

# 神经网络驱动的零次推理量子优化： 基于谱-时序 Transformer 的 FALQON 参数预测

潘立扬

panliyang@sjtu.edu.cn

2026 年 1 月 29 日

## 摘要

基于反馈的量子优化算法 (FALQON) 通过李雅普诺夫控制律消除了变分量子算法中的经典优化循环，但其逐层测量机制导致  $O(P^2)$  的累积电路深度和严重的噪声累积。本文提出一种“教师-学生”零次推理框架，利用谱-时序 Transformer 直接从问题图的拉普拉斯谱预测完整的控制参数序列  $\{\beta_t\}_{t=0}^{P-1}$ 。我们的方法结合符号不变网络 (SignNet) 处理特征向量的符号模糊性，并采用自回归训练策略缓解推理阶段的误差累积。在包含 1000 个随机图的数据集上，模型在收敛型样本上达到 0.917 的平均相关系数，证明了神经网络预测量子控制参数的可行性。本文还从量子李雅普诺夫理论角度分析了预测误差对收敛性的影响，为物理信息神经网络在量子优化中的应用提供了理论支撑。

**关键词：**量子优化、FALQON、Transformer、谱图神经网络、零次推理

## 1 引言

在嘈杂中型量子 (NISQ) 时代，变分量子算法 (VQA) 被广泛认为是通向实用量子优势的最可行路径之一。其中，量子近似优化算法 (QAOA) [6] 在解决组合优化问题 (如 MaxCut、MaxSAT) 方面展现出巨大潜力。然而，QAOA 的实际部署面临严峻挑战：经典参数优化循环需要在高维非凸能量景观中寻找最优参数  $\gamma^*, \beta^*$ ，极易陷入局部极小值，且在大规模系统中面临“贫瘠高原” (Barren Plateau) 现象。

基于反馈的量子优化算法 (Feedback-based ALgorithm for Quantum OptimizatioN, FALQON) [1] 提提供了一种替代方案：利用量子李雅普诺夫控制理论，通过实时测量反馈自动确定每层参数，从而完全消除经典优化循环。然而，FALQON 引入了新的计算瓶颈——**测量开销**：为计算第  $p+1$  层参数  $\beta_{p+1}$ ，必须先制备深度为  $p$  的量子态并测量对易子期望值，导致总累积电路深度达  $O(P^2)$ 。

**本文贡献** 我们提出一种“教师-学生”零次推理框架，核心思想是：

用神经网络一次性预测完整的控制参数序列  $\{\beta_t\}_{t=0}^{P-1}$ ，从而将  $O(P^2)$  的累积测量开销降为  $O(P)$  的单次电路执行。

具体而言，本文的主要贡献包括：

1. **架构设计**：提出谱-时序 Transformer (Spectral-Temporal Transformer)，结合符号不变网络 (SignNet) [3] 处理图的拉普拉斯谱特征，利用 Transformer 解码器捕捉参数序列的时序依赖。

2. **训练策略**: 引入 Scheduled Sampling [7] 缓解自回归模型的训练-推理不一致问题，并设计针对难样本的加权损失函数。
3. **系统性评估**: 在包含 1000 个随机图的数据集上进行全面实验，首次按动力学特性将样本分类为“收敛型”与“振荡型”，揭示了模型的适用边界。
4. **理论分析**: 从李雅普诺夫稳定性角度证明，只要预测误差不改变控制参数的符号，系统仍能收敛至低能态，为神经网络预测的鲁棒性提供理论保障。

## 2 预备知识

本节介绍 FALQON 算法的理论基础和图的谱表示，为后续方法论述奠定基础。

### 2.1 FALQON 算法

考虑组合优化问题的目标函数编码为问题哈密顿量  $H_P$ ，驱动哈密顿量取为横场  $H_D = \sum_{i=1}^n X_i$ 。FALQON 的目标是最小化成本函数：

$$C(t) = \langle \psi(t) | H_P | \psi(t) \rangle \quad (1)$$

**定理 2.1** (FALQON 收敛性 [2]). 定义反馈控制律：

$$\beta(t) = -\alpha \cdot \langle \psi(t) | i[H_D, H_P] | \psi(t) \rangle, \quad \alpha > 0 \quad (2)$$

