针难以区分不同的拓扑相, 也无法准确定位相变点. 为了将我们的分析框架推广至超越LGW范式的物理问题, 我们必须突破 $\sqrt{\text{Var}(r_H)}$ 的限制, 转而从完整的距离分布 $P(r)$ 出发, 发掘并构建能够捕捉全局或非局域特征的新观测

变量。

5.2 全局信息度量: 熵与费雪信息

事实上, 在第3节中, 我们已引入能够刻画 $P(r_H)$ 分布形态的全局性度量——参与熵 S_{r_H} . 作为信息论中的标准度量之一, 它依赖于概率分布 $P(r)$ 的整体结构, 而非局限于某个特定矩(如方差). 正如图5所示, S_r 的极值点能够精准地标识有限尺寸系统中的临界点, 表现出良好的标度行为. 这一结果初步表明, 通过挖掘概率分布的全局统计信息, 是揭示和定位量子多体系统临界行为的有效途径。

一个更为深刻的理解源于信息几何的视角. 我们可以将由系统参数(如横场 h)所参数化的概率分布族 $P(r_H; h)$ 视作一个统计流形. 在物理上, 相变点标志着系统状态发生突变的临界位置; 而在几何意义上, 这对应于该流形上出现几何奇异性的点. 刻画这种奇异行为的核心工具之一是费雪信息(Fisher Information) $I(h)$, 其定义为

$$I(h) = \sum_{r_H} P(r_H; h) \left[\frac{\partial \ln P(r_H; h)}{\partial h} \right]^2. \quad (27)$$

费雪信息量化了两个无穷接近的概率分布 $P(r_H; h)$ 与 $P(r_H; h + dh)$ 之间的“可区分度”. 当系统趋近临界点时, 概率分布对参数的微小变化变得极度敏感, 导致费雪信息在此处达到峰值. 因此, 费雪信息的极大值或许能精确地标识出系统的临界区域, 成为探测相变的有力几何探针。

图8的结果有力地支持了这一观点. 在一维横场伊辛模型中, 在 δ^x 基下测量的 $\sqrt{\text{Var}(r_H)}$ 无法提供完整的临界信号(见图7), 表现出对测量基底的敏感性. 相比之下, 费雪信息 $I(h)$ 在其峰值处展现出清晰且良好的标度行为, 能够精确确定量子相变点. 这一对比凸显了费雪信息在该框架中作为一种更强大、更具鲁棒性的全局探针的优势. 其核心优势在于不依赖于特定测量基的选择, 而是直接量化构型空间几何结构随系统参数变化的演化速率, 这一特性源于对概率分布流形本征几何的敏感性, 比分布静态信息更为根本. 正因如此, 费雪信息在序参量难以定义或未知的复杂系统中, 展现出识别和定位临界点的巨大潜力。

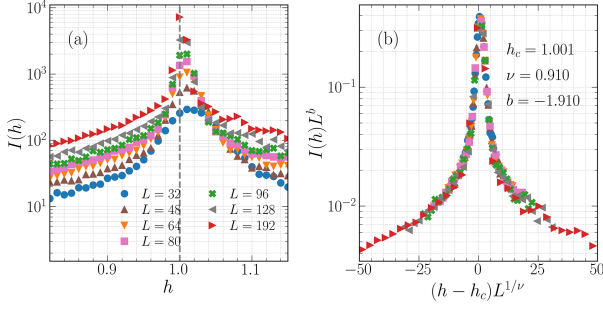


图 8 (a) 一维横场伊辛模型在 σ^x 基矢下的费雪信息 $I(h)$; (b) 其有限尺寸数据塌缩. 图片改编自文献[33]

Figure 8 (a) The Fisher information $I(h)$; (b) its finite-size data collapse for 1D TFIM in the σ^x basis. The figure is adapted from ref. [33].

5.3 探索非局域序: 宇称指数的构造与应用

费雪信息作为一种全局性的信息度量, 在“探测”相变方面展现出高度的普适性和灵敏性. 然而, 要深入“描述”相的本质特征, 尤其是诸如拓扑序这类具有非局域特性的量子序, 仅依赖费雪信息是不足的. 因此, 有必要从概率分布 $P(r_H)$ 出发, 构建一个全新的观测量, 使其能够有效捕捉和反映系统的全局拓扑结构.

