arXiv:2601.00284v1 [stat.ME] 1 Jan 2026

大域上函数型时间序列谱密度的深度学习估计

Neda Mohammadi SohamSarkar PiotrKokoszka*德克萨斯大学埃尔帕索分校 印度统计研究所 科罗拉多州立大学

2026年1月5日

摘要

我们推导了一种函数型时间序列谱密度估计量，该时间序列由多层感知机神经网络的输出构成。提出这一估计量的动机，是在非常大的网格上定义的函数型时间序列，其现有谱密度估计量的计算存在困难；这类时间序列例如见于气候计算模型和医学扫描。现有估计量使用表示为大型 $G \times G$ 矩阵的自协方差核，其中 G 是函数取值所在的网格点数。在许多近期应用中，函数定义在二维和三维域上，且 G 可能达到 $G \sim 10^5$的量级，这使得自协方差核的计算量极大，甚至无法完成。我们利用谱函数主成分理论推导出深度学习估计量，并证明在一般假设下，它是谱密度的通用逼近器。训练该估计量无需计算自协方差核，而且可以并行化，从而比现有方法更快地给出估计结果。我们通过模拟和一项针对功能性磁共振成像图像的应用验证了其性能。

关键词：函数型时间序列；多层感知机；空间域；谱密度。

1 引言

关于深度学习应用与改进的研究数量激增。绝大多数研究都聚焦于新的应用、架构、训练方式及类似问题，而这些问题最适合通过实验和大量数值研究来解决。深度学习的一般数学基础

*通讯地址：Piotr Kokoszka，美国科罗拉多州立大学，科罗拉多州柯林斯堡，邮编 80523-1877。电子邮箱：Piotr.Kokoszka@colostate.edu

早在 20 世纪 90 年代便已建立，其中包括各种通用逼近和算法收敛结果，但在数理统计框架下研究深度学习的成果仍相对较少。我们提出一种用于估计函数型时间序列谱密度的深度学习估计量，并在一般假设下论证其应用的合理性。

标量二阶平稳时间序列由均值和所有自协方差描述，而独立同分布样本的二阶参数则是均值和方差。对于标量平稳观测值 X_t，谱密度是自协方差序列的傅里叶变换，即定义为

$$f(\theta) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} c_h e^{-\mathrm{i}h\theta}, \quad \theta \in [-\pi, \pi],$$

其中 $c_h = \mathrm{Cov}(X_t, X_{t+h})$。谱密度包含了关于二阶依赖结构的所有信息，因为自协方差可以通过逆傅里叶变换得到。

这项研究的一个重要动机是，高分辨率功能性磁共振成像扫描是在包含 10^4-10^5 个点的三维网格上生成的，因此，即使只是估计协方差也很困难，因为这需要对 10^8-10^{10} 对点进行运算。我们的深度网络直接在网格值上训练，无需样本自协方差。计算机气候模型会在非常大的空间域上生成函数型时间序列，例如北美洲大陆及其周边海域。比如，平均温度表面可以按月建模，网格分辨率为几公里，对应约10^5 个点。

本文对深度学习与函数型时间序列分析的交叉领域作出了贡献。近年来，基于深度学习的推断研究日益增多，其中大多数针对函数的独立同分布样本。我们将在本节后文对这些研究进行回顾。本文的主要贡献是推导出用于估计函数型时间序列谱密度的神经网络。我们通过若干普适逼近结果，在与本文研究情境相关的度量下，论证了在自协方差算子衰减条件较弱时应用此类网络的合理性。我们的方法基于 Panaretos and Tavakoli（2013b）和 H¨ormann 等（2015）各自独立提出的函数型时间序列频域主成分分析。这些论文提出的估计方法需要计算许多滞后阶数下的自协方差函数；如果函数定义域由数十万个稠密网格点构成，这一计算可能并不可行。任何可分的 $L^2(\mathcal{Q})$ 空间都与 $L^2([0,1])$ 同构，但实践中，过大的定义域 $\mathcal{Q}$ 会使得即便是协方差核的估计也在计算上颇具挑战性，正如 Sarkar 和 Panaretos（2022）所解释的那样。我们所研究的数据是在 $\mathcal{Q}$ 上定义的函数，其取值在稠密网格上观测得到。本文的研究动机源自这样一些应用：其中可能有数万或数十万个网格点，因此，基于自协方差加权和估计谱密度在计算上具有挑战性，甚至目前尚不可行。我们将说明如何克服这一困难。本文的一项主要贡献是说明如何

通过构造合适的输出层，将函数型时间序列的频域主成分分析与深度学习相结合。在数值研究中，我们对深层采用特定架构，即 Sarkar 和 Panaretos（2022）提出的架构；不过，在不影响理论的前提下，这些架构可以有多种修改方式。我们基于线性滤波器和网络的傅里叶变换建立了一个理论框架，以证明我们的方法适用于非常一般的假设。除 Panaretos and Tavakoli（2013b）、H¨ormann 等（2015）和 Sarkar 和 Panaretos（2022）之外，其他密切相关的论文包括 Panaretos and Tavakoli（2013a），他们为函数型时间序列的谱分析推导了一个数学框架；以及 Kartsioukas 等（2023），他们关注希尔伯特空间中连续域平稳过程谱密度的估计。Kartsioukas 等（2023）针对网格上观测的数据推导了收敛速率和极限分布，这正是我们数值研究所考虑的情形。

本节最后，我们简要回顾深度学习在函数型数据分析中的应用研究进展。我们的目标是概述整体情况，而非罗列所有重要成果。Wang 等（2024）提供了一篇颇具参考价值的综述。目前的大多数应用都通过基展开将函数转换为向量。展开系数组成的向量可作为学习网络的输入。向量输出可以再转换回函数，也可以直接用于分类或聚类等其他目的。针对这一方法，人们已取得一些有益进展。Yao 等（2021）展示了如何构建一个微型神经网络，并将其嵌入更大的架构中，使其能够根据任务自适应地学习基。研究背景是标量响应 y_i 取决于函数型回归变量 $x_i(u)$，$u \in [0,1]$。模型基于独立同分布实现样本（x_i，y_i）进行训练。Thind 等（2023）研究了一个包含多个函数型协变量和标量协变量的模型。其他利用深度学习方法推进函数型回归的研究包括 Rao and Reimherr（2023a，2023b）以及 Wu等（2023）。Hong 等（2024）研究了如何根据带噪声且稀疏观测的实现 $x(u_{ij}) + \varepsilon_{ij}, j = 1, \ldots\ldots, n_i, i = 1, \ldots\ldots, n$重建潜在轨迹 $x_i(u)$，$u \in [0,1]$。他们开发了 GeLU 激活的变换器，并加入增强模块；这些模块本质上是定制的附加输出层，可生成可微函数 x_i。Wu等（2024）和 Wang and Cao（2024）研究了函数型数据的表示，包括多元函数。上述论文研究的是定义在紧区间上的独立同分布函数，其网格点数约为 10^2 。关于函数型时间序列的应用，Wang and Cao（2023）提出了一种能够改善预测效果的输出层，并展示了其在空气质量、电价和死亡率曲线预测中的应用。Ma 等（2024）在交通流量预测的背景下研究了函数型时间序列。

本文的结构安排如下。第 2 节介绍函数型时间序列的设定及其谱分析。在第 3 节中，我们推导出适用于逼近谱密度的函数型多层感知机，并给出一般性的普适逼近结果。我们运用第 3 节的理论，在第 4 节中推导出一种估计算法。第 5 节包含一项模拟研究，第 6 节则介绍其在功能性磁共振成像脑部扫描中的应用。补充材料包含证明和额外的模拟结果。

2 预备知识

本节介绍我们提出的方法和理论所依托的框架，并确定相关符号。

回顾一下，$\mathcal{Q}$ 是 $\mathbb{R}^d$ 的一个紧子集，并定义 $L^2(\mathcal{Q})$为 $\mathcal{Q}$ 上平方可积复值函数构成的希尔伯特空间，其内积为$\langle f,g\rangle = \int_{\mathcal{Q}} f(u)\overline{g(u)}du$，由此诱导的范数为 $\|f\| = \sqrt{\langle f,f\rangle}$。这里，$\overline{\cdot}$表示复共轭。对于 $f,g \in L^2(\mathcal{Q})$，算子 $f \otimes g$定义为（$f \otimes g$）$(h) = \langle h,g\rangle f$。算子 $f \otimes g$是一个核算子，其核为 $f \otimes g(u,v) = f(u)\overline{g(v)}$，即（$f \otimes g$）$(h)(u) = \int_{\mathcal{Q}} f(u)\overline{g(v)}h(v)dv$。我们用 $f \otimes g$同时表示该算子及其核。注意，$\|f \otimes g\|_{\mathcal{S}} = \|f \otimes g(\cdot,\cdot)\|_{L^2(\mathcal{Q}\times\mathcal{Q})}$，其中左侧表示该算子的希尔伯特–施密特范数，右侧表示其核的 $L^2(\mathcal{Q}\times\mathcal{Q})$范数。希尔伯特–施密特内积记为 $\langle\cdot,\cdot\rangle_{\mathcal{S}}$。$L^2(\mathcal{Q})$上的积分希尔伯特–施密特算子可以与其在 $L^2(\mathcal{Q}\times\mathcal{Q})$中的核相对应。因此，为简洁起见，提及核时，我们可以用 $\|\cdot\|_{\mathcal{S}}$ 代替$\|\cdot\|_{L^2(\mathcal{Q}\times\mathcal{Q})}$。Hsing和Eubank（2015）对希尔伯特空间中算子理论作了详尽阐述。

假设 $\{X_t\}_{t\in\mathbb{Z}}$ 是取值于 $L^2(\mathcal{Q})$ 的弱平稳过程，即 $\mathbb{E}X_t = \mu$ 且 $\mathbb{E}[(X_{t+h}-\mu)\otimes(X_t-\mu)] = C_h$，对所有 $t,\ h \in \mathbb{Z}$ 均成立。在大多数实际数据情形中，随机场取实值，这是我们采用的假设。不过，由于我们是在频域中进行研究，采用复向量空间会更有利。下面的假设 2.1 保证谱密度算子的存在。

Assumption 2.1 过程 $\{X_t\}_{t\in\mathbb{Z}}$ 在空间 $L^2(\mathcal{Q})$ 中弱平稳，均值为零，且其自协方差算子满足 $\sum_{h\in\mathbb{Z}}\|C_h\|_{\mathcal{S}} < \infty$。（我们假设对每个 $u \in \mathcal{Q}$，$X_t(u)$ 都是实数。）

在实践中，我们通过减去样本均值来对数据进行中心化。众所周知，在相当一般的假设下，样本均值以 $\sqrt{N}$ 的速率收敛到真实均值（N 为样本量），例如参见 Horv´ath 等（2013）。因此，其估计的影响渐近可忽略不计；为避免偏离主要贡献，本文不予考虑。

在假设 2.1 下，我们将频率$\theta \in [-\pi,\pi]$ 处的谱密度算子 $F^X(\theta)$ 定义为

$$F^X(\theta) = \frac{1}{2\pi}\sum_{h\in\mathbb{Z}} C_h \exp(-\mathrm{i}h\theta),$$

其中，$\mathrm{i} = \sqrt{-1}$ 是虚数单位，且该收敛在希尔伯特–施密特范数下成立。对于每个 $\theta \in [-\pi,\pi]$，谱密度算子 $F^X(\theta)$ 都是半正定的希尔伯特–施密特自伴算子。对于每个 h，交叉协方差算子 C_h 是一个核为 $c_h(u,\ v) = \mathbb{E}[X_h(u)\overline{X_0(v)}]$,的积分算子，即 $C_h(f)(u) = \int_{\mathcal{Q}} c_h(u,\ v)f(v)dv$，对所有 $f \in L^2(\mathcal{Q})$ 都成立。这意味着，对于每个 $\theta \in [-\pi,\pi]$，谱密度算子

$F^X(\theta)$ 是一个核为以下形式的积分算子

$$f^X(\theta)(u,v) = \frac{1}{2\pi}\sum_{h\in\mathbb{Z}} c_h(u,v)\exp(-\mathrm{i}h\theta), \quad u,v\in\mathcal{Q}, \tag{2.1}$$

其中收敛在 $L^2(\mathcal{Q}\times\mathcal{Q})$ 中成立。与相干性分析中一样，我们可以写成

$$f^X(\theta)(u,v) = p^X(\theta)(u,v) - \mathrm{i}q^X(\theta)(u,v),$$

其中

$$p^X(\theta)(u,v) = \frac{1}{2\pi}\sum_{h\in\mathbb{Z}} c_h(u,v)\cos(h\theta) \tag{2.2}$$

是余谱，而

$$q^X(\theta)(u,v) = \frac{1}{2\pi}\sum_{h\in\mathbb{Z}} c_h(u,v)\sin(h\theta) \tag{2.3}$$

是四谱。注意，根据我们的假设，$X_t(u)$ 为实值，因此余谱和四谱都是实值函数。由此，复值核函数 f^X 的估计可归结为对两个实值核的估计。

现在我们给出函数型时间序列的频域主成分分析的关键结果。由于对于每个 $\theta\in[-\pi,\pi]$，$F^X(\theta)$ 都是半正定、希尔伯特–施密特且自伴的，因此有如下谱分解

$$f^X(\theta)(u,v) = \sum_{m\geq 1}\lambda_m(\theta)\varphi_m^\dagger(\theta)(u)\overline{\varphi}_m^\dagger(\theta)(v), \quad u,v\in\mathcal{Q}, \tag{2.4}$$

其中特征值 $\lambda_m(\theta)$ 非负，特征函数 $\varphi_m^\dagger(\theta)$ 在 $L^2(\mathcal{Q})$ 中构成一组标准正交系。各对 $(\lambda_m(\theta)\ \varphi_m^\dagger(\theta))$ 按特征值递减的顺序排列。

考虑以下函数

$$\varphi_{m,h}(u) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\exp(-\mathrm{i}h\theta)\varphi_m^\dagger(\theta)(u)d\theta, \quad u\in\mathcal{Q}, \quad h\in\mathbb{Z}.$$

于是，我们有

$$\lim_{L\to\infty}\int_{-\pi}^{\pi}\left\|\sum_{h=-L}^{L}\exp(\mathrm{i}h\theta)\varphi_{m,h} - \varphi_m^\dagger(\theta)\right\|^2_{L^2(\mathcal{Q})} d\theta = 0, \quad m\geq 1, \tag{2.5}$$

参见 H¨ormann 等人(2015) 的第 3.3 小节。随机场 $\{X_t\}$ 可通过以下方式恢复

$$X_t = \sum_{m\geq 1}\sum_{h\in\mathbb{Z}} Y_{m,t+h}\varphi_{m,h}, \tag{2.6}$$

其中 $Y_{m,\ t} = \sum_{h\in\mathbb{Z}}\langle X_{t-h},\ \varphi_{m,\ h}\rangle$，且该收敛在均方意义下成立。当 $m\neq m'$ 时，序列 $\{Y_{m,t}\}$ 和 $\{Y_{m',t}\}$ 在所有滞后阶数上均不相关。