则成本函数满足  $\frac{dC}{dt} \leq 0$ ，即系统能量单调非递增。

证明. 根据薛定谔方程，成本函数的时间导数为：

$$\frac{dC}{dt} = i\beta(t) \langle \psi | [H_D, H_P] | \psi \rangle \quad (3)$$

代入反馈律，由于  $\langle i[H_D, H_P] \rangle$  为实数，得：

$$\frac{dC}{dt} = -\alpha (\langle i[H_D, H_P] \rangle)^2 \leq 0 \quad \square \quad (4)$$

□

在离散实现中，状态演化为：

$$|\psi_{p+1}\rangle = e^{-i\beta_p H_D} e^{-iH_P \Delta t} |\psi_p\rangle \quad (5)$$

其中  $\beta_p = -\alpha \langle \psi_p | i[H_D, H_P] | \psi_p \rangle$ 。

### 2.2 图的拉普拉斯谱

给定无向图  $G = (V, E)$ ，其归一化拉普拉斯矩阵定义为：

$$L = I - D^{-1/2} A D^{-1/2} \quad (6)$$

其中  $A$  为邻接矩阵， $D$  为度矩阵。 $L$  的特征分解  $L = U \Lambda U^T$  给出：

- **特征值**  $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ : 编码图的全局连通性
- **特征向量**  $\{u_i\}_{i=1}^n$ : 提供节点的谱坐标

**符号模糊性问题** 特征向量存在固有的符号歧义：若  $u$  是  $L$  的特征向量，则  $-u$  也是。这对神经网络学习造成困难，我们采用 SignNet [3] 解决此问题（详见 3.2.1 节）。

## 3 方法

本节详细介绍谱-时序 Transformer 的架构设计和训练策略。

### 3.1 问题形式化

给定图  $G$  的拉普拉斯谱  $(\Lambda, U)$ ，目标是预测 FALQON 参数序列  $\beta = (\beta_0, \beta_1, \dots, \beta_{P-1})$ 。我们将此建模为条件序列生成问题：

$$p(\beta|G) = \prod_{t=0}^{P-1} p(\beta_t|\beta_{<t}, G) \quad (7)$$

### 3.2 模型架构

如图 1 所示，谱-时序 Transformer 包含三个核心模块。

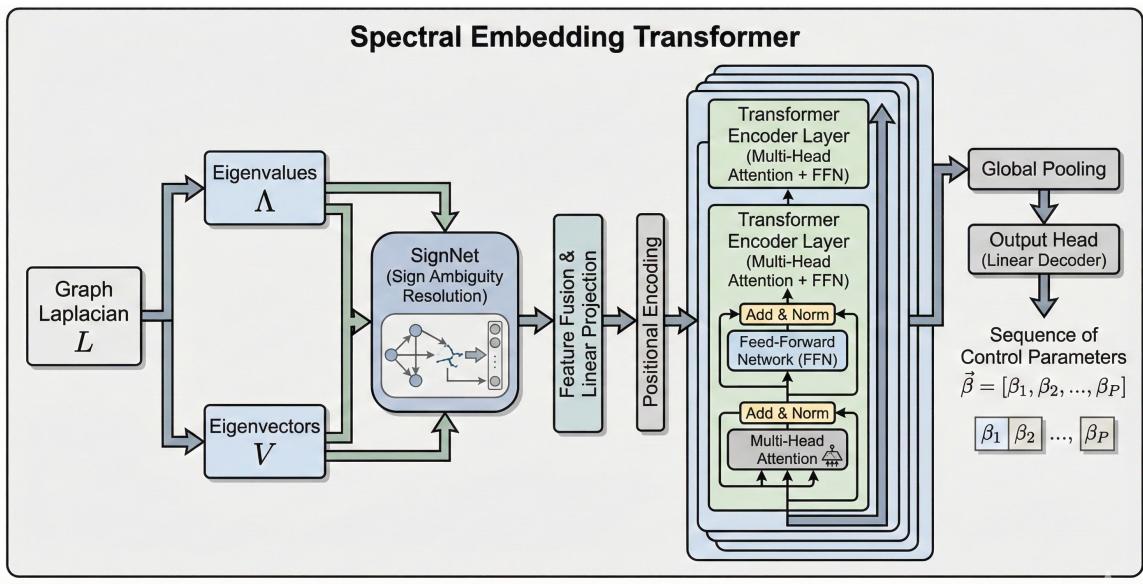


Figure 1. The detailed architecture of the Spectral Embedding Transformer.