5.3.1 局限性: 更高阶矩与特征函数谱并非理想选择

在探讨如何从 $P(r_H)$ 中提取非局域信息之前, 值得指出的是, 直接计算更高阶的统计矩, 例如三阶矩(偏度, Skewness)或四阶矩(峰度, Kurtosis)虽看似是一个自然的选择, 却不足以捕捉真正的非局域特征. 关键在于, 对于任意有限的 k , 距离分布的 k 阶矩所携带的物理信息最多仅涵盖系统中的 k 点局域关联函数, 其本质上仍局限于探测局域关联. 这意味着, 仅仅通过计算有限阶的矩, 我们无法从根本上跳出局域观测量的范畴. 此外, 从数值计算的角度看, 高阶矩的估计量对分布尾部的采样噪声极为敏感, 这会导致其统计误差显著增大, 降低了作为可靠探针的稳健性.

我们也考察了直接分析完整的特征函数谱 $\Phi(k) = \langle e^{ikr_H} \rangle$ 的路径, 但发现它在实践中应用困难, 并非理想的无监督探针. 为方便本节的讨论, 我们这里采用了未归一化的构型距离 $R_H = Nr_H$. 诚然, $\Phi(k)$ 在理论上能反映出 $P(R_H)$ 的周期性结构. 例如, 在 6 态时钟模型的低温有序相中, 其 $\mathbb{Z}_6$ 对称性会导

致 $P(R_H)$ 在 $R_H = 0, N/3, 2N/3, N$ 等处呈现周期性峰值, 这将在 $\Phi(k)$ 中表现为 $k_m = 6m\pi/N$ (m 为整数) 处的一系列布拉格峰.

然而, 试图利用这些布拉格峰来识别对称性的做法在实践中困难重重.

(1) 信号稳健性差: 在有限尺寸和热涨落的影响下, $P(R_H)$ 的峰会不可避免地发生展宽和移动. 这两种效应共同导致高阶布拉格峰信号被急剧削弱, 极易淹没在数值噪声中, 难以被可靠地识别.

(2) 依赖先验知识: 在上述例子中, 我们之所以在 $k_m = 6m\pi/N$ 处寻找信号, 是因为我们已知系统具有 $\mathbb{Z}_6$ 对称性. 对于一个未知系统, 我们无法先验地判断谱中的哪个峰对应真实的物理对称性, 这违背了我们寻求无偏、无监督方法的初衷.

5.3.2 对称性保护的拓扑态宇称指数: 一个非局域观测量的构造

即使不考虑完整的、复杂的 $\Phi(k)$ 谱, 一些特殊 k 值也有其深刻的物理意义. 以 $k = \pi$ 为例, 其定义了一个宇称指数 (Parity Index):

$$\begin{aligned} \chi &= \sum_{R_H=0}^N P(R_H) e^{i\pi R_H} \\ &= \sum_{R_H=0}^N P(R_H) (-1)^{R_H} \\ &= P(R_H = \text{even}) - P(R_H = \text{odd}). \end{aligned} \quad (28)$$

这一构造保留了特征函数的全局性, 以一种非平庸的方式整合了所有阶的关联信息; 且其是一个计算简单的实数, 避免了分析复数谱的复杂性; 同时, 这是一个无偏的观测量, 其计算不依赖于任何关于系统对称性的先验知识.

5.3.3 探索性应用: SPT 相变

我们将这一新构造的探针应用于一个典型的非 LGW 相变模型——一维交错键海森堡链, 其哈密顿量为

$$H = J_1 \sum_{i \in \text{odd}} \mathbf{S}_i \cdot \mathbf{S}_{i+1} + J_2 \sum_{i \in \text{even}} \mathbf{S}_i \cdot \mathbf{S}_{i+1}. \quad (29)$$

该模型在 $J_2/J_1 = 1$ 处经历一个从拓扑平庸相到拓扑非平庸相的 SPT 相变 [36,37]; 相变点两侧系统的体态完全相同, 这一相变无法由任何局域序参量来描述.