我们的方法采用适当构造的深度网络，进行与 (2.5) 类似的逼近。因此，下面的定义起着关键作用。

Definition 2.1 （傅里叶可变换性）如果存在一个族 $\{g^{\dagger}(\theta)\}_{\theta} \subset L^2(\mathcal{Q})$，使得我们称序列 $\{g_h\}_h \subset L^2(\mathcal{Q})$ 具有傅里叶可变换性

$$\lim_{L\to\infty} \int_{-\pi}^{\pi} \left\| \sum_{h=-L}^{L} \exp(\mathrm{i}h\theta) g_h - g^{\dagger}(\theta) \right\|_{L^2(\mathcal{Q})}^2 d\theta = 0. \tag{2.7}$$

由于 $g^{\dagger}(\theta) \in L^2(\mathcal{Q})$，$\int_{-\pi}^{\pi} \left\|g^{\dagger}(\theta)\right\|_{L^2(\mathcal{Q})}^2 d\theta = \int_{-\pi}^{\pi} \int_{\mathcal{Q}} |g^{\dagger}(\theta)(u)|^2 du d\theta$ 是$L^2([-\pi,\pi]\times\mathcal{Q})$中的平方范数，而该空间是完备空间，因此我们得出 $\int_{-\pi}^{\pi} \left\|g^{\dagger}(\theta)\right\|_{L^2(\mathcal{Q})}^2 d\theta < \infty$。

我们将使用由 m 和 θ 定义的函数类 $\mathcal{A}$，其定义如下

$$\mathcal{A} = \{\eta_{\cdot}(\cdot) : \mathbb{D}\times[-\pi,\pi] \to [0,\infty), \quad \text{for some } \mathbb{D}\subseteq\mathbb{N},\ \sup_{m,\theta} \eta_m(\theta) < \infty\}. \tag{2.8}$$

等价地，$\mathcal{A} = \bigcup_{\mathbb{D}\subseteq\mathbb{N}} \mathcal{A}_{\mathbb{D}}$，其中 $\mathcal{A}_{\mathbb{D}} = \{\eta_{\cdot}(\cdot) : \mathbb{D}\times[-\pi,\pi] \to [0,\infty),\ \sup_{m,\theta}\eta_m(\theta) < \infty\}$, $\mathbb{D}\subseteq\mathbb{N}$。$\mathcal{A}$ 中的每个函数均为非负函数，且都有上界。下述引理可由 H¨ ormann 等人（2015）的命题 7 直接推出。

Lemma 2.1 假设假设 *2.1*成立。那么，对每个 m，函数 $\theta \mapsto \lambda_m(\theta)$都是连续的。特别地，$\lambda_{\cdot}(\cdot) \in \mathcal{A}_{\mathbb{N}} \subset \mathcal{A}$，因为 $\Lambda^* := \sup_{m,\theta} \lambda_m(\theta) = \sup_{\theta} \lambda_1(\theta) < \infty$。

本节最后列出本文中经常用到的函数，以及它们的定义域和值域。最下面三行中的函数将在第 3 节介绍。我们用哥特字体表示网络。

$$\begin{aligned}
f^X &: [-\pi,\pi] \to L^2(\mathcal{Q}\times\mathcal{Q}), \quad \text{or equivalently} \quad F^X : [-\pi,\pi] \to \mathcal{S};\\
\mathfrak{f} &: [-\pi,\pi] \to L^2(\mathcal{Q}\times\mathcal{Q}), \quad \text{or equivalently} \quad \mathfrak{f} : [-\pi,\pi] \to \mathcal{S};\\
\varphi^{\dagger}, \mathfrak{g}^{\ddagger} &: [-\pi,\pi] \to L^2(\mathcal{Q});\\
\varphi, \mathfrak{g} &: \mathcal{Q} \to \mathbb{C} \quad \text{such that } \varphi, \mathfrak{g} \in L^2(\mathcal{Q}).
\end{aligned}$$

我们用 f^X表示核，用 F^X表示相应的算子。与此相反，我们用相同的记号 $\mathfrak{f}$ 表示算子及其核。回顾一下，空间$L^2(\mathcal{Q})$由复值函数构成。

3 深度网络的谱密度逼近

本节将说明如何通过深度网络逼近谱密度算子的数学机制。我们将借助与神经网络通用逼近结果思路相近的定理来阐述。首先介绍浅层网络，它们将作为输出层的组成部分。深层则由标准多层感知机构成，并且可以采用共享参数。我们所考虑的基本网络与 Sarkar 和 Panaretos（2022）提出的网络类似；其关键进展在于

展示如何对这些网络进行变换并将其组合起来，从而构造谱密度算子的逼近。

考虑定义在 $\mathcal{Q}$ 上的以下复值浅层神经网络类:

$$\begin{aligned}\mathcal{C}^{\text{sh}} =&\{\mathfrak{g}(\cdot)=\sum_{r=1}^{R} c_r\sigma(w_r^\top\cdot+b_r),\quad R\in\mathbb{N}, w_r\in\mathbb{R}^d, b_r\in\mathbb{R}, c_r\in\mathbb{C}\}\\ =:&\{\mathfrak{g}=\sum_{r=1}^{R} g_r,\quad R\in\mathbb{N}\}.\end{aligned}$$

激活函数 $\sigma:\mathbb{R}\to\mathbb{R}$ 始终逐元素应用。给定超参数 $R\in\mathbb{N}$ 和 $\sigma(\cdot)$，这些网络中需要学习的参数为$w_r\in\mathbb{R}^d, b_r\in\mathbb{R}, c_r\in\mathbb{C}$。

接下来，我们介绍深层共享神经网络类。对于正整数$J, d_1, d_2, \ldots, d_J$、矩阵 $W_1\in\mathbb{R}^{d_1\times d}$, $W_2\in\mathbb{R}^{d_2\times d_1},\ldots,W_J\in\mathbb{R}^{d_J\times d_{J-1}}$、向量 $B_1\in\mathbb{R}^{d_1},\ldots,B_J\in\mathbb{R}^{d_J}, w_r\in\mathbb{R}^{d_J}$、标量 $b_r\in\mathbb{R}$ 和 $c_r\in\mathbb{C}$，定义

$$\begin{aligned}u_1 =&\sigma(W_1u+B_1),\quad u\in\mathcal{Q},\\ u_{j+1} =&\sigma(W_{j+1}u_j+B_{j+1}),\quad j=1,2,\ldots,J-1,\\ g_r(u) =&c_r\sigma(w_r^\top u_J+b_r),\quad r=1,\ldots R.\end{aligned}\tag{3.1}$$

在这种架构中，我们有深度为 J、宽度为 $\max\{d_1,\ldots d_J\}$ 的神经网络 $g_r(\cdot)$。我们假设前 $J-1$ 层由 $g_r(\cdot)$、$r=1,\ldots R$ 共享，只有最后一层随 r变化。这定义了定义在 $\mathcal{Q}$ 上的以下深层共享神经网络类:

$$\mathcal{C}^{\text{ds}} =\{\mathfrak{g}=\sum_{r=1}^{R} g_r,\quad R\in\mathbb{N},\ \text{the } g_r \text{ are of the form of (3.1)}\}.$$

放宽共享权重和偏置的假设，我们定义通用深层神经网络

$$\begin{aligned}u_{1,r} =&\sigma(W_{1,r}u+B_{1,r}),\quad u\in\mathcal{Q},\\ u_{j+1,r} =&\sigma(W_{j+1,r}u_{j,r}+B_{j+1,r}),\quad j=1,2,\ldots,J-1,\\ g_r(u) =&c_r\sigma(w_r^\top u_{J,r}+b_r),\quad r=1,\ldots R.\end{aligned}\tag{3.2}$$

在这一构造中，除了允许最后一层的参数变化外，我们还允许其他所有隐藏层的矩阵和向量参数随 r 变化。这定义了以下定义在 $\mathcal{Q}$ 上的深度神经网络类:

$$\mathcal{C}^{\text{d}} =\{\mathfrak{g}=\sum_{r=1}^{R} g_r,\quad R\in\mathbb{N},\ \text{the } g_r \text{ are in the form of (3.2)}\}.$$

为简洁起见，我们用记号 $\mathcal{C}^{\text{nn}}$ 表示类 $\mathcal{C}^{\text{sh}}$、 $\mathcal{C}^{\text{ds}}$ 或 $\mathcal{C}^{\text{d}}$ 中的任意一个。 $\mathcal{C}^{\text{nn}}$ 的一般元素记为 $\mathfrak{g}$。我们强调， $\mathcal{C}^{\text{nn}}$ 中的每个元素都是 $L^2(\mathcal{Q})$ 中具有特定形式的函数，这种形式称为多层神经网络。

我们现在定义这类最终为零的神经网络序列 :

$$\mathcal{C} = \{\{\mathfrak{g}_h\}_{h\in\mathbb{Z}} : \ \mathfrak{g}_h \in \mathcal{C}^{\text{nn}} \text{ if } |h| \le L, \ \mathfrak{g}_h = 0 \text{ if } |h| > L, \text{ for some } L \in \mathbb{N}\}, \tag{3.3}$$

也就是说， $\mathcal{C}$ 包含只有有限个非零元素的复值神经网络序列。特别地，这些序列满足定义 2.1，即显然可进行傅里叶变换。下文中，我们使用记号 $\mathfrak{g}^{\ddagger}$ 而非 $\mathfrak{g}^{\dagger}$，以强调对于网络而言无需取极限。因此，我们定义类 $\mathcal{D}$，它包含 $\mathcal{C}$中序列的傅里叶变换:

$$\mathcal{D} = \{\mathfrak{g}^{\ddagger} : [-\pi,\pi] \to L^2(\mathcal{Q}), \ \mathfrak{g}^{\ddagger} : \theta \mapsto \sum_{h\in\mathbb{Z}} \exp(\mathrm{i}h\theta)\mathfrak{g}_h, \text{ for some } \{\mathfrak{g}_h\} \in \mathcal{C}\}. \tag{3.4}$$

最后，回顾类 $\mathcal{A}$ 在 (2.8) 中定义，并定义类 $\mathcal{E}$:

$$\begin{aligned}\mathcal{E} = &\{\mathfrak{f} : [-\pi,\pi] \to \mathcal{S}, \ \mathfrak{f} : \theta \mapsto \sum_{m=1}^{M} \eta_m(\theta)\mathfrak{g}_m^{\ddagger}(\theta) \otimes \mathfrak{g}_m^{\ddagger}(\theta), \\ &\text{for some } M \in \mathbb{N}, \eta_\cdot(\cdot) \in \mathcal{A}, \ \mathfrak{g}_m^{\ddagger} \in \mathcal{D}, m = 1,\dots M\}.\end{aligned} \tag{3.5}$$

注意，若 $\mathfrak{f} \in \mathcal{E}$，则对于每个 $\theta \in [-\pi,\pi]$， $\mathfrak{f}(\theta)$都是非负的自伴算子，这与频率 θ处的谱密度 $F^X(\theta)$一样。我们将在第 4 节中证明，无需估计 (2.1) 中出现的自协方差$c_h(u, v)$，即可训练网络 $\mathfrak{f}(\theta)$。下面的定理 3.1 指出，在以下一般假设下，每个谱密度都可以由 $\mathcal{E}$ 中的神经网络以积分希尔伯特-施密特范数进行逼近。

Assumption 3.1 激活函数 $\sigma(\cdot)$ 满足：对于任意 $\epsilon > 0$ 和任意 $\varphi \in L^2(\mathcal{Q})$，存在一个网络 $\mathfrak{g}$，属于 $\mathcal{C}^{\text{nn}}$，使得 $\|\varphi - \mathfrak{g}\|_{L^2(\mathcal{Q})} < \epsilon$。

我们在补充材料的 A 节中验证，所有实际使用的激活函数都满足假设 3.1。

Theorem 3.1 假设 *2.1* 和假设 *3.1* 成立。则对于任意 $\epsilon > 0$，存在$\mathfrak{f} \in \mathcal{E}$，使得

$$\int_{-\pi}^{\pi} \|f^X(\theta) - \mathfrak{f}(\theta)\|_{\mathcal{S}} d\theta < \epsilon.$$

定理 3.1 不能直接用于构造深度学习估计量。因此，我们对定理 3.1 中表述的通用逼近进行修改，并通过一列网络的傅里叶变换陈述一个类似的结果。这为

第 4 节中估计器的构造铺平了道路。根据 (2.6)，对于 M，$L \in \mathbb{N}$，考虑以下随机场平稳序列

$$\widetilde{\mathfrak{X}}_t = \sum_{h=-L}^{L} \sum_{m=1}^{M} \xi_{m,t+h} \mathfrak{g}_{m,h}, \tag{3.6}$$

其中，对于每一对 $(m,\ h)$，$\mathfrak{g}_{m,\ h} \in \mathcal{C}^{\mathrm{nn}}$ 以及 $\{\xi_t = (\xi_{1,\ t}、\ \ldots\ldots、\ \xi_{M,\ t})\}$ 都是一个 M 维均值为零的平稳随机过程，且在所有滞后阶数上各分量彼此不相关。特别地，谱密度算子 $F^{\xi}(\theta)$ 是对角 $M \times M$ 矩阵，且其元素非负。下文的定理 3.2 指出，式 (3.5) 中定义的类 $\mathcal{E}$ 的元素可视为式 (3.6) 中定义的平稳序列的谱密度核。式 (3.6) 形式的序列可视为带有额外输出层的网络，该输出层由 $\xi_{m,\ t+h}$ 参数化。为便于参考，我们提出如下假设。

~Assumption 3.2 神经随机场 $\mathfrak{X}_t$ 具有式 (3.6) 的形式，其中参数为某个 M，$L \in \mathbb{N}$；对于每一对 $(m,\ h)$，$\mathfrak{g}_{m,\ h} \in \mathcal{C}^{\mathrm{nn}}$，且序列 $\{\xi_t = (\xi_{1,\ t}、\ \ldots\ldots、\ \xi_{M,\ t})\}$ 是一个 M 维均值为零的平稳随机过程，在所有滞后阶数上各分量彼此不相关，并且其自协方差矩阵绝对可和。特别地，长期方差矩阵 F^{ξ} 是一个对角 $M \times M$ 矩阵，且其元素非负。

~Theorem 3.2 设 $\{\mathfrak{X}_t\}$ 是满足假设 *3.2* 的随机场序列。则，~$\{\mathfrak{X}_t\}$ 是平稳的，其谱密度核可表示为

$$F^{\widetilde{\mathfrak{X}}}(\theta) = \sum_{m=1}^{M} F^{\xi}_{m,m}(\theta) \mathfrak{g}^{\ddagger}_m(\theta) \otimes \mathfrak{g}^{\ddagger}_m(\theta), \tag{3.7}$$