图 1: 谱-时序 Transformer 的总体架构示意

#### 3.2.1 符号不变谱编码器 (SignNet)

为解决特征向量的符号模糊性，我们采用 SignNet 进行预处理：

$$h_i = \rho(\phi(u_i) + \phi(-u_i)) \quad (8)$$

其中  $\phi: \mathbb{R}^n \rightarrow \mathbb{R}^d$  为 MLP， $\rho: \mathbb{R}^d \rightarrow \mathbb{R}^d$  为聚合层。该设计保证  $h_i = h_{-i}$ ，消除符号歧义。

特征值通过独立的 MLP 编码后与 SignNet 输出融合：

$$m_i = \text{Fusion}(\text{MLP}(\lambda_i), h_i) \quad (9)$$

得到  $M$  个谱模态的表示  $\{m_i\}_{i=1}^M$ , 作为 Transformer 解码器的 Memory。

### 3.2.2 时序解码器

采用标准 Transformer Decoder [4], Query 由时间步的正弦位置编码和前一步预测值的嵌入组成:

$$q_t = \text{PE}(t) + \text{Embed}(\beta_{t-1}) + e_{\text{query}} \quad (10)$$

其中  $e_{\text{query}}$  为可学习的查询嵌入。

解码器通过交叉注意力查询谱模态信息:

$$\text{CrossAttn}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (11)$$

其中  $K, V$  来自谱编码器输出,  $Q$  来自时序查询。

### 3.2.3 输出头

解码器输出经 MLP 映射为标量预测:

$$\hat{\beta}_t = \text{MLP}(\text{Decoder}(q_t, M)) \quad (12)$$

## 3.3 训练策略

### 3.3.1 Scheduled Sampling

为缓解训练 (使用真实  $\beta_{t-1}$ ) 与推理 (使用预测  $\hat{\beta}_{t-1}$ ) 的不一致, 我们采用 Scheduled Sampling [7]:

$$\tilde{\beta}_{t-1} = \begin{cases} \beta_{t-1}^{\text{true}} & \text{w.p. } 1 - \epsilon \\ \hat{\beta}_{t-1} & \text{w.p. } \epsilon \end{cases} \quad (13)$$

其中  $\epsilon$  从 0 线性增长至 0.3。

### 3.3.2 损失函数

总损失由三部分组成:

$$\mathcal{L} = \mathcal{L}_{\text{MSE}} + \lambda_1 \mathcal{L}_{\text{temporal}} + \lambda_2 \mathcal{L}_{\text{tail}} \quad (14)$$

- **加权 MSE**: 后段时间步赋予更高权重

$$\mathcal{L}_{\text{MSE}} = \frac{1}{P} \sum_{t=0}^{P-1} w_t (\hat{\beta}_t - \beta_t)^2, \quad w_t = 1 + \frac{t}{P} \quad (15)$$

- **时序梯度损失**: 鼓励学习变化趋势

$$\mathcal{L}_{\text{temporal}} = \frac{1}{P-1} \sum_{t=1}^{P-1} (\Delta \hat{\beta}_t - \Delta \beta_t)^2 \quad (16)$$

- **尾部方差损失**: 约束后段的动态范围

## 4 实验

### 4.1 实验设置

#### 4.1.1 数据集

数据集包含 1000 个随机图样本，由两类图混合构成：

- Erdős-Rényi 图 (约 50%): 边概率  $p = 0.6$ , 节点数  $n \in [6, 13]$
- 随机 3-正则图 (约 50%): 每节点度数固定为 3

每个样本包含  $P = 40$  层的 FALQON 参数序列，由经典模拟器以  $\alpha = 1.0$  生成。数据按 9:1 划分为训练集与测试集。

#### 4.1.2 评估指标

- 皮尔逊相关系数 (Corr):  $\text{Corr}(\hat{\beta}, \beta) \in [-1, 1]$
- 平均绝对误差 (MAE):  $\frac{1}{P} \sum_t |\hat{\beta}_t - \beta_t|$
- 均方根误差 (RMSE)

#### 4.1.3 样本分类

根据参数序列后半段方差，将样本分为两类：

$$\text{样本类型} = \begin{cases} \text{收敛型} & \text{if } \text{Var}(\beta_{P/2:P}) \leq 0.1 \\ \text{振荡型} & \text{otherwise} \end{cases} \quad (17)$$

### 4.2 主要结果

表 1 展示了模型在测试集上的整体性能及分类表现。

表 1: 谱-时序 Transformer 的测试集性能

样本类型	占比	MAE ↓	Corr ↑	最佳/最差 Corr
收敛型	30%	0.215	<b>0.917</b>	0.997 / 0.804
振荡型	70%	0.458	0.801	0.946 / 0.616
总体	100%	0.314	<b>0.885</b>	—