而基于构型距离分布 $P(r_H)$ 的统计几何分析却能清晰刻画这一拓扑相变. 在拓扑平庸相中, 我们观察到 $P(r_H)$ 的数值热图呈现出明显的条纹状结构, 如图 9(a) 所示, 这直观地反映了距离为偶数和奇数的概率存在显著的不平衡. 此时, χ 表现为一个有效的拓扑序参量. 如图 9(b) 所示, 在平庸相 ($J_2/J_1 < 1$) 中, χ 的值接近于一个较大的有限值; 而在跨越相变点进入非平庸相 ($J_2/J_1 > 1$) 后, χ 则迅速趋于零.

这一探索性研究表明, 我们的构型空间几何框架具有巨大的潜力. 完整的距离分布 $P(r_H)$ 中确实蕴含着超越传统两点关联的非局域信息. 通过构造如宇称指数 χ 这样的全局观测量, 我们有望将这套方法的应用领域从传统的对称性破缺相变, 拓展到更前沿的拓扑相变研究中.

6 总结与展望

6.1 总结

本文系统性地回顾了我们的课题组在构建“构型空间统计几何”这一相变研究新范式方面所开展的一系列工作. 研究的起点源于一项新颖的数值发现: 在笼目自旋冰系统中, 构型距离分布 $P(r_H)$ 的形态能够有效刻画其复杂的 BKT 相变行为. 通过系统的数值模拟与分析, 我们进一步将这一观察提炼为一条普适的经验标

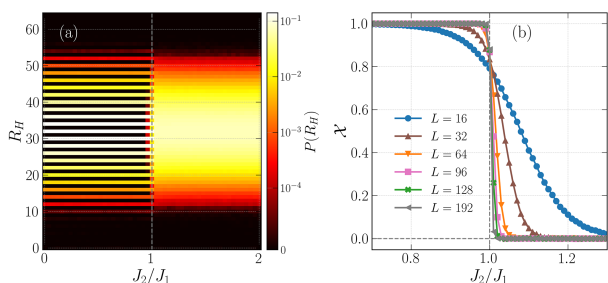


图 9 一维交错键海森堡链在开边界条件下的构型距离分析. (a) $P(r_H)$ 的演化热图; (b) 宇称指数 χ 随 J_2/J_1 的变化. 图片数据改编自文献 [33]

Figure 9 Configuration distance analysis of the one-dimensional staggered-bond Heisenberg chain under open boundary conditions. (a) A heat map showing the evolution of $P(r_H)$; (b) the variation of the parity index χ with J_2/J_1 . The data are adapted from ref. [33].

度律——即 $\sqrt{\text{Var}(r_H)}$ 的有限尺寸标度指数与临界反常维度 η 完全等价.

这一经验规律的发现促使我们在面对量子相变带来的数值挑战时, 发展出一个坚实的解析理论框架. 该框架首次从第一性原理出发, 揭示了 $\text{Var}(r_H)$ 与实空间关联函数之间的深刻联系, 从而为前述标度律提供了严谨的理论解释. 在此基础上, 我们将这一方法推广至更具普适性的全局探针 (如费雪信息 $\mathcal{I}(h)$) 和非局域探针 (如宇称指数 χ), 并初步展示了这些工具在处理拓扑相变等超越传统 LGW 范式的复杂问题中的潜力.

综上所述, 我们的工作建立了一套利用构型空间几何表征相变的研究范式. 其核心贡献在于, 明确揭示并严格证明了宏观尺度上的临界普适性如何被深刻编码于微观构型集合的统计几何结构之中. 我们提出的方法是一种“白箱”式的、可解析的无监督数据驱动路径, 既保持了传统物理学的理论严谨性, 又融合了现代数据科学的强大分析能力, 为在缺乏先验知识的前提下探索复杂相变现象提供了崭新的理论视角与实用工具.

6.2 展望

从希尔伯特空间, 特别是构型距离分布的角度出发来刻画相变, 是一种新颖且富有潜力的相变研究范式. 我们认为以下几个方向尤为值得关注.

