arXiv:2305.00997v1 [hep-th] 1 May 2023

经典神经网络与量子神经网络在纠缠熵上的表达性

Chih-Hung Wu[a] **和 Ching-Che Yen**[b]

[a]美国加利福尼亚州圣巴巴拉市 *93106*，加利福尼亚大学圣塔芭芭拉分校物理系[b]联发科技股份有限公司

电子邮箱：chih-hungwu@physics, ucsb, edu, johnson.yan@mediatek.com

Abstract: 在量子场论中，将冯·诺依曼熵从 R´enyi 熵解析延拓是一项具有挑战性的任务。虽然可以利用量子场论路径积分表示中的复制方法计算第 n 阶 R´enyi 熵，但目前只能针对一些简单系统逐案实现解析延拓。本文提出了一个利用经典和量子神经网络进行监督学习的通用框架，以解决这一问题。我们首先研究若干已知冯·诺依曼熵的例子，其中输入数据通过用生成函数表示 $\mathrm{Tr}\,\rho_A^n$ 生成。我们采用 KerasTuner，在数据有限的情况下确定最优网络架构和超参数。此外，我们还针对量子机器学习模型构建了一个类似问题，并证明了量子模型能够以部分傅里叶级数的形式表达纠缠熵。我们提出的方法能够准确预测冯·诺依曼熵和 R´enyi 熵的数值，展现了深度学习技术在解决量子信息论问题方面的潜力。

目录

1 引言 **1**

2 冯·诺依曼熵的解析延拓，源自 R´enyientropies **4**

3 深度学习冯·诺依曼熵 **7**

3.1 模型架构与训练策略 8 3.2 单个区间的纠缠熵 11 3.3 两个不相交区间的纠缠熵 16

4 R´enyientropies：序贯深度学习 **23**

4.1 模型架构与训练策略 24 4.2 序贯模型示例 27

5 《Quantum》神经网络与冯·诺依曼熵 **29**

5.1 变分量子机器学习模型的傅里叶级数 29 5.2 生成函数作为傅里叶级数 32 5.3 量子模型对纠缠熵的表达性 34 5.4 恢复冯·诺依曼熵 39

6 讨论 **43**

生成函数的傅里叶级数表示 **46** **B Gegenbauer 多项式与吉布斯现象** **49**

1 引言

《style id='1'>冯·诺依曼熵 被广泛认为是衡量量子纠缠的有效指标，通常也称为《style id='3'>纠缠熵。对纠缠熵的研究带来了许多有价值的应用，尤其是在量子信息和量子引力领域（综述见 [1, 2]）。然而，对于一般系统而言，在量子场论中将《style id='6'>R´《style id='9'>ényi 熵 解析延拓至冯·诺依曼熵仍然是一项挑战。我们利用经典

和量子神经网络考察其在纠缠熵方面的表达能力，以及从 R´ényi 熵重构冯·诺依曼熵是否能更为简便。

量子场论（QFT）提供了一种高效方法，可计算整数 $n > 1$ 下的第 n 阶 R´enyi 熵，其定义为 [3]

$$S_n(\rho_A) \equiv \frac{1}{1-n} \ln \mathrm{Tr}(\rho_A^n). \tag{1.1}$$

该计算通过将约化密度矩阵 ρ_A 的路径积分表示 n 次复制来完成。这一步并非易事；不过，我们主要关注能够得到 Rényi 熵显式解析表达式的例子，尤其是二维共形场论（CFT_2）中的情形 [4–7]。随后，对 $n \to 1$ 进行解析延拓，得到冯·诺依曼熵

$$S(\rho_A) = \lim_{n \to 1} S_n(\rho_A). \tag{1.2}$$

解析延拓可以视为一个独立于计算第 n 阶 R´enyi 熵的问题。尽管 Carlson 定理保证了通过延拓得到的 $S(\rho_A)$ 的唯一性，但目前大多数情况下尚无闭式解析表达式。

此外，虽然 $S_n(\rho_A)$ 在整数和非整数 n下都有定义，但仅确定一组整数值 $n > 1$ 下的结果并不足够。为了得到冯·诺依曼熵，我们还必须令 $n \to 1$ 经过一段实数 $n > 1$ 范围取极限。因而，Rényi 熵与冯·诺依曼熵之间的关系颇为复杂，而且要精确数值近似 $S(\rho_A)$，所需的 n 值尚不明确。

沿着这一思路，我们受到 [8], 中所提出的另一种方法的启发，该方法能够让我们“累积地”研究高阶 R´ényi 熵与冯·诺依曼熵之间的联系。该方法依赖于定义一个表现为泰勒级数的生成函数

$$G(w; \rho_A) = \sum_{k=1}^{\infty} \frac{\tilde{f}(k)}{k} w^k, \quad \tilde{f}(k) = \mathrm{Tr}[\rho_A (1-\rho_A)^k]. \tag{1.3}$$

对 k 显式求和可得到一个绝对收敛级数；随着 $w \to 1$，该级数对冯·诺依曼熵的近似准确率不断提高。这种方法兼具数值和解析方面的优势，显式示例见 [8]。请注意，我们近似冯·诺依曼熵时所能达到的准确率取决于 k中部分和的截断方式；该方式因具体情形而异，且

难以评估。在共形场论的一般双区间情形下，计算高阶黎曼-西格尔θ函数尤其具有挑战性$_2$ [8], ，这仍然是一个尚未解决的问题。

另一方面，深度学习技术凭借其通用逼近性质，已成为解决解析延拓问题 [9–14], 的强大工具。通用逼近定理指出，在温和的假设条件下，人工神经网络能够逼近任意连续函数 [15], ，而冯·诺依曼熵也不例外。神经网络通过已知函数值的数据集进行训练，目标是学习一个潜在流形，使其能够在已知参数空间内逼近原函数。训练完成后，可通过对训练好的网络进行外推，在该空间之外作出预测。目标是尽量减小模型输出与实际函数值之间的预测误差。在本研究中，我们以两种不同方式构建监督学习任务：第一种方法使用稠密连接神经网络预测冯·诺依曼熵；第二种方法则利用序贯学习模型提取更高阶的 Rényi 熵。

我们没有采用静态的“定义后运行”方案（即预先定义模型结构，并使其在整个训练过程中保持固定），而是选择了动态的“运行时定义”方法。我们的目标是根据输入的验证数据自动确定最佳模型复杂度和超参数。为此，我们使用 KerasTuner [16] 和贝叶斯优化，通过交叉验证对不同的神经网络配置进行训练和评估，从而高效探索超参数空间。KerasTuner 利用结果更新超参数空间的概率模型，再据此建议下一组待评估的超参数，以期最大限度地提升预期性能。

一个类似的问题可以明确地从量子机器学习的角度来提出：我们可以通过量子门对数据输入和可训练权重进行编码，利用可训练的量子电路模拟神经网络。这种方法有许多不同的名称 [17–22], ，但我们将其称为量子神经网络。与经典神经网络不同，量子神经网络由一系列定义明确的酉变换构成，而非通过数值优化权重来实现目标值与数据之间的非线性映射。这就引出了量子计算实践者面临的一个根本问题：任意酉变换都能实现吗？还是说可学习函数类具有某种特定的刻画方式？换言之，量子模型能否普遍地表达给定数据输入所对应的任意函数？回答这些问题不仅有助于设计未来的算法，也能让我们更深入地了解量子模型如何实现通用逼近[23, 24]。

近期关于量子神经网络的研究进展表明，数据编码策略对其表达能力起着关键作用。数据编码问题已成为大量理论和数值研究的主题 [25–28]。在这项工作中，我们以 [29, 30], 提出的思路为基础；该工作展示了量子模型作为部分傅里叶级数的表达性。通过将冯·诺依曼熵的生成函数改写为傅里叶级数，我们同样可以确立量子神经网络的表达性。然而，傅里叶级数中的吉布斯现象给冯·诺依曼熵的恢复带来了挑战。为克服这一问题，我们将傅里叶级数展开到 Gegenbauer 多项式基底中，以此重构熵。

本文结构如下。第2节简要概述了在量子场论框架下，如何从 R´enyi 熵解析延拓得到冯·诺依曼熵。此外，我们还介绍了全文采用的生成函数法。第3节中，我们使用 KerasTuner 配合稠密连接神经网络，从若干已知解析表达式的示例中提取冯·诺依曼熵。第4节中，我们采用序贯学习模型来提取更高阶的 R´enyi 熵。第5节专门研究量子神经网络逼近冯·诺依曼熵的表达能力。第6节总结了我们的研究发现，并讨论了该方法的可能应用。附录A详细介绍了如何将生成函数改写为部分傅里叶级数，而附录B 则利用 Gegenbauer 多项式讨论吉布斯现象。

2 从 Rényi 熵解析延拓冯·诺依曼熵

下面讨论如何计算量子场论中的冯·诺依曼熵 [31–34]。假设我们从一个定义在 d 维闵可夫斯基时空上的量子场论开始，其希尔伯特空间定义在该时空的柯西切片 Σ 上。不失一般性，我们可以将 Σ 划分为两个互不相交的子区域 $\Sigma = A \cup A^c$。这里，A^c 表示 A 的补子区域。因此，希尔伯特空间也可以分解为张量积 $\mathcal{H}_\Sigma = \mathcal{H}_A \otimes \mathcal{H}_{A^c}$。随后，我们从 Σ 上的一个纯态定义约化密度矩阵 ρ_A；该约化密度矩阵因此为混态，用于描述两个区域之间的纠缠。冯·诺依曼熵 $S(\rho_A)$ 可以用来量化这种纠缠

$$S(\rho_A) \equiv -\operatorname{Tr}(\rho_A \ln \rho_A) = \frac{\operatorname{Area}(\partial A)}{\epsilon^{d-2}} + \cdots . \tag{2.1}$$

此外，它还具有诸多良好性质，例如在酉变换下保持不变、对纯态具有互补性，以及在纯态和最大

混合态之间平滑过渡，因此它是衡量 A 与 A^c之间纠缠程度的精细指标。第二个等式适用于场论，此时我们需要一个长度尺度 ϵ 来调节短程关联中包含的紫外发散。领先阶发散由纠缠面的面积 ∂A所刻画，这是量子场论 [35]的普适特征。[1]

人们一直在努力更深入地理解量子场论中的纠缠结构，包括自由理论 [36], 热核 [37, 38], 共形场论技术 [39] 以及基于 AdS/CFT 的全息方法 [40, 41]。然而，在实际操作中，对一般的相互作用量子场论进行冯·诺依曼熵的解析或数值计算，仍然是一项艰巨的挑战。综述请参见 [1]。

路径积分提供了一种求取 S（ρ_A）的一般方法。该方法从 R´enyi 熵 [3]开始

$$S_n(\rho_A) = \frac{1}{1-n} \ln \operatorname{Tr} \rho_A^n, \tag{2.2}$$

对于实数 $n > 1$。如前所述，通过对 n 进行解析延拓来求得冯·诺依曼熵，需要完成两个关键步骤。首先，必须从底层场论推导出 n阶 R´enyi 熵的解析形式；然后，我们需要将其向 $n \to 1$进行解析延拓。这两个步骤是相互独立的问题，通常需要采用不同的技术。下面我们将简要讨论这两个步骤。

计算 $\operatorname{Tr} \rho_A^n$ 并不容易；因此，复制方法应运而生。复制方法的早期形式在 [34], 中得到发展，后来被用于计算共形场论$_2$ [4–7], 中的各种例子，并可与全息例子 [42] 进行比较。复制方法背后的思想是考虑场论的 n份复制品的轨形，从而计算正整数 n 下的 $\operatorname{Tr} \rho_A^n$ 。该计算归结为在 n层黎曼曲面上求配分函数，也可以通过 n 份复制品中的扭结算符的关联函数来计算。有关共形场论中构造的更多细节，请参见 [4–7]。如果我们能够对任意正整数 $n \geq 1$计算 $\operatorname{Tr} \rho_A^n$ ，那么我们就有

$$S(\rho_A) = \lim_{n \to 1} S_n(\rho_A) = -\lim_{n \to 1} \frac{\partial}{\partial n} \operatorname{Tr} \rho_A^n. \tag{2.3}$$

对于某些特殊状态和区域，例如 CFT$_d$真空态中的球形区域，这一量可以计算。然而，在 CFT$_2$, 中，由于其无限维对称性足以确定低阶关联函数，我们能够在若干情况下计算 $\operatorname{Tr} \rho_A^n$。

[1]而在 CFT 中$_2$, ，真空态下无限直线上长度为 ℓ 的单区间 A 的主要发散项是长度的对数函数，这是我们稍后将讨论的最简单示例。

对 $n \to 1$ 进行解析延拓则更为微妙。要确保存在唯一的解析延拓，使其从整数 n 取值延伸出去，通常需要应用 Carlson 定理。该定理保证从 R´enyi 熵延拓到冯·诺依曼熵的解析延拓具有唯一性，前提是我们能够找到某个局部全纯函数 $\mathcal{S}_\nu$，使其满足 $\nu \in \mathbb{C}$，且 $\mathcal{S}_n = S_n(\rho)$ 对所有整数 $n > 1$均成立，并在 $\nu \to \infty$中具有适当的渐近行为。于是我们得到唯一的 $S_\nu(\rho) \quad = \mathcal{S}_\nu$ [43, 44]。Carlson 定理不仅解决了解析延拓唯一性的问题，也解决了如何延拓至非整数 Rényi 熵取值的问题。´。

在弦理论和 AdS/CFT 的背景下，还有其他方法可以计算 $S(\rho_A)$；例如参见 [45–50]。本文将重点讨论一种在 [8] 中阐述的有效方法，该方法适用于数值计算。在 [8], 中，使用以下生成函数对 n 进行解析延拓，其中变量为 z

$$G(z;\rho_A) \equiv -\operatorname{Tr}\left(\rho_A \ln \frac{1-z\rho_A}{1-z}\right) = \sum_{n=1}^{\infty} \frac{z^k}{k}\Big(\operatorname{Tr}(\rho_A^{k+1}) - 1\Big). \tag{2.4}$$

这个显式的泰勒级数在单位圆盘内绝对收敛，其中 $|z| < 1$。通过将对数的支割线选在正实轴上，我们可以将该函数从单位圆盘解析延拓为$\mathbb{C} \setminus [1,\infty)$内的全纯函数。极限 $z \to -\infty$ 位于全纯区域内，恰好对应于我们得到冯·诺依曼熵的位置

$$S(\rho_A) = \lim_{z\to-\infty} G(z;\rho_A). \tag{2.5}$$

然而，通过对变量进行 Möbius 变换，引入新变量 w，可以得到一种更实用的形式

$$G(w;\rho_A) = -\operatorname{Tr}\Big(\rho_A \ln\{1 - w(1-\rho_A)\}\Big), \quad w = \frac{z}{z-1}. \tag{2.6}$$

它再次呈现为泰勒级数

$$G(w;\rho_A) = \sum_{k=1}^{\infty} \frac{\tilde{f}(k)}{k} w^k, \tag{2.7}$$

其中

$$\tilde{f}(k) = \operatorname{Tr}[\rho_A(1-\rho_A)^k] = \sum_{m=0}^{k} \frac{(-1)^m k!}{m!(k-m)!} \operatorname{Tr}\left(\rho_A^{m+1}\right). \tag{2.8}$$

我们再次得到一个以 $\mathrm{Tr}\rho_A^n$ 表示的级数，它在单位圆盘 $|w| < 1$ 内绝对收敛。使用 w 的便利之处在于，取 $w \to 1$ 后，我们便得到冯·诺依曼熵。

$$S(\rho_A) = \lim_{w\to 1} G(w;\rho_A) = \sum_{k=1}^{\infty} \frac{\tilde{f}(k)}{k}. \tag{2.9}$$

从已知的 $\mathrm{Tr}\rho_A^n$ 表达式出发，可以得到 $S(\rho_A)$ 的精确表达式。通过计算 k 中的部分和，我们可以数值求得 $S(\rho_A)$ 的准确值。该方法保证，随着求和至足够大的 k，所得结果会以越来越高的准确率逼近冯·诺依曼熵。

然而，难点在于，为了达到 $k \sim 10^3$ 精度，我们通常需要对 10^{-3} 求和至 [8]。对于 $\mathrm{Tr}\,\rho_A^n$较为复杂的某些情形，计算成本会很高。因此，神经网络框架的一项优势在于，只需有限的数据便能给出准确的预测，使其成为一种效率更高的方法。

本文聚焦于 CFT_2 中的一些实例，这些实例的 $\mathrm{Tr}\,\rho_A^n$ [6] 具有已知的解析表达式。我们使用生成函数 $G(w;\,\rho_A)$ 为神经网络生成所需的训练数据集。

3 深度学习冯·诺依曼熵

本节旨在利用深度神经网络，通过监督学习方法预测冯·诺依曼熵。借助神经网络基于梯度的学习原理，我们期望找到输入数据与输出目标值之间的非线性映射。在从 n 阶 R´enyi 熵到冯·诺依曼熵的解析延拓问题中，这种非线性映射自然存在。因此，我们将 $S_n(\rho_A)$（等价地，$\mathrm{Tr}\,\rho_A^n$ 和生成函数）视为输入数据，并将$S(\rho_A)$视为训练过程的目标函数。作为监督学习，我们将考虑两侧解析表达式均已知的示例。最终，我们将利用训练好的模型，在各种物理参数区间内预测冯·诺依曼熵，以展示该方法的有效性和稳健性。

使用深度神经网络的主要优势在于，它们提高了用于计算冯·诺依曼熵的生成函数的准确率。正如我们所提到的，这种方法的准确率取决于部分和截断的位置，而且通常需要对大量 k 项求和（2.9），这在数值计算上颇为困难。从某种意义上说，这需要了解更多信息，例如级数中由 $\mathrm{Tr}´$ 所表示的高阶 Rényi 熵。训练好的神经网络能够预测 ρ_A^n

在输入数据项数少得多的情况下，更准确地预测冯·诺依曼熵。我们甚至可以预测其他参数空间下的冯·诺依曼熵，而无需借助生成函数的任何数据。

此外，深度神经网络所揭示的非线性映射可用于研究神经网络在冯·诺依曼熵方面的表达能力。此外，它们还可用于研究解析延拓未知的情形，以及其他需要解析延拓的纠缠度量。

在接下来的小节中，我们将详细介绍数据准备和训练策略，然后通过具体示例进行说明。

3.1 模型架构与训练策略

生成合适的训练数据集并设计灵活的深度学习模型是由经验驱动的。在本小节中，我们将概述这两个方面所采用的策略。

数据准备

为准备训练数据集，我们考虑若干个已知 $S(\rho_A)$ 的示例。对于每个示例，我们使用生成函数 $G(w;\rho)$，该函数可由 Tr ρ_A^n 计算得到。这等价于在选择不同物理参数的情况下计算更高阶 R´enyi 熵，因为可用的“信息”始终是 Tr ρ_A^n。然而，需要注意的是，所有更高阶 R´enyi 熵都是不同的信息。因此，相较于直接使用 $S_n(\rho_A)$，采用生成函数更为合适，因为随着精度提高，它会趋近于冯·诺依曼熵，从而使比较更加清晰。

我们生成 $N = 10000$ 一系列物理参数取值范围固定的输入数据集，每组包含 $k_{\max} = 50$ 项（2.9）；对应的冯·诺伊曼熵将作为目标值。我们限制数据量，以模拟使用生成函数所需的计算成本。我们随机打乱输入数据集，然后将数据划分为 80% 用于训练、10% 用于验证集，另外 10% 用作测试数据集。此外，我们使用训练好的神经网络，对另一组包含 10000 个测试数据集、且物理参数范围不同的数据进行预测，并将预测结果与正确值进行比较，作为针对每个示例的一项具有实质意义的测试。

模型设计

为防止过拟合并增强模型的泛化能力，我们在神经网络设计中采用了多种技术。整个部分均使用 ReLU激活函数。训练过程中，我们采用 Adam优化器 [51]，并以均方误差（MSE）作为损失函数。

我们考虑一个由若干隐藏Dense层组成的神经网络，这些层的单元数各不相同，并使用 TensorFlow-Keras [52, 53]。在这种情况下，层中的每个神经元