其中 $F^{\xi}_{m,\ m}(\theta)$ 是对角矩阵 $F^{\xi}(\theta)$ 的 (m,m) 元素，而 $\mathfrak{g}^{\ddagger}_m$ 是有限级数 $\{\mathfrak{g}_{m,h}\}_{-L \le h \le L} \in \mathcal{C}$ 的傅里叶变换，其中 $m = 1$，……，M。此外，类 $\mathcal{E}$ 可表示为

$$\mathcal{E} = \{\{f^{\widetilde{\mathfrak{X}}}(\theta)\}_{\theta \in [-\pi,\pi]},\ \widetilde{\mathfrak{X}}_t = \sum_{h=-L}^{L} \sum_{m=1}^{M} \xi_{m,t+h} \mathfrak{g}_{m,h},\ as\ in\ Assumption\ 3.2\}. \tag{3.8}$$

Remark 3.1 由于随机场 X_t 是实值的，每个特征函数 $\varphi^{\dagger}_m(\theta)$ 都是厄米的，即 $\varphi^{\dagger}_m(\theta) = \varphi^{\dagger}_m(-\theta)$。这意味着 $\varphi_{m,\ \overline{h}} = \varphi_{m,\ h}$，进而意味着式 (2.6) 中出现的得分 $Y_{m,\ t}$ 是实数。因此，式 (3.6) 中定义的网络 $\mathfrak{g}_{m,\ h}$ 是实值的。这将类 $\mathcal{D}$ 限定为厄米函数，并将 $\xi_{m,\ t}$ 限定为实数。

现在，我们讨论如何用网络场 $\mathfrak{X}_t$ 的自协方差的有限加权~和来逼近谱密度核 f^X。我们考虑三种特定的核，其中一些核要求加强假设 2.1 中的自协方差可和性条件。

可以提出抽象假设，但采用便于直接应用且能使证明更加清晰的具体假设会更有用。

我们考虑三种常用的核：截断核

$$\omega(s) = \begin{cases} 1, & |s| \le 1, \\ 0, & |s| > 1, \end{cases}$$

巴特利特核

$$\omega(s) = \begin{cases} 1 - |s|, & 0 \le |s| \le 1, \\ 0, & |s| > 1, \end{cases}$$

和帕曾核

$$\omega(s) = \begin{cases} 1 - 6|s|^2 + 6|s|^3, & |s| < \frac{1}{2}, \\ 2(1 - |s|)^3, & \frac{1}{2} \le |s| \le 1, \\ 0, & |s| > 1. \end{cases}$$

我们采用以下假设。

Assumption 3.3 假设 $\omega(\cdot)$ 是截断核、巴特利特核或帕曾核。如果使用巴特利特核或帕曾核，假设对于某个 $0 < \alpha \le 1$，$\sum_{h\in\mathbb{Z}} |h|^\alpha \left\|C_h^X\right\|_{\mathcal{S}} < \infty$。假设 $q \to \infty$ 和 $q/N \to 0$，当 $N \to \infty$ 时。

Theorem 3.3 设 $\{X_t\}_{t\in\mathbb{Z}}$ 是满足假设 *2.1* 的平稳过程。此外，假设 *3.1* 和 *3.3* 成立。则对于任意 $\epsilon > 0$，存在 $q \ge 1$ 和满足假设 *3.2* 的网络$\{\mathfrak{X}_t\}$，使得

$$\int_{-\pi}^{\pi} \left\| f^X(\theta) - \sum_{|h|\le q} \omega\left(\frac{h}{q}\right) C_h^{\mathfrak{X}} \exp(-\mathrm{i}h\theta) \right\|_{\mathcal{S}} d\theta < \epsilon.$$

Remark 3.2 检查定理 3.3 的证明可知，还可以推导出其他逼近结果：可以使用平方范数，和/或用 $\sum_{|h|\le q} \omega\,(h/q)\, C_h^X \exp(-\mathrm{i}h\theta)$ 代替 $f^X(\theta)$。为节省篇幅，我们不再列出这些变体。

4 网络估计器的构造

假设我们观测到一个实现 $\{X_1, \ldots, X_N\}$。对于每个 $t = 1, \ldots, N$，我们用一个由式 (3.6) 给出的网络 $\mathfrak{X}_t$ 逼近X_t，其中系数 $\xi_{m,t} \in \mathbb{R}$ 和网络 $\mathfrak{g}_{m,h} \in \mathcal{C}^{\mathrm{nn}}$ 需要学习。本节将 $\xi_{m,t}$ 视为未知参数，而非随机序列。

鉴于定理 3.1 和定理 3.3 以及注释 3.2，我们选择损失函数

$$\ell := \int_{-\pi}^{\pi} \left\| \hat{f}^X(\theta) - \hat{f}^{\widetilde{\mathfrak{X}}}(\theta) \right\|_{\mathcal{S}} d\theta = \int_{-\pi}^{\pi} \left\| \sum_{|h|\le q} \omega\left(\frac{h}{q}\right) \left[\widehat{C}_h^X - \widehat{C}_h^{\mathfrak{X}}\right] \exp(-\mathrm{i}h\theta) \right\|_{\mathcal{S}} d\theta. \tag{4.1}$$

在此损失函数中，$\hat{}\ f^X$和$\hat{}f^{\tilde{\mathfrak{x}}}$分别是谱密度 f^X和 $f^{\tilde{\mathfrak{x}}}$的估计器，因此该优化问题关于 $\mathcal{Q}$ 中观测到场 X_t 的网格点数量（记为 G）是线性的。本节中的计算将展示如何构建这样的训练算法。

考虑经验自协方差算子

$$\widehat{C}_h^X = \frac{1}{N}\sum_{k=1}^{N-h} X_{h+k}\otimes X_k, \quad h\geq 0; \qquad\qquad \widehat{C}_h^X = \frac{1}{N}\sum_{k=1}^{N-|h|} X_k\otimes X_{|h|+k}, \quad h<0.$$

以及基于观测场 X_1、.....、 X_N的滞后窗估计量，定义如下:

$$\widehat{F}^X(\theta) = \frac{1}{2\pi}\sum_{|h|\leq q}\omega\left(\frac{h}{q}\right)\widehat{C}_h^X\exp(-\mathrm{i}h\theta) = \sum_{|h|\leq q}\tilde{\omega}\left(\frac{h}{q}\right)\left(\frac{1}{N}\sum_{k=1}^{N} X_{h+k}\otimes X_k\right), \tag{4.2}$$

将下标不合理的项设为零，并

$$\tilde{\omega}(h,\theta) := \frac{1}{2\pi}\omega(h/q)\exp(-\mathrm{i}h\theta).$$

~同样，对于式 (3.6) 中的近似网络场 $\{\mathfrak{x}_t\}_{t=1}^N$，我们定义

$$\widehat{F}^{\tilde{\mathfrak{x}}}(\theta) = \sum_{|h|\leq q}\tilde{\omega}(h,\theta)\widehat{C}_h^{\tilde{\mathfrak{x}}}, \tag{4.3}$$

其中

$$\widehat{C}_h^{\tilde{\mathfrak{x}}} = \frac{1}{N}\sum_{k=1}^{N}\widetilde{\mathfrak{x}}_{h+k}\otimes\widetilde{\mathfrak{x}}_k = \frac{1}{N}\sum_{k=1}^{N}\sum_{j,j'=-L}^{L}\sum_{m,m'=1}^{M}\left(\xi_{m,h+k+j}\xi_{m',k+j'}\right)\mathfrak{g}_{m,j}\otimes\mathfrak{g}_{m',j'},$$

将下标不合理的项设为零。

如第 2 节所述，算子 $\widehat{F}^X(\theta)$ 和 $\widehat{F}^{\tilde{\mathfrak{x}}}(\theta)$ 分别与其核 $\hat{}\ f^X(\theta)$、$\hat{}f^{\tilde{\mathfrak{x}}}(\theta)$ 一一对应。计算这些核需要与 G^2 成正比的运算次数，因此不会在训练算法中直接使用。不过，二者之差的希尔伯特–施密特范数的平方可以写成四项之和，每一项都可以通过 $O(G)$ 次运算计算得到。为此，注意到

$$\left\|\widehat{F}^X(\theta)-\widehat{F}^{\tilde{\mathfrak{x}}}(\theta)\right\|_{\mathcal{S}}^2 = \frac{1}{N^2}\left\langle\sum_{h=-q}^{q}\sum_{k=1}^{N}\tilde{\omega}(h,\theta)\left(X_{h+k}\otimes X_k-\widetilde{\mathfrak{x}}_{h+k}\otimes\widetilde{\mathfrak{x}}_k\right),\right.$$
$$\left.\sum_{h=-q}^{q}\sum_{k=1}^{N}\tilde{\omega}(h,\theta)\left(X_{h+k}\otimes X_k-\widetilde{\mathfrak{x}}_{h+k}\otimes\widetilde{\mathfrak{x}}_k\right)\right\rangle_{\mathcal{S}}$$

$$= \frac{1}{N^2} \sum_{h=-q}^{q} \sum_{k=1}^{N} \sum_{h'=-q}^{q} \sum_{k'=1}^{N} \tilde{\omega}(h,\theta)\overline{\tilde{\omega}(h',\theta)} \left\langle \left(X_{h+k} \otimes X_k - \widetilde{\mathfrak{X}}_{h+k} \otimes \widetilde{\mathfrak{X}}_k \right), \left(X_{h'+k'} \otimes X_{k'} - \widetilde{\mathfrak{X}}_{h'+k'} \otimes \widetilde{\mathfrak{X}}_{k'} \right) \right\rangle_{\mathcal{S}}.$$

因此,

$$\begin{aligned} (4.4) \qquad & \left\| \widehat{F}^{X}(\theta) - \widehat{F}^{\widetilde{\mathfrak{X}}}(\theta) \right\|_{\mathcal{S}}^{2} \\ &= \frac{1}{N^2} \sum_{h=-q}^{q} \sum_{k=1}^{N} \sum_{h'=-q}^{q} \sum_{k'=1}^{N} \tilde{\omega}(h,\theta)\overline{\tilde{\omega}(h',\theta)} \langle X_{h+k}, X_{h'+k'} \rangle \langle X_k, X_{k'} \rangle \\ &\quad - \frac{1}{N^2} \sum_{h=-q}^{q} \sum_{k=1}^{N} \sum_{h'=-q}^{q} \sum_{k'=1}^{N} \tilde{\omega}(h,\theta)\overline{\tilde{\omega}(h',\theta)} \langle X_{h+k}, \widetilde{\mathfrak{X}}_{h'+k'} \rangle \langle X_k, \widetilde{\mathfrak{X}}_{k'} \rangle \\ &\quad - \frac{1}{N^2} \sum_{h=-q}^{q} \sum_{k=1}^{N} \sum_{h'=-q}^{q} \sum_{k'=1}^{N} \tilde{\omega}(h,\theta)\overline{\tilde{\omega}(h',\theta)} \langle \widetilde{\mathfrak{X}}_{h+k}, X_{h'+k'} \rangle \langle \widetilde{\mathfrak{X}}_k, X_{k'} \rangle \\ &\quad + \frac{1}{N^2} \sum_{h=-q}^{q} \sum_{k=1}^{N} \sum_{h'=-q}^{q} \sum_{k'=1}^{N} \tilde{\omega}(h,\theta)\overline{\tilde{\omega}(h',\theta)} \langle \widetilde{\mathfrak{X}}_{h+k}, \widetilde{\mathfrak{X}}_{h'+k'} \rangle \langle \widetilde{\mathfrak{X}}_k, \widetilde{\mathfrak{X}}_{k'} \rangle. \end{aligned}$$

注意

$$\tilde{\omega}(h,\theta)\overline{\tilde{\omega}(h',\theta)} = \frac{1}{4\pi^2} \omega\left(\frac{h}{q}\right) \omega\left(\frac{h'}{q}\right) \left\{ \cos\left([h-h']\theta\right) - \mathrm{i}\sin\left([h-h']\theta\right) \right\}.$$

~根据注3.1， $\mathfrak{X}_t$ 是实数，因此在优化算法中，式（4.4）右侧的乘积 $\tilde{\omega}(h,\ \theta)\overline{\tilde{\omega}(h',\ \theta)}$ 可以替换为实数

$$r(h,h';\theta) = \frac{1}{4\pi^2} \omega\left(\frac{h}{q}\right) \omega\left(\frac{h'}{q}\right) \cos\left([h-h']\theta\right).$$

表示式（4.4）表明，计算 $\|\widehat{F}^{X}(\theta) - \widehat{F}^{\widetilde{\mathfrak{X}}}(\theta)\|_{\mathcal{S}}^2$ 只涉及关于 $\mathcal{Q}$ 中观测到场 X_t 的网格点数 G 的线性表达式，从而避免了关于 G 的二次计算。这同样适用于其平方根 $\|\widehat{F}^{X}(\theta) - \widehat{F}^{\widetilde{\mathfrak{X}}}(\theta)\|_{\mathcal{S}}$，它出现在损失函数（ 4.1 ）中，并对大域上定义的随机场时间序列的谱密度估计问题在计算上的可行性产生关键影响。这些计算引出了以下算法。

Algorithm 1（谱密度估计）：

~**步骤 1：** 根据式 (3.6) 构造 $\{\mathfrak{X}_t\}_{t=1}^{N}$ ， 其中 ξ_m 为实数， $_{,\ h}$ 为实数值网络 $\mathfrak{g}_{m,\ h}$。给定超参数 M、 L 和 q，参数向量为

$$\vartheta = \{\xi_{m,h} \text{ and the parameters of } \mathfrak{g}_{m,h},\ 1 \le m \le M,\ |h| \le L+q\}.$$

˜**步骤 2：** 使用构造的 $\{\mathfrak{X}_t\}_{t=1}^N$ 和观测到的 $\{X_t\}_{t=1}^N$，在 $[-\pi,\pi]$ 和 $\mathcal{Q}$ 上使用离散网格计算式 (4.4)。

步骤 3： 对 θ 上的 (4.4) 的平方根进行数值积分。这ˆ 得到损失函数 ℓ 的数值版本，该函数在 (4.1) 中定义，我们将其记为 ℓ。

步骤 4： ˆ ˆ 对步骤 1 中指定的参数最小化 ℓ。将该极小化子记为 ϑ .