(1) 理论层面: 构型空间几何结构的信息边界. 当前的研究主要聚焦于由两构型距离分布 $P(r_H)$ 所刻画的几何特性. 然而, $P(r_H)$ 本质上源于对量子态在计算基底下测量概率 $|\langle s|\Psi\rangle|^2$ 的统计汇总, 这一过程不可避免地抹去了波函数中蕴含的全部相对相位信息. 由此引出一个深刻的理论问题: 我们在经典构型空间中所揭示的几何结构, 与完整保有量子相干性和相位关联的希尔伯特空间几何之间, 究竟存在何种内在对应关系? 换言之, 基于经典采样所得的统计几何, 在多大程度上能够反映或逼近真实的量子几何? 进一步厘清这类经典化分析方法的适用范围与理论极限, 不仅有助于评估现有构型空间研究范式的有效性, 更可能为突破经典采样瓶颈指明方向. 因此, 理解构型空间几何与量子几何之间的映射关系, 并发展能部分恢复相干性信息的理论工具, 将成为该领域未来最关键的理论挑战之一.

(2) 应用层面: 解决序参量未知的疑难问题. 将希尔伯特空间中的测量应用于当前凝聚态物理中那些序参量未知或存在争议的疑难问题, 是一个充满前景的研究方向. 一个典型的例子是阻挫磁体中的量子自旋液体(Quantum Spin Liquid, QSL), 其核心特征正是缺乏传统的磁长程序. 我们的理论已经揭示, $\text{Var}(r_H)$ 正比于静态结构因子 $S(q)$ 在整个布里渊区的积分. 因此, 从磁序相到 QSL 相的转变, 必然伴随着 $\text{Var}(r_H)$ 有限尺寸标度指数的质变, 这为实验和数值上精确绘制如 QSL 等奇异物相的相图边界提供了一个全新的、可操作的判据. 更进一步, 该框架在处理其他超越朗道对称性破缺范式的复杂问题方面也展现出巨大潜力. 例如, 在去禁闭量子临界点(Deconfined Quantum Critical Point, DQCP)的研究中, 系统在两个具有完全不同传统序参量的相之间发生连续相变, 传统的序参量描述面临根本性挑战. 而我们的方法不依赖于任何先验的序参量假设, 有望通过揭示构型空间几何结构的演化规律, 为这类非平凡相变提供一种统一且无偏的刻画方式. 此外, 该方法在自旋玻璃等无序系统中的应用也极具吸引力. 这类系统的物理本质源于高度复杂的构型空间结构和崎岖的能量景观. 我们的探针能够直接、定量地反映这些几何特征, 从而为深入探测无序系统中能量景观的拓扑与层级结构开辟了新的路径. 总体而言, 这一基于几何响应与构型空间分析的框架, 为解决一系列长期悬而未决的强关联与无序系统难题提供了强有力的工具和全新的视角.

(3) 实验层面: 连接量子模拟实验. 我们的理论框架与当前前沿的量子模拟实验具有天然的高度契合性. 在里德堡原子阵列、超导量子比特等主流实验平台上, 最直接获取的测量结果是通过量子系统进行重复投影所获得的大量构型“快照(Snapshots)”. 传统

数据分析方法通常需要从这些快照中重构特定的(有时是高阶的)关联函数, 这一过程不仅数据需求量大, 且在实验上面临显著挑战. 相比之下, 我们提出的方法——特别是对 $\text{Var}(r_H)$ 和 χ 的计算——仅依赖于对原始快照数据的全局统计分析, 无需精确测量复杂的多体关联函数. 这大幅降低了数据后处理的复杂度, 同时减少了对特定物理观测量的依赖. 更重要的是, 该方法提供了一种全新、物理意义清晰且基于第一性原理的数据分析范式, 能够高效处理实验产生的海量快照数据. 通过这一途径, 有望直接从实验数据中提取出临界指数等普适的临界行为信息, 而无需预先设定序参量或依赖相变的具体对称性破缺机制.

(4) 更广阔的视角: 在可解析的数据驱动方法中的定位. 最后, 值得将我们的工作置于当前“可解析的”数据驱动方法探索的更广阔背景中. 当前的一个重要趋势是从“黑箱”式的机器学习应用, 转向发展物理意义更透明的“白箱”方法. 这方面的典型例子包括主成分分析——其第一、第二主成分已被证明在许多模型中与物理量直接相关^[16–18], 和从抽样的微观构型出发的微观本征态类方法^[38,39]等. 相比之下, 一些更强大的流形学习方法, 如扩散图嵌入^[19,20], 虽然与我们的工作类似, 同样从构型间的距离矩阵出发, 但其后续引入的非线性步骤, 使得建立其输出与传统物理量之间的解析联系成为一个巨大的、悬而未决的挑战. 我们的“构型空间几何”框架则独辟蹊径: 它比主成分分析更普适, 同时又不像扩散图嵌入那样难以捉摸, 其核心观测量与基础物理量之间存在明确的解析联系, 在通用性与可解释性之间取得了精妙的平衡. 因此, 深入探索这类可解析的数据驱动方法, 理解不同方法之间的内在联系与各自的适用边界, 将是未来计算物理领域一个极具价值的研究方向.