都会接收来自前一层所有神经元的输入。Dense连接使模型能够找到输入与输出之间的非线性关系，这正是解析延拓所涉及的情况。最后一层是仅包含一个单元的Dense层，它会为每个训练数据集输出一个唯一值，该值预期对应于冯·诺依曼熵。作为示例，我们在图1中展示了一个包含3个隐藏Dense层的神经网络，每层有8个单元。

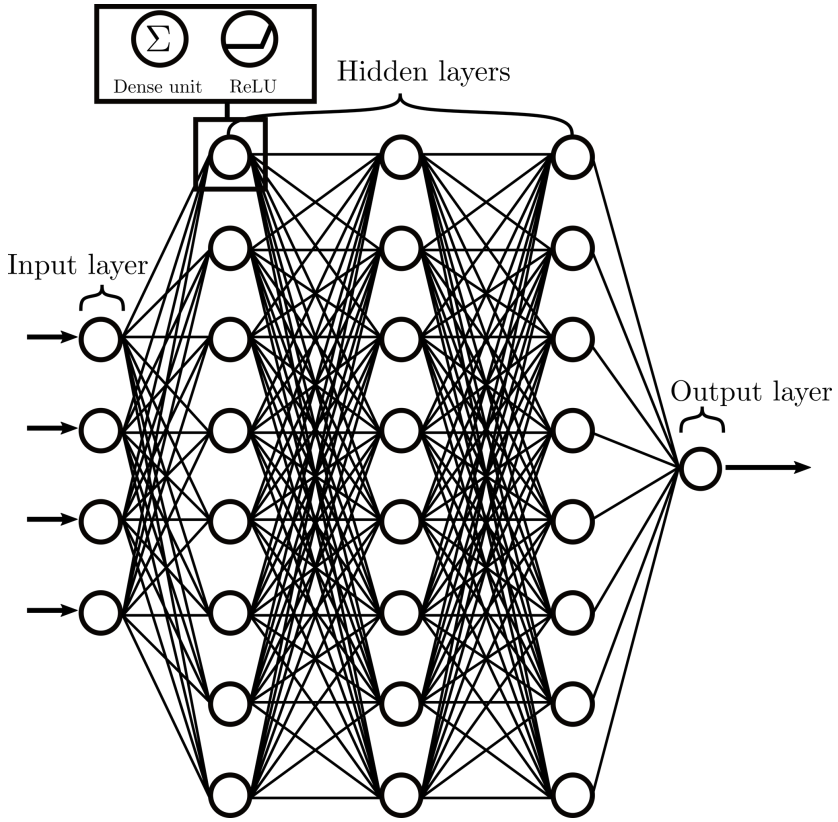

图 1：由 3 个稠密连接层组成的架构，每层有 8 个单元。最终输出层是一个单独的 Dense 单元，其输出唯一对应于冯·诺依曼熵。

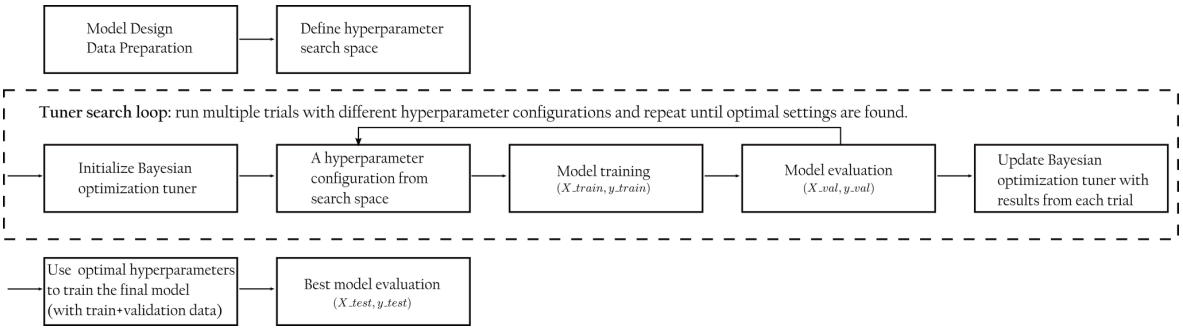

图 2：展示 KerasTuner 贝叶斯优化步骤的流程图。贝叶斯优化是一种针对给定数据集，通过从目标函数的先验分布迭代构建概率模型，并利用该模型指导搜索，以寻找最佳设计方案和超参数组合的方法。调参器搜索循环完成后，我们在最终训练阶段同时纳入训练数据和验证数据，提取最佳模型。

为了确定神经网络的最佳设置，我们采用 KerasTuner[16], 这一强大的工具，可以针对给定任务探索不同的模型复杂度、深度和超参数组合。KerasTuner 的流程示意图见图 2。我们采用贝叶斯优化，并调整以下设计方案和超参数：

- 我们最多使用 4 个 Dense层。对于每一层，允许的单元数范围为 16 至 128，步长为 16。每一层的单元数彼此独立。

- 我们将是否在稠密层之后添加 BatchNormalization 层作为一个布尔选项，以提升泛化能力并发挥正则化作用。

- 我们还将是否添加最后一个 Dropout 层作为一个布尔选项，其 Dropout 率通过对数采样从 0.1 到 0.5 的范围内选取。

- 对于 Adam优化器，我们仅通过对数采样调整学习率，取值范围为 3×10^{-3} 至 9×10^{-3}。其他所有参数均采用 TensorFlow-Keras 中的默认值。我们还将是否使用该算法的 AMSGrad [54] 变体作为一个布尔选项。

我们使用 KerasTuner 进行 100 次试验，每次试验执行 2 次，并通过耐心值为 8 的 EarlyStopping 监控验证损失。训练完成后，由于我们不会再进行任何超参数调整，因此不再评估

模型在验证数据上的表现。一种常见做法是使用 KerasTuner 找到的最佳模型设计来初始化新模型，同时将验证数据也纳入训练数据。事实上，我们选出排名前 5 的最佳设计，并分别使用耐心值为 8 的 EarlyStopping 训练 20 次。我们从 5×20 模型中选取相对目标值误差最小的模型作为最终模型。我们将 KerasTuner 和最终训练的批大小都设为 512。

在接下来的两个小节中，我们将考察来自 CFT 的实例$_2$ 及其 $\mathrm{Tr}\,\rho_A^n$以及相应的冯·诺伊曼熵 $S(\rho_A)$ [4–8]。这些实例各不相同，出于多种原因值得研究。它们具有不同的数学结构，并且尽管都涉及某些配分函数的计算，但从场论角度来看，它们的推导过程缺乏共同的模式。此外，每种情形下的解析延拓都十分复杂，这有力地表明有必要针对不同情形独立设计模型。

3.2 单区间的纠缠熵

在下文中，我们只给出 $\mathrm{Tr}\,\rho_A^n$ 的解析表达式，因为它是生成函数的唯一输入。我们也会在公式中明确保留 UV 截断 ϵ 。

单区间

最简单的例子对应于无限直线上的 CFT_2 真空态中的一个长度为 ℓ 的单区间 A 。在这种情况下，$\mathrm{Tr}\,\rho_A^n$ 和 $S(\rho_A)$ 的解析形式都是已知的 [4], ，其中 $S(\rho_A)$ 简化为一个依赖于 ℓ 的简单对数函数。我们得到如下中心荷为 c 时的解析形式。

$$\mathrm{Tr}\,\rho_A^n = \left(\frac{\ell}{\epsilon}\right)^{\frac{c}{6}(\frac{1}{n}-n)}, \tag{3.1}$$

定义了 $G(w;\rho_A)$。相应的冯·诺依曼熵为

$$S(\rho_A) = \frac{c}{3}\ln\frac{\ell}{\epsilon}. \tag{3.2}$$

我们在准备数据集时固定了中心荷 $c = 1$ 和紫外截断 $\epsilon = 0.1$ 。我们从 $\ell = 1$ 到 50 生成了 10000 组数据，用于训练-验证-测试集划分，每一步之间的增量为 $\Delta\ell = 5 \times 10^{-3}$，直至 $k = 50$，取值于 $G(w;\ \rho_A)$。为进一步验证我们的模型，我们针对以下物理参数额外生成了 10000 个测试数据集： $\ell = 51$ 到 100，步长为 $\Delta\ell = 5 \times 10^{-3}$。关于数据分布相对于目标冯·诺伊曼熵的密度图，请参见图 3。

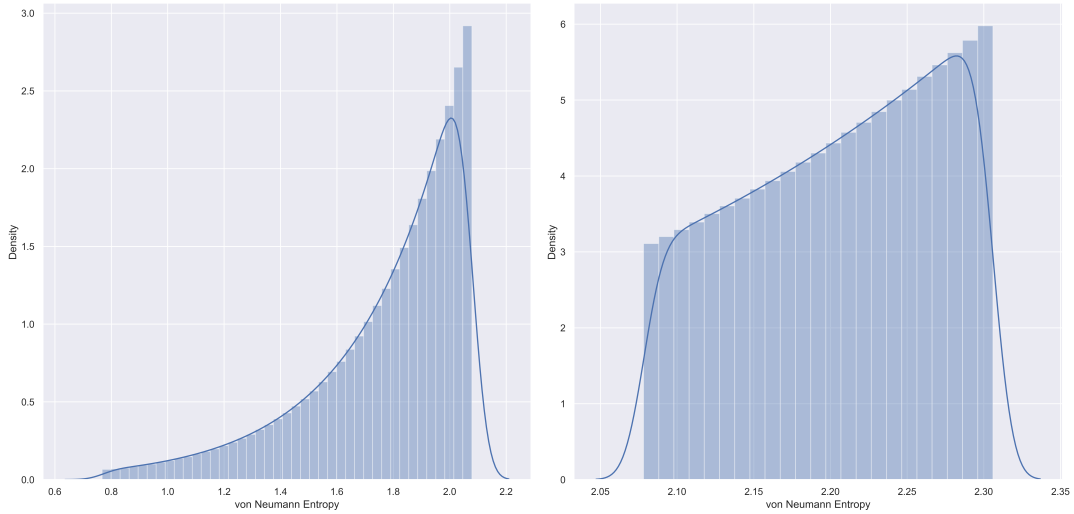

图 3：单区间情况下的数据分布，其中我们绘制了密度随由（3.2）计算的冯·诺依曼熵变化的关系，且 ℓ 取不同值。左图表示训练-验证-测试集划分中的 10000 个数据集，右图则对应处于不同物理参数范围的另外 10000 个测试数据集。

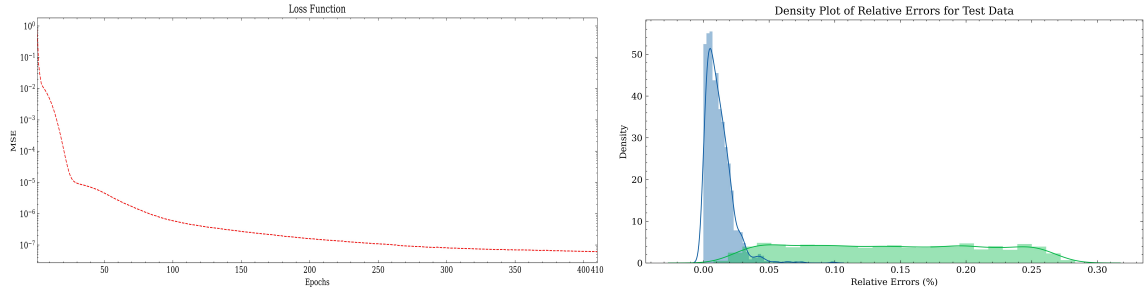

图 4：左：均方误差损失函数随训练轮次的变化。我们使用 EarlyStopping 监控损失函数；对于此实例，在训练轮次 410 时达到最小损失，损失为 $\approx 10^{-7}$ 。右：模型预测值与目标值之间相对误差的密度图。请注意，蓝色对应初始训练-验证-测试集划分中的测试数据集，绿色对应额外的测试数据集。可以清楚地看到，对于这两个数据集，我们都取得了较高的准确率，相对误差为 $\lesssim 0.30\%$。

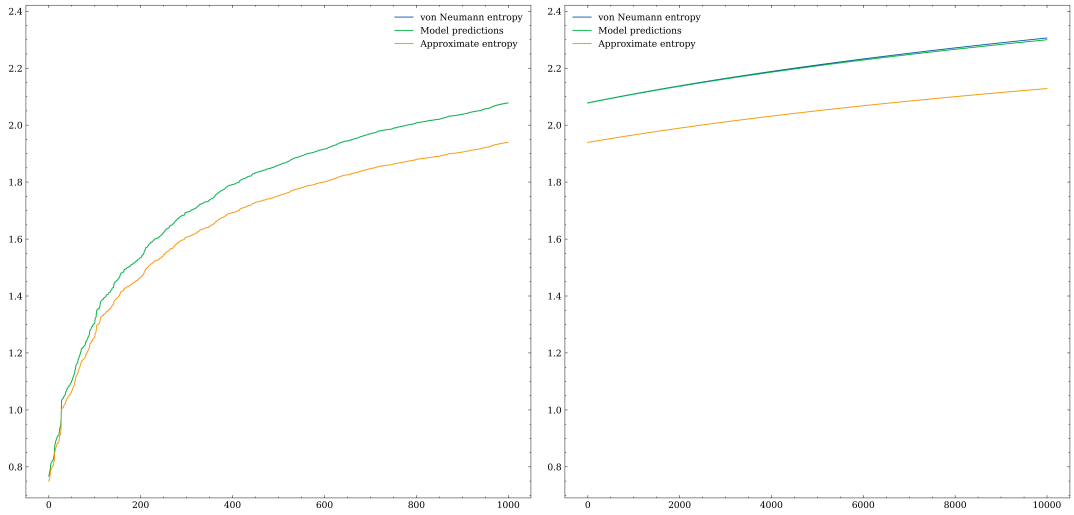

图 5：我们绘出了模型的预测结果，以及通过（3.2）计算出的解析冯·诺依曼熵，分别对应训练-验证-测试划分中的 1000 个测试数据集（左）和额外的 10000 个测试数据集（右）；两幅图使用相同的刻度。正确的冯·诺依曼熵与模型的预测结果精确重合。我们还通过对生成函数中的 $k = 50$ 项求和，纳入了近似熵。

图 4表明，上一小节概述的方法能够有效地将测试数据预测的相对误差降至极低水平。此外，模型在预测额外测试数据集时也能达到同样低的相对误差，进一步证实了其有效性。从图 5可以看出，模型对两个测试数据集的预测准确率远高于通过对生成函数的前 50 项求和得到的近似熵。我们强调，要使生成函数达到与深度神经网络相同的准确率，通常需要对（ $k \geq 400$ （2.9） [8]中的项求和。以下所有示例均适用这一点。

在这个例子中，冯·诺依曼熵是一个简单的对数函数，因此深度学习模型相对容易对其进行解析。不过，我们现在将转向一个更具挑战性的例子。

有限温度和有限长度下的单区间

我们将单区间情形推广到有限温度和有限长度，此时 $\mathrm{Tr}\,\rho_A^n$会成为逆温度 $\beta = T^{-1}$ 和长度ℓ的复杂函数。对于圆上的二维自由狄拉克费米子，R´enyi 熵的解析表达式最初由 [55] 通过玻色化推导得出。我们可以施加周期

边界条件，对应有限尺寸和有限温度。为简便起见，我们令总空间尺寸 L 为 1，并用 ℓ 表示区间长度。在这种情况下，我们有 [55]

$$\mathrm{Tr}\,\rho_A^n = \prod_{k=-\frac{n-1}{2}}^{\frac{n-1}{2}} \left|\frac{2\pi\epsilon\eta(\tau)^3}{\theta_1(\ell|\tau)}\right|^{\frac{2k^2}{n^2}} \frac{|\theta_\nu(\frac{k\ell}{n}|\tau)|^2}{|\theta_\nu(0|\tau)|^2}, \tag{3.3}$$

其中 ϵ 是紫外截断。我们研究 $\nu = 3$的情况，即 Neveu-Schwarz (NS-NS) 扇区。此时有如下 Dedekind η 函数 $\eta(\tau)$，以及 Jacobi theta 函数 $\theta_1(z|\tau)$ 和 $\theta_3(z|\tau)$

$$\eta(\tau) \equiv q^{\frac{1}{24}} \prod_{n=1}^{\infty}(1-q^n), \tag{3.4}$$

$$\theta_1(z|\tau) \equiv \sum_{n=-\infty}^{n=\infty} (-1)^{n-\frac{1}{2}} e^{(n+\frac{1}{2})^2 i\pi\tau} e^{(2n+1)\pi i z}, \qquad \theta_3(z|\tau) \equiv \sum_{n=-\infty}^{n=\infty} e^{n^2 i\pi\tau} e^{2n\pi i z}. \tag{3.5}$$

此前，经过解析延拓（3.3）后的冯·诺依曼熵仅在高温和低温情形下已知 [55]。事实上，只有无限长度或零温度下的部分才具有普适性。然而，最近 [56, 57], 求出了适用于所有温度的解析冯·诺依曼熵，现列于下文

$$S(\rho_A) = \frac{1}{3}\log\frac{\sigma(\ell)}{\epsilon} + 4i\ell\int_0^\infty dq \frac{\zeta(iq\ell + 1/2 + i\beta/2) - \zeta(1/2) - \zeta(i\beta/2)}{e^{2\pi q} - 1}. \tag{3.6}$$

这里，σ 和 ζ 分别是周期为 1 和$i\beta$的 Weierstrass σ 函数和 ζ 函数。我们可以清楚地看到，与前一个例子相比，$\mathrm{Tr}\,\rho_A^n$ 和$S(\rho_A)$ 的解析表达式大不相同。

在准备数据集时，我们固定区间长度 $\ell = 0.5$ 和紫外截断$\epsilon = 0.1$。我们从 $\beta = 0.5$ 到 1.0 生成了 10000 组数据，用于训练-验证-测试集划分，相邻两步的增量为 $\Delta\beta = 5\times10^{-5}$，直至 $k = 50$，取值为 $G(w;\ \rho_A)$。由于 β 对应逆温度，因此这是一个自然的变化参数，因为公式（3.6）对所有温度均成立。为进一步验证模型，我们针对以下物理参数额外生成了 10000 个测试数据集：$\beta = 1.0$到 1.5，其中 $\Delta\beta = 5\times10^{-5}$。关于冯·诺依曼熵的数据密度图见图6。如图7 和图8所示，我们的模型在预测这两个测试数据集时表现出良好效果；对于这个高度非平凡的例子，模型给出了准确结果。

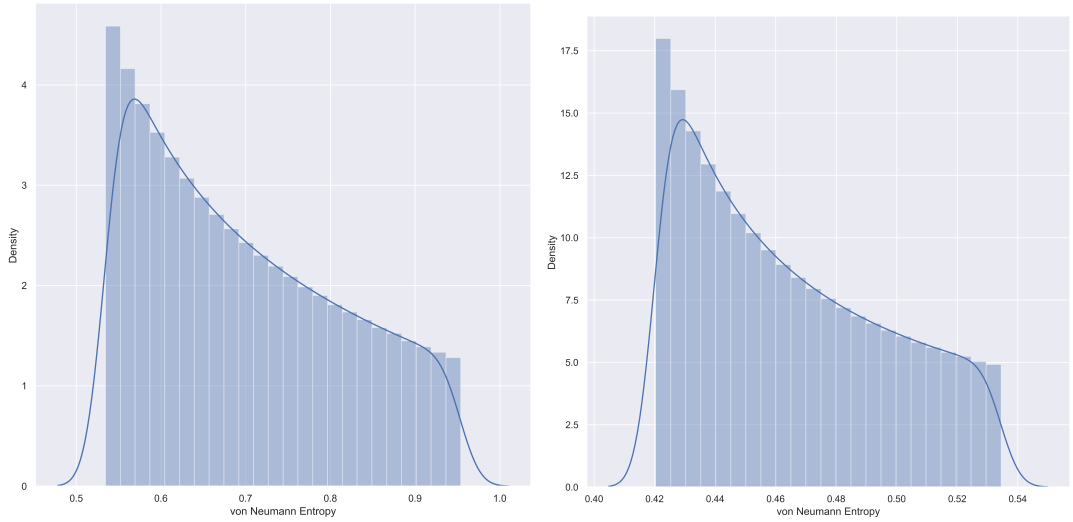

图 6：单区间处于有限温度且长度固定情况下两个测试数据集的分布，其中我们绘制了密度随冯·诺依曼熵的变化关系，该熵由（3.6）计算得出，且 β 取不同值。

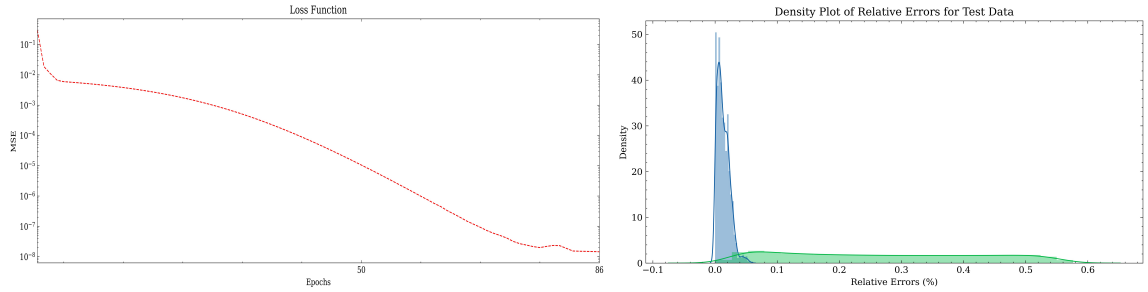

图 7：左：均方误差损失函数随训练轮次的变化关系。在此实例中，第 86 个训练轮次时的最小损失接近 10^{-8} 。右：两个测试数据集的模型预测值与目标值之间的相对误差；相对误差为 $\lesssim 0.6\%$，准确率很高。