ˆ**步骤 5：** 代入极小化子 ϑ，得到

$$\widehat{C}_h^{\widetilde{\mathfrak{X}}}(\hat{\vartheta}) = \frac{1}{N}\sum_{k=1}^{N}\sum_{j,j'=-L}^{L}\sum_{m,m'=1}^{M}\left(\hat{\xi}_{m,h+k+j}\hat{\xi}_{m',k+j'}\right)\hat{\mathfrak{g}}_{m,j}\otimes\hat{\mathfrak{g}}_{m',j'},$$

其中，ξ 上方的帽号以及网络表示它们在优化后的取值处的评估。

步骤 6： 计算估计的余谱

$$\hat{p}^X(\theta)(u,v) = \frac{1}{2\pi}\sum_{|h|\le q}\omega\left(\frac{h}{q}\right)\hat{c}_h^{\widetilde{\mathfrak{X}}}(\hat{\vartheta})(u,v)\cos(h\theta)$$

以及四谱

$$\hat{q}^X(\theta)(u,v) = \frac{1}{2\pi}\sum_{|h|\le q}\omega\left(\frac{h}{q}\right)\hat{c}_h^{\widetilde{\mathfrak{X}}}(u,v)\sin(h\theta).$$

（坐标（u，v）仅出现在训练后的网络 ˆ$\mathfrak{g}_{m,\ j}$ 中。）

步骤 4 是我们所考虑的前馈网络的学习过程。相关细节，包括超参数的选择，将在第 5 节中讨论。我们将算法 1 所定义的估计量称为谱神经网络估计量。

5 数值实现和模拟

我们在前几节中已经证明，在关于网络和核/带宽的假设下，谱神经网络估计量是函数型时间序列谱密度的通用逼近器，而这些假设在实际中几乎总是成立的，即假设 3.1 和 3.3。本节通过模拟，将谱神经网络估计量与滞后窗口估计量（ 4.2 ）进行比较；后者已在以往研究中得到探讨和应用，例如 H¨ormann等人（2015）以及 Kuenzer 等人（2021）。我们将看到，谱神经网络估计量几乎从不逊色，往往表现好得多；当网格点数量非常大时，它甚至是唯一能够实际计算的估计量。与所有模拟研究一样，这类结论无法像数学结果那样以一般性得到确立，但能提供有益的见解。我们方法的实现代码可在https://github.com/sohamsarkar1991/spectral-NN获取。

我们考虑最简单的函数型时间序列模型，$X_t = \gamma X_{t-1} + Z_t$, $\gamma \in (-1, 1)$。首先，我们生成白噪声（创新）过程 $Z_t = \{Z_t(u) \ u \in [0,1]^d\}$, $t = 1, 2, \ldots\ldots, N$的离散观测值。Sarkar 和 Panaretos（2022）对此进行了全面讨论，他们在网格上为 $d = 2, 3$生成相互独立的高斯随机场。初始值 X_0 的分布设为与 Z_1相同。为确保近似平稳性，我们生成长度为 $N + N_0$的时间序列，并舍弃前 N_0 个元素（在模拟中，我们使用$N_0 = 100$ ）。

此外，为了将谱密度算子 $F^X(\theta)$ 与估计得到的对象进行比较，我们还需要获得其闭式表达。对于创新过程 $\{Z_t\}$，$F^Z(\theta) = \frac{1}{2\pi} C_0^Z$，$\theta \in [-\pi, \pi]$，其中 C_0^Z 是 $\{Z_t\}$ 的滞后 0 协方差算子。该过程的因果表示 $\{X_t\}$ 蕴含

$$\begin{aligned} F^X(\theta) &= \left(\sum_{h=0}^{\infty} \gamma^h \exp(-\mathrm{i}h\theta) \right) \frac{1}{2\pi} C_0^Z \left(\sum_{h=0}^{\infty} \gamma^h \exp(\mathrm{i}h\theta) \right) \\ &= (1 - \gamma \exp(-\mathrm{i}\theta))^{-1} \frac{1}{2\pi} C_0^Z \, (1 - \gamma \exp(\mathrm{i}\theta))^{-1} \\ &= \frac{1}{2\pi} \frac{C_0^Z}{1 + \gamma^2 - 2\gamma \cos(\theta)}. \end{aligned}$$

对于滞后 0 协方差算子 C_0^Z（等价地，协方差核 c_0^Z），我们仿照 Sarkar 和 Panaretos（2022）作出三种选择，即

（i）布朗片：$c_0^Z(u, v) = \min\{u_1, v_1\} \times \cdots \times \min\{u_d, v_d\}$, $u, v \in [0,1]^d$。对于 $d = 1$，这就退化为标准布朗运动。（ii）积分布朗片：$c_0^Z(u, v) = c_{\mathrm{ibm}}(u_1, v_1) \times \cdots \times c_{\mathrm{ibm}}(u_d, v_d)$ $u, v \in [0,1]^d$，其中 c_{ibm} 是积分布朗运动的协方差核，定义为 $c_{\mathrm{ibm}}(u, v) = \int_0^u \int_0^v \min\{s, t\} ds dt$。（iii）Mat´ern：$c_0^Z(u, v) = 2^{1-\nu}/\Gamma(\nu)$（$\sqrt{2\nu} \, \|u - v\|_d$）$^\nu K_\nu(\sqrt{2\nu} \, \|u - v\|_d)$ $u, v \in [0,1]^d$，其中 Γ 是伽马函数，K_ν 是第二类修正贝塞尔函数，$\|\cdot\|_d$ 是 $\mathbb{R}^d$ 上的欧几里得距离。Mat´ern 协方差模型由光滑度参数 $\nu > 0$ 索引。在模拟研究中，我们使用 $\nu = 0.001$、0.01、0.1 和 1。

这些协方差模型会使生成的随机场呈现丰富多样的光滑性结构。特别是，使用布朗片生成的随机场连续但处处不可微，而使用积分布朗片生成的随机场则连续可微。对于 Mat´ern 协方差，ν 的值越大，随机场就越光滑；详情请参见 Sarkar 和 Panaretos（2022）。

我们考虑 $d = 1$、2 和 3 维的模拟，分别称为一维、二维和三维。随机场在 $[0,1]^d$ 上的 $K \times \cdots \times K$ 规则网格上生成。我们需要选择样本量 N、网格尺寸参数 K 和自回归系数 γ。我们通过固定其中两个参数并

改变第三个参数来考虑三种不同的设置。为了评估谱神经网络估计量的性能，我们考察相对估计误差

$$\frac{\int_{-\pi}^{\pi} \|F^X(\theta) - \widehat{F}^{\widetilde{\mathfrak{X}}}(\theta)\|_{\mathcal{S}} d\theta}{\int_{-\pi}^{\pi} \|F^X(\theta)\|_{\mathcal{S}} d\theta},$$

其中， $\widehat{F}^{\widetilde{\mathfrak{X}}}(\cdot)$ 是使用神经网络模型估计的谱密度。由于这些积分无法以闭式形式计算，我们采用蒙特卡洛近似来计算积分。具体而言，我们从 $[-\pi, \pi]$ 上的均匀分布中生成随机样本 $\theta_1, \ldots, \theta_I$，并对于每个 $i = 1, \ldots, I$，从 $[0,1]^d \times [0,1]^d$ 上的均匀分布中生成随机样本 $(u_{i1}, v_{i1}) \ldots, (u_{iJ}, v_{iJ})$，以近似

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \|F^X(\theta) - \widehat{F}^{\widetilde{\mathfrak{X}}}(\theta)\|_{\mathcal{S}} d\theta \approx \frac{1}{I} \sum_{i=1}^{I} \left[\frac{1}{J} \sum_{j=1}^{J} \left\{ f^X(\theta_i)(u_{ij}, v_{ij}) - \widehat{f}^{\widetilde{\mathfrak{X}}}(\theta_i)(u_{ij}, v_{ij}) \right\}^2 \right]^{1/2},$$

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \|F^X(\theta)\|_{\mathcal{S}} d\theta \approx \frac{1}{I} \sum_{i=1}^{I} \left[\frac{1}{J} \sum_{j=1}^{J} \left\{ f^X(\theta_i)(u_{ij}, v_{ij}) \right\}^2 \right]^{1/2}.$$

最后，这两个量的比值给出了相对误差的一个逼近。在我们的模拟中，我们使用了 $I = 100$ 和 $J = 10000$。我们采用类似的方法计算滞后窗估计量（4.2）的相对估计误差。在下表中，我们分别将这两个估计器称为神经网络和经验值。

在所有模拟中，我们都使用深度谱神经网络估计量。我们还需要选择几个超参数： M、L，以及神经网络的深度、宽度和激活函数$\mathfrak{g}_m$、$_h$。对于这两种估计量，我们都需要选择截断阶数 q 和权重核 ω。在整个模拟过程中，我们都使用 sigmoid 函数 $\sigma(t) = (1 + e^{-t})^{-1}$ 作为激活函数。我们针对不同的 M(5、10、20)、 L(5、10、20)、深度（2、3、4、5、6）、宽度（10、20、30、40、50）、 q(5、10、20、40)以及ω （截断、Bartlett核、Parzen核、Tukey-Hanning核、二次谱核）进行了若干次试验模拟。在这些模拟中，我们发现，除了 q 和 ω之外，结果并未受到这些超参数选择的显著影响。使用 $q = 20$、40时的结果远优于使用$q = 5$、10时的结果。不过，增大 M、 L 和 q会增加计算时间。根据我们的经验，在整个模拟研究中，我们采用 $M = L = 10$、深度=4、宽度=20和 $q = 20$ 。对于权重函数，我们的试验研究表明，Parzen核效果最好，因此整个过程中都使用该核。为确保比较公平，我们也对经验（滞后窗）估计量使用$q = 20$ 和Parzen核。

我们对每种模拟设置重复进行 25 次，并报告平均相对误差及相应的标准误。表 1–3 仅报告二维设置的结果。一维和三维设置的结果与此类似，见补充材料（参见附录 D ）。

表1报告了 $K = 50$、 $\gamma = 0.5$ 以及 N 的不同取值下的结果。在这种设置下，结果符合常见规律：随着 N 增大，两种估计器的误差都会减小。

表1：在不同二维示例中，经验谱密度估计量（经验值）和谱神经网络估计量（神经网络）的相对误差率（%）。结果对应于固定的自回归系数 $\gamma = 0.5$、固定分辨率 $K = 50$ 和不同的样本量 N。数值为基于 25 次模拟运行的平均值。相应的标准误在下一行以斜体和较小字号列出。

	布朗运动片		积分布朗运动片		Matérn $\nu = 0.001$		Matérn $\nu = 0.01$		Matérn $\nu = 0.1$		Matérn $\nu = 1$	
N	经验值	神经网络	经验值	神经网络	经验值	神经网络	经验值	神经网络	经验值	神经网络	经验值	神经网络
100	47。95	43。32	32。44	34。10	1966。41	52。66	1766。47	56。67	209。21	67。71	43。32	42。38
	1.74	*1.80*	*1.82*	*1.77*	*4.41*	*2.72*	*4.43*	*1.61*	*1.36*	*1.60*	*1.16*	*1.17*
200	33.07	30.01	23.77	24.74	1416.19	46.68	1279.16	52.39	151。35	52。46	30。86	30。58
	0.70	*0.68*	*1.10*	*1.04*	*1.66*	*1.28*	*2.04*	*1.62*	*0.47*	*0.69*	*0.94*	*0.96*
400	23.76	22.38	16.13	16.53	1016.09	43.25	916.12	47.97	108。31	39。07	23。56	23。36
	0.59	*0.79*	*0.89*	*0.94*	*1.19*	*1.04*	*1.19*	*1.01*	*0.27*	*0.47*	*0.79*	*0.84*
800	17。78	17。41	11.84	12.69	730.87	42.32	659.79	47.70	77。95	31。72	16.45	16.52
	0.52	*0.54*	*0.64*	*0.83*	*0.61*	*0.77*	*0.56*	*0.99*	*0.16*	*0.33*	*0.56*	*0.56*
1600	13.13	13.86	8.44	9.83	531。17	41。83	479.56	45.87	56。32	26。59	11。49	11。98
	0.45	*0.44*	*0.43*	*0.63*	*0.47*	*0.67*	*0.41*	*0.58*	*0.17*	*0.28*	*0.45*	*0.45*

表 2：不同二维示例中经验谱密度估计量（经验值）和谱神经网络估计量（神经网络）的相对误差率（%）。结果对应于固定的自回归系数 $\gamma = 0.5$、固定样本量 $N = 250$ 和变化的分辨率 K。数值是基于 25 次模拟运行的平均值。相应的标准误显示在下一行，采用斜体和较小字号。短横线（—）表示程序因内存不足而失败。

	布朗片		积分布朗运动片		Matérn $\nu = 0.001$		Matérn $\nu = 0.01$		Matérn $\nu = 0.1$		Matérn $\nu = 1$	
K	经验值	神经网络	经验值	神经网络	Emp	NN	Emp	NN	经验值	神经网络	经验值	神经网络
10	33。53	31。58	26。47	22。51	1413。66	172。78	1276.75	222.37	147。18	82。57	29。47	29。15
	0.97	*0.95*	*0.99*	*1.00*	*4.15*	*6.59*	*4.01*	*7.93*	*1.04*	*1.59*	*1.06*	*0.97*
20	31.19	29.06	21。60	20。53	1301.69	80.31	1176.46	84.45	138。13	57。98	28.47	28.08
	0.76	*0.84*	*0.97*	*1.06*	*2.83*	*3.20*	*2.61*	*3.02*	*0.78*	*0.90*	*0.81*	*0.82*
40	29。11	26。60	18。28	18。40	1275.89	47.07	1153.01	51.27	135。11	47。21	26。70	26。29
	0.76	*0.79*	*0.98*	*1.09*	*1.64*	*1.39*	*1.53*	*1.21*	*0.53*	*0.76*	*0.81*	*0.84*
80	30.72	28.34	20。07	20。26	1267.63	42.38	1145.32	47.85	135.58	45.96	28.30	27.85
	0.80	*0.87*	*0.96*	*1.10*	*1.12*	*1.17*	*0.99*	*0.94*	*0.54*	*0.75*	*0.88*	*0.87*
160	—	26.76	—	19.27	—	37.58	—	43.96	—	43.53	—	26.04
		0.79		*1.12*		*0.64*		*0.53*		*0.60*		*0.78*

对于较粗糙的布朗片示例，谱神经网络估计量的表现略优于经验估计量。在更平滑的积分型

表3：不同二维示例中经验谱密度估计量（经验值）和谱神经网络估计量（神经网络）的相对误差率（%）。结果采用固定样本量 $N = 250$、固定分辨率 $K = 50$ 以及变化的自回归系数 γ。数值为基于25次模拟运行次数计算的平均值。相应的标准误在下一行以斜体和较小字号显示。