参考文献

- 1 Cardy J. Scaling and Renormalization in Statistical Physics. Cambridge: Cambridge University Press, 1996
- 2 Landau L D, Lifshitz E M, Pitaevskii E M. Statistical Physics. New York: Butterworth-Heinemann, 1999
- 3 Sachdev S. Quantum Phase Transitions, 2nd ed. Cambridge: Cambridge University Press, 2011
- 4 Wilson K. The renormalization group and the expansion. *Phys Rep*, 1974, 12: 75–199
- 5 Balents L. Spin liquids in frustrated magnets. *Nature*, 2010, 464: 199–208
- 6 Thouless D J, Kohmoto M, Nightingale M P, et al. Quantized Hall conductance in a two-dimensional periodic potential. *Phys Rev Lett*, 1982, 49: 405–408

- 7 Kane C L, Mele E J. Z_2 topological order and the quantum spin Hall effect. *Phys Rev Lett*, 2005, 95: 146802
- 8 Nandkishore R, Huse D A. Many-body localization and thermalization in quantum statistical mechanics. *Annu Rev Condens Matter Phys*, 2015, 6: 15–38
- 9 Abanin D A, Altman E, Bloch I, et al. Colloquium: Many-body localization, thermalization, and entanglement. *Rev Mod Phys*, 2019, 91: 021001
- 10 Logan D E, Welsh S. Many-body localization in Fock space: A local perspective. *Phys Rev B*, 2019, 99: 045131
- 11 Sutradhar J, Ghosh S, Roy S, et al. Scaling of the fock-space propagator and multifractality across the many-body localization transition. *Phys Rev B*, 2022, 106: 054203
- 12 Roy S, Logan D E. The fock-space landscape of many-body localisation. *J Phys: Condens Matter*, 2024, 37: 073003
- 13 Guo Q, Cheng C, Li H, et al. Stark many-body localization on a superconducting quantum processor. *Phys Rev Lett*, 2021, 127: 240502
- 14 Yao Y, Xiang L, Guo Z, et al. Observation of many-body Fock space dynamics in two dimensions. *Nat Phys*, 2023, 19: 1459–1465
- 15 Carrasquilla J, Melko R G. Machine learning phases of matter. *Nat Phys*, 2017, 13: 431–434
- 16 Hu W, Singh R R P, Scalettar R T. Discovering phases, phase transitions, and crossovers through unsupervised machine learning: A critical examination. *Phys Rev E*, 2017, 95: 062122
- 17 Kiwata H. Deriving the order parameters of a spin-glass model using principal component analysis. *Phys Rev E*, 2019, 99: 063304
- 18 Qi N, Wang S J, Jin T, et al. Theoretical derivation of principal component analysis results for the two-dimensional Ising model. *Sci Sin-Phys Mech Astron*, 2024, 54: 230512
- 19 Rodriguez-Nieva J F, Scheurer M S. Identifying topological order through unsupervised machine learning. *Nat Phys*, 2019, 15: 790–795
- 20 Lidiak A, Gong Z. Unsupervised machine learning of quantum phase transitions using diffusion maps. *Phys Rev Lett*, 2020, 125: 225701
- 21 Facco E, d'Errico M, Rodriguez A, et al. Estimating the intrinsic dimension of datasets by a minimal neighborhood information. *Sci Rep*, 2017, 7: 12140
- 22 Mendes-Santos T, Turkeshi X, Dalmonte M, et al. Unsupervised learning universal critical behavior via the intrinsic dimension. *Phys Rev X*, 2021, 11: 011040
- 23 Mendes-Santos T, Angelone A, Rodriguez A, et al. Intrinsic dimension of path integrals: Data-mining quantum criticality and emergent simplicity. *PRX Quantum*, 2021, 2: 030332
- 24 Chern G W, Tchernyshyov O. Magnetic charge and ordering in kagome spin ice. *Philos Trans R Soc A: Math Phys Eng Sci*, 2012, 370: 5718–5737
- 25 Lapilli C M, Pfeifer P, Wexler C. Universality away from critical points in two-dimensional phase transitions. *Phys Rev Lett*, 2006, 96: 140603
- 26 Hwang C O. Six-state clock model on the square lattice: Fisher zero approach with Wang-Landau sampling. *Phys Rev E*, 2009, 80: 042103
- 27 Kosterlitz J M, Thouless D J. Ordering, metastability and phase transitions in two-dimensional systems. *J Phys C-Solid State Phys*, 1973, 6: 1181–1203
- 28 Tobochnik J. Properties of the q -state clock model for $q = 4, 5$, and 6 . *Phys Rev B*, 1982, 26: 6201–6207
- 29 Takagi T, Mekata M. Magnetic ordering of Ising spins on kagomé lattice with the 1st and the 2nd neighbor interactions. *J Phys Soc Jpn*, 1993, 62: 3943–3953
- 30 Wills A S, Ballou R, Lacroix C. Model of localized highly frustrated ferromagnetism: The kagomé spin ice. *Phys Rev B*, 2002, 66: 144407
- 31 Su W Y, Hu F, Cheng C, et al. Berezinskii-Kosterlitz-Thouless phase transitions in a kagome spin ice by a quantifying Monte Carlo process: Distribution of Hamming distances. *Phys Rev B*, 2023, 108: 134422
- 32 Su W Y, Liu Y J, Ma N, et al. Probing phase transitions with correlations in configuration space. *Phys Rev B*, 2024, 110: 195108
- 33 Liu Y J, Su W Y, Yang Y F, et al. On the criticality of the configuration-space statistical geometry. arXiv: 2508.00787
- 34 Sandvik A W. Stochastic series expansion method for quantum Ising models with arbitrary interactions. *Phys Rev E*, 2003, 68: 056701
- 35 Pfeuty P. The one-dimensional Ising model with a transverse field. *Ann Phys*, 1970, 57: 79–90
- 36 Su W P, Schrieffer J R, Heeger A J. Solitons in polyacetylene. *Phys Rev Lett*, 1979, 42: 1698–1701
- 37 Verresen R, Moessner R, Pollmann F. One-dimensional symmetry protected topological phases and their transitions. *Phys Rev B*, 2017, 96: 165124
- 38 Hu G K, Liu T, Liu M X, et al. Condensation of eigen microstate in statistical ensemble and phase transition. *Sci China-Phys Mech Astron*, 2019, 62: 990511
- 39 Tang Y Y, Zhong Y, Lu H. Characterizing spin ordering via maximal row correlation in classical spin models. *Phys Rev E*, 2025, 112: 034123