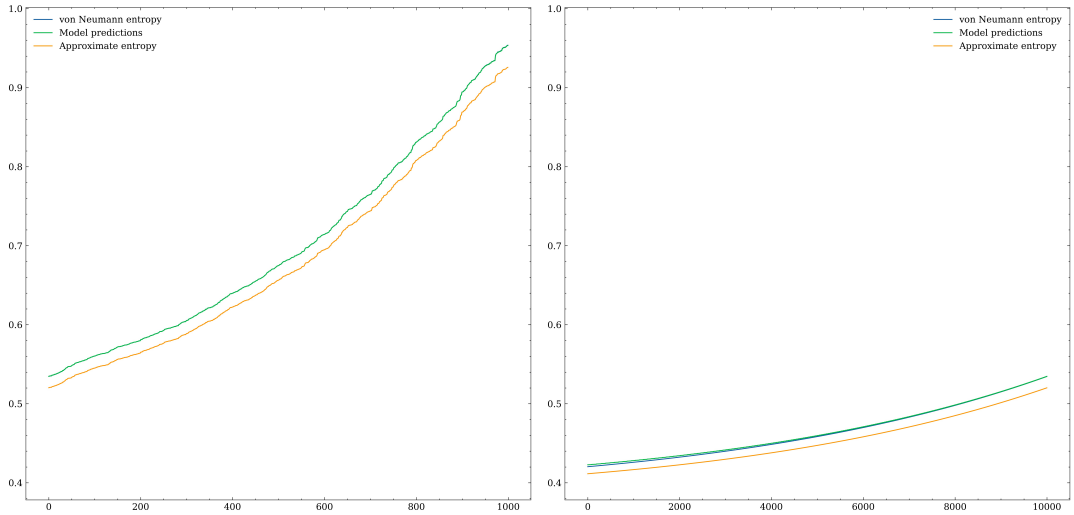

图 8：我们将模型的预测与通过（3.6）计算得到的解析冯·诺依曼熵进行比较，结果针对两个测试数据集。与前面一样，也包含了对生成函数中的 $k = 50$ 项求和得到的近似熵。

3.3 两个不相交区间的纠缠熵

现在我们转而讨论无限直线上两个区间并集的冯·诺依曼熵。在这种情况下，可以推导出 Rényi 熵和冯·诺依曼熵的若干解析表达式。我们将考虑的理论是中心荷为 $c = 1$ 的自由玻色子 CFT_2，其冯·诺依曼熵由两个参数表征：交比 x 和普适临界指数 η。后者与紧化半径的平方成正比。

为建立系统，我们将两个区间的并集定义为 $A \cup B$ with$A = [x_1, x_2]$ 和 $B = [x_3, x_4]$。交比定义为

$$x = \frac{x_{12}x_{34}}{x_{13}x_{24}}, \quad x_{ij} = x_i - x_j. \tag{3.7}$$

根据这个定义，我们可以写出具有有限 x 和 η [5] 的自由玻色子共形场论中两个区间的生成函数

$$\mathrm{Tr}(\rho^n) = c_n \left(\frac{\epsilon^2 x_{13}x_{24}}{x_{12}x_{34}x_{14}x_{23}} \right)^{\frac{1}{6}(n-\frac{1}{n})} \mathcal{F}_n(x, \eta), \tag{3.8}$$

其中 ϵ 是紫外截断，c_n 是依赖于模型的系数 [6]，为简洁起见，我们将其设为 $c_n = 1$。$\mathcal{F}_n(x, \eta)$ 的精确表达式为

$$\mathcal{F}_n(x, \eta) = \frac{\Theta(0|\eta\Gamma)\Theta(0|\Gamma/\eta)}{[\Theta(0|\Gamma)]^2}, \tag{3.9}$$

对于整数 $n \geq 1$。这里的 $\Theta(z|\Gamma)$ 是定义为以下形式的黎曼-西格尔θ函数。

$$\Theta(z|\Gamma) \equiv \sum_{m\in\mathbb{Z}^{n-1}} \exp[i\pi m^t \cdot \Gamma \cdot m + 2\pi i m^t \cdot z], \tag{3.10}$$

其中，Γ 是一个 $(n-1) \times (n-1)$ 矩阵，其元素为

$$\Gamma_{rs} = \frac{2i}{n}\sum_{k=1}^{n-1} \sin\left(\frac{\pi k}{n}\right)\beta_{k/n}\cos\left[\frac{2\pi k}{n}(r-s)\right], \tag{3.11}$$

and

$$\beta_y = \frac{F_y(1-x)}{F_y(x)}, \qquad F_y(x) \equiv {}_2F_1(y, 1-y; 1; x), \tag{3.12}$$

其中，${}_2F_1$ 是超几何函数。该例的一个性质是，(3.9) 在 $\eta \leftrightarrow 1/\eta$ 下显然保持不变。

通向冯·诺伊曼熵的解析延拓尚未得到，因此无法直接使用监督学习研究这一例子。尽管生成函数的泰勒级数保证了当部分和中的 k 足够大时向真实冯·诺伊曼熵收敛，但计算更高维的黎曼-西格尔θ函数会变得越来越困难。关于这方面的研究，请参见 [58, 59]。不过，在下一节讨论序列模型时，我们将重新考察这一例子。

不过，有两个极限情形可以进行解析微扰展开，并可得到冯·诺伊曼熵的近似解析延拓。第一个极限对应于交比 x 取较小值的情形，此时冯·诺伊曼熵已被解析地计算到 x 的二阶。第二个极限是去紧化极限，即取 $\eta \to \infty$。在此极限下，冯·诺伊曼熵存在一个近似表达式。

小交比下的两个区间

让我们考虑 $\mathcal{F}_n(x, \eta)$ 在某个 $\eta \neq 1$ 下当 x 较小时的如下展开式。

$$\mathcal{F}_n(x,\eta) = 1 + \left(\frac{x}{4n^2}\right)^{\alpha} s_2(n) + \left(\frac{x}{4n^2}\right)^{2\alpha} s_4(n) + \cdots, \tag{3.13}$$

其中我们可以考察一阶贡献，其形式为

$$s_2(n) \equiv \mathcal{N}\frac{n}{2}\sum_{j=1}^{n-1}\frac{1}{[\sin(\pi j/n)]^{2\alpha}}. \tag{3.14}$$

自由玻色子的系数 α 由 $\alpha = \min[\eta, 1/\eta]$ 给出。 $\mathcal{N}$ 是最低维算符的重数，对于自由玻色子，我们有 $\mathcal{N} = 2$。在此阶数以内，解析冯·诺依曼熵由以下公式给出。

$$S(\rho_{AB}) = \frac{1}{3}\ln\left(\frac{x_{12}x_{34}x_{14}x_{23}}{\epsilon^2 x_{13}x_{24}}\right) - \mathcal{N}\left(\frac{x}{4}\right)^{\alpha}\frac{\sqrt{\pi}\Gamma(\alpha+1)}{4\Gamma\left(\alpha+\frac{3}{2}\right)} - \cdots . \tag{3.15}$$

我们可以通过设定 $|x_{12}| = |x_{34}| = r$，并将 A 和 B 的中心间距设为 L，来进行数值计算；那么交比就很简单地表示为

$$x = \frac{x_{12}x_{34}}{x_{13}x_{24}} = \frac{r^2}{L^2}. \tag{3.16}$$

类似地，我们可以表示 $|x_{14}| = L + r = L(1+\sqrt{x})$ 和 $|x_{23}| = L - r = L(1-\sqrt{x})$。这样就可以将所有内容都用 x 和 L表示。

对于数据集，我们固定了 $L = 14$、 $\alpha = 0.5$ 和 $\epsilon^2 = 0.1$。我们从 $x = 0.05$ 到 0。1 生成了 10000 组数据，用于训练-验证-测试集划分，每一步的增量为 $\Delta x = 5\times10^{-6}$，直至 $k = 50$，取值范围为 $G(w;\ \rho_A)$ 。为了进一步验证模型，我们针对以下物理参数额外生成了 10000 个测试数据集： $x = 0.1$ 到 0。15，步长为 $\Delta x = 5 \times 10^{-6}$。图 9 展示了相对于冯·诺依曼熵的数据密度图。学习结果的清晰展示请参见图 10 和图 11 。

利用生成函数法对 x 进行二阶研究的内容见 [8], ，也可通过全息方法获得 [60]。此外，针对一般 CFT_2 ，关于 x 的冯·诺依曼熵二阶解析延拓可见 [61]。尽管这是一项次领头修正，但也可以用我们的方法进行研究。

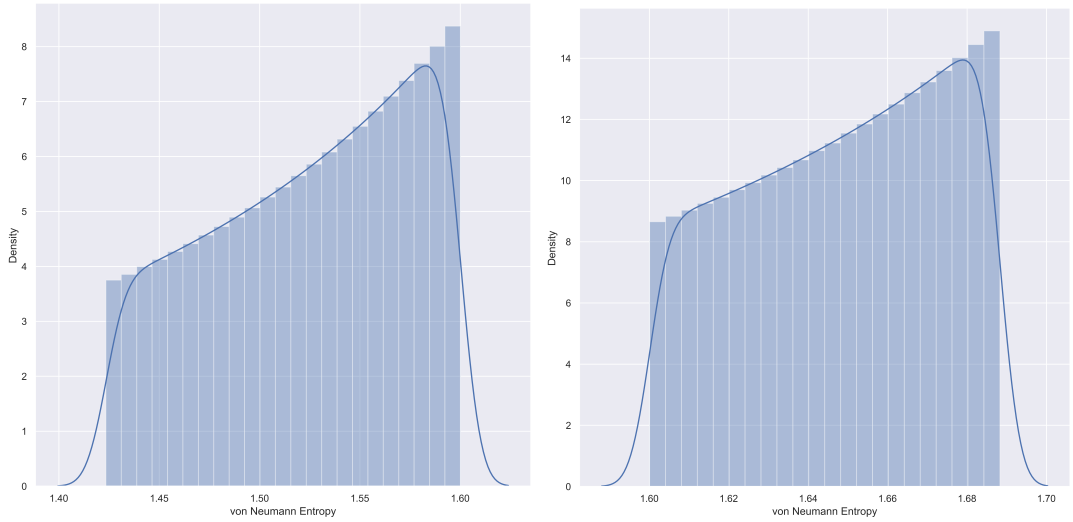

图 9：交比很小时两个区间情况下，两个测试数据集的分布。图中绘出了密度随根据（3.15）计算的冯·诺依曼熵变化的函数，其中 x。

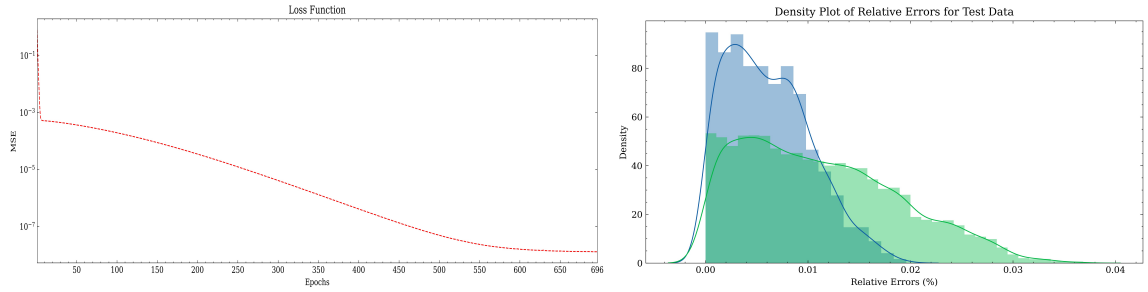

图 10：左：均方误差损失函数随训练轮次变化的曲线。对于此实例，在第 696 个训练轮次时，损失达到接近 10^{-8} 的最小值。右：两个测试数据集的模型预测值与目标值之间的相对误差；相对误差为 $\lesssim 0.03\%$，准确率很高。

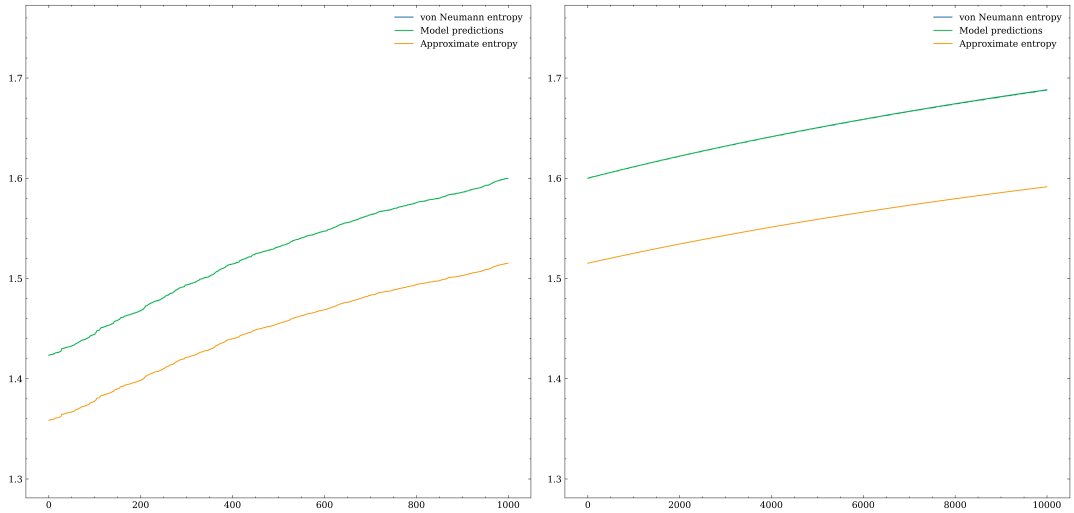

图 11：我们将模型的预测结果与通过（3.15）计算得到的解析冯·诺依曼熵进行比较，数据来自两个测试数据集。我们还通过对生成函数中的 $k = 50$ 项求和，纳入了近似熵。

去紧化极限下的两个区间

除了小交比极限，还可以考虑另一种极限，在这种极限下可以得到近似的解析 R´enyi 熵。这称为去紧化极限，此时我们取 $\eta \to \infty$，于是对每个固定的 x，有 $\mathcal{F}(x,\ \eta)$，其中

$$\mathcal{F}_n(x,\eta) = \left[\frac{\eta^{n-1}}{\prod_{k=1}^{n-1} F_{k/n}(x)F_{k/n}(1-x)}\right]^{\frac{1}{2}}, \tag{3.17}$$

其中 ${}_2F_1$ 是超几何函数。方程（3.17）在 $\eta \leftrightarrow 1/\eta$变换下保持不变，因此我们将改用 $\eta \ll 1$对应的结果

$$\mathcal{F}_n(x,\eta) = \left[\frac{\eta^{-(n-1)}}{\prod_{k=1}^{n-1} F_{k/n}(x)F_{k/n}(1-x)}\right]^{\frac{1}{2}}. \tag{3.18}$$

在这种情况下，冯·诺依曼熵的精确解析延拓尚不为人所知，不过可以通过展开得到一个近似结果

$$S(\rho_{AB}) \simeq S^W(\rho_{AB}) + \frac{1}{2}\ln\eta - \frac{D_1'(x) + D_1'(1-x)}{2} + \cdots, \quad (\eta \ll 1) \tag{3.19}$$

其中 $S^W(\rho_{AB})$ 是由不含特殊函数 ´ 的 Rényi 熵计算得出的冯·诺依曼熵 $\mathcal{F}_n(x,\eta)$，见 (3.8)。注意

$$D_1'(x) = -\int_{-i\infty}^{i\infty} \frac{dz}{i}\frac{\pi z}{\sin^2(\pi z)}\ln F_z(x). \tag{3.20}$$

这种近似冯·诺依曼熵已在先前的研究中经过充分检验 [5, 8]，我们将采用它作为深度学习模型的目标值。

对于这些数据集，我们固定 $L = 14$、 $x = 0.5$ 和 $\epsilon^2 = 0.1$。我们生成了 10000 组数据，用于训练-验证-测试集划分，参数从 $\eta = 0.1$ 到 0。2，每一步的增量为 $\Delta\eta = 10^{-5}$，直到 $k = 50$。为了进一步验证模型，我们针对以下物理参数生成了另外 10000 个测试数据集： $\eta = 0.2$到 0。3，其中$\Delta\eta = 10^{-5}$。数据相对于冯·诺依曼熵的密度图见图12。有关学习结果的清晰展示，请再次参阅图13和图14。

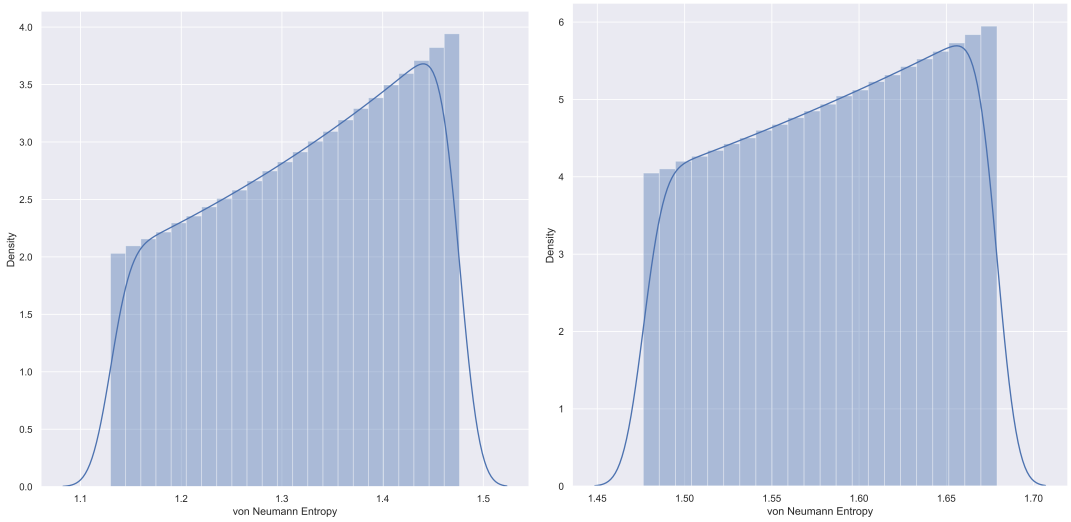

图 12：去紧化极限下两个区间情形的两个测试数据集分布图，其中密度是冯·诺依曼熵的函数，该熵由（3.19）计算，并令 η。

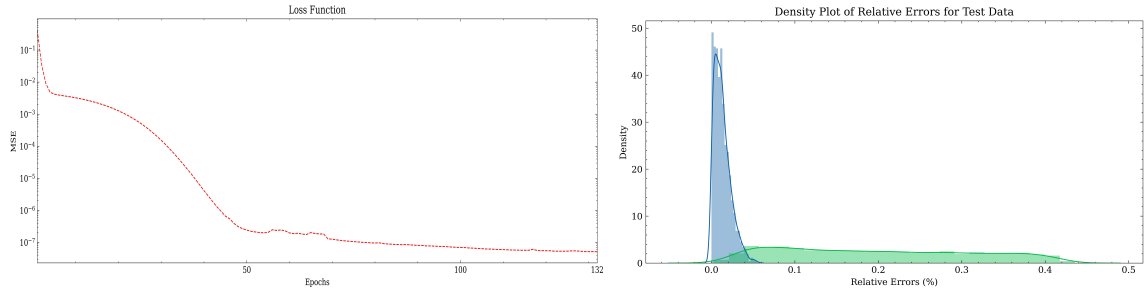

图 13：左：均方误差损失函数随训练轮次变化的曲线。对于此实例，约在 10^{-7} 时达到最小损失，对应的训练轮次为 132。右：两个测试数据集的模型预测值与目标值之间的相对误差；我们取得了较高的准确率，相对误差为 $\lesssim 0.4\%$。