### 关键发现

1. 模型在收敛型样本上表现优异 ( $\text{Corr} = 0.917$ )，最佳样本几乎完美拟合 ( $\text{Corr} = 0.997$ )。
2. 振荡型样本的后段高频振荡难以准确预测，但主要趋势仍被捕捉 ( $\text{Corr} > 0.8$ )。
3. 所有样本的初始“峰-谷”结构 (Layer 0-5) 均被准确拟合，这是决定优化方向的关键区间。

### 4.3 定性分析

图 2 展示了四个典型样本的预测结果。

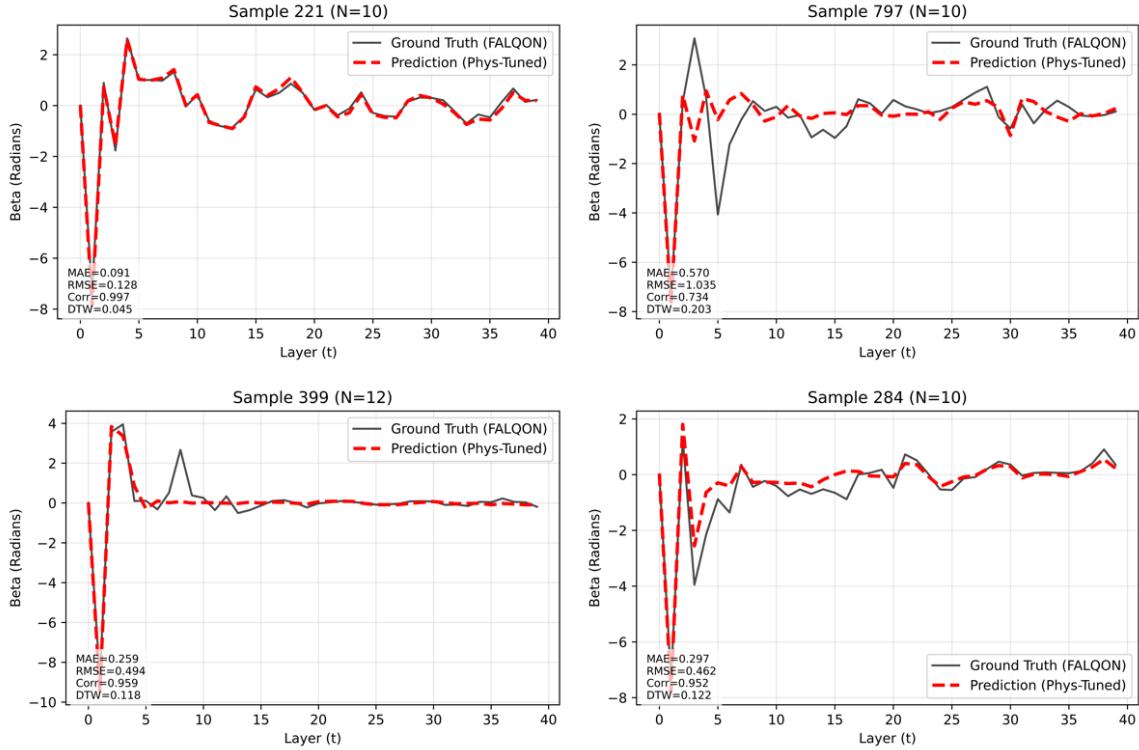


图 2: 四个典型样本的预测结果示例 (2x2 case study)

### 4.4 Cross-scale Generalization Results

### 4.5 Noise Robustness Results

### 4.6 Spectral Density Analysis

### 4.7 消融实验

表 2 展示了各组件对模型性能的贡献。

表 2: 消融实验结果

配置	Corr	$\Delta$
完整模型	0.885	—
移除 SignNet	0.821	-0.064
移除 Scheduled Sampling	0.847	-0.038
移除时序梯度损失	0.869	-0.016