	布朗片		积分布朗片		Matérn							
					$\nu = 0.001$		$\nu = 0.01$		$\nu = 0.1$		$\nu = 1$	
γ	经验值	神经网络	经验值	神经网络	经验值	神经网络	经验值	神经网络	经验值	神经网络	经验值	神经网络
0.1	30.64	28.24	18。24	18。49	1269。82	45。74	1147.81	51.61	135。23	44。88	27.82	27.41
	0.66	*0.72*	*0.71*	*0.69*	*0.85*	*1.01*	*1.09*	*1.31*	*0.29*	*0.51*	*0.65*	*0.62*
0.25	30.15	27.44	19.57	19.49	1268。87	46。81	1146。76	50。32	135.50	46.16	28.35	28.10
	0.71	*0.71*	*1.01*	*0.91*	*1.07*	*1.67*	*1.01*	*1.28*	*0.46*	*0.56*	*0.96*	*1.02*
0.5	30.38	27.97	20.75	20.90	1272。52	45。03	1150.08	52.82	136.49	48.56	28.12	27.81
	0.84	*0.83*	*1.23*	*1.21*	*1.87*	*1.52*	*1.54*	*1.47*	*0.60*	*0.95*	*0.82*	*0.78*
0.75	35。00	32。82	23.36	24.09	1301.26	49.80	1176.66	53.70	138。13	52。46	33.49	33.08
	1.34	*1.40*	*1.72*	*1.62*	*5.00*	*1.91*	*5.65*	*2.11*	*0.77*	*0.77*	*1.41*	*1.46*
0.9	55.16	52.94	53.31	54.33	1414.54	63.75	1316。36	79。09	163.79	77.81	52.28	52.12
	2.30	*1.80*	*3.67*	*3.47*	*34.76*	*2.47*	*34.54*	*3.58*	*5.35*	*3.09*	*1.85*	*1.91*

布朗片示例。Mat´ern 协方差的情况则截然不同，尤其是 ν 值较小时，对应的曲面也更粗糙。在这些示例中，经验估计量无法捕捉到潜在的谱密度。另一方面，谱神经网络估计量即使面对这些粗糙的观测值，也能成功检测出其潜在结构。有趣的是，经验估计量的误差可能达到谱神经网络估计量的 20 倍（甚至更高）。

表 2 报告了 $N = 250$、 $\gamma = 0.5$ 以及不同 K 下的结果。此设置下的结果在定性上与前一设置相似。对于布朗片和积分布朗片，误差不受分辨率影响。然而，对于 Mat´ern 协方差，尽管谱神经网络估计量的误差随着分辨率提高而迅速降低，经验估计量却并非如此。这再次表明，粗糙度会对经验估计量产生不利影响，而采用所提出的神经网络结构可以缓解这一问题。

由于内存不足，无法以 160×160 的分辨率计算经验估计量。这是因为经验估计量对内存的需求很高：它需要计算经验自协方差，而这类对象的维度很高。具体而言，对于位于 $K \times K$ 网格上的观测值，经验自协方差是 K^4 维对象。此外，计算并存储经验估计量还需要多个（准确地说，是 $2q + 1$ 个）此类自协方差，即使 K 的值适中，也可能导致内存需求高得难以承受。

为进一步说明这一点，表 4 报告了不同二维示例中经验估计量和谱神经网络估计量的最大内存需求，其中

表 4：不同二维示例中经验谱密度估计量（经验值）和谱神经网络估计量（神经网络）的平均计算时间（秒）和最大内存使用量（MB）。对于神经网络，使用图形处理器时的计算时间以斜体显示在下一行。代码在一台配备 64 GiB 内存、AMD Ryzen 9 5900X（3.7 GHz）中央处理器、NVIDIA GeForce RTX 3090 GPU 和 Ubuntu 24.04.2 LTS（64 位）操作系统的计算机上运行。破折号（—）表示程序因内存不足而运行失败。

固定分辨率为 50×50，样本量可变（N）。

N	100		200		400		800		1600	
	Emp	NN	Emp	NN	Emp	NN	Emp	NN	Emp	NN
Fit	2.86	288.07	5.51	370.31	10.69	483.51	20.42	773.32	40。91	2116。57
		250.64		*271.51*		*318.35*		*403.86*		*626.89*
Eval	428.48	105.83	425.26	104.99	421.25	104.87	428.28	101.25	425.37	102.28
		1.85		*1.85*		*1.85*		*1.85*		*1.89*
总计	431.34	393.90	430.77	475.30	431.94	588.38	448.70	824.56	466。28	2218。85
		252.49		*273.36*		*320.20*		*405.71*		*628.72*
内存	1184	700	1185	679	1185	692	1187	709	1188	694

固定样本量为 250，分辨率可变。

K	10		20		40		80		160	
	Emp	NN	Emp	NN	Emp	NN	Emp	NN	Emp	NN
Fit	0.22	165.71	0.33	198.36	0.92	296.76	346.41	746.18	—	2698.37
		278.12		*279.04*		*278.55*		*288.59*		*315.84*
Eval	48.86	103.24	95.02	98.18	281.34	98.27	1092.12	98.96	—	99.50
		1.86		*1.85*		*1.86*		*1.86*		*1.86*
总计	49.08	268.96	95.35	296.54	282.25	395.03	1438.53	845.14	—	2797.87
		279.98		*280.89*		*280.41*		*290.45*		*317.70*
内存	120	696	150	696	564	700	7143	839	—	2224

在一台配备 64 GiB 内存、AMD Ryzen 9 5900X（3.7 GHz）中央处理器、NVIDIA GeForce RTX 3090 GPU，以及 Ubuntu 24.04.2 LTS 64 位操作系统的计算机上，N 和 K 的取值各不相同。从表中可以看出，经验估计量的最大内存需求随 K 呈指数级增长。对于 $K = 160$，经验估计量需要超过 64 吉字节的内存，而谱神经网络估计量仅需 2.2 吉字节。在三维情况下，情况可能糟糕得多：即使分辨率适中的 $25 \times 25 \times 25$，经验估计量也可能无法运行（见附录 D 中的表 9）。

在表 4 中，我们还报告了两种估计器的平均运行时间。对于这两种估计器，计算过程都包含两个部分：拟合模型，以及评估模型以计算误差。对于经验估计量，拟合包括计算并存储自协方差；而对于谱神经网络估计量，则包括估计模型参数。经验估计量的拟合部分耗时相对较少，但评估部分的计算开销很大，尤其是在 K 较大时。另一方面，谱神经网络估计量的拟合部分可能相当耗时。不过，模型拟合完成后，评估速度非常快。这表明，与经验估计量相比，所提出的估计量在适用性方面具有优势。

此外，使用 GPU 计算可以显著缩短谱神经网络估计量的计算时间，尤其是在 K 或 N（或两者）较大时。事实上，表 4 显示，对于 $N = 1600$，计算时间几乎缩短至原来的四分之一；对于 $K = 160$，则缩短至原来的九分之一以下。采用小批量学习等其他现代机器学习技术，还可以进一步缩短计算时间，不过我们在模拟中并未实现这一方法。

在表3中，我们报告了 $N = 250$, $K = 50$ 以及自回归系数 γ 取不同值时的结果。在这种设定下，γ 的值越大，问题就越难，不过这两种估计器的相对表现仍然相近。

6 脑部扫描时间序列的应用

为进一步展示谱神经网络估计量的实用性，我们将其应用于一组三维功能性磁共振成像数据。我们考察了来自 Beijin 的受试者 `sub69518` 的脑部扫描，这些数据来自 *1000* 功能连接组项目（https://www.nitrc.org/projects/fcon_1000/）。数据由三维脑部扫描组成，分辨率为 $64 \times 64 \times 33$，共包含225个时间点，时间点之间相隔2秒。这些数据集此前由 Aston 和 Kirch（2012）以及 Stoehr等人（2021）进行分析，他们得出结论，经过标准的逐体素预处理后，数据具有平稳性。Sarkar 和 Panaretos（2022）使用他们的 CovNet 方法估计这些数据的协方差，并将其视为独立同分布的函数型观测。然而，这些数据实际上是函数型时间序列，因为间隔2秒的扫描很可能存在依赖性。

应用估计量之前，我们对数据进行了预处理，移除了最初的 5 个时间点。为减轻边缘效应，我们还移除了 x轴和 y轴上最初的三个和最后的三个体素，以及 z轴上最初的两个和最后的两个体素。这样，我们得到了一组包含 220 次三维扫描的时间序列，每次扫描的分辨率为 $59 \times 59 \times 29$。按照 Aston和Kirch（2012）以及 Sarkar 和 Panaretos（2022）的建议，我们从每个体素中去除了 3 阶多项式趋势，并对数据进行了缩放，使每个体素的方差均为单位方差。

在这个例子中，谱神经网络估计量 $\widehat{f}$ 是定义在 $[-\pi, \pi]$ 上的函数。为了直观展示该估计量，我们计算了它在不同频率 $\theta \in [-\pi, \pi]$ 处的模。也就是说，对于 $\theta \in [-\pi, \pi]$，我们计算 $\|\widehat{F}^{\tilde{x}}(\theta)\|_{\mathcal{S}} = \sqrt{\langle \widehat{f}^{\tilde{x}}(\theta),\ \widehat{f}^{\tilde{x}}(\theta) \rangle_{\mathcal{S}}}$。这些模显示在图 1 中。图中显示，估计谱密度的模随 θ变化。如果这些扫描构成函数型白噪声，那么图 1 中的曲线应当（近似）为一条水平直线。这表明数据中确实存在时间依赖性。其总体形状与第 5 节中使用的 AR(1)模型一致，但在 $\pi/2$处有一个凸起。图中仅显示了范数，但我们可以看到，谱神经网络估计量是分析大域上函数型时间序列的一个有前景的工具。

Acknowledgement

本研究部分得到了美国国家科学基金会项目 DMS–2412408 的资助。Soham Sarkar 的研究部分得到了

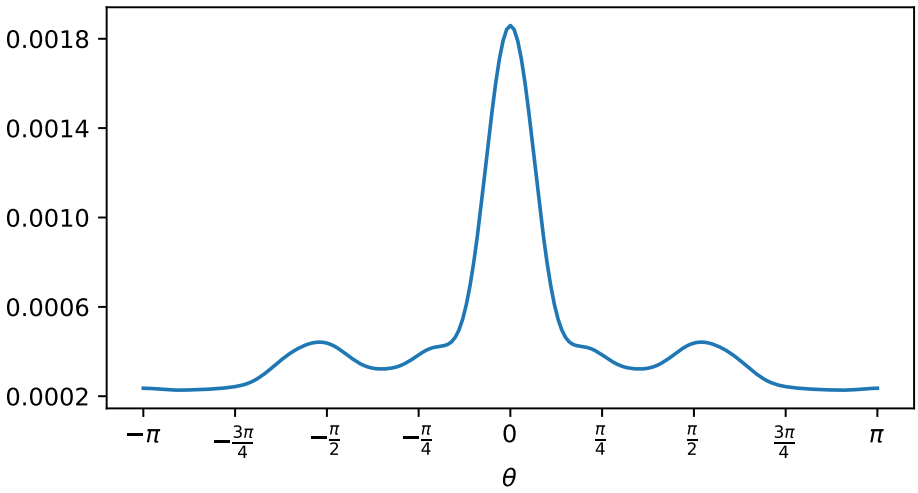

图 1：三维功能性磁共振成像数据的谱神经网络估计量拟合结果的幅值。谱神经网络模型使用 $M = L = 10$、深度$= 4$、宽度$= 20$ 和 $q = 20$进行拟合。

印度政府科学技术部颁发的 INSPIRE 教师奖学金资助。

Supplementary material

补充材料包含证明和额外的模拟结果。

参考文献

Aston, JohnAD 和 Kirch, Claudia (2012)。使用变点备选方案评估平稳性及其在功能性磁共振成像数据中的应用。《应用统计年鉴》，1906–1948。

Bishop, C. 和 Bishop, H. (2024)。深度学习。Springer。

Brockwell, P. J. 和 Davis, R. A. (1991)。时间序列：理论与方法。Springer，纽约。

Hong, J-S., Yao, J., Mueller, J. 和 Wang, J-L.（2024）。SAND：利用 Transformer 网络对稀疏含噪函数型数据进行平滑插补。载于第 *38* 届神经信息处理系统大会论文集（*NeurIPS 2024*），第 1–12 页。NeurIPS 基金会。

H¨ormann, S., Kidzinski, L. 和 Hallin, M.（2015）。动态函数型主成分。《皇家统计学会杂志》。系列B，**77**，第 2 期，319–348。

Horv´ath, L., Kokoszka, P. 和 Reeder, R.（2013）。函数型时间序列的均值估计与双样本问题。《皇家统计学会杂志》（*B*），**75**，103–122。

Hsing, T. 和 Eubank, R.（2015）。《函数型数据分析的理论基础：线性算子导论》。Wiley。

Kartsioukas, R., Stoev, S. 和 Hsing, T.（2023）。函数值空间过程的谱密度估计。*arXiv:2302.02247* 1–84。

Kokoszka, P. 和 Mohammadi, N.（2020）。函数型时间序列的频域理论：方差分解与不变性原理。*Bernoulli*，**26,** 第3期，2383–2399。Kuenzer, T., H¨ormann, S. 和 Kokoszka, P.（2021）。空间索引函数的主成分分析。美国统计学会会刊，**116,** 第535期，1444–1456。Leshno, M., Lin, V., Pinkus, A. 和 Schocken, S.（1993）。具有非多项式激活函数的多层前馈网络能够逼近任意函数。神经网络，**6,** 861–867。Ma, T., Yao, F. 和 Zhou, Z.（2024）。网络级交通流预测：函数型时间序列与函数型神经网络方法。《应用统计年鉴》，**18,** 424–444。Panaretos, V. M. 和 Tavakoli, S.（2013a）。函数空间中平稳时间序列的傅里叶分析。*Ann. Stat.*，**41,** 568–603。Panaretos, V. M. 和 Tavakoli, S.（2013b）。Cram´er–Karhunen–Lo´eve 表示与函数型时间序列的谐波主成分分析。随机过程及其应用，**123,** 2779–2807。Rao, A. R. 和 Reimher, M.（2023a）。使用神经网络进行非线性函数型建模。计算与图形统计学杂志，**32,** 1248–1257。Rao, A. R. 和 Reimher, M.（2023b）。现代非线性函数对函数回归。统计与计算，**33,** 130。Sarkar, S. 和 Panaretos, V. M.（2022）。CovNet：多维域上函数型数据的协方差网络。英国皇家统计学会会刊*B*辑：统计方法学，**84,** 1785–1820。Stoehr, Christina, Aston, John A D 和 Kirch, Claudia（2021）。检测函数型时间序列协方差结构的变化，并应用于功能性磁共振成像数据。计量经济学与统计学，**18,** 44–62。Tavakoli, S.（2014）。函数型时间序列的傅里叶分析及其在DNA动力学中的应用。博士论文。洛桑联邦理工学院。Thind, B., Multani, K. 和 Cao, J.（2023）。具有函数型输入的深度学习。计算与图形统计学杂志，**32,** 171–180。Wang, H. 和 Cao, J.（2023）。函数型时间序列的非线性预测。环境计量学，**34,** e2792。Wang, H. 和 Cao, J.（2024）。函数型非线性学习。计算与图形统计学杂志，**33,** 181–191。Wang, S., Zhang, W., Cao, G. 和 Huang, Y.（2024）。使用深度神经网络进行函数型数据分析。*WIREs*计算统计学，**16,** e70001。Wu, S., Beaulac, C. 和 Cao, J.（2023）。用于标量输入和函数型输出的神经网络。统计与计算，**33,** 文章编号118。Wu, S., Beaulac, C. 和 Cao, J.（2024）。用于平滑和表征学习的函数型自编码器。统计与计算，**34,** 文章编号203。