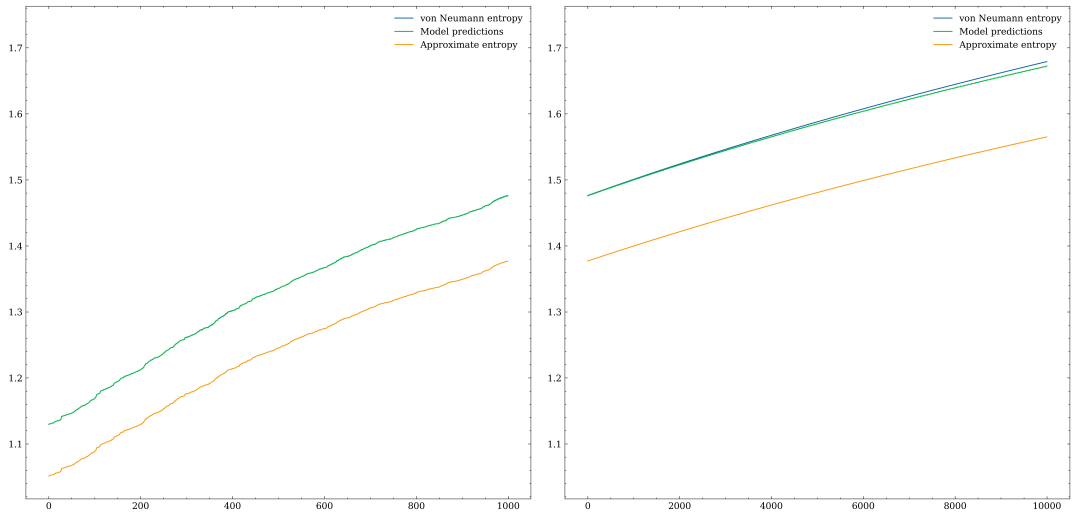

图 14：我们绘制了模型的预测结果，以及通过（3.19）计算得到的解析冯·诺依曼熵，所针对的是两个测试数据集。我们还通过对生成函数中的 $k = 50$ 项求和，纳入了近似熵。

我们已经看到，将深度神经网络用于监督学习时，即使在超出训练阶段所覆盖参数范围的区域内，也能准确预测冯·诺依曼熵。然而，深度神经网络的潜力或许不止于此。

众所周知，解析延拓必须针对具体情况逐一求解（参见 [4–7]中的示例），甚至可能取决于我们采用的方法 [8]。发现解析延拓中的一般规律仍是一个开放问题。尽管这仍是一项艰巨的任务，神经网络所揭示的非线性映射将

使我们能够研究深度神经网络在冯·诺依曼熵解析延拓问题上的表达能力。

我们的方法还开辟了使用深度神经网络研究解析延拓未知情形的可能性，例如一般双区间情形。此外，这或许还能使我们研究其他遵循类似模式或需要解析延拓的纠缠度量。我们将这些问题留待未来研究。

4 Rényi 熵作为序列深度学习

本节中，我们聚焦于利用序贯学习模型研究更高阶的黎曼熵′。研究依赖于 Tr′ 的更高阶黎曼熵，等价于研究生成函数（2.9）泰勒级数表示中的高阶项。这样做主要有几个动机。首先，尽管生成函数可用于计算高阶项，但对于更复杂的例子，效率会变低。此外，在一般双区间情形下计算（3.8）中的 Tr′ 涉及黎曼-西格尔θ函数，这给计算更高阶的黎曼熵 [8, 58, 59]带来了挑战。另一方面，所有更高阶的黎曼熵′都应被视为相互独立，无法通过线性方式获得。它们都可用于预测冯·诺依曼熵，但在泰勒级数展开式（2.9）中，已知更高阶的黎曼熵′等价于得到更精确的冯·诺依曼熵。由于我们无法简单地对级数进行外推，采用序贯学习方法是一种在统计上稳健的识别底层模式的方式。

循环神经网络 （RNNs）是一类功能强大的神经网络，凭借其“记忆”特性 [62], 适合处理序列。RNNs 通过内部循环遍历序列中的各个元素，同时维护一个状态，其中包含截至目前已观察到的信息。这一特性使 RNNs 无论模式在序列中的什么位置，都能识别出序列中的模式。训练 RNN 时，我们先初始化一个任意状态，并对大小为 (steps, input features) 的二阶张量进行编码，在多个时间步中循环处理。每一步中，网络都会在 k 处将当前状态与输入结合起来，得到 $k+1$处的输出，并将其作为下一次迭代的状态。

循环神经网络结合了前馈网络和随时间反向传播（BPTT） [63, 64], ，其中“时间”在我们的案例中表示步骤 k 。这类网络将全连接层的输出连接到同一层的输入，这些输入称为隐藏状态。输入接收上一步的输出值；神经元的输入数量由该层的输入数量和层内神经元数量共同决定，这种结构称为循环

连接。计算输出时，需要迭代地输入某一步的输入向量、计算隐藏状态，再提供下一步的输入向量以计算新的隐藏状态。

循环神经网络适用于根据序列数据进行预测，也就是“序列回归”，因为它们会从过去的步骤中学习模式，以预测下一步最可能出现的值。

4.1 模型架构与训练策略

在本小节中，我们将讨论如何把 R´enyi 熵（生成函数的泰勒级数）作为序列模型来处理。

数据准备

为了模拟无法高效计算序列中 $k_{\max}$ 的情形，我们针对不同的物理参数生成 $N = 10000$ 个数据集，每个数据集的序列最多包含 $k_{\max} = 50$ 步。我们还会打乱这些 N 个数据集，因为物理参数相近的样本，其大部分取值都相同。在这些 N 个数据集中，我们只取其中一部分 $p < N$ 用于训练-验证-测试集划分。其余部分 $q = N - p$ 将全部用作已训练模型的测试数据。这是对我们找到的序列模型的一项关键检验。理想情况下，我们只需使用少量 p 数据集，就能在 q 数据集上取得准确的表现。

由于可用的步数相当少，我们可以采用 TensorFlow-Keas 中的 SimpleRNN 结构[2]，而不是 LSTM 或 GRU 网络等更复杂的结构 [66, 67]。

我们还需要谨慎处理训练-验证-测试集划分过程。对于这类问题，验证集和测试数据应比训练数据更新，这一点非常重要。这是因为目标是根据过去的步骤预测下一个值，而数据划分应体现这一点。此外，通过增加近期数据的权重，可以缓解 BPTT 早期阶段可能出现的梯度消失（记忆丢失）问题。在本研究中，前 60% 的步骤（$k = 1 \sim 30$）用于训练，中间 20%（$k = 31 \sim 40$）用于验证集，最后 20%（$k = 41 \sim 50$）用于测试。

[2]SimpleRNN 在学习长期依赖关系时会遇到梯度消失问题 [65]。即使使用不会导致梯度消失的 ReLU，通过时间反向传播结合权重共享，仍可能导致不同步之间的梯度消失。然而，由于在我们的情况下可用的最大步数有限，序列长度较短，我们发现 SimpleRNN 通常比其变体表现更好。

我们按以下方式划分数据集：对于每一步中的单个数据集，我们使用由[3], 指定的固定数量的过去步骤 ℓ来预测下一个值。这将从每个数据集生成（步骤数 $-\ell$）个序列，总共生成（步骤数 $-\ell$）个 $\times p$序列，用于数据集的训练-验证-测试集划分。使用固定的序列长度 p 可以让网络专注于预测下一个值时最相关且最新的信息，同时简化输入大小，使其与我们的网络架构更加兼容。我们取 ℓ ， $p = 1000$、 $q = 9000$和 $\ell = 5$。图15展示了我们数据准备策略的示意图。

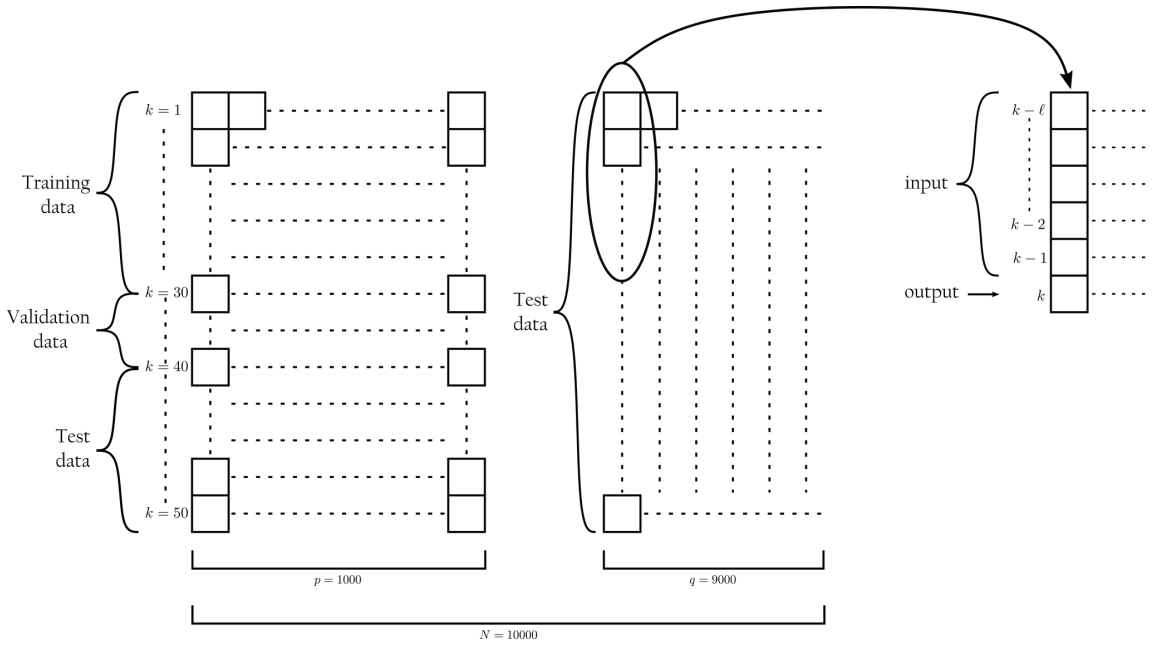

图 15：序贯模型的数据准备过程。总共将 N 个数据集分为两部分： p 个数据集用于初始的训练-验证-测试集划分，而 q 个数据集则仅作为测试数据集。右侧的放大图展示了如何生成单个示例序列，其中使用了固定数量的过去时间步 $\ell = 5$。请注意，对于额外的 q 个测试数据集，总共生成了 (steps $-\ell$) $\times q = 405000$ 个序列。

模型设计

完成数据预处理后，我们转向模型设计。在本节中，我们使用 ReLU激活函数和 Adam 优化器，并采用 MSE 作为损失函数。

[3]我们也可以尽可能多地纳入过去的时间步，但我们发现这样效果不佳。这可能是由于我们选择的网络架构，以及可用的最大时间步数相当有限。

在 KerasTuner 中，我们通过调整几个关键超参数和设计来进行贝叶斯优化。总结如下：

- 我们引入一层或两层 SimpleRNN，并选择是否使用循环丢弃。第一层的单元数从 64 到 256，步长为 16。如果使用第二层，单元数则从 32 到 128，步长为 8。循环丢弃的 Dropout 率设为 0.1 到 0.3，并采用对数采样。

- 我们将层归一化作为一个布尔选项，以提升训练稳定性，即便使用浅层网络也是如此。如果只有一层，则在 SimpleRNN 层之后添加层归一化；如果有两个 SimpleRNN 层，则在两层之间添加。

- 我们允许在循环层之后可选地添加一个 Dense层，其单元数范围为 16 至 32，步长为 8，作为回归器。

- 最后添加一个 Dropout，并将 Dropout 率在 0.2 到 0.5 的范围内进行对数采样，作为布尔选项。

- 在 Adam优化器中，我们仅通过对数采样来调整学习率，采样范围为 10^{-5} 至 10^{-4}。所有其他参数均采用 TensorFlow-Keras 中的默认值。我们将此算法的 AMSGrad [54] 变体作为一个布尔选项。

KerasTuner 运行 300 次试验，每次试验执行 2 次。在此过程中，我们使用耐心值为 8 的 EarlyStopping 监控验证损失。根据验证数据确定最佳超参数和模型架构后，我们使用相同的设计初始化一个新模型，并同时使用训练数据和验证数据。该新模型训练 30 次，同时使用耐心值为 10 的 EarlyStopping 监控训练损失。最终预测结果通过对若干个相对误差较小且彼此接近的案例取平均值获得。取平均值而非选择损失最小的案例，旨在平滑可能出现的离群值。我们将 KerasTuner 和最终训练的批大小均设为 2048。

我们还将使用训练好的模型对 q 测试数据进行预测，并将预测结果与正确值进行比较，作为达到基准的验证。

4.2 序贯模型示例

我们将通过两个示例来展示所提出的方法。第一个示例是单区间的一个简单代表性情形（3.1）；第二个示例则是去紧化极限下更具挑战性的双区间情形（3.19），在该情形中，生成函数中的高阶项无法高效计算。此外，我们还将简要评论一般双区间情形中最具非平凡性的示例。

单区间

在本例中，我们使用了与第 N 节中单区间相同的3.2数据集。按照我们刚才概述的数据拆分策略，值得注意的是，训练数据占整个数据集的比例相对较小。我们绘制了表现最佳的三个模型的损失，以及两个测试数据集相对误差的密度图，如图16所示。令人惊讶的是，即使训练数据的比例较小，我们仍然能够在额外的测试数据集上实现较小的相对误差。

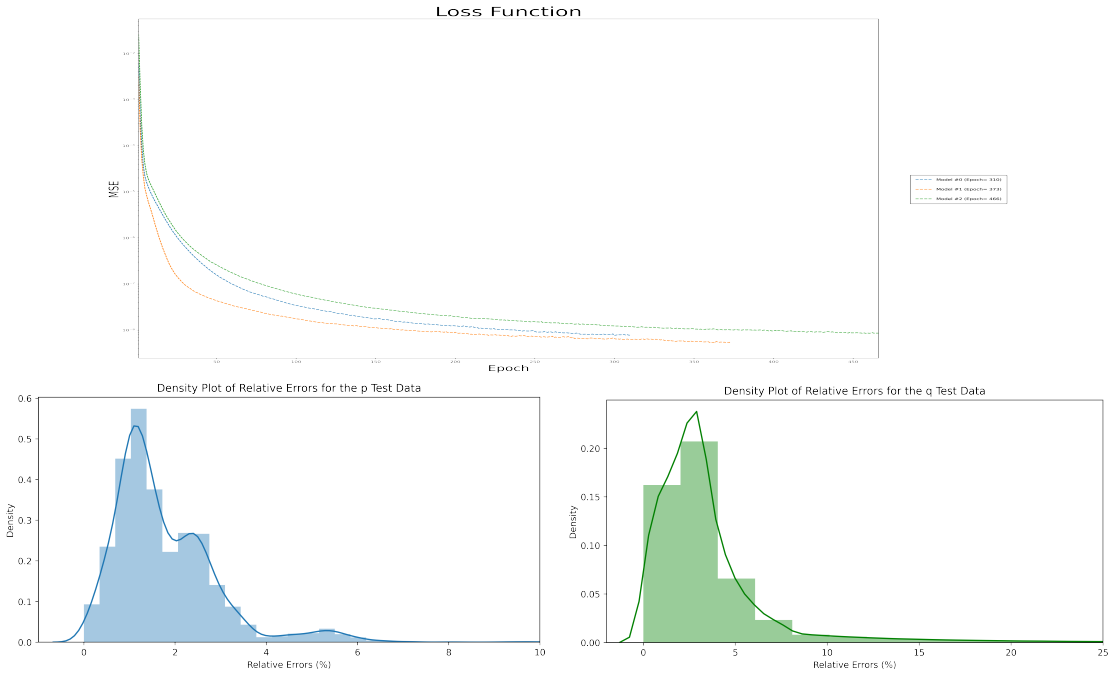

图 16：上图：最佳 3 个模型的损失函数随训练轮次的变化。我们使用 EarlyStopping 监控损失函数，图例括号中标明了不同模型的最小损失大约出现在第 10^{-8} 轮。下图：两个测试数据集的相对误差密度图。 p 测试数据集的相对误差集中在 1% 左右；而额外的 q测试数据集的相对误差集中在 2.5% 左右，离群值比例极低。

去紧化极限下的两个区间

同样，我们在 N 两个区间中使用了相同的 $\eta \to \infty$极限情况数据集，与第3.3节中的情况相同。在图17中，我们绘制了表现最佳的四个模型的损失曲线，以及两个测试数据集相对误差的密度图。在此示例中，KerasTuner 确定了一个相对较小的学习率，因此，由于我们已经达到了所需的准确率，便将训练截断在最多 1500 个训练轮次处。在这种情况下，预测具有较高的准确率，基本不存在离群值。

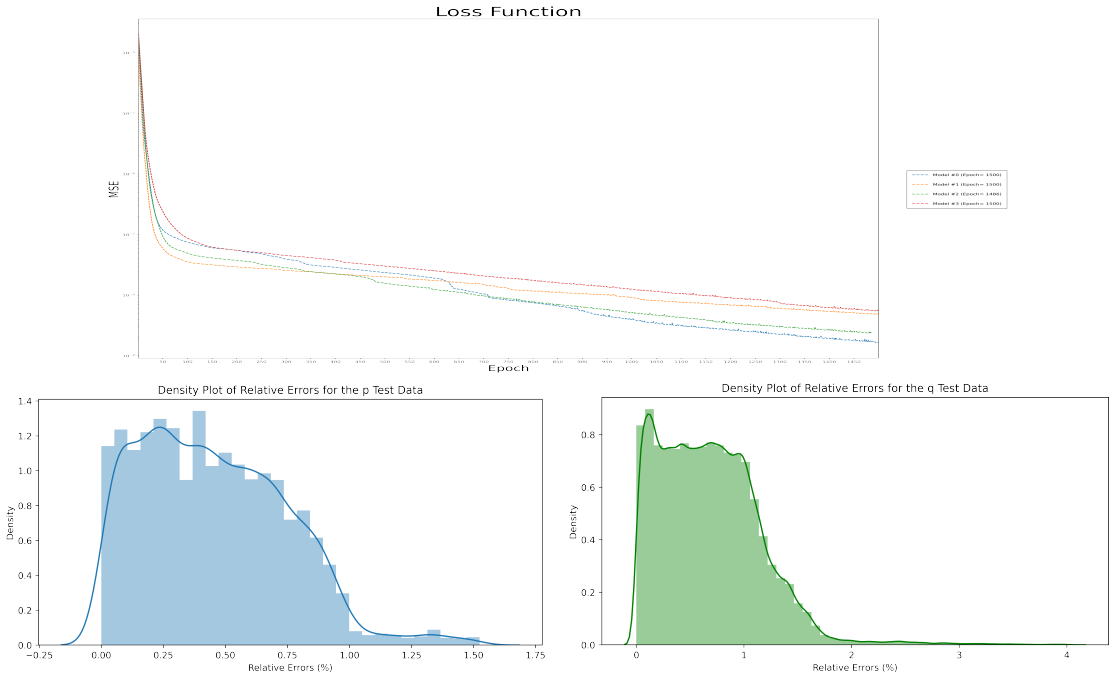

图 17：上图：表现最佳的 4 个模型的损失函数随训练轮次的变化。我们使用 EarlyStopping 监控损失函数。下图：两个测试数据集的相对误差密度图。 p 测试数据集的相对误差均在 $\lesssim 1.5\%$ 以内；而额外的 q 测试数据集的相对误差均在$\lesssim 2\%$ 以内。