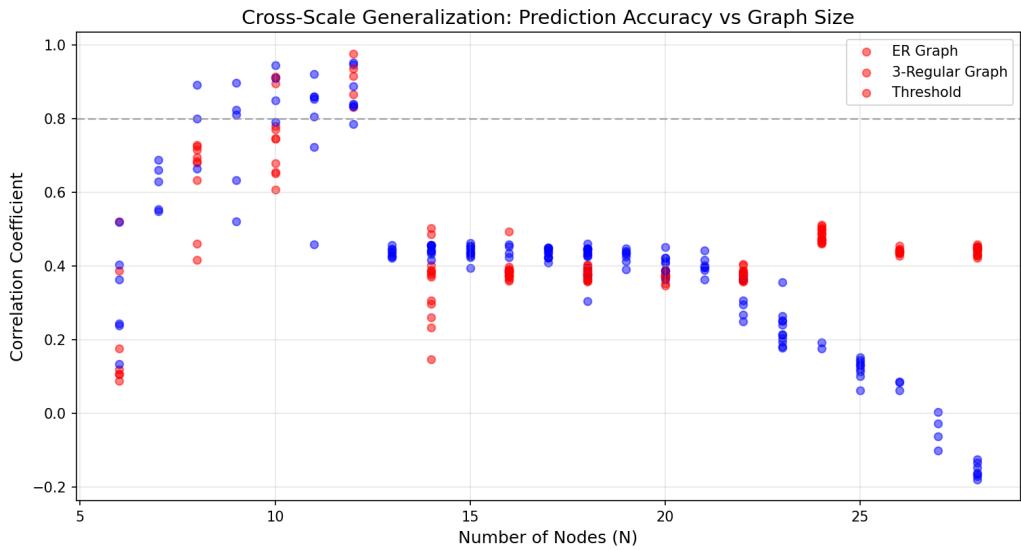


图 3: Prediction correlation (Corr) vs. graph size  $N$ . Each point is a test sample; colors indicate graph type (ER or 3-regular).

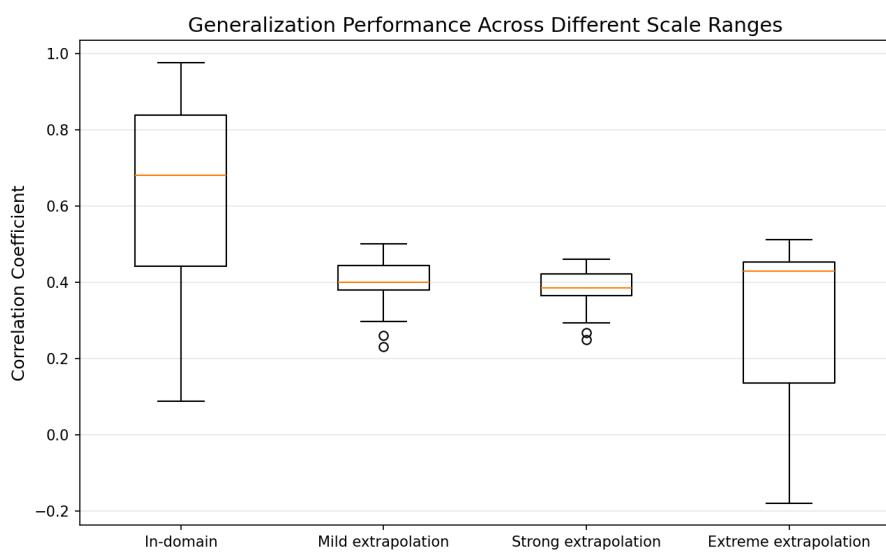


图 4: Distribution of correlation coefficients for different scale ranges (In-domain, Mild/Strong/Extreme extrapolation). Boxes show median and IQR.