Yao, J.、Mueller, J. 和 Wang, J-L.（2021年）。基于自适应基函数层的函数型数据分析深度学习。《机器学习研究论文集》，**139,** 11898–11908。

补充材料

空间 $L^2(\mathcal{Q})$ 中的普适逼近

例如，Bishop 和 Bishop（2024）第6.2.3节介绍了常用的激活函数，包括 ReLU、带泄漏的 ReLU、硬 tanh、tanh、softplus 和逻辑 sigmoid。它们都是连续函数：要么是分段线性函数，且有一两个点处导数不存在；要么是无限次可微的非多项式函数。因此，我们可以运用 Leshno等人（1993）的结果，得到以下命题。回顾一下， $\mathcal{Q}$ 是 $\mathbb{R}^d$的一个紧子集。

PROPOSITION A.1 如果激活函数 σ 不是多项式，那么每个类 $\mathcal{C}^{nn}$ 在 $L^2(\mathcal{Q})$ 中都是稠密的。

PROOF. 每个类 $\mathcal{C}^{\text{nn}}$ 都包含类 $\mathcal{C}^{\text{sh}}$，后者与 Leshno等人（1993）定理 1 中考察的类 Σ_d 相同；不同之处在于，Leshno等人（1993）考察的是实值函数，而我们考察的是复值函数。他们的结果可应用于实部和虚部。因此，Leshno等人（1993）的命题 1 表明，只要 μ 关于 $\mathbb{R}^d$上的勒贝格测度是绝对连续的， $\mathcal{C}^{\text{sh}}$ 就在任意空间 $L^p(\mu)$、 $1 \leq p < \infty$中稠密。特别地， $\mathcal{C}^{\text{sh}}$ 在 $L^2(\mathcal{Q})$中稠密。

∎

B 预备引理

为便于查阅，我们在此陈述两个在证明中经常用到的引理。第一个引理直接由 Kokoszka 和 Mohammadi（2020）中的定理 2.1 得出，第二个则由 $\varphi_m^\dagger(\theta)$ 是正交归一的这一事实以及引理 B.1 得出。

LEMMA B.1 假设假设 *2.1* 成立，则

$$\infty > \mathbb{E}\|X_0\|^2 = \sum_{m\geq 1}\int_{-\pi}^{\pi}\lambda_m(\theta)d\theta =: \sum_{m\geq 1}\Lambda_m =: \Lambda.$$

LEMMA B.2 令

$$f_M^X(\theta)(u,v) := \sum_{m=1}^{M}\lambda_m(\theta)\varphi_m^\dagger(\theta)(u)\overline{\varphi}_m^\dagger(\theta)(v), \quad u,v\in\mathcal{Q}.$$

在假设 *2.1* 下，对于任意 $\epsilon > 0$，存在 M 使得

$$\int_{-\pi}^{\pi}\|f^X(\theta) - f_M^X(\theta)\|_{\mathcal{S}}d\theta < \epsilon.$$

C 第 3 节结果的证明

在继续进行证明之前，我们先回顾所需的背景信息。关于更全面的讨论，请参阅 H¨ormann 等 (2015) 的第 3.3 小节。回顾谱密度分解 (2.4)，我们可以将其写为

$$F^X(\theta) = \sum_{m\geq 1}\lambda_m(\theta)\varphi_m^\dagger(\theta)\otimes\varphi_m^\dagger(\theta). \tag{C.1}$$

由于 对于每个 θ，函数 $\varphi_m^\dagger(\theta)$、 $m \geq 1$ 均为正交归一的， $\|\varphi_m^\dagger(\theta)\| = 1$，因此

$$\int_{\mathcal{Q}} \int_{-\pi}^{\pi} \left|\varphi_m^\dagger(\theta)(u)\right|^2 d\theta du = \int_{-\pi}^{\pi} \int_{\mathcal{Q}} \left|\varphi_m^\dagger(\theta)(u)\right|^2 du d\theta = 2\pi < \infty. \tag{C.2}$$

因此，对于几乎所有的 $u \in \mathcal{Q}$, $\int_{-\pi}^{\pi} \left|\varphi_m^\dagger(\theta)(u)\right|^2 d\theta < \infty$。记 $\mathcal{L}eb(\cdot)$ 为 $\mathbb{R}^d$ 上的勒贝格测度，则存在 $A_m \subseteq \mathcal{Q}$，满足 $\mathcal{L}eb(A_m) = \mathcal{L}eb(\mathcal{Q}) < \infty$，使得对所有 $u \in A_m$，都有 $\int_{-\pi}^{\pi} \left|\varphi_m^\dagger(\theta)(u)\right|^2 d\theta < \infty$。定义

$$\varphi_{m,l}(u) = \begin{cases} \frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi_m^\dagger(\theta)(u) \exp(-\mathrm{i}l\theta) d\theta, & u \in A_m, \\ 0, & u \notin A_m. \end{cases} \tag{C.3}$$

引入集合 A_m 仅仅是为了使 $\varphi_{m,\,l}(u)$ 在每个 $u \in \mathcal{Q}$ 处都有定义；它们不会影响任何均方收敛结果。特别地，反演公式 (2.6) 仍然成立（在均方意义下），即

$$X_t = \sum_{m=1}^{\infty} \sum_{l \in \mathbb{Z}} Y_{m,t+l} \varphi_{m,l}, \quad \text{where } Y_{m,t} = \sum_{l \in \mathbb{Z}} \langle X_{t-l}, \varphi_{m,l} \rangle. \tag{C.4}$$

此外，对于每个 m, $\varphi_m^\dagger(\theta)$ 和 $\varphi_{m,\,h}$ 都通过定义 2.1 相互关联。

Proof of Theorem 3.1 . 固定 $\epsilon > 0$。利用三角不等式，对于 $\mathfrak{f} \in \mathcal{E}$ 和正整数 M，我们有

$$\begin{aligned} &\int_{-\pi}^{\pi} \|f^X(\theta) - \mathfrak{f}(\theta)\|_{\mathcal{S}} d\theta \\ &\leq \int_{-\pi}^{\pi} \|f^X(\theta) - f_M^X(\theta)\|_{\mathcal{S}} d\theta + \int_{-\pi}^{\pi} \|f_M^X(\theta) - \mathfrak{f}(\theta)\|_{\mathcal{S}} d\theta, \end{aligned} \tag{C.5}$$

其中 $f_M^X(\theta)$ 在引理 B.2 中定义，这意味着存在一个充分大的 M，使得

$$\int_{-\pi}^{\pi} \|f^X(\theta) - f_M^X(\theta)\|_{\mathcal{S}} d\theta < \epsilon/2. \tag{C.6}$$

接下来，我们固定这个 M，并着重考察涉及网络逼近的第二项。我们将找到网络 $\mathfrak{g}_{m,\,h}$, $m = 1,\ 2,\ \ldots,\ M$，使得 $\{\mathfrak{g}_{m,h}\}_h \in \mathcal{C}$，从而令 (C.5) 中的第二项小于 $\epsilon/2$。回顾一下， $\varphi_m^\dagger(\theta)$ 和 $\varphi_{m,\,h}$ 通过关系 (2.7) 相互关联。对于每个 $m = 1,\ 2,\ \ldots,\ M$ 和正整数 L，定义 $c = c(L) = 2^{L-1}$，并选取神经网络 $\mathfrak{g}_{m,\,h}(\cdot)$，使得

$$\|\varphi_{m,h} - \mathfrak{g}_{m,h}\|_{L^2(\mathcal{Q})}^2 \leq \frac{\tilde{\epsilon}}{6\pi c 2^{|h|}}, \quad -L \leq h \leq L,\ m = 1, 2 \ldots, M. \tag{C.7}$$

请注意，假设 3.1 保证了此类网络的存在。对于每个 m =1，2，.....，M 以及任意正整数 L，有限序列 $\{\mathfrak{g}_{m,h}\}_{-L \leq h \leq L}$ 通过将其余元素设为零，扩展为 $\mathcal{C}$ 中的无限序列 $\{\mathfrak{g}_{m,h}\}_{h \in \mathbb{Z}}$。级数 $\{\mathfrak{g}_{m,h}\}_{h \in \mathbb{Z}}$ 的傅里叶变换记为 $\mathfrak{g}_m^\ddagger(\theta)$。我们使用记号 $\cdot^\ddagger$ 以强调这确实是

有限级数的傅里叶变换。特别地， $\mathfrak{g}_m^{\ddagger}(\theta)$ 和 $\mathfrak{g}_{m,\ h}$ 通过定义 2.1 联系起来。注意，对于每个 $m=1$，2，.....， M 以及正整数 L，我们有

$$\frac{1}{4}\int_{-\pi}^{\pi}\left\|\varphi_m^{\dagger}(\theta)-\mathfrak{g}_m^{\ddagger}(\theta)\right\|_{L^2(\mathcal{Q})}^2 d\theta \tag{C.8}$$

$$\leq\int_{-\pi}^{\pi}\left\|\varphi_m^{\dagger}(\theta)-\sum_{h=-L}^{L}\exp(\mathrm{i}h\theta)\varphi_{m,h}\right\|_{L^2(\mathcal{Q})}^2 d\theta \tag{C.9}$$

$$+\int_{-\pi}^{\pi}\left\|\sum_{h=-L}^{L}\exp(\mathrm{i}h\theta)\varphi_{m,h}-\sum_{h=-L}^{L}\exp(\mathrm{i}h\theta)\mathfrak{g}_{m,h}\right\|_{L^2(\mathcal{Q})}^2 d\theta \tag{C.10}$$

$$+\int_{-\pi}^{\pi}\left\|\sum_{h=-L}^{L}\exp(\mathrm{i}h\theta)\mathfrak{g}_{m,h}-\mathfrak{g}_m^{\ddagger}(\theta)\right\|_{L^2(\mathcal{Q})}^2 d\theta. \tag{C.11}$$

现在证明，存在足够大的 L，使得上述被加数可以被任意小的 $\tilde{\epsilon}>0$ 从上方界定。首先注意，关系式 (2.7) 保证存在足够大的 $L=L(M)$，使得 (C.9) 被 $\tilde{\epsilon}$ 界定，对于 $m=1,\ 2\ldots,\ M$。

对于此 L，不等式 (C.7) 意味着 (C.10) 的上界为

$$\begin{aligned}
&\int_{-\pi}^{\pi}\left\|\sum_{h=-L}^{L}\exp(\mathrm{i}h\theta)\left(\varphi_{m,h}-\mathfrak{g}_{m,h}\right)\right\|_{L^2(\mathcal{Q})}^2 d\theta\\
&\qquad\leq c\int_{-\pi}^{\pi}\sum_{h=-L}^{L}\left\|\varphi_{m,h}-\mathfrak{g}_{m,h}\right\|_{L^2(\mathcal{Q})}^2 d\theta\\
&\qquad\leq c\int_{-\pi}^{\pi}\sum_{h=-L}^{L}\frac{\tilde{\epsilon}}{6\pi c2^{|h|}}d\theta\\
&\qquad\leq 2\pi c\sum_{h=-\infty}^{\infty}\frac{\tilde{\epsilon}}{6\pi c2^{|h|}}=\frac{2\pi c\tilde{\epsilon}}{6\pi c}\times 3=\tilde{\epsilon}.
\end{aligned}$$

根据构造，(C.11) 等于零。

综上所述，关系式 (2.7) 意味着存在足够大的正整数 L，使得 (C.9) 可由任意小的 $\tilde{\epsilon}$ 界定。对于这个有限 L，存在满足 (C.7) 的神经网络 $\{\mathfrak{g}_{m,h}\}_{-L\leq h\leq L}$ 有限序列。这意味着 (C.10) 可由 $\tilde{\epsilon}$ 从上方界定。有限序列 $\{\mathfrak{g}_{m,h}\}_{-L\leq h\leq L}$ 被扩展为无限序列$\{\mathfrak{g}_{m,h}\}_{h\in\mathbb{Z}}$，使得 (C.11) 等于零。因此，对于 $m=1,\ \ldots M$，存在足够大的 L和$\mathfrak{g}_m^{\ddagger}(\theta)\in\mathcal{D}$，使得 (C.8) 满足

$$\int_{-\pi}^{\pi}\left\|\varphi_m^{\dagger}(\theta)-\mathfrak{g}_m^{\ddagger}(\theta)\right\|_{L^2(\mathcal{Q})}^2 d\theta\leq 8\tilde{\epsilon},\quad m=1,2,\ldots,M. \tag{C.12}$$

现在注意到

$$\left\|\sum_{m=1}^{M}\lambda_m(\theta)\varphi_m^{\dagger}(\theta)\otimes\varphi_m^{\dagger}(\theta)-\sum_{m=1}^{M}\lambda_m(\theta)\mathfrak{g}_m^{\ddagger}(\theta)\otimes\mathfrak{g}_m^{\ddagger}(\theta)\right\|_{\mathcal{S}}$$

$$\leq \sum_{m=1}^{M} 2\lambda_m(\theta) \left\|\varphi_m^\dagger(\theta) - \mathfrak{g}_m^\ddagger(\theta)\right\|_{L^2(\mathcal{Q})} + \sum_{m=1}^{M} \lambda_m(\theta) \left\|\varphi_m^\dagger(\theta) - \mathfrak{g}_m^\ddagger(\theta)\right\|_{L^2(\mathcal{Q})}^2 \tag{C.13}$$

$$\leq 2\Lambda^* \sum_{m=1}^{M} \left\|\varphi_m^\dagger(\theta) - \mathfrak{g}_m^\ddagger(\theta)\right\|_{L^2(\mathcal{Q})} + \Lambda^* \sum_{m=1}^{M} \left\|\varphi_m^\dagger(\theta) - \mathfrak{g}_m^\ddagger(\theta)\right\|_{L^2(\mathcal{Q})}^2, \tag{C.14}$$