我们简要讨论本文所述最具挑战性的例子，即一般双区间情形（3.8），在这种情形下，冯·诺依曼熵没有可用的解析表达式。这个例子中，我们只知道 $\mathrm{Tr}\rho_A^n$；由于它涉及黎曼-西格尔θ函数，在部分和中针对较大的 k 计算生成函数几乎不可行。因此，我们引入的序贯学习模型是在这种情形下提取有用信息的最可行方法。

在这种情形下，由生成函数只能高效计算 $k_{\max} \approx 10$ ，因此序贯学习模型的步长要短得多。我们使用 $N = 10000$ 数据集和 $k_{\max} = 10$测试了上述流程，但平均相对误差只能达到5%。采用更大的数据集并延长训练时间或许能带来改进，这将留待今后研究。

一般而言，序贯学习模型有望为高效计算生成函数中的高阶项提供解决方案。为了将我们的方法扩展到超出 $k_{\max}$ 步的更长序列，我们可以将该问题视为自监督学习。不过，这可能需要更精细的模型设计，以防止误差传播。尽管如此，探索更长的序列有助于更全面地理解冯·诺依曼熵的行为及其与 R´enyi 熵的关系。

5 量子神经网络与冯·诺依曼熵

在本节中，我们将量子电路视为将数据输入映射为预测的模型，并据此探讨一个类似的监督学习任务；这会影响量子电路作为函数近似器的表达能力。

5.1 F变分量子机器学习模型的傅里叶级数

我们将聚焦于量子神经网络能够显式实现的一类特定函数，即一种简单的傅里叶型求和 [29, 30]。在将其与冯·诺依曼熵联系起来之前，我们先概述 [30]领域中的开创性研究。

考虑以下形式的一般傅里叶型求和

$$f_{\theta_i}(\vec{x}) = \sum_{\vec{\omega} \in \Omega} c_{\vec{\omega}}(\theta_i) e^{i\vec{\omega}\cdot\vec{x}}, \tag{5.1}$$

其频谱由 $\Omega \subset \mathbb{R}^N$ 指定。注意， $c_{\vec{\omega}}(\theta_i)$ 是（复数）傅里叶系数。我们需要设计一个量子模型，使其能够通过对频谱和傅里叶系数的控制来学习该求和的特征。

现在，我们将量子机器学习模型定义为以下期望值

$$f_{\theta_i}(x) = \langle 0|U^\dagger(x,\theta_i) M U(x,\theta_i)|0\rangle, \tag{5.2}$$

其中， $|0\rangle$ 被视为量子计算机的某个初始态。 M 则表示物理可观测量。请注意，我们省略了向量符号以及算符上方的帽子标记，具体含义应可根据上下文明确判断。其核心组成部分是$U(x,\theta_i)$，这是一个依赖于数据输入 x 和可训练参数的量子电路。

参数 θ_i ，包含 L 层。每一层都包含一个数据编码电路块 $S(x)$ 和一个可训练电路块 $W(\theta_i)$。从示意图上看，其形式为

$$U(x,\theta_i) = W^{(L+1)}(\theta_i)S(x)W^{(L)}(\theta_i)\cdots W^{(2)}(\theta_i)S(x)W^{(1)}(\theta_i), \tag{5.3}$$

这里我们参考图18 以清晰地说明。

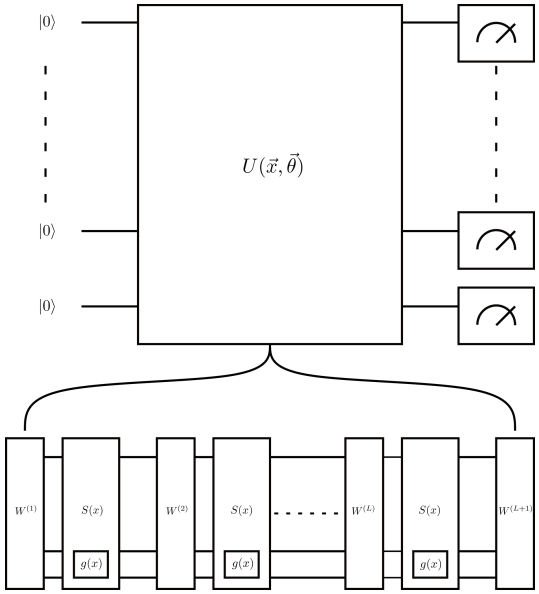

图18：带有重复数据编码电路块 $S(x)$（其量子门的形式为 $g(x) = e^{-ixH}$）和可训练电路块 $W_{(i)}$的量子神经网络。数据编码电路块决定了 $\vec{\omega}$可用的频谱，而其余部分决定了傅里叶系数 $c_{\vec{\omega}}$。

下面我们来讨论量子电路的三个主要组成部分：

- 重复的数据编码电路块 $S(x)$会制备一个初始态，对（一维）输入数据 x 进行编码；由于没有自由参数，该电路块不可训练。它由将经典数据嵌入量子态的特定量子门表示，这些量子门的形式为 $g(x) = e^{-ixH}$，其中 H 是编码哈密顿量，可以是任意酉算子。在这项工作中，我们使用泡利 X 旋转门，而 $S(x)$中的编码哈密顿量将决定可用频谱 Ω。

- 可训练电路块 $W(\theta_i)$ 由一组自由参数$\theta_i = (\theta_1, \theta_2, ...)$参数化。这里没有特殊假设，因此我们可以将这些可训练电路块视为任意酉变换。可训练参数将影响系数 c_ω。

- 最后一部分是在 M 输出端测量一个物理可观测量。该可观测量具有一般性，可以是电路中每条线路或线路子集上的局域可观测量。

我们的目标是证明 $f(x)$ 可以写成部分傅里叶级数 [29, 30]

$$f_{\theta_i}(x) = \langle 0|U^\dagger(x,\theta_i)MU(x,\theta_i)|0\rangle = \sum_{n\in\Omega} c_n e^{inx}. \tag{5.4}$$

请注意，为简单起见，这里我们将频率 $\Omega \subset \mathbb{Z}^N$ 取为整数。训练过程如下：我们对量子模型进行采样，模型具有 $U(x, \theta_i)$，然后将均方误差定义为损失函数。要优化损失函数，我们需要调整自由参数 $\theta = (\theta_1, \theta_2, ...)$。优化由经典优化算法执行，该算法会查询量子设备；在此过程中，我们可以将量子过程视为黑箱，只检查经典数据输入和测量输出。量子模型的输出是泡利-Z测量的期望值。

我们使用单量子比特泡利旋转门作为编码 $g(x)$ [30]。频率谱 Ω 由编码哈密顿量决定。确定可用频率可以考虑两种情形：数据重上传 [68] 和并行编码 [69] 模型。在前一种情形中，我们依次重复执行 r 次泡利旋转门，这意味着我们作用于同一个量子比特，但使用多个层 $r = L$；而在后一种情形中，我们在 r 个不同的量子比特上并行执行类似操作，但只使用一层 $L = 1$。这些模型使量子电路能够访问越来越丰富的频率，其中 $\Omega = \{-r, ..., -1, 0, 1, ..., r\}$ 包含最高次数为 r 的整数频率谱。这对应于我们想要计算的部分傅里叶级数的最高次数。

根据上述讨论，可以直接推导出此类量子模型 [30]所能达到的最高频率。但在实际应用中，如果目标函数的次数大于层数（例如，在单量子比特的情况下），拟合的准确度将会低得多。[4] 增加 L 的值通常需要更多的训练轮次，才能在相同的学习率下收敛。

这与一个更难的问题相关：在训练过程中如何控制傅里叶系数，因为所有模块 $W^{(i)}(\theta_i)$ 和测量观测量都会对“每个”傅里叶系数产生影响。然而，这些系数是自由度有限的量子电路的函数。这意味着，

[4]某些初始权重采样甚至可能无法收敛到令人满意的解。这与随机初始化的变分量子电路中普遍存在的贫瘠高原问题 [70]有关，类似于经典的梯度消失问题。

具有特定结构的量子电路，即使自由度足够，也只能实现所有可能傅里叶系数的一个子集。虽然目前还没有系统性的理解，但可参见一项探索哪些傅里叶系数能够实现的模拟 [30]。事实上，一个尚未解决的问题是：当L渐近趋于无穷大时，单量子比特模型能否通过构造任意傅里叶系数来逼近任意函数。

5.2 作为傅里叶级数的生成函数

鉴于量子模型的框架及其与部分傅里叶级数的关系，一个自然的问题是：纠缠熵能否在这一设置中实现。为探讨这一问题，有必要重新审视冯·诺依曼熵的生成函数

$$G(z;\rho_A) \equiv -\operatorname{Tr}\left(\rho_A \ln \frac{1-z\rho_A}{1-z}\right) = \sum_{k=1}^{\infty} \frac{f(k)}{k} z^k, \tag{5.5}$$

将其明确表示为泰勒级数。目标是用部分傅里叶级数来改写生成函数。因此，我们将能够确定冯·诺依曼熵和 R´恩yi 熵是否属于量子神经网络能够描述的函数类。需要注意的是，我们将只关注低深度或低宽度电路的小规模测试，因为电路的深度或宽度将与傅里叶级数中可近似的阶数完全对应。

但我们不能简单地将原始生成函数或其泰勒级数形式转换为傅里叶级数。这样做通常会涉及 ρ_A 中的特殊函数，而我们将无法用 $\operatorname{Tr}\rho_A^n$ 对其进行明确表达。因此，必须得到一种傅里叶级数表达式，使我们能够利用 $\operatorname{Tr}\rho_A^n$ 计算不同阶数下对应的傅里叶系数；对于 $\operatorname{Tr}\rho_A^n$，我们已从共形场论中知晓其解析形式。

<escape>这确实可以实现，详见附录 A，其中给出了详细推导。生成函数在区间 $[w_1, w_2]$ 上、周期为$T = w_2 - w_1$ 的傅里叶级数表示为</escape>

$$\begin{aligned} G(w;\rho) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \Bigg\{ & \sum_{m=0}^{\infty} \frac{\tilde{f}(m)}{m} C_{cos}(n,m) \cos\left(\frac{2\pi n w}{T}\right) \\ & + \sum_{m=0}^{\infty} \frac{\tilde{f}(m)}{m} C_{sin}(n,m) \sin\left(\frac{2\pi n w}{T}\right) \Bigg\}, \end{aligned} \tag{5.6}$$

其中， C_{cos} 和 C_{sin} 是定义为如下形式的一些特殊函数。

$$C_{cos}(n,m) = \frac{2}{(m+1)T}\left[{}_pF_q\left(\frac{m+1}{2};\frac{1}{2},\frac{m+3}{2};-\frac{n^2\pi^2t_2^2}{T^2}\right)t_2^{m+1} - {}_pF_q\left(\frac{m+1}{2};\frac{1}{2},\frac{m+3}{2};-\frac{n^2\pi^2t_2^2}{T^2}\right)t_1^{m+1}\right], \tag{5.7}$$

$$C_{sin}(n,m) = \frac{4n\pi}{(m+2)T^2}\left[{}_pF_q\left(\frac{m+2}{2};\frac{3}{2},\frac{m+4}{2};-\frac{n^2\pi^2t_2^2}{T^2}\right)t_2^{m+2} - {}_pF_q\left(\frac{m+2}{2};\frac{3}{2},\frac{m+4}{2};-\frac{n^2\pi^2t_1^2}{T^2}\right)t_1^{m+2}\right], \tag{5.8}$$

其中， ${}_pF_q$ 为广义超几何函数。另请注意，

$$\tilde{f}(m) \equiv \sum_{k=0}^{m}\frac{(-1)^{2m-k+1}m!}{k!(m-k)!}\,\mathrm{Tr}\,(\rho_A^{k+1}). \tag{5.9}$$

类似地，零阶傅里叶系数为

$$a_0 = \sum_{m=0}^{\infty}\frac{\tilde{f}(m)}{m}C_{cos}(0,m) = \sum_{m=0}^{\infty}\frac{\tilde{f}(m)}{m}\frac{2(w_2^{m+1}-w_1^{m+1})}{(m+1)T}. \tag{5.10}$$

请注意，对 $m = 10$ 求和即可满足我们的目的，而 n 中的求和对应于傅里叶级数的次数。请注意，我们在模拟中使用的复值傅里叶系数 c_n 可以很容易地从表达式 ˜ 中重构。因此，计算傅里叶级数所需的唯一输入是 $f(m)$，其中 $\mathrm{Tr}\rho_A^{k+1}$ 已明确给出。这正是我们所预期的，也使得与泰勒级数形式的比较十分直接。

请注意，傅里叶级数的区间并非任意选取。我们将取区间$[w_1, w_2]$为 $[-1, 1]$，这是傅里叶级数（5.6）收敛的最大区间。此外，我们预期当 $w \to 1$ 由（5.6）可得冯·诺依曼熵，即

$$S(\rho_A) = \lim_{w\to 1} G(w;\rho_A). \tag{5.11}$$

然而，正如我们在图19中所见，傅里叶级数在区间端点附近会发生快速振荡。这种“跳跃不连续”的出现，是使用傅里叶级数逼近不连续函数或非周期函数时的一种普遍现象，称为吉布斯现象。由于我们要取边界点 $w \to 1$处的极限，这种现象会给准确求得冯·诺依曼熵带来严重问题。我们将在第5.4节再讨论这一问题。

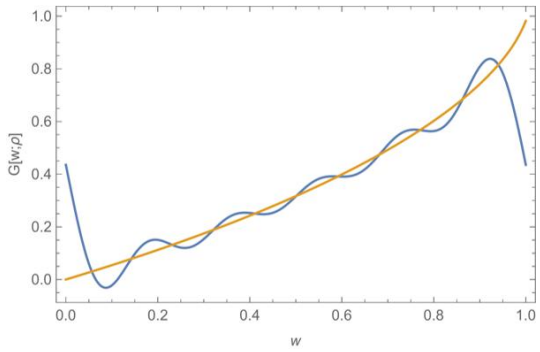

图 19: $w \to 1$ 端点附近傅里叶级数的吉布斯现象。我们采用单区间示例，其中黄色曲线表示泰勒级数形式的生成函数，蓝色曲线表示生成函数的傅里叶级数近似。

5.3 量子模型在纠缠熵方面的表达性

本小节将通过共形场论中的示例，展示部分傅里叶级数的量子模型的表达性。我们将重点讨论两个具体示例：小交比 x下的单区间和两个区间。虽然这些示例足以满足我们的目的，但值得注意的是，一旦使用（5.6）中的表达式推导出傅里叶级数表示，就可以研究所有具有已知解析形式的 Tr ρ_A^n 示例。

该演示使用 PennyLane [71] 完成。我们采用 Adam优化器，学习率为 0.005，批大小为 100，其中 MSE 是损失函数。请注意，与 [30] 相比，我们选择了更小的学习率，并使用 EarlyStopping 进行监控。对于所研究的两个示例，我们在训练中同时考虑了串行（数据重上传）模型和并行（并行编码）模型。请注意，在并行模型中，我们使用了 PennyLane 中的 StronglyEntanglingLayers，其自身包含 3 个用户定义层。在每种情况下，我们首先使用 300 个样本点随机初始化一个量子模型，以拟合目标函数

$$f(x) = \sum_{n=-k}^{n=k} c_n e^{-inx}. \tag{5.12}$$

其中，复值傅里叶系数由式（5.6）中的实系数计算得出。对于单区间和双区间示例，我们选取了 $k = 4$ ，并指定了物理参数。因此，我们需要使串行和并行模型中的 r 大于 $k = 4$。我们对每种情况进行了多次试验，并

在图 $\lesssim$ 3中列出最成功的结果，并将最大相对误差控制在 20∼23%。

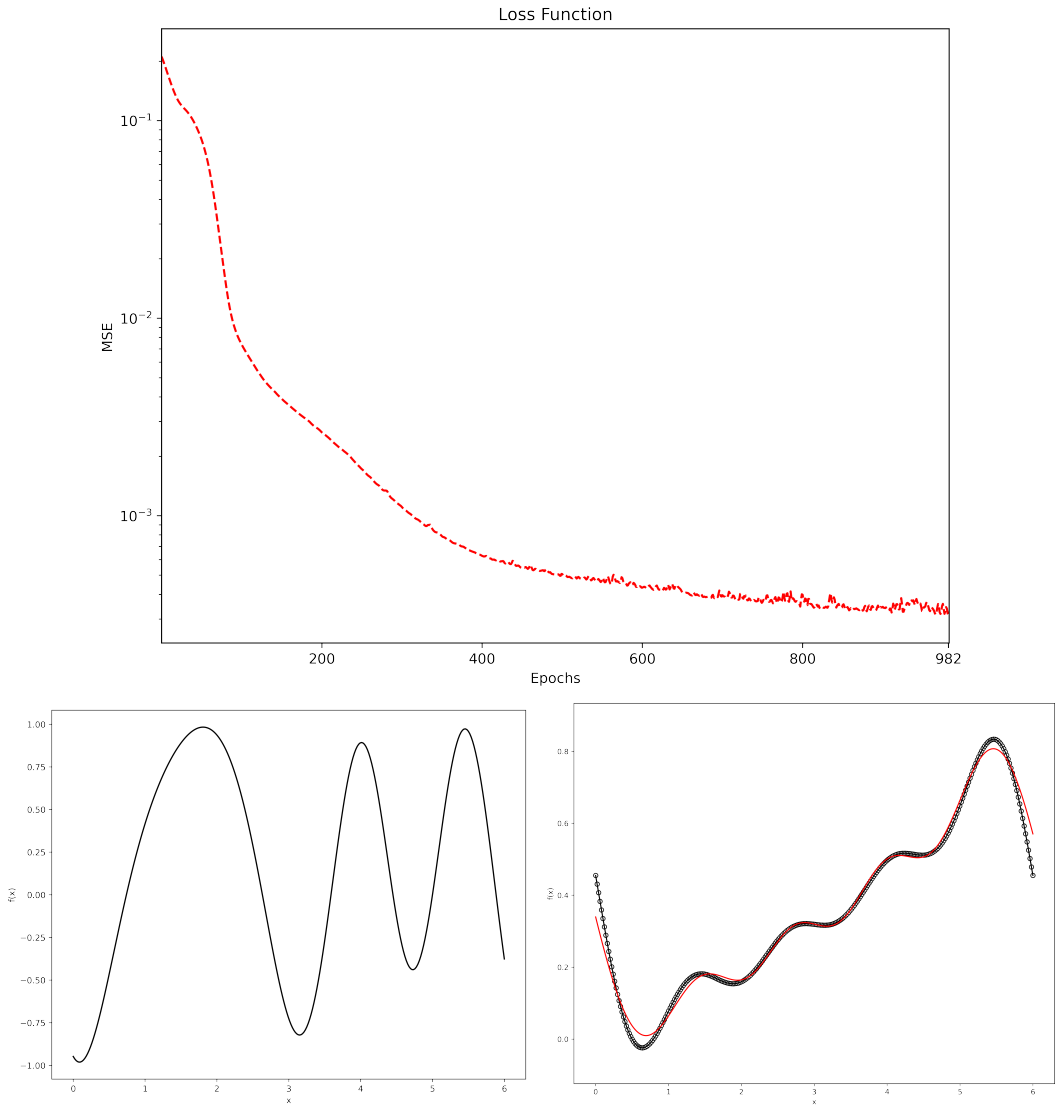

图20：使用数据样本训练随机串行量子模型，以拟合单区间情形的目标函数。上图：均方误差损失函数随训练轮次变化的曲线，最小损失在第 982 轮训练时达到。左下图：随机初始化的串行量子模型，包含 $r = 6$ 个连续重复的泡利编码门。右下图：圆表示单区间傅里叶级数的 300 个数据样本，其中 $\ell = 2$ 且 $\epsilon = 0.1$，对于（3.2）。红色曲线表示训练后的量子模型。