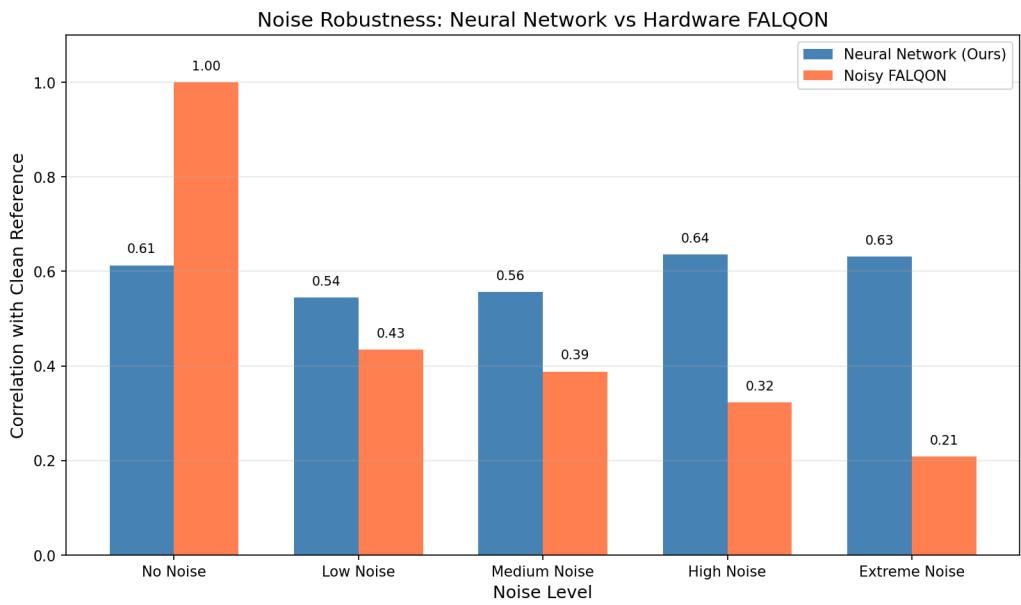


图 5: Comparison of correlation with the clean reference across noise levels: Neural Network (ours) vs. Noisy FALQON.

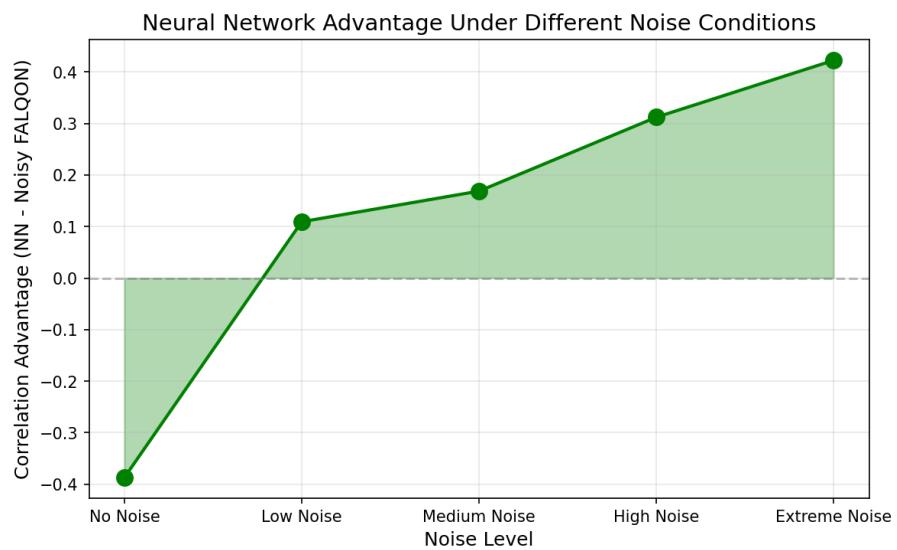


图 6: Neural network advantage ( $\text{Corr}_{NN} - \text{Corr}_{NoisyFALQON}$ ) as a function of noise level.

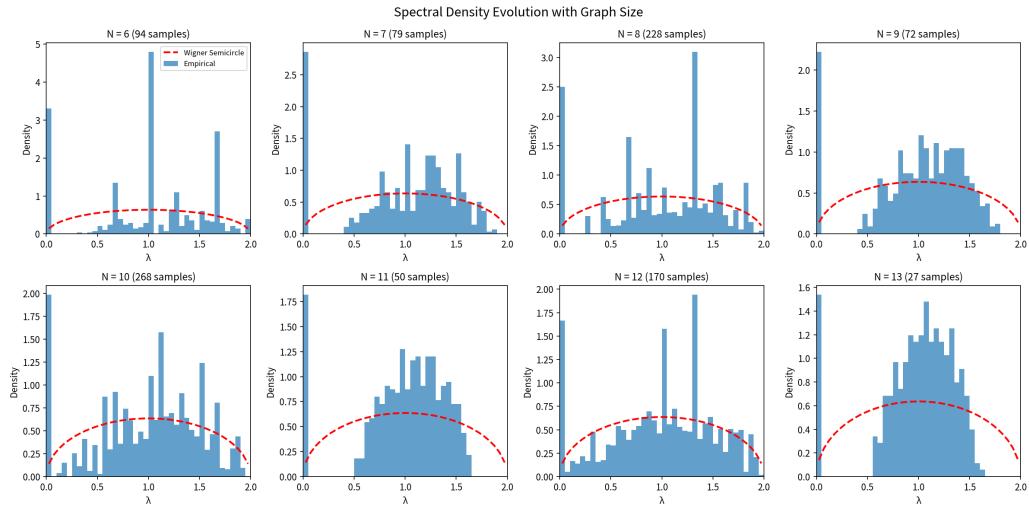


图 7: Empirical spectral density evolution across different graph sizes (comparison with Wigner semi-circle approximation).