Characterizing phase transitions from statistics of configuration-space distances

LIU YuJing¹, SU WenYu¹, MA NvSen² & CHENG Chen^{1*}

¹*Gansu Provincial Research Center for Basic Disciplines of Quantum Physics, Key Laboratory of Theoretical Physics of Gansu Province, Key Laboratory of Quantum Theory and Applications of Ministry of Education, Lanzhou Center for Theoretical Physics, School of Physical Science and Technology, Lanzhou University, Lanzhou 730000, China*

²*School of Physics, Beihang University, Beijing 102206, China*

**Corresponding author (email: chengchen@lzu.edu.cn)*

This paper systematically reviews a series of works of our research group in the new paradigm of phase transition research, namely “statistical geometry of configuration space”. This method identifies phase transitions from the raw data in an unsupervised manner and extracts universal critical behaviors by analyzing the distance distribution between microscopic configurations sampled by Monte Carlo. We demonstrate how this paradigm has evolved from a numerical exploration of complex models to a universal empirical scaling law, and, finally, to an analytical theoretical framework that connects geometric quantities in configuration space with correlation functions in real space. This work reveals how macroscopic criticality is encoded in the geometry of microscopic configurations, providing a brand-new, “white-box” and analytically tractable data-driven perspective for studying complex phase transition problems with unknown order parameters.

phase transitions, lattice models, configuration space, critical phenomena, Monte Carlo simulation

PACS: 05.50.+q, 05.70.Fh, 05.10.Ln

doi: [10.1360/SSPMA-2025-0549](https://doi.org/10.1360/SSPMA-2025-0549)