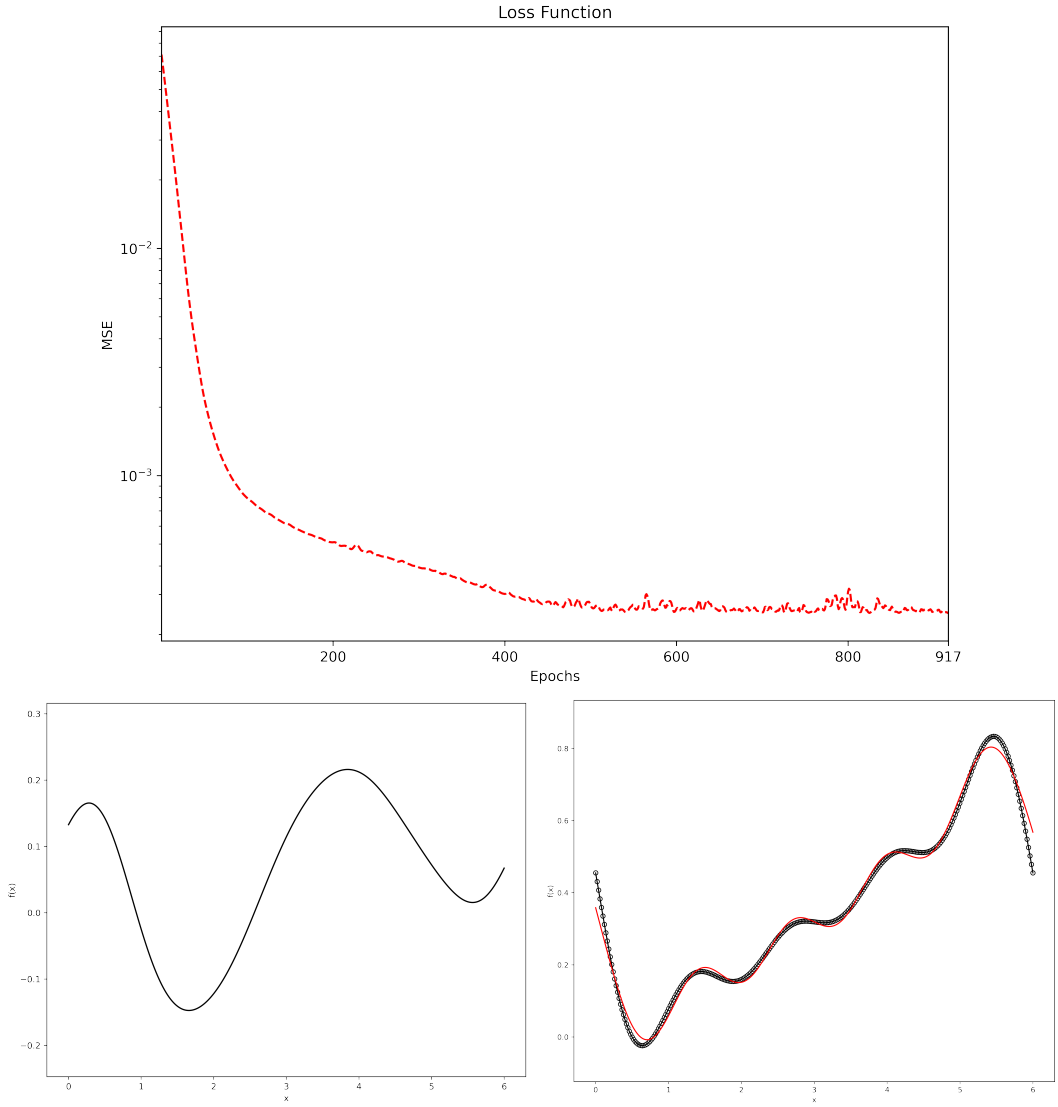

图 21：单区间情形的随机并行量子模型。上图：损失函数在第 917 个训练轮次达到最小值。下图：量子模型的一种随机初始化，其中包含 $r = 5$ 个并行重复的泡利编码门，并且已取得良好的拟合效果。

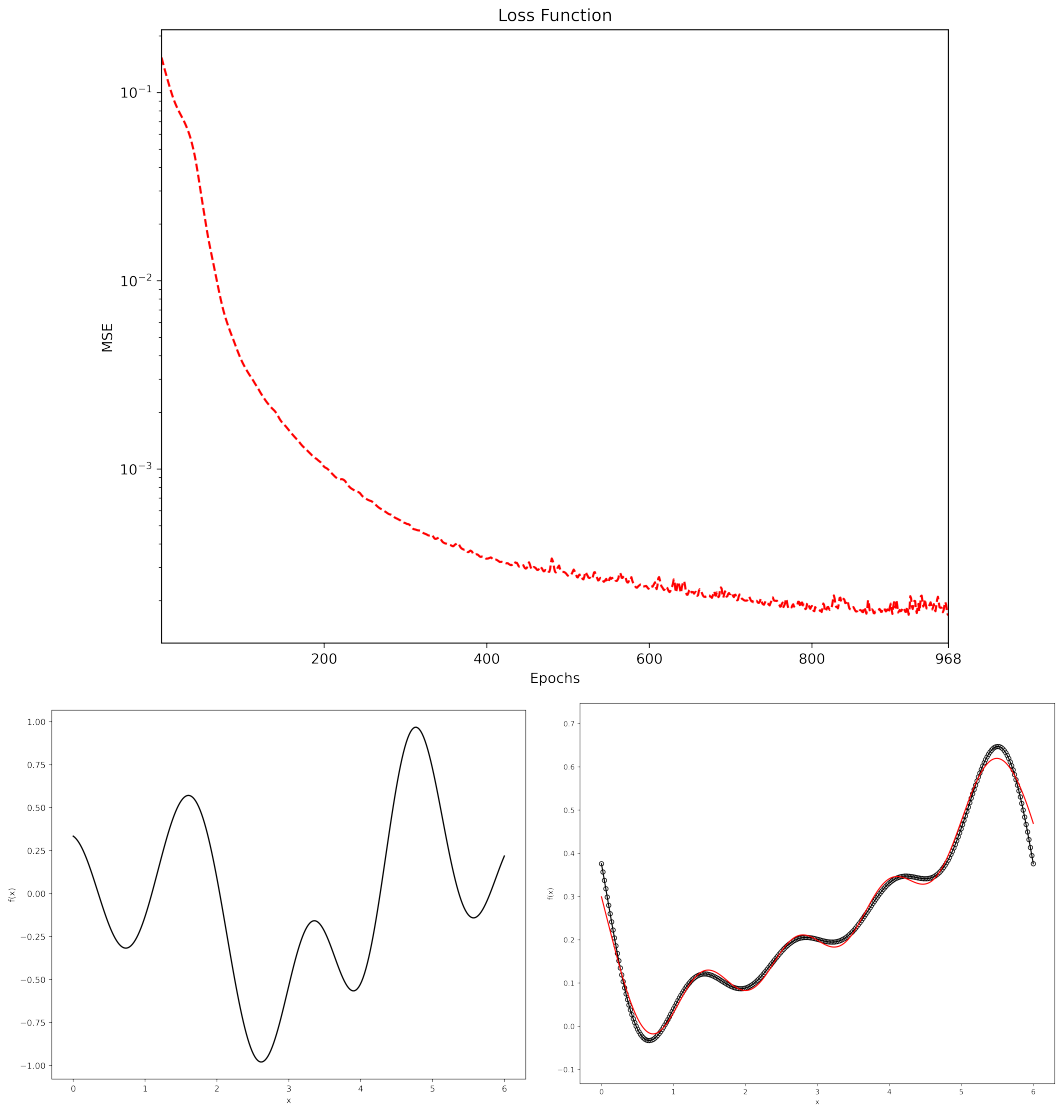

图22：使用数据样本训练随机串行量子模型，使其拟合交比较小的双区间系统的目标函数。上图：损失函数在第968轮训练时达到最小值。左下：具有 $r = 6$ 次泡利编码门连续重复的串行量子模型的随机初始化。右下：圆表示双区间傅里叶级数的300个数据样本，其中 $x = 0.05$，$\alpha = 0.1$，且 $\epsilon = 0.1$ 对于（3.15）。红色曲线表示训练后的量子模型。

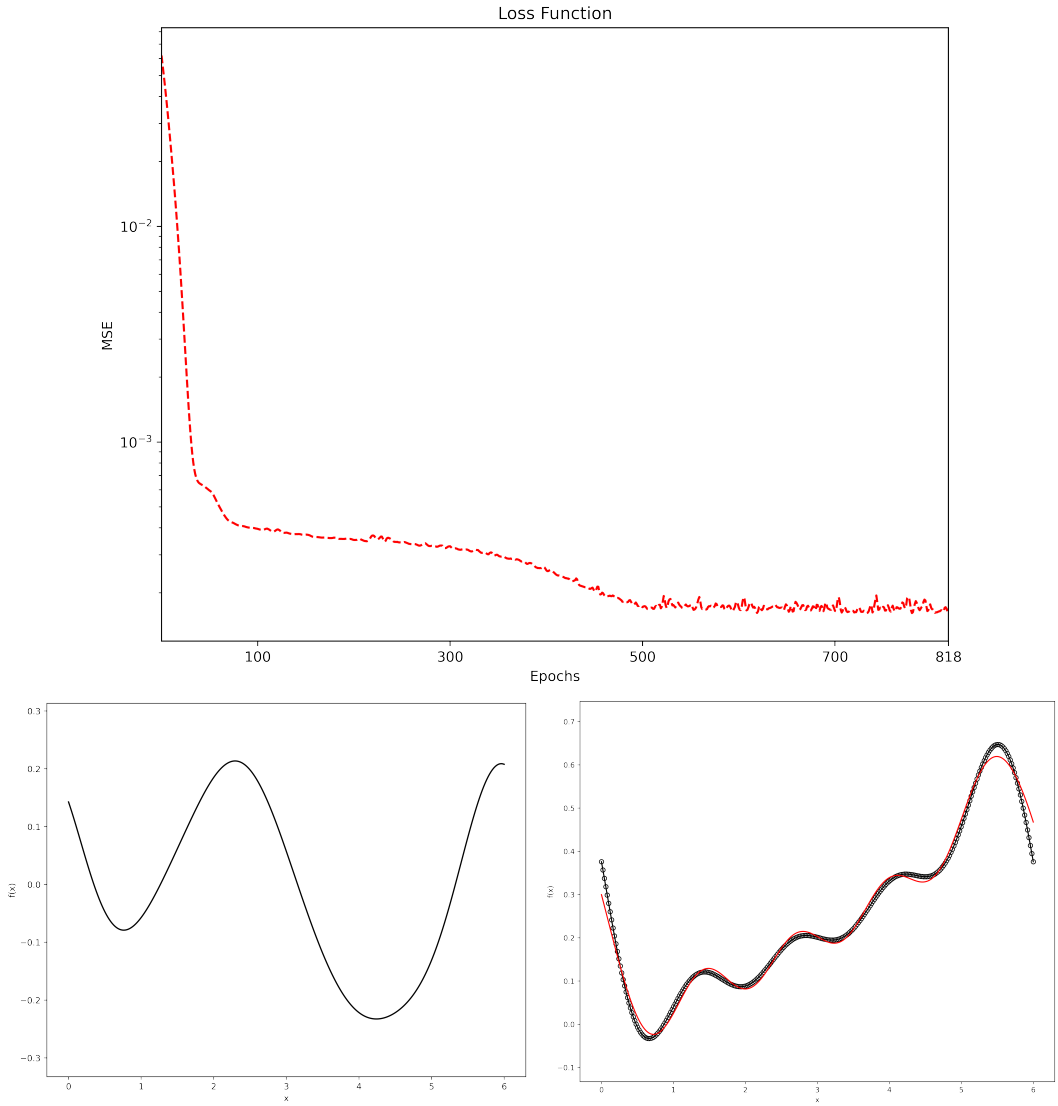

图 23：双区间情形下的随机并行量子模型。上图：损失函数在第 818 个训练轮次达到最小值。下图：量子模型经过 $r = 5$ 次并行重复的泡利编码门随机初始化后，已实现良好拟合。

如图所示 20∼23，必须对数据进行重新缩放，才能在量子模型与我们示例的傅里叶谱之间实现精确匹配。由于全局相位无法观测，这种重新缩放是可行的 [30]，，这会在数据编码中引入歧义。考虑我们的量子模型

$$f_\theta(x) = \langle 0|U^\dagger(x,\theta)MU(x,\theta)|0\rangle = \sum_{\omega\in\Omega} c_\omega(\theta)e^{i\omega x}, \tag{5.13}$$

其中，我们考虑单量子比特 $L = 1$ 的情况，然后

$$U(x) = W^{(2)} g(x) W^{(1)}. \tag{5.14}$$

请注意，频谱 Ω 由数据编码哈密顿量的特征值决定，该哈密顿量由算符给出

$$g(x) = e^{-ixH}. \tag{5.15}$$

H 有两个特征值（λ_1，λ_2），但我们可以将能谱重新缩放为（$-\gamma$，γ），因为全局相位不可观测（例如，对于泡利旋转，我们有 $\gamma = \frac{1}{2}$）。我们可以通过重新缩放，将 H 特征值中的 γ 吸收到数据输入中，具体方法是

$$\tilde{x} = \gamma x. \tag{5.16}$$

因此，我们可以假设 H 的特征值为其他一些值。具体而言，我们在训练过程中选择了 $\gamma = 6$，其中 x 中的区间从 $[0, 1]$ 扩展至$[0, 6]$,，如图 20∼23所示。

我们应当强调，我们并未重新缩放原始目标数据，而是重新缩放数据的编码方式。实际上，我们重新缩放的是量子模型本身的频率。有趣的是，作用于量子态的算符的全局相位偏移无法被观测到，却会影响量子模型的表达能力。这可以理解为对数据进行预处理；据称，这种预处理能够扩展量子模型可表示 [30] 的函数类。

这表明，可以考虑将重新缩放参数 γ 作为可训练参数 [68]。这样一来，缩放便会转变为自适应的“频率匹配”过程，从而有可能提高量子模型的表达性。这里我们仅将 γ 视为可调超参数。缩放无需与数据相匹配，但找到合适的缩放参数对模型训练至关重要。

5.4 恢复冯·诺依曼熵

到目前为止，我们已成功将生成函数改写为定义在区间 $w \in [-1, 1]$ 上的 $f_N(w)$ 阶部分傅里叶级数，其次数为 N。借助变分量子电路，我们能够准确再现该级数的傅里叶系数。原则上，通过采用适当的数据编码和重新缩放策略，增加量子模型的深度或宽度，就能使我们以任意次数 N 捕捉该级数。因此，R´enyi 熵的表达性可以

通过量子模型来确立。然而，仍有一个关键问题：我们需要在极限 $w \to 1$ 下恢复冯·诺依曼熵。

$$\lim_{w\to 1} G(w;\rho_A) = S(\rho_A), \tag{5.17}$$

其中，极限点恰好位于我们进行近似的区间边界处。然而，正如我们从图24中清楚看到的那样，取这样的极限会天真地得出一个与真实冯·诺依曼熵相比极不准确的值。即使增大 N，以便相较于泰勒级数形式获得更好的级数近似，如图24所示，这种效应也不会消失。这是因为傅里叶级数近似在端点处总是会发生振荡；在用傅里叶级数近似不连续函数或非周期函数时，这种普遍现象称为吉布斯现象。

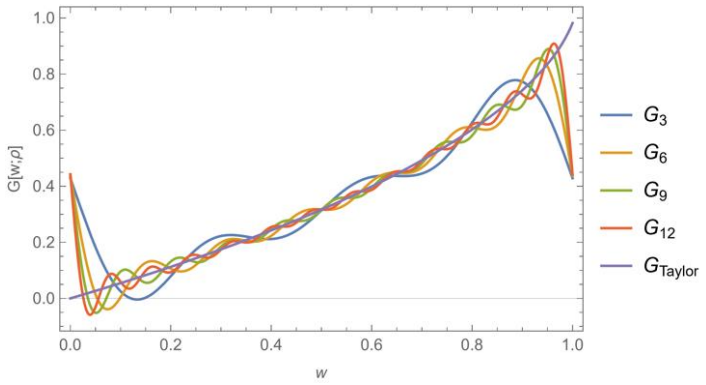

图24：我们绘制了单区间示例，其中 $L = 2$ 和 $\epsilon = 0.1$ 对应于（3.2）。这里，图例 G_N 表示生成函数的傅里叶级数，展开至 N次，即在（ $m = 10$ ）中求和至5.6。G_{Taylor} 表示生成函数的泰勒级数形式（2.9），求和至 $k = 100$。

先验地，只要函数 $f(x)$ 平滑且具有周期性，函数 $f(x)$ 的部分傅里叶级数就是一种非常精确的点值重构方式。此外，如果 $f(x)$ 是解析且具有周期性的，那么随着 N 的增大，部分傅里叶级数 f_N 将以指数速度收敛到 $f(x)$。然而，如果 $f(x)$ 不连续或不具有周期性，那么一般而言， $f_N(x)$ 并不能准确近似 $f(x)$。不仅收敛速度缓慢，而且在区间边界附近还会出现超调。理解这一现象有许多不同的方法。总的来说，困难在于我们试图从傅里叶系数的整体性质中获取精确的局部信息，而这些系数是通过对区间积分定义的；这似乎从根本上是不可能的。

从数学上看，吉布斯现象的出现可以通过狄利克雷核的振荡特性轻松理解；当傅里叶级数写成卷积形式时，就会出现狄利克雷核。具体而言，傅里叶部分和可以写作

$$s_n(x) = \frac{1}{\pi}\int_{-\pi}^{\pi} f(\xi) D_n(\xi - x) d\xi, \tag{5.18}$$

其中，狄利克雷核 $D_n(x)$ 定义为

$$D_n(x) = \frac{\sin\left(n + \frac{1}{2}\right)x}{2\sin\frac{x}{2}}. \tag{5.19}$$

该函数在正值和负值之间振荡。因此，傅里叶级数在边界处的跳跃间断点附近会出现吉布斯现象。

因此，我们的问题可以准确地表述如下：给定 $2N+1$ˆ 生成函数的傅里叶系数 f_k (5.6)，其中 $-N \leq k \leq N$，生成函数定义在区间 $w \in [-1,1]$内，我们需要重构函数在极限 $w \to 1$处的点值。生成函数在该极限处的点值恰好对应于冯·诺依曼熵。尤其是，我们需要重构结果以指数速度收敛到生成函数的正确点值，其中 N ，即

$$\lim_{w\to 1} |G(w;\rho_A) - f_N(w)| \leq e^{-\alpha N}, \quad \alpha > 0. \tag{5.20}$$

这是为了使量子模型具有实际应用价值；目前，部分傅里叶级数可近似的阶数 N 受量子电路的深度或宽度限制。

我们需要一种能够减弱振荡，甚至更理想地将其彻底消除的操作。为改善振荡，人们开发了多种滤波方法，包括非负且衰减的 Fej´er 核，它能平滑整个区间上的傅里叶级数；以及引入 Lanczos σ因子，它能在边界附近局部减弱振荡。关于吉布斯现象和这些滤波方法的全面讨论，请参见 [72]。不过，我们强调，这些方法都不令人满意，因为它们仍然无法在边界附近恢复函数 $f(x)$ 的精确点值。

因此，我们需要一种更有效的方法来彻底消除吉布斯现象。在这里，我们将采用一种强大的方法，重新展开部分傅里叶彩票论坛

展开为 Gegenbauer 多项式基。[5] 这是 20 世纪 90 年代一系列奠基性研究提出的方法 [74–79], 我们也推荐参阅 [80, 81] 以了解较新的综述。