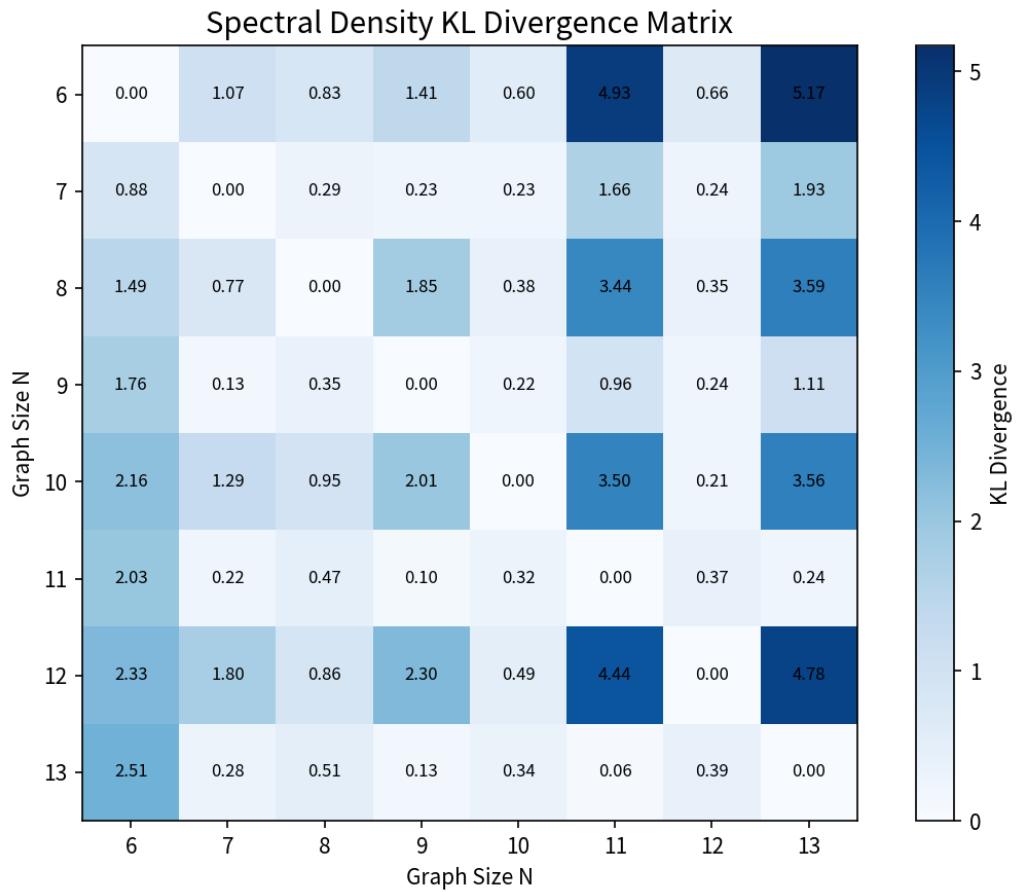


图 8: KL divergence matrix between spectral histograms at different graph sizes.