其中 $\Lambda^* = \sup \lambda_m(\theta) < \infty$ 在引理 2.1 中定义，而不等式 (C.13) 是由$_{m,\theta}\|f \otimes f - g \otimes g\|_{L^2(}$ $_{\mathcal{Q}\times\mathcal{Q})} \leq 2\|f\|\|f-g\| + \|f-g\|^2$得出的。于是，对于足够小的 $\tilde{\epsilon}$，(C.14) 和 (C.12) 蕴含

$$\begin{aligned}
&\int_{-\pi}^{\pi} \left\| \sum_{m=1}^{M} \lambda_m(\theta)\varphi_m^\dagger(\theta) \otimes \varphi_m^\dagger(\theta) - \sum_{m=1}^{M} \lambda_m(\theta)\mathfrak{g}_m^\ddagger(\theta) \otimes \mathfrak{g}_m^\ddagger(\theta) \right\|_{\mathcal{S}} d\theta \\
\leq & 2\Lambda^* \sum_{m=1}^{M} \left[\int_{-\pi}^{\pi} \left\|\varphi_m^\dagger(\theta) - \mathfrak{g}_m^\ddagger(\theta)\right\|_{L^2(\mathcal{Q})}^2 d\theta \right]^{1/2} \\
& + \Lambda^* \sum_{m=1}^{M} \int_{-\pi}^{\pi} \left\|\varphi_m^\dagger(\theta) - \mathfrak{g}_m^\ddagger(\theta)\right\|_{L^2(\mathcal{Q})}^2 d\theta \\
\leq & 3\Lambda^* M (8\tilde{\epsilon})^{1/2}.
\end{aligned} \tag{C.15}$$

令 $\mathfrak{f}(\theta) = \sum_{m=1}^{M} \lambda_m(\theta)\mathfrak{g}_m^\ddagger(\theta) \otimes \mathfrak{g}_m^\ddagger(\theta) \in \mathcal{E}$ 并选择 $\tilde{\epsilon}$ 足够小，使其取决于ϵ，则 (C.15) 意味着

$$\int_{-\pi}^{\pi} \left\| f_M^X(\theta) - \mathfrak{f}(\theta) \right\|_{\mathcal{S}} d\theta \leq \epsilon/2. \tag{C.16}$$

结合不等式 (C.5)、(C.6) 和 (C.16)，即可得到所需的通用逼近。
■

Proof of Theorem 3.2. **步骤 1：** 在这一步中，我们证明了定理 3.2 的陈述中定义的~平稳过程 $\{\mathfrak{X}_t\}$ 的谱密度核具有表示式 (3.7)。~为此，将 $\mathfrak{X}_t$ 改写为以下形式

$$\begin{aligned}
\widetilde{\mathfrak{X}}_t &= \sum_{h=-L}^{L} (\mathfrak{g}_{1,h}, \ldots, \mathfrak{g}_{M,h}) (\xi_{1,t+h}, \ldots, \xi_{M,t+h})^\top \\
&=: \sum_{h=-L}^{L} \mathfrak{g}_h \xi_{t+h}^\top.
\end{aligned}$$

~平稳随机过程 $\{\mathfrak{X}_t\}$ 的谱密度算子具有如下形式

$$\begin{aligned}
F^{\widetilde{\mathfrak{X}}}(\theta) =& \frac{1}{2\pi} \sum_{h\in\mathbb{Z}} C_h^{\widetilde{\mathfrak{X}}} \exp(-\mathrm{i}h\theta) \\
=& \frac{1}{2\pi} \sum_{h\in\mathbb{Z}} \mathrm{Cov}(\widetilde{\mathfrak{X}}_h, \widetilde{\mathfrak{X}}_0) \exp(-\mathrm{i}h\theta)
\end{aligned}$$

$$=\frac{1}{2\pi}\sum_{h\in\mathbb{Z}}\mathrm{Cov}(\sum_{s=-L}^{L}\mathfrak{g}_s\xi_{h+s}^{\top},\sum_{s'=-L}^{L}\mathfrak{g}_{s'}\xi_{s'}^{\top})\exp(-\mathrm{i}h\theta).$$

这意味着

$$\begin{aligned}F^{\widetilde{\mathfrak{x}}}(\theta)=&\frac{1}{2\pi}\sum_{h\in\mathbb{Z}}\sum_{s=-L}^{L}\sum_{s'=-L}^{L}\mathrm{Cov}(\mathfrak{g}_s\xi_{h+s}^{\top},\mathfrak{g}_{s'}\xi_{s'}^{\top})\exp(-\mathrm{i}h\theta)\\=&\frac{1}{2\pi}\sum_{h\in\mathbb{Z}}\sum_{s=-L}^{L}\sum_{s'=-L}^{L}\mathfrak{g}_s\mathrm{Cov}(\xi_{h+s}^{\top},\xi_{s'}^{\top})\overline{\mathfrak{g}}_{s'}^{\top}\exp(-\mathrm{i}h\theta-\mathrm{i}s\theta+\mathrm{i}s'\theta+\mathrm{i}s\theta-\mathrm{i}s'\theta).\end{aligned}$$

根据可和性条件 $\sum_{h\in\mathbb{Z}}\|C_h^{\xi}\|_{\mathcal{S}}<\infty$，我们有

$$\begin{aligned}F^{\widetilde{\mathfrak{x}}}(\theta)=&\frac{1}{2\pi}\sum_{s\in\mathbb{Z}}\sum_{s'\in\mathbb{Z}}\sum_{h\in\mathbb{Z}}\mathfrak{g}_s\mathrm{Cov}(\xi_{h+s}^{\top},\xi_{s'}^{\top})\overline{\mathfrak{g}}_{s'}^{\top}\exp(-\mathrm{i}h\theta-\mathrm{i}s\theta+\mathrm{i}s'\theta+\mathrm{i}s\theta-\mathrm{i}s'\theta)\\=&\frac{1}{2\pi}\sum_{s=-L}^{L}\sum_{s'=-L}^{L}\mathfrak{g}_sF^{\xi}(\theta)\overline{\mathfrak{g}}_{s'}^{\top}\exp(\mathrm{i}s\theta-\mathrm{i}s'\theta).\end{aligned}$$

由此得到所需的形式 (3.7)。

步骤 2： 在此步骤中，我们证明在 (3.5) 中定义的类 $\mathcal{E}$ 可以写成形式 (3.8)。考虑由以下形式给出的、属于 (3.5) 中定义的类 $\mathcal{E}$ 的一般元素 $\mathfrak{f}$

$$\mathfrak{f}(\theta)=\sum_{m=1}^{M}\eta_m(\theta)\mathfrak{g}_m^{\ddagger}(\theta)\otimes\mathfrak{g}_m^{\ddagger}(\theta),\quad\theta\in[-\pi,\pi],$$

对某些 $M\in\mathbb{N}$、$\eta_\cdot(\cdot)\in\mathcal{A}_{\{1,\ldots,M\}}\subset\mathcal{A}$、$\mathfrak{g}_m^{\ddagger}\in\mathcal{D}$、$m=1,\ldots M$。根据步骤 1，只需证明存在一个 M 维随机过程 $\{\xi_t=(\xi_{1,t},\ldots,\xi_{M,t})\}$，其谱密度算子为 diag $(\eta_1(\theta)\ldots,\eta_M(\theta)$。考虑具有独立分量且其分量具有以下协方差结构的高斯 M 维随机过程 $\{\xi_t=(\xi_{1,t},\ldots,\xi_{M,t})\}$：

$$c_h^m=\int_{-\pi}^{\pi}\eta_m(\theta)\exp(\mathrm{i}h\theta)d\theta,\quad m=1,\ldots M.$$

由于 $\eta_\cdot(\cdot)\in\mathcal{A}$，上述协方差定义良好。由于它们构成一个正定族，因此高斯过程 ξ_t 存在。参见 Brockwell 和 Davis（1991）第 1 章。至此，证明完成。 ∎

REMARK C.1 定理 3.2 证明中的步骤 1 也可以由 Tavakoli (2014) 中的定理 2.5.5 推导得出。Tavakoli (2014) 中的定理 2.5.5 提出了两个假设：对某个 $p\in[1,\infty$，其条件 2.4.1(p)，以及极限关系 (2.5.12)。在我们的情形中，我们有可和性假设 $\sum_{h\in\mathbb{Z}}\|C_h^{\xi}\|_{\mathcal{S}}<\infty$。该可和性条件意味着 $\sum_{h\in\mathbb{Z}}\|C_h^{\xi}\|_{\mathcal{N}}<\infty$，其中 $\|\cdot\|_{\mathcal{N}}$表示核范数。这是因为有限维拓扑向量空间中定义的所有范数都等价。因此，$\sum_{h\in\mathbb{Z}}\|C_h^{\xi}\|_{\mathcal{N}}<\infty$ 蕴含 Tavakoli (2014) 中的条件 2.3.3 和条件 2.3.4。根据他们的注 2.4.2，我们得出条件 2.4.1 对于 $p=\infty$ 成立。此外，在我们的定理 3.2 中，我们处理有限序列 $\{\mathfrak{g}_h\}_{-L\le h\le L}$

并且特别地，可和性条件 $\sum_{h\in\mathbb{Z}}\|\mathfrak{g}_{m,\ h}\|_{L^2\ (\mathcal{Q})}\ <\infty$ 成立。根据 Tavakoli (2014) 中的注 2.5.6，这意味着其公式 (2.5.12) 成立。总之，Tavakoli (2014) 中定理 2.5.5 的假设在我们的情形下均成立。因此，谱密度$\widetilde{}$算子 $\{\mathfrak{X}_t\}$的形式是 Tavakoli (2014) 中定理 2.5.5 的推论。

在开始证明定理 3.3 之前，我们先回顾并根据本文的需要修改有关函数型滤波过程的背景知识。更全面的讨论请参见 H¨ormann等人(2015) 的 A.3 节和 A.4 节。回想 C 节开头的讨论。以下引理可由 H¨ormann等人(2015) 的 A.4.1 小节中的计算推出。

LEMMA C.1 假设假设 *2.1* 成立，并考虑一组函数 $\gamma_{m,\ l}\in L^2(\mathcal{Q})$ （一列线性滤波器），使得

$$\forall\ m\geq 1,\ \ \sum_{l\in\mathbb{Z}}\|\gamma_{m,l}\|<\infty.$$

对于 *(C.3)* 中的$\varphi_{m,\ l}$ 以及 *(C.4)* 中的$Y_{m,\ t}$，令

$$\gamma_{M,t}^{(Y)}=\sum_{m=1}^{M}\sum_{l\in\mathbb{Z}}Y_{m,t+l}\gamma_{m,l}.$$

那么级数 $\gamma_{M\,t}^{(Y)}$，在 $L^2(\mathcal{Q})$ 中定义良好，并且

$$\mathbb{E}\left\|X_t-\gamma_{M,t}^{(Y)}\right\|^2=\int_{-\pi}^{\pi}\left\|\sqrt{F^X(\theta)}-\Gamma_M(\theta)\sqrt{F^X(\theta)}\right\|_{\mathcal{S}}^2 d\theta,$$

其中，$\sqrt{F^X(\theta)}$ 表示 $F^X(\theta)$ 的平方根，并且

$$\Gamma_M(\theta)=\sum_{m=1}^{M}\left[\sum_{l}\gamma_{m,l}\exp\left(\mathrm{i}l\theta\right)\right]\otimes\varphi_m^{\dagger}(\theta).$$

现在我们转而讨论 X_t的均方网络逼近。

PROPOSITION C.1 假设假设 *2.1* 和 *3.1* 成立。那么，存在如假设 *3.2* 所述的网络 $\{\mathfrak{X}_t\}$，其由 M、L 索引，使得对于每个 $t\in\mathbb{Z}$,

$$\mathbb{E}\|X_t-\widetilde{\mathfrak{X}}_t\|^2\to 0,\ \ as\ M,L\to\infty.$$

PROOF. 回顾谱密度分解 (C.1)。方程 (C.2) 和 (C.3) 表明，对于每一对 (m,l)，$\varphi_{m,l}\in L^2(\mathcal{Q})$。假设 3.1 进而表明，存在逼近神经网络 $\mathfrak{g}_{m,l}\in\mathcal{C}^{\mathrm{nn}}$，可任意接近 $\varphi_{m,l}$ （按 $L^2(\mathcal{Q})$ 距离衡量）。对于稍后定义的有限正整数 L 和 M，选取网络 $\mathfrak{g}_{m,l}$，使得 (C.7) 成立。同样，令 $\mathfrak{g}_m^{\ddagger}(\theta)$ 为有限序列 $\{\mathfrak{g}_{m,l}\}_{-L\leq l\leq L}$ 的傅里叶变换。对于 $M,L\geq 1$，定义

$$\widetilde{\mathfrak{X}}_t=\widetilde{\mathfrak{X}}_t^{M,L}=\sum_{m=1}^{M}\sum_{l=-L}^{L}Y_{m,t+l}\mathfrak{g}_{m,l},$$

其中 $Y_{m,t}$ 在 (C.4) 中定义。只需证明，对于任意小的 $\epsilon > 0$，存在足够大的 M 和 L，使得

$$\mathbb{E}\|X_t - \widetilde{\mathfrak{X}}_t\|^2 < \epsilon. \tag{C.17}$$

为得到 (C.17)，我们将应用引理 C.1。首先注意到

$$\begin{aligned}\Gamma_M(\theta) &= \sum_{m=1}^{M}\left[\sum_{l=-L}^{L}\mathfrak{g}_{m,l}\exp(\mathrm{i}l\theta)\right]\otimes\varphi_m^\dagger(\theta)\\ &= \sum_{m=1}^{M}\mathfrak{g}_m^\ddagger(\theta)\otimes\varphi_m^\dagger(\theta).\end{aligned}$$

回想起，$\mathfrak{g}_m^\ddagger(\theta)$ 和 $\mathfrak{g}_{m,\ l}$ 通过定义 2.1 相连。现在注意到

$$\begin{aligned}\Gamma_M(\theta)\sqrt{F^X(\theta)} &= \left[\sum_{m=1}^{M}\mathfrak{g}_m^\ddagger(\theta)\otimes\varphi_m^\dagger(\theta)\right]\left[\sum_{m\geq 1}\sqrt{\lambda_m(\theta)}\varphi_m^\dagger(\theta)\otimes\varphi_m^\dagger(\theta)\right]\\ &= \sum_{m=1}^{M}\sum_{m'\geq 1}\sqrt{\lambda_{m'}(\theta)}\langle\varphi_{m'}^\dagger,\varphi_m^\dagger\rangle\mathfrak{g}_m^\ddagger(\theta)\otimes\varphi_{m'}^\dagger(\theta)\\ &= \sum_{m=1}^{M}\sqrt{\lambda_m(\theta)}\mathfrak{g}_m^\ddagger(\theta)\otimes\varphi_m^\dagger(\theta).\end{aligned}$$