Gegenbauer 展开方法仅通过对傅里叶系数求和少数几项，即可实现指数级准确率的精确表示。给定区间 $[-1,1]$（或子区间 $[a,b]\subset[-1,1]$）上的一个解析且非周期函数 $f(x)$，其傅里叶系数

$$\hat{f}_k=\frac{1}{2}\int_{-1}^{1}f(x)e^{-ik\pi x}dx, \tag{5.21}$$

以及部分傅里叶级数

$$f_N(x)=\sum_{k=-N}^{N}\hat{f}_k e^{ik\pi x}. \tag{5.22}$$

以下 Gegenbauer 展开表示我们想要用傅里叶信息近似的原始函数

$$S_{N,M}(x)=\sum_{n=0}^{M}g_{n,N}^{\lambda}C_n^{\lambda}(x), \tag{5.23}$$

其中，$g_{n,N}^{\lambda}$ 是 Gegenbauer 展开系数，$C_n^{\lambda}(x)$ 是 Gegenbauer 多项式。[6] 注意，我们有以下积分公式可用于计算 $g_{n,N}^{\lambda}$

$$\frac{1}{h_n^{\lambda}}\int_{-1}^{1}(1-x^2)^{\lambda-\frac{1}{2}}e^{in\pi x}C_n^{\lambda}(x)dx=\Gamma(\lambda)\left(\frac{2}{\pi k}\right)^{\lambda}i^n(n+\lambda)J_{n+\lambda}(\pi k), \tag{5.26}$$

then

$$g_{n,N}^{\lambda}=\delta_{0,n}\hat{f}(0)+\Gamma(\lambda)i^n(n+\lambda)\sum_{k=-N,k\neq0}^{N}J_{n+\lambda}(\pi k)\left(\frac{2}{\pi k}\right)^{\lambda}\hat{f}_k, \tag{5.27}$$

[5]请注意，还存在其他方法，通过对函数进行周期性扩展，以便在感兴趣的定义域内准确表示函数；这些方法涉及基于切比雪夫多项式重构函数 [73]。然而，本文不探讨这种方法。[6]

格格 nbauer 展开系数 $g_{n,\ N}^{\lambda}$ 由部分傅里叶级数 $f_N(x)$ 定义为

$$g_{n,N}^{\lambda}=\frac{1}{h_n^{\lambda}}\int_{-1}^{1}(1-x^2)^{\lambda-\frac{1}{2}}f_N(x)C_n^{\lambda}(x)dx,\quad 0\le n\le M. \tag{5.24}$$

对于 $\lambda\ge0$，次数为 n 的 Gegenbauer 多项式定义为满足

$$\int_{-1}^{1}(1-x^2)^{\lambda-\frac{1}{2}}C_k^{\lambda}(x)C_n^{\lambda}(x)dx=0,\quad k\neq n. \tag{5.25}$$

我们参见附录。B 了解有关 Gegenbauer 展开的性质的更详细说明。

ˆ 其中，我们只需要傅里叶系数 f_k。

事实上，Gegenbauer 展开是一个由两个参数构成的函数族，其特征由 λ 和 M决定。已有研究表明，令 $\lambda = M = \beta\epsilon N$ 其中$\epsilon = (b-a)/2$，并且在傅里叶情形下取 $\beta < \frac{2\pi e}{27}$，该展开便能达到指数级准确率，且 N。注意， M 决定 Gegenbauer 多项式的次数，因此我们应允许原始傅里叶级数的次数随 M增大。图 25 清楚展示了 Gegenbauer 展开如何根据傅里叶数据逼近生成函数。最终，随着展开阶数的增加，我们将能够重构 $w \to 1$ 附近冯·诺依曼熵的点值。关于指数级准确率的更精确表述，见附录B。该方法确实是一个以指数级准确率从全局信息重构局部信息的过程，从而有效消除吉布斯现象。

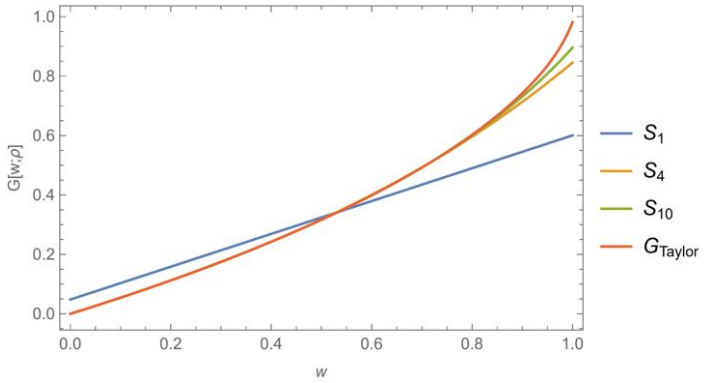

图 25：根据傅里叶信息构造的 Gegenbauer 展开。其中， S_M指的是 M 阶 Gegenbauer 多项式。注意，我们设定了 $\beta\epsilon = 0.25$，然后$\lambda = M = 0.25N$。因此，为了构造 M 阶多项式，我们需要截至 $N = 4M$ 阶的傅里叶系数信息。

6 讨论

本文中，我们探讨了一种利用经典神经网络和量子神经网络研究冯·诺依曼熵从 R´enyi 熵解析延拓的新方法。我们通过将 $\mathrm{Tr}\rho_A^n$ 写成一个生成函数的形式来处理解析延拓问题，该生成函数表现为泰勒级数（2.9）。我们证明，深度学习模型只需有限个 R´enyi 熵即可实现这一目标。

我们不再为经典神经网络采用静态模型设计，而是使用 KerasTuner 来寻找最优的模型架构和超参数。我们考虑了两种监督学习场景：一种是利用稠密连接神经网络，在已知 R´enyi 熵的情况下预测冯·诺依曼熵；另一种是使用循环神经网络，将更高阶的 R´enyi 熵作为序贯深度学习任务。在这两种情况下，我们都实现了对相应目标值的高准确率预测。

对于量子神经网络，我们将类似的监督学习问题表述为从输入到预测的映射。这使我们能够研究量子神经网络作为函数逼近器的表达能力，尤其是对冯·诺依曼熵的表达能力。我们研究了能够将生成函数显式表示为部分傅里叶级数的量子模型。然而，吉布斯超调阻碍了冯·诺依曼熵准确点值的恢复。为解决这一问题，我们用 Gegenbauer 多项式重新展开级数，从而实现指数收敛并提高准确率。

我们的方法带来了若干相关问题和潜在改进方向：

- 在使用 KerasTuner 之前，选择合适的架构至关重要，例如第3 节中的稠密连接层和第4节中的循环神经网络。因为这些架构是先验地针对特定任务构建的。KerasTuner 仅作为一种有效方法，用于确定模型训练的最佳复杂度和超参数。然而，由于 CFT_2 中的示例在冯·诺依曼熵和 R´enyi 熵方面具有不同的解析结构，因此探究不同超参数与各示例之间的相关性将很有意义。

- 尽管效率较高，第3.1 节和第4.1节中概述的参数空间并不能保证包含最优设置，而 KerasTuner 搜索的参数空间也可能存在更好的方案。

- 我们可以通过固定不同的物理参数来生成数据集，例如式（3.6）中的温度或式（x 中的交比3.15）。虽然我们考虑了需要变化的自然参数，但探索不同的参数可能会带来更强的表示能力。有可能找到一个稠密模型，使其在所有参数范围内都能提供可行的预测，但这可能需要模型集成。

- 正则化方法（例如 K 折交叉验证）有可能在保持性能不变的情况下缩减模型规模或数据集。确定所需的最小数据集规模，或探究低复杂度模型是否仍具备学习纠缠熵所需的相同表征能力，将很有价值。

• 另一方面，使用更多数据和资源训练模型，是提升模型性能最有效的方法。还可以扩大 KerasTuner 中的搜索规模，或使用集成方法组合它找到的模型。

• 对于量子神经网络，需要注意的是，正如我们在第5.1节中所述，我们的方法并不能保证收敛到正确的傅里叶系数。研究各种预处理或数据编码策略，或许有助于提高部分傅里叶级数的近似精度 r。

还有一些值得探索的未来方向，我们将在此简要讨论：

• **互信息：**我们可以将研究扩展到两个不相交区间 A 和 B的互信息，它是一种与冯·诺依曼熵相关的纠缠度量，定义为

$$I(A:B) \equiv S(\rho_A) + S(\rho_B) - S(\rho_{A\cup B}). \tag{6.1}$$

特别地，[8]，中有一种关于生成函数的猜想形式，其中 Tr ρ_A^n 被 Tr ρ_A^n Tr ρ_B^n/ Tr $\rho_{A\cup B}^n$ 取代。值得使用这一生成函数来探究经典神经网络和量子神经网络的表达能力，尤其是互信息能够消除紫外发散，并可与一些现实的模拟（例如自旋链模型 [82]）进行比较。

• **用于更高 Rényi 熵的自监督学习´：尽管我们已在第 4节证明，RNN 架构在序列学习问题中很有效，但也值得考虑其他可能带来更佳表现的架构。例如，时间延迟神经网络、深度可分离卷积神经网络或 Transformer 可能适用于某些类型的数据。在扩展提取更高阶 R´ényi 熵这一任务时，或许值得探索这些架构，并将其作为自监督学习方法，特别是对于无法进行解析延拓的情形。

• **来自解析延拓的其他纠缠度量：**其他重要的纠缠度量，例如相对熵或纠缠负性，可能需要进行解析延拓，并且可以基于神经网络进行数值研究。我们还可以考虑纠缠熵或纠缠谱，这些量可以在源自凝聚态物理或全息系统的特定模型中进行模拟。

• **经典和量子神经网络的表达性：**我们以生成函数为媒介，研究了经典神经网络和量子神经网络对于冯·诺依曼熵和 R´enyi 熵的表达性。这或许有助于我们为其他适用于神经网络的纠缠度量设计优良的生成函数。此外，了解其他纠缠度量是否也属于量子神经网络能够实现的函数类，同样值得探究。

致谢

我们感谢 Xi Dong 对这项工作的鼓励。C-H.W. 的研究部分得到了美国能源部 DE-SC0023275 号拨款和台湾教育部的资助。本研究还获得了空军科学研究办公室 FA9550-19-1-0360 号项目的支持。

生成函数的傅里叶级数表示

假设生成函数可表示为如下傅里叶级数，该生成函数见式（2.4）

$$G(z;\rho_A)=\sum_{n=-\infty}^{\infty}c_n e^{inz}. \tag{A.1}$$

我们的思路是，仅凭关于 $G(z;\ \rho)$ 或 $\mathrm{Tr}\,\rho_A^n$ 的信息来计算傅里叶系数。对于一般周期 T，我们可以利用实值系数 a_n 和 b_n 计算复值傅里叶系数 c_n，其中

$$G(z;\rho_A)=\frac{a_0}{2}+\sum_{n=1}^{\infty}a_n\cos\left(\frac{2\pi nz}{T}\right)+b_n\sin\left(\frac{2\pi nz}{T}\right). \tag{A.2}$$

注意

$$a_n=\frac{2}{T}\int_{z_1}^{z_2}G(z;\rho_A)\cos\left(\frac{2\pi nz}{T}\right)dz, \tag{A.3}$$

$$b_n=\frac{2}{T}\int_{z_1}^{z_2}G(z;\rho_A)\sin\left(\frac{2\pi nz}{T}\right)dz, \tag{A.4}$$

其中，我们只需利用 $\mathrm{Tr}\rho_A^n$ 的生成函数计算两个傅里叶系数。然而，上述积分通常难以求解。我们将转而证明，a_n 和 b_n 都可以写成以下级数

$$a_n=\sum_{m=0}^{\infty}\frac{G(0;\rho)^{(m)}}{m!}C_{cos}(n,m), \tag{A.5}$$

$$b_n = \sum_{m=0}^{\infty} \frac{G(0;\rho)^{(m)}}{m!} C_{sin}(n,m). \tag{A.6}$$

其中 $C_{cos}(n,\ m)$ 和 $C_{sin}(n,\ m)$ 涉及某些特殊函数。$(0;\rho_A)^{(m)}$ 的定义始于下面这个关于 w （2.6）

$$G(w;\rho_A) = -\operatorname{Tr}\left(\rho_A \ln\left[1-w(1-\rho_A)\right]\right), \tag{A.7}$$

其中，关于 $w \to 0$ 的 m 阶导数

$$\begin{aligned} G(0;\rho_A)^{(m)} &= -\operatorname{Tr}[(-1)^{m+1}(m-1)!\rho_A(\rho_A-1)^m] \\ &= -(m-1)!\sum_{k=0}^{m} \frac{(-1)^{2m-k+1}m!}{k!(m-k)!} \operatorname{Tr}\left(\rho_A^{k+1}\right). \end{aligned} \tag{A.8}$$

注意，我们必须定义 $m=0$，使得

$$G(0;\rho_A)^{(0)} = -\operatorname{Tr}(\rho_A \ln 1) = 0. \tag{A.9}$$

那么，生成函数在周期为 $T = w_2 - w_1$ 的区间 $[w_1, w_2]$ 上的傅里叶级数表示为

$$\begin{aligned} G(w;\rho_A) = \frac{a_0}{2} &+ \sum_{n=1}^{\infty}\Bigg\{\sum_{m=0}^{\infty}\frac{\tilde{f}(m)}{m}C_{cos}(n,m)\cos\left(\frac{2\pi n w}{T}\right) \\ &+ \sum_{m=0}^{\infty}\frac{\tilde{f}(m)}{m}C_{sin}(n,m)\sin\left(\frac{2\pi n w}{T}\right)\Bigg\}, \end{aligned} \tag{A.10}$$

其中我们定义了

$$\tilde{f}(m) \equiv -\sum_{k=0}^{m} \frac{(-1)^{2m-k+1}m!}{k!(m-k)!} \operatorname{Tr}\left(\rho_A^{k+1}\right). \tag{A.11}$$

表达式中显然出现了 $\mathbf{Tr}\,\rho_A^{k+1}$。

现在我们需要求出 $C_{cos}(n,\ m)$ 和 $C_{sin}(n,\ m)$。首先，让我们考虑一般情形

$$a_n = \frac{2}{T}\int_{t_1}^{t_2} f(t)\cos\left(\frac{2\pi n t}{T}\right)dt, \tag{A.12}$$

为简便起见，我们将 $G(w;\ \rho_A)$ 写作 $f(t)$。我们可以写出这两部分的泰勒级数。

$$f(t) = \sum_{j=0}^{\infty}\frac{f^{(j)}(0)}{j!}t^j, \quad \cos\left(\frac{2\pi n t}{T}\right) = \sum_{k=0}^{\infty}\frac{(-1)^k}{(2k)!}\left(\frac{2\pi n t}{T}\right)^{2k}, \tag{A.13}$$

考虑以下函数

$$T_{cos}(t) \equiv f(t)\cos\left(\frac{2\pi nt}{T}\right) = \left[\sum_{j=0}^{\infty}\frac{f^{(j)}(0)}{j!}t^j\right]\left[\sum_{k=0}^{\infty}\frac{(-1)^k}{(2k)!}\left(\frac{2\pi nt}{T}\right)^{2k}\right], \tag{A.14}$$

然后让我们按 t 的阶数收集各项

$$\begin{aligned} T_{cos}(t) = f(0) + f^{(1)}(0)t &+ \left(\frac{1}{2}f^{(2)}(0) - 2f(0)\left(\frac{\pi n}{T}\right)^2\right)t^2 \\ &+ \left(\frac{1}{6}f^{(3)}f(0) - 2f^{(1)}(0)\left(\frac{\pi n}{T}\right)^2\right)t^3 \\ &+ \left(\frac{1}{24}f^{(4)}(0) - f^{(2)}\left(\frac{\pi n}{T}\right)^2 + \frac{2}{3}f(0)\left(\frac{\pi n}{T}\right)^4\right)t^4 \\ &+ \cdots, \end{aligned} \tag{A.15}$$

则积分变为

$$\begin{aligned} \int_{t_1}^{t_2} T_{cos}(t)dt = f(0)(t_2 - t_1) &+ \frac{1}{2}f^{(1)}(0)(t_2^2 - t_1^2) \\ &+ \frac{1}{3}\left(\frac{1}{2}f^{(2)}(0) - 2f(0)\left(\frac{\pi n}{T}\right)^2\right)(t_2^3 - t_1^3) \\ &+ \frac{1}{4}\left(\frac{1}{6}f^{(3)}f(0) - 2f^{(1)}(0)\left(\frac{\pi n}{T}\right)^2\right)(t_2^4 - t_1^4) \\ &+ \frac{1}{5}\left(\frac{1}{24}f^{(4)}(0) - f^{(2)}\left(\frac{\pi n}{T}\right)^2 + \frac{2}{3}f(0)\left(\frac{\pi n}{T}\right)^4\right)(t_2^5 - t_1^5) \\ &+ \cdots. \end{aligned} \tag{A.16}$$

现在我们 希望重新排列此表达式，将关于 $f^{(m)}(0$ 的各项收集起来)

$$\begin{aligned} \int_{t_1}^{t_2} T_{cos}(t)dt = f(0)&\left((t_2 - t_1) - \frac{2}{3}\left(\frac{\pi n}{T}\right)^2(t_2^3 - t_1^3) + \frac{2}{15}\left(\frac{\pi n}{T}\right)^4(t_2^5 - t_1^5) + \cdots\right) \\ &+ f^{(1)}(0)\left(\frac{1}{2}(t_2^2 - t_1^2) - \frac{1}{2}\left(\frac{\pi n}{T}\right)^2(t_2^4 - t_1^4) + \cdots\right) \\ &+ f^{(2)}(0)\left(\frac{1}{24}(t_2^4 - t_1^4) + \cdots\right) + \cdots. \end{aligned} \tag{A.17}$$

乘以因子 $2/T$ 后，可将其写为

$$a_n = \frac{2}{T}\int_{t_1}^{t_2} T_{cos}(t)dt = \sum_{m=0}^{\infty}\frac{f^{(m)}(0)}{m!}C_{cos}(n,m), \tag{A.18}$$

其中

$$C_{cos}(n,m) = \sum_{p=0}^{\infty}\left[\frac{(-1)^p 2^{(2p+1)} n^{2p}\pi^{2p}(t_2^{(2p+m+1)} - t_1^{(2p+m+1)})}{(2p+m+1)(2p)!T^{2p+1}}\right]$$
$$= \frac{2}{(m+1)T}\left[{}_pF_q\left(\frac{m+1}{2};\frac{1}{2},\frac{m+3}{2};-\frac{n^2\pi^2 t_2^2}{T^2}\right)t_2^{m+1}\right.$$
$$\left.-{}_pF_q\left(\frac{m+1}{2};\frac{1}{2},\frac{m+3}{2};-\frac{n^2\pi^2 t_2^2}{T^2}\right)t_1^{m+1}\right]. \tag{A.19}$$

接下来，我们考虑 $C_{sin}(n,\ m)$ 的情况，其中我们需要推导出

$$b_n = \frac{2}{T}\int_{t_1}^{t_2} f(t)\sin\left(\frac{2\pi nt}{T}\right)dt, \tag{A.20}$$

再者，我们知道

$$\sin\left(\frac{2\pi nt}{T}\right) = \sum_{k=0}^{\infty}\frac{(-1)^k}{(2k+1)!}\left(\frac{2\pi nt}{T}\right)^{(2k+1)}, \tag{A.21}$$

然后我们定义

$$T_{sin}(t) \equiv f(t)\sin\left(\frac{2\pi nt}{T}\right) = \left[\sum_{j=0}^{\infty}\frac{f^{(j)}(0)}{j!}t^j\right]\left[\sum_{k=0}^{\infty}\frac{(-1)^k}{(2k+1)!}\left(\frac{2\pi nt}{T}\right)^{2k+1}\right], \tag{A.22}$$

唯一的区别在于分母变为 $(2k)! \to (2k+1)!$，而 $\frac{2\pi nt}{T}$ 的幂变为 $2k+1$。接着 ——

$$C_{sin}(n,m) = \sum_{p=0}^{\infty}\left[\frac{(-1)^p 2^{(2p+2)} n^{2p+1}\pi^{2p+1}(t_2^{(2p+m+2)} - t_1^{(2p+m+2)})}{(2p+m+2)(2p+1)!T^{2p+2}}\right]$$
$$= \frac{4n\pi}{(m+2)T^2}\left[{}_pF_q\left(\frac{m+2}{2};\frac{3}{2},\frac{m+4}{2};-\frac{n^2\pi^2 t_2^2}{T^2}\right)t_2^{m+2}\right.$$
$$\left.-{}_pF_q\left(\frac{m+2}{2};\frac{3}{2},\frac{m+4}{2};-\frac{n^2\pi^2 t_1^2}{T^2}\right)t_1^{m+2}\right]. \tag{A.23}$$