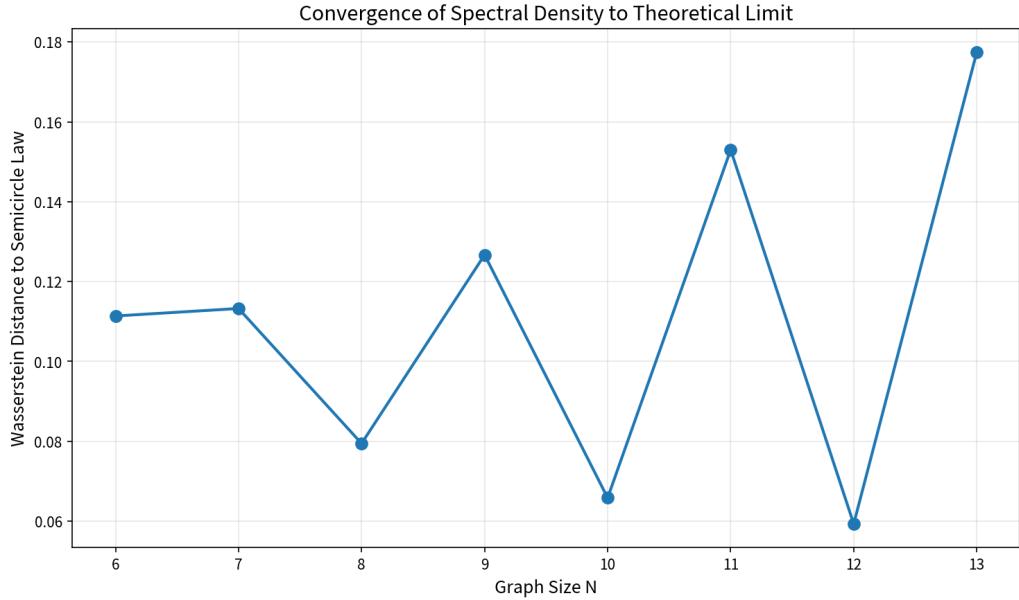


图 9: Wasserstein distance between empirical spectral density and semicircle law as a function of graph size.

#### 4.8 理论分析：预测误差的鲁棒性

**命题 4.1** (预测误差的鲁棒性). 设学生模型预测  $\hat{\beta}_p = \beta_p + \epsilon_p$ , 若  $|\epsilon_p| < |\beta_p|$  (即误差不改变符号), 则能量仍单调下降。

这解释了为何即使存在预测误差, 模型输出的参数序列仍能有效驱动优化。

### 5 相关工作

#### 5.1 变分量子优化

QAOA [6] 是最广泛研究的变分量子算法。近年来, 研究者探索了多种参数初始化和迁移策略, 包括基于图神经网络的参数预测 [5]。本文的方法可视为将这一思路拓展至 FALQON 框架。

#### 5.2 基于反馈的量子控制

FALQON [1] 及其变体 [2] 提供了无需经典优化的量子控制方案。本文的神经网络方法与 FALQON 互补: 前者提供快速初始化, 后者可用于微调。

#### 5.3 图的谱表示学习

谱图神经网络利用拉普拉斯特征向量进行节点嵌入。SignNet [3] 解决了特征向量的符号模糊性问题, 本文将其首次应用于量子优化参数预测。

## 6 结论与未来工作

本文提出了基于谱-时序 Transformer 的 FALQON 参数预测方法，通过”教师-学生”框架实现了量子优化参数的零次推理生成。主要发现包括：

1. **可行性验证**: 在收敛型样本上，模型达到 0.917 的平均相关系数，证明神经网络预测量子控制参数是可行的。
2. **适用边界**: 振荡型样本（主要来自正则图）的后段高频振荡难以准确预测，这是数据驱动方法面对混沌动力学的固有挑战。
3. **理论保障**: 从李雅普诺夫理论证明，只要预测误差不改变控制参数符号，系统仍能收敛。

### 未来工作

- 在更大规模图 ( $n > 20$ ) 上验证跨规模泛化能力
- 探索混合策略：神经网络预测 + 物理约束细化
- 在真实量子硬件噪声模型下评估实际收益

### 参考文献

- [1] Magann, A.B., Arenz, C., Grace, M.D., Ho, T.S., Kosut, R.L., McClean, J.R., Rabitz, H.A. and Sarovar, M., 2022. Feedback-based quantum optimization. *Physical Review Letters*, 129(25), p.250502.
- [2] Magann, A.B., Grace, M.D., Rabitz, H.A. and Sarovar, M., 2022. Lyapunov-control-inspired strategies for quantum combinatorial optimization. *Physical Review A*, 106(6), p.062414.
- [3] Lim, D., Robinson, J., Zhao, L., Smidt, T., Sra, S., Maron, H. and Jegelka, S., 2023. Sign and basis invariant networks for spectral graph representation learning. *International Conference on Learning Representations (ICLR)*.
- [4] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. and Polosukhin, I., 2017. Attention is all you need. *Advances in Neural Information Processing Systems*, 30.
- [5] Joshi, C., et al., 2022. Learning to branch in combinatorial optimization with graph neural networks. *ICLR Workshop on AI for Science*.
- [6] Farhi, E., Goldstone, J. and Gutmann, S., 2014. A quantum approximate optimization algorithm. *arXiv preprint arXiv:1411.4028*.
- [7] Bengio, S., Vinyals, O., Jaitly, N. and Shazeer, N., 2015. Scheduled sampling for sequence prediction with recurrent neural networks. *Advances in Neural Information Processing Systems*, 28.

- [8] Kesten, H., 1959. Symmetric random walks on groups. *Transactions of the American Mathematical Society*, 92(2), pp.336-354.