B Gegenbauer 多项式与吉布斯现象

在附录中，我们简要讨论用于消除第5.4节吉布斯现象的 Gegenbauer 多项式的定义和性质。

次数为 n 的 Gegenbauer 多项式 $C_n^{\lambda}(x)$，对于 $\lambda \geq 0$，由以下积分定义

$$\int_{-1}^{1}(1-x^2)^{\lambda-\frac{1}{2}}C_k^{\lambda}(x)C_n^{\lambda}(x)dx = 0, \quad k \neq n. \tag{B.1}$$

其归一化方式如下

$$C_n^\lambda(1) = \frac{\Gamma(n+2\lambda)}{n!\Gamma(2\lambda)}. \tag{B.2}$$

注意，这些多项式并非正交归一， $C_n^\lambda(x)$ 的范数为

$$\int_{-1}^{1}(1-x^2)^{\lambda-\frac{1}{2}}(C_n^\lambda(x))^2dx = h_n^\lambda, \tag{B.3}$$

其中

$$h_n^\lambda = \pi^{\frac{1}{2}}C_n^\lambda(1)\frac{\Gamma(\lambda+\frac{1}{2})}{\Gamma(\lambda)(n+\lambda)}. \tag{B.4}$$

设函数 $f(x)$ 定义在区间 $[-1,1]$ (或子区间 $[a,b] \subset [-1,1]$) 上，相应的 Gegenbauer 系数 $\hat{} f^\lambda(l)$ 为

$$\hat{f}^\lambda(l) = \frac{1}{h_n^\lambda}\int_{-1}^{1}(1-x^2)^{\lambda-\frac{1}{2}}f(x)C_l^\lambda(x)dx, \tag{B.5}$$

那么，取前 $m+1$ 项的截断 Gegenbauer 展开为

$$f_m^\lambda(x) = \sum_{l=0}^{m}\hat{f}^\lambda(l)C_l^\lambda(x). \tag{B.6}$$

这里我们将简要说明，正如第5.4节所述，Gegenbauer 展开如何消除吉布斯现象。事实上，可以证明，目标函数 $f(x)$ 与 m 次 Gegenbauer 多项式之间具有指数收敛。我们仅简要概述证明思路，详细内容请参阅 [79]中的综述。

通过证明展开为 Gegenbauer 多项式后第 N 个傅里叶系数的误差可以呈指数级减小，即可确立指数收敛。设 $f_N^m(x)$ 表示将 $f_N(x)$ 展开为 m 次 Gegenbauer 多项式， $f^m(x)$ 表示将 $f(x)$ 展开为 m 次 Gegenbauer 多项式。那么有如下关系，其中用 $f_N^m(x)$ 近似 $f(x)$ 所产生的误差，显然不超过 $f(x)$ 与 $f^m(x)$ 之间的误差和 $f^m(x)$ 与 $f_N^m(x)$ 之间的误差

$$||f(x) - f_N^m(x)|| \leq ||f(x) - f^m(x)|| + ||f^m(x) - f_N^m(x)||. \tag{B.7}$$

在不等式右侧，我们将第一个范数称为正则化误差，而将第二个范数称为截断误差。注意，我们取范数来

为区间 $[-1,1]$ 上的最大范数。更准确地说，我们可以将截断误差写作

$$||f^m - f_N^m|| = \max_{-1\le x\le 1}\left|\sum_{k=0}^{m}(\hat{f}_k^\lambda - \hat{g}_k^\lambda)C_k^\lambda(x)\right|, \tag{B.8}$$

其中，我们将 $\hat{}\, f_k^\lambda$ 视为函数 $f(x)$ 的未知 Gegenbauer 系数。若 λ 和 m 均随 N 线性增长，则可以证明该误差呈指数级小。另一方面，正则化误差可写为

$$||f - f^m|| = \max_{-1\le 1}\left|f(x) - \sum_{k=0}^{m}\hat{f}_k^\lambda C_k^\lambda(x)\right|. \tag{B.9}$$

还可以证明，当 $\lambda = \gamma m$ 具有正的常数 γ 时，该误差呈指数级小。由于在给定条件下，正则化误差和截断误差都可以做到指数级小，Gegenbauer 展开因而达到一致指数精度，并消除了傅里叶数据中的吉布斯现象。

参考文献

[1] *T. Faulkner*、*T. Hartman*、*M. Headrick*、*M. Rangamani* 和 *B. Swingle*，《*Snowmass* 白皮书：量子场论和量子引力中的量子信息》，载于《*2022* 年 *Snowmass* 夏季研讨会》，第 *3* 卷，*2022*年，*2203.07117*。[2] *R. Bousso*、*X. Dong*、*N. Engelhardt*、*T. Faulkner*、*T. Hartman*、*S. H. Shenker* 等，《*Snowmass* 白皮书：黑洞的量子方面与时空的涌现》，*2201.03096*。[3] *A. R´enyi*，《熵和信息的度量》，载于《第四届伯克利数学统计与概率研讨会论文集》，第 *1* 卷：统计理论的贡献，第 *547–561* 页，加利福尼亚大学出版社，*1961*年，*http://projecteuclid.org/euclid.bsmsp/1200512181*。[4] *P. Calabrese* 和 *J. L. Cardy*，《纠缠熵与量子场论》，《统计力学杂志》*0406 (2004) P* 06002 [*hep-th/*0405152]。[5] *P. Calabrese*、*J. Cardy* 和 *E. Tonni*，《共形场论中两个不相交区间的纠缠熵》，《统计力学杂志》*0911 (2009) P*11001 [0905。2069]。[6] *P. Calabrese* 和 *J. Cardy*，《纠缠熵与共形场论》，《物理学杂志 *A*》*42 (2009)* 504005 [0905。4013]。[7] *P. Calabrese*、*J. Cardy* 和 *E. Tonni*，《共形场论中两个不相交区间的纠缠熵 *i*

n

共形场论 *II*，《统计力学杂志》**1101** (2011)P01021 [1011。5482]。

[8] *E. D'Hoker*、*X. Dong* 和 *C.-H. Wu*，《从 *R´enyi* 熵中提取冯·诺依曼熵的另一种方法》，*JHEP 01*（*2021*） 042 [2008。10076]。[9] *H. Yoon*、*J.-H. Sim* 和 *M. J. Han*，《通过无领域知识机器学习进行解析延拓》，《物理评论*B*》*98*（*2018*）*245101*。[10] *R. Fournier*、*L. Wang*、*O. V. Yazyev* 和 *Q. Wu*，《解析延拓问题的人工神经网络方法》，《物理评论快报》*124*（*2020*）。[11] *X. Xie*、*F. Bao*、*T. Maier* 和 *C. Webster*，《使用 *Adams-Bashforth ResNet* 对噪声数据进行解析延拓》，*arXiv* 预印本 *arXiv:1905.10430*（*2019*）。[12] *T. Song*、*R. Valenti* 和 *H. Lee*，《通过机器学习技术对自能进行解析延拓》，*arXiv* 预印本 *arXiv:2007.13610*（*2020*）。[13] *D. Huang* 和 *Y.-f. Yang*，《用于解析延拓的学习型优化器》，《物理评论*B*》*105*（*2022*）*075112*。[14] *K.-W. Sun* 和 *F. Wang*，《蒙特卡罗的神经网络解析延拓：通过统计误差进行改进》，*arXiv* 预印本 *arXiv:2302.11317*（*2023*）。[15] *G. Cybenko*，《通过 *S* 形函数的叠加进行逼近》，《控制、信号与系统数学》（*MCSS*）*2*（*1989*）*303*。[16] *T. O'Malley*、*E. Bursztein*、*J. Long*、*F. Chollet*、*H. Jin*、*L. Invernizzi* 等，《*KerasTuner*》。*https://github.com/keras-team/keras-tuner*，*2019*。[17] *J. R. McClean*、*J. Romero*、*R. Babbush* 和 *A. Aspuru-Guzik*，《变分混合量子-经典算法理论》，《新物理学杂志》*18*（*2016*）*023023*。[18] *J. Romero* 和 *A. Aspuru-Guzik*，《变分量子生成器：面向连续分布的生成对抗量子机器学习》，《先进量子技术》*4*（*2021*）*2000003*。[19] *K. Mitarai*、*M. Negoro*、*M. Kitagawa* 和 *K. Fujii*，《量子电路学习》，《物理评论*A*》*98*（*2018*）。[20] *E. Farhi* 和 *H. Neven*，《近期量子处理器上的量子神经网络分类》，*arXiv* 预印本 *arXiv:1802.06002*（*2018*）。[21] *J. R. McClean*、*S. Boixo*、*V. N. Smelyanskiy*、*R. Babbush* 和 *H. Neven*，《量子神经网络训练景观中的贫瘠高原》，《自然·通讯》*9*（*2018*）。[22] *M. Benedetti*、*E. Lloyd*、*S. Sack* 和 *M. Fiorentini*，《作为机器学习模型的参数化量子电路》，《量子科学与技术》*4*（*2019*）*043001*。[23] *A. P´erez-Salinas*、*D. L´opez-N´u˜nez*、*A. Garc´ıa-S´aez*、*P. Forn-D´ıaz* 和 *J. I. Latorre*，《单量子比特作为通用逼近器》，《物理评论*A*》*104*（*2021*）*012405*。

[24] T. Goto、Q. H. Tran 和 K.Nakajima：量子增强特征空间中量子机器学习模型的通用逼近性质，《物理评论快报》**127** (2021年) 090506。[25] B. Y. Gan、D. Leykam 和 D.G.Angelakis：量子机器学习模型的福克态增强表达性，*EPJ Quantum Technology* **9** (2022年) 16。[26] C.-C. Chen、M. Watabe、K.Shiba、M.Sogabe、K.Sakamoto 和 T. Sogabe：关于量子电路学习的可表达性与过拟合，*ACM Transactions on Quantum Computing* **2** (2021年)1。[27] S. Shin、Y. Teo 和 H.Jeong：量子监督学习的指数级数据编码，*arXiv* 预印本 *arXiv:2206.12105* (2022年) 。[28] M. C. Caro、E. Gil-Fuster、J.J.Meyer、J.Eisert 和 R.Sweke：参数化量子电路的依赖编码泛化界，《*Quantum*》**5** (2021年) 582。[29] F. J. Gil Vidal 和 D. O. Theis：参数化量子电路的输入冗余

，

《物理学前沿》*8 (2020) 297*。[30] *M. Schuld*、*R. Sweke* 和 *J. J. Meyer*，数据编码对变分量子机器学习模型表达能力的影响，《物理评论*A*》*103 (2021)*。[31] *R. D. Sorkin*，关于纠缠熵的 *1983* 年论文：“视界外真空的熵”，载于第十届广义相对论与引力国际会议，第 *2* 卷，第 *734–736* 页，*1984* 年，*1402.3589*。[32] *L. Bombelli*、*R. K. Koul*、*J. Lee* 和 *R. D. Sorkin*，黑洞熵的量子来源，*Phys. Rev. D 34 (1986) 373*。[33] *M. Srednicki*，熵与面积，物理评论快报 *71 (1993)* 666 [*hep-th/*9303048].[34] *C. Holzhey*、*F. Larsen* 和 *F. Wilczek*，共形场论中的几何熵与重整化熵，*Nucl. Phys. B424 (1994)* 443 [*hep-th/*9403108].[35] *E. Witten*，杰出研究成就 *APS* 奖章：关于量子场论纠缠性质的特邀文章，*Rev. Mod. Phys. 90 (2018) 045003*[1803.04993].[36] *H. Casini* 和 *M. Huerta*，自由量子场论中的纠缠熵，《物理学杂志 *A*》*42 (2009)* 504007 [0905.2562].[37] *S. N. Solodukhin*，纠缠熵、共形不变性与外在几何久久综合网

，

物理快报*B* **665** (2008年) 305 [0802.3117].[38] M. P. Hertzberg 和 F. Wilczek，纠缠熵的一些可计算贡献，物理评论快报 **106** (2011年) 050404 [1007.0993].[39] V. Rosenhaus 和 M. Smolkin，纠缠熵：微扰计算，*JHEP* **12** (2014年) 179 [1403.3733].

[40] *R. C. Myers* 和 *A. Sinha*，任意维度中的全息 *c* 定理，*JHEP 01 (2011*年*)* 125 [1011.5819].[41] *H. Liu* 和 *M. Mezei*，纠缠熵及自由度数量的改进，*JHEP 04 (2013*年*)* 162 [1202.2070].[42] *T. Faulkner*，*AdS/CFT* 中不相交区间的纠缠 *Rényi* 熵，*1303.7221*.[43] *R. P. Boas*，《整函数》。*Academic Press*，纽约，*1954*年。[44] *E. Witten*，*Rindler* 视界上的开弦，*JHEP 01 (2019*年*)* 126 [1810.11912].[45] *A. Dabholkar*，圆锥上的弦与黑洞熵，*Nucl. Phys. B 439 (1995) 650*[*hep-th/*9408098].[46] *E. Witten*，*Rindler* 视界上的开弦，*JHEP 01 (2019*年*)* 126 [1810.11912].[47] *C. A. Agon*、*M. Headrick*、*D. L. Jafferis* 和 *S. Kasko*，麦克斯韦场的圆盘纠缠熵，*Phys. Rev. D 89 (2014*年*)* 025018 [1310.4886].[48] *A. Lewkowycz* 和 *J. Maldacena*，广义引力熵，*JHEP 08 (2013*年*)*090 [1304.4926].[49] *C. Akers* 和 *G. Penington*，量子极值曲面方案的首阶修正，*JHEP 04 (2021*年*)* 062 [2008.03319].[50] *X. Dong*、*X.-L. Qi* 和 *M. Walter*，全息纠缠负性与副本对称性破缺，*JHEP 06 (2021*年*)* 024 [2101.11029].[51] *D. P. Kingma* 和 *J. Ba*，*Adam*：一种随机优化方法，*arXiv* 预印本 *arXiv:1412.6980 (2014*年*)*。[52] *M. Abadi*、*A. Agarwal*、*P. Barham*、*E. Brevdo*、*Z. Chen*、*C. Citro* 等，*TensorFlow*：异构系统上的大规模机器学习，*2015*年。[53] *F. Chollet* 等，*“Keras”*。*https://keras.io*，*2015*年。[54] *S. J. Reddi*、*S. Kale* 和 *S. Kumar*，论 *Adam* 及其改进方法的收敛性，*arXiv* 预印本 *arXiv:1904.09237 (2019*年*)*。[55] *T. Azeyanagi*、*T. Nishioka* 和 *T. Takayanagi*，通过 *AdS(2)/CFT(1)* 将近极值黑洞熵视为纠缠熵，*Phys. Rev. D77 (2008*年*) 064005*[0710.2956].[56] *D. Blanco*、*A. Garbarz* 和 *G. P´erez-Nadal*，环面上手征费米子的纠缠，*JHEP 09 (2019*年*)* 076 [1906.07057].[57] *P. Fries* 和 *I. A. Reyes*，环面上手征费米子的纠缠与相对熵，*Phys. Rev. D 100 (2019*年*)* 105015 [1906.02207]。

[58] B. Deconinck、M. Heil、A. Bobenko、M. Van Hoeij 和 M. Schmies，计算黎曼 *theta* 函数，《计算数学》**73**(2004) 1417 [nlin/0206009]

[59] *J. Frauendiener*、*C. Jaber* 和 *C. Klein*，多维 *theta* 函数的高效计算，《几何与物理学杂志》*141*（*2019*年） 147 [1701.07486].[60] *T. Barrella*、*X. Dong*、*S. A. Hartnoll* 和 *V. L. Martin*，超越经典引力的全息纠缠，*JHEP 09*（*2013*） 109 [1306.4682].[61] *E. Perlmutter*，关于 AdS_3/CFT_2，中 *Rényi* 熵的评论，*JHEP 05*（*2014*）*052*[1312.5740].[62] *D. E. Rumelhart*、*G. E. Hinton* 和 *R. J. Williams*，通过误差反向传播学习表征，《自然》*323*（*1986*）*533*。[63] *D. Rumelhart*、*G. Hinton*、*R. Williams* 和加州大学认知科学研究所，通过误差传播学习内部表征，认知科学研究所报告。加州大学圣迭戈分校，*1985*年。[64] *P. J. Werbos*，将反向传播推广并应用于循环天然气市场模型，《神经网络》*1*（*1988*）*339*。[65] *Y. Bengio*、*P. Simard* 和 *P. Frasconi*，使用梯度下降学习长期依赖关系颇具难度，《*IEEE* 神经网络汇刊》*5*（*1994*）*157*。[66] *S. Hochreiter* 和 *J. Schmidhuber*，长短期记忆，《神经计算》*9*（*1997*）*1735*。[67] *K. Cho*、*B. van Merri¨enboer*、*D. Bahdanau* 和 *Y. Bengio*，关于神经机器翻译的性质：编码器—解码器方法，.[68] *A. P´erez-Salinas*、*A. Cervera-Lierta*、*E. Gil-Fuster* 和 *J. I. Latorre*，用于通用量子分类器的数据重上传，《*Quantum*》*4*（*2020*年）*226*。[69] *P. Rebentrost*、*M. Mohseni* 和 *S. Lloyd*，用于大数据分类的量子支持向量机，《物理评论快报》*113*（*2014*）*130503*。[70] *J. R. McClean*、*S. Boixo*、*V. N. Smelyanskiy*、*R. Babbush* 和 *H. Neven*，量子神经网络训练景观中的贫瘠高原，《自然·通讯》*9*（*2018*年）*4812*。[71] *V. Bergholm* 等，*PennyLane*：量子—经典混合计算的自动微分，*1811.04968*。[72] *A. J. Jerri*，傅里叶分析、样条和小波逼近中的吉布斯现象，《数学及其应用》*446*（*1998*年）。[73] *D. Huybrechs*，非周期函数的傅里叶延拓，《*SIAM* 数值分析杂志》*47*（*2010*）*4326*。

[74] *D. Gottlieb*、*C.-W. Shu*、*A. Solomonoff* 和 *H. Vandeven*，论吉布斯现象 *I*：从非周期解析函数的傅里叶部分和中恢复指数精度，《计算与应用数学杂志》*43*（*1992*）*81*。[75] *D. Gottlieb* 和 *C.-W. Shu*，不连续波的傅里叶方法分辨率特性，《计算力学与工程方法》*116*（*1994*）*27*。[76] *D. Gottlieb* 和 *C.-W. Shu*，论吉布斯现象 *III*：从分段解析函数的谱部分和中恢复子区间内的指数精度，《*SIAM* 数值分析杂志》*33*（*1996*）*280*。[77] *D. Gottlieb* 和 *C.-W. Shu*，论吉布斯现象 *IV*：从分段解析函数的 *Gegenbauer* 部分和中恢复子区间内的指数精度，《计算数学》*64*（*1995*）*1081*。[78] *D. Gottlieb* 和 *C.-W. Shu*，论吉布斯现象 *V*：从分段解析函数的配置点值中恢复指数精度，《数值数学》*71*（*1995*）*511*。[79] *D. Gottlieb* 和 *C.-W. Shu*，论吉布斯现象及其消解，《*SIAM* 评论》*39*（*1997*）*644*。[80] *A. Gelb* 和 *S. Gottlieb*，傅里叶谱方法中吉布斯现象的消解，载于《吉布斯现象进展》。*Sampling Publishing*，纽约州波茨坦（*2007*）。[81] *S. Gottlieb*、*J.-H. Jung* 和 *S. Kim*，*David Gottlieb* 关于吉布斯现象消解的研究综述，《计算物理通讯》*9*（*2011*）*497*。[82] *S. Furukawa*、*V. Pasquier* 和 *J. Shiraishi*，一维 $c=1$ 临界相中的互信息与紧化半径，物理评论快报 *102*（*2009*）*170602*[0809.5113]。