因此，

$$\begin{aligned}\sqrt{F^X(\theta)} - \Gamma_M(\theta)\sqrt{F^X(\theta)} &= \sum_{m\geq 1}\sqrt{\lambda_m(\theta)}\varphi_m^\dagger(\theta)\otimes\varphi_m^\dagger(\theta) - \sum_{m=1}^{M}\sqrt{\lambda_m(\theta)}\mathfrak{g}_m^\ddagger(\theta)\otimes\varphi_m^\dagger(\theta)\\ &= \sum_{m\geq 1}\sqrt{\lambda_m(\theta)}\varphi_m^\dagger(\theta)\otimes\varphi_m^\dagger(\theta) - \sum_{m=1}^{M}\sqrt{\lambda_m(\theta)}\varphi_m^\dagger(\theta)\otimes\varphi_m^\dagger(\theta)\\ &\quad + \sum_{m=1}^{M}\sqrt{\lambda_m(\theta)}\varphi_m^\dagger(\theta)\otimes\varphi_m^\dagger(\theta) - \sum_{m=1}^{M}\sqrt{\lambda_m(\theta)}\mathfrak{g}_m^\ddagger(\theta)\otimes\varphi_m^\dagger(\theta).\end{aligned}$$

因此，

$$\begin{aligned}\frac{1}{2}\left\|\sqrt{F^X(\theta)} - \Gamma_M(\theta)\sqrt{F^X(\theta)}\right\|_{\mathcal{S}}^2 &\leq \left\|\sum_{m>M}\sqrt{\lambda_m(\theta)}\varphi_m^\dagger(\theta)\otimes\varphi_m^\dagger(\theta)\right\|_{\mathcal{S}}^2\\ &\quad + \left\|\sum_{m=1}^{M}\sqrt{\lambda_m(\theta)}\left[\varphi_m^\dagger(\theta) - \mathfrak{g}_m^\ddagger(\theta)\right]\otimes\varphi_m^\dagger(\theta)\right\|_{\mathcal{S}}^2\\ &= \sum_{m>M}\lambda_m(\theta) + \sum_{m=1}^{M}\lambda_m(\theta)\|\varphi_m^\dagger(\theta) - \mathfrak{g}_m^\ddagger(\theta)\|^2.\end{aligned}$$

这意味着

$$\begin{aligned}\frac{1}{2}\mathbb{E}\|X_t-\widetilde{\mathfrak{X}}_t\|^2 &\le \int_{-\pi}^{\pi}\sum_{m>M}\lambda_m(\theta)d\theta+\Lambda^*\sum_{m=1}^{M}\int_{-\pi}^{\pi}\left\|\varphi_m^{\dagger}(\theta)-\mathfrak{g}_m^{\ddagger}(\theta)\right\|^2 d\theta\\ &=\sum_{m>M}\Lambda_m+\Lambda^*\sum_{m=1}^{M}\int_{-\pi}^{\pi}\left\|\varphi_m^{\dagger}(\theta)-\mathfrak{g}_m^{\ddagger}(\theta)\right\|^2 d\theta\\ &=:S_1+S_2.\end{aligned}$$

其中 $\Lambda^*=\sup\ \lambda_m(\theta)<\infty$ 由引理 2.1 定义。引理 B.1 保证存在一个$m,\ \theta$足够大的 M，使得 $S_1<\epsilon/4$。对于这个 M，通过类似于推导式 (C.12) 的论证可知，对于足够大的 L，有 $S_2<\epsilon/4$。证明完毕。 ∎

Lemma C.2 考虑命题 *C.1* 中的情形。回顾一下，平稳过程 $\{\widetilde{\mathfrak{X}}_t^{M,L}\}$ 和 $\{X_t\}$ 的滞后 h 自协方差算子分别记为 $C_h^{\widetilde{\mathfrak{X}}}$ 和 C_h^X。于是，

$$\lim_{M,L\to\infty}\sup_{h\in\mathbb{Z}}\left\|C_h^{\widetilde{\mathfrak{X}}}-C_h^X\right\|_{\mathcal{S}}^2=0. \tag{C.18}$$

Proof of Lemma C.2注意到

$$\begin{aligned}C_h^{\widetilde{\mathfrak{X}}}-C_h^X&=\mathbb{E}[\widetilde{\mathfrak{X}}_h\otimes\widetilde{\mathfrak{X}}_0]-\mathbb{E}[X_h\otimes X_0]\\ &=\mathbb{E}[(\widetilde{\mathfrak{X}}_h-X_h)\otimes(\widetilde{\mathfrak{X}}_0-X_0)]+\mathbb{E}[(\widetilde{\mathfrak{X}}_h-X_h)\otimes X_0]+\mathbb{E}[X_h\otimes(\widetilde{\mathfrak{X}}_0-X_0)].\end{aligned}$$

这意味着

$$\left\|C_h^{\widetilde{\mathfrak{X}}}-C_h^X\right\|_{\mathcal{S}}\le\mathbb{E}\left\|\widetilde{\mathfrak{X}}_h-X_h\right\|\left\|\widetilde{\mathfrak{X}}_0-X_0\right\|+\mathbb{E}\left\|\widetilde{\mathfrak{X}}_h-X_h\right\|\|X_0\|+\mathbb{E}\|X_h\|\left\|\widetilde{\mathfrak{X}}_0-X_0\right\|.$$

因此，

$$\begin{aligned}&\frac{1}{4}\left\|C_h^{\widetilde{\mathfrak{X}}}-C_h^X\right\|_{\mathcal{S}}^2\\ &\quad\le\mathbb{E}\left\|\widetilde{\mathfrak{X}}_h-X_h\right\|^2\mathbb{E}\left\|\widetilde{\mathfrak{X}}_0-X_0\right\|^2+\mathbb{E}\left\|\widetilde{\mathfrak{X}}_h-X_h\right\|^2\mathbb{E}\|X_0\|^2+\mathbb{E}\|X_h\|^2\mathbb{E}\left\|\widetilde{\mathfrak{X}}_0-X_0\right\|^2,\end{aligned}$$

于是，由命题 C.1 可得，

$$\lim_{M,L\to\infty}\sup_{h\in\mathbb{Z}}\left\|C_h^{\widetilde{\mathfrak{X}}}-C_h^X\right\|_{\mathcal{S}}^2=0$$

如所期望的。 ∎

Proof of Theorem 3.3. 利用三角不等式，对于任意 $q\ge1$，我们有

$$\begin{aligned}&\int_{-\pi}^{\pi}\|f^X(\theta)-\sum_{|h|\le q}\omega\left(\frac{h}{q}\right)C_h^{\mathfrak{X}}\exp(-\mathrm{i}h\theta)\|_{\mathcal{S}}d\theta\\ \le&\int_{-\pi}^{\pi}\|f^X(\theta)-\sum_{|h|\le q}C_h^X\exp(-\mathrm{i}h\theta)\|_{\mathcal{S}}d\theta\end{aligned}$$

$$
\begin{aligned}
&+\int_{-\pi}^{\pi}\|\sum_{|h|\leq q}C_h^X\exp(-\mathrm{i}h\theta)-\sum_{|h|\leq q}\omega\left(\frac{h}{q}\right)C_h^X\exp(-\mathrm{i}h\theta)\|_{\mathcal{S}}d\theta\\
&+\int_{-\pi}^{\pi}\|\sum_{|h|\leq q}\omega\left(\frac{h}{q}\right)C_h^X\exp(-\mathrm{i}h\theta)-\sum_{|h|\leq q}\omega\left(\frac{h}{q}\right)C_h^{\mathfrak{X}}\exp(-\mathrm{i}h\theta)\|_{\mathcal{S}}d\theta\\
=:&A_1+A_2+A_3.
\end{aligned}
$$

现在只需证明每个被加数都可以用 $\epsilon/3$ 作为上界。对于 A_1，注意到根据 H¨older 不等式和希尔伯特–施密特范数的定义，我们有

$$
\begin{aligned}
A_1^2&\leq\int_{-\pi}^{\pi}\|f^X(\theta)-\sum_{|h|\leq q}C_h^X\exp(-\mathrm{i}h\theta)\|_{\mathcal{S}}^2d\theta\\
&=\int_{-\pi}^{\pi}\|\sum_{|h|>q}C_h^X\exp(-\mathrm{i}h\theta)\|_{\mathcal{S}}^2d\theta\\
&=\int_{-\pi}^{\pi}\iint_{\mathcal{Q}\times\mathcal{Q}}\left|\sum_{|h|>q}c_h^X(u,v)\exp(-\mathrm{i}h\theta)\right|^2dudvd\theta
\end{aligned}
$$

应用帕塞瓦尔等式可得

$$
\begin{aligned}
A_1^2\leq&2\pi\sum_{|h|>q}\iint_{\mathcal{Q}\times\mathcal{Q}}\left|c_h^X(u,v)\right|^2dudv\\
=&2\pi\sum_{|h|>q}\|C_h^X\|_{\mathcal{S}}^2.
\end{aligned}
$$

根据假设 2.1，存在一个足够大的 q，使得上述求和有界于 $\epsilon^2/9$，即 $A_1<\epsilon/3$。

关于 A_2 的论证取决于所用的核，但显然它适用于实践中使用的任何核。注意到

$$
A_2=\int_{-\pi}^{\pi}\left\|\sum_{|h|\leq q}\left(1-\omega(h/q)\right)C_h^X\exp(-\mathrm{i}h\theta)\right\|_{\mathcal{S}}d\theta\leq2\pi\sum_{|h|\leq q}\left|1-\omega(h/q)\right|\left\|C_h^X\right\|_{\mathcal{S}}.
$$

因此，使用截断核可得 $A_2=0$。使用巴特利特核

$$
\begin{aligned}
\frac{1}{2\pi}A_2&\leq\sum_{|h|\leq q}\left|1-\left(1-\frac{|h|}{q})\right)\right|\left\|C_h^X\right\|_{\mathcal{S}}=\sum_{|h|\leq q}\frac{|h|}{q}\left\|C_h^X\right\|_{\mathcal{S}}\\
&\leq\sum_{|h|\leq q}\frac{|h|^\alpha}{q^\alpha}\left\|C_h^X\right\|_{\mathcal{S}}\leq\frac{1}{q^\alpha}\sum_{h\in\mathbb{Z}}|h|^\alpha\left\|C_h^X\right\|_{\mathcal{S}},
\end{aligned}
$$

其中，根据假设 3.3，$\sum_{h\in\mathbb{Z}}|h|^\alpha\left\|C_h^X\right\|_{\mathcal{S}}<\infty$。假设 3.3 还意味着 q^α 趋于无穷大，因此当 q 足够大时，项 A_2 有界于 $\epsilon/3$。使用帕曾核，我们得到

$$
\frac{1}{2\pi}A_2\leq\sum_{|h|=0}^{\frac{q}{2}-1}\left|1-1+6\frac{|h|^2}{q^2}-6\frac{|h|^3}{q^3}\right|\left\|C_h^X\right\|_{\mathcal{S}}+\sum_{|h|=\frac{q}{2}}^{q}\left|1-2\left(1-\frac{|h|}{q}\right)^3\right|\left\|C_h^X\right\|_{\mathcal{S}}
$$

$$
\begin{aligned}
\leq & 6 \sum_{|h|=0}^{\frac{q}{2}-1} \left| \frac{|h|^2 q^3 - |h|^3 q^2}{q^5} \right| \left\| C_h^X \right\|_{\mathcal{S}} + \sum_{|h|=\frac{q}{2}}^{q} (1+2) \left\| C_h^X \right\|_{\mathcal{S}} \\
\leq & \frac{6}{q^\alpha} \sum_{h \in \mathbb{Z}} |h|^\alpha \left\| C_h^X \right\|_{\mathcal{S}} + 3 \sum_{|h|=\frac{q}{2}}^{\infty} \left\| C_h^X \right\|_{\mathcal{S}}.
\end{aligned}
$$

根据假设 3.3，第一项可以任意小；根据假设 2.1，第二项也可以任意小。

现在来看最后一项 A_3，对此我们只需知道 ω 有界即可。这由假设 3.3 中函数的特定选取方式可知；更一般地，例如，对于任意连续且紧支撑的核，这一结论也成立。选取足够大的带宽 q，使得 A_1 和 A_2 都被 $\epsilon/3$ 控制。由于权重函数 $\omega(\cdot)$ 被某个有限常数 c 控制，

$$
\begin{aligned}
A_3 \leq & \int_{-\pi}^{\pi} \sum_{|h| \leq q} \left| \exp(-\mathrm{i}h\theta) \omega\left(\frac{h}{q}\right) \right| \| C_h^X - C_h^{\mathfrak{X}} \|_{\mathcal{S}} d\theta \\
\leq & c \int_{-\pi}^{\pi} \sum_{|h| \leq q} \| C_h^X - C_h^{\mathfrak{X}} \|_{\mathcal{S}} d\theta \\
\leq & 2\pi c \sum_{|h| \leq q} \left\| C_h^X - C_h^{\widetilde{\mathfrak{X}}} \right\|_{\mathcal{S}}.
\end{aligned}
$$

关系式 (C.18) 表明，对于足够大的 M 和 L，项 A_3 被 $\epsilon/3$ 控制。证明完毕。 ∎

D 其他模拟结果

本节报告第 5 节中提到的一维和三维设置下的模拟结果。表 5 和表 6 报告了一维情况下的估计误差。三维情况下的估计误差见表 8。我们还在表 7 中报告了一维情况下、在表 9 中报告了三维情况下两种估计器的计算时间和最大内存需求。对于谱神经网络估计量，我们还报告了采用 GPU 计算时的计算时间。时间和内存复杂度基于一台配备 64 GiB 内存、AMD Ryzen 9 5900X（3.7 GHz）中央处理器、NVIDIA GeForce RTX 3090 GPU 和 Ubuntu 24.04.2 LTS（64 位）操作系统的计算机运行结果得出。

我们在一维和三维设置中的观测结果与二维设置中的结果相似。随着样本量和分辨率的增加，两种估计器的估计误差均有所降低。随着自回归系数 γ的增大，问题变得更加困难。此外，谱神经网络估计量的表现与经验估计量相当或更优。当底层函数型观测不平滑时（布朗片和较低值的 Mat´ern ν），这种改进尤为明显。在三维示例中，这些差异更加突出。

从计算时间和内存需求来看，我们再次看到谱神经网络估计量的实用性，尤其是在数据分辨率较高时，三维示例中更是如此。即使在分辨率为 $15 \times 15 \times 15$ 的中等分辨率下，经验估计量所需的内存也几乎是谱神经网络估计量的三倍。经验估计量的计算时间也呈指数级增长，在分辨率为 $25 \times 25 \times 25$ 时，几乎是谱神经网络估计量的三倍。此外，我们观察到，使用 GPU 计算可以大幅缩短谱神经网络的计算时间，尤其是在分辨率较高时。在分辨率为

表5：不同一维示例中经验谱密度估计量（经验值）和谱神经网络估计量（神经网络）的相对误差率（%）。数值为基于25次模拟运行得出的平均值。相应的标准误列于下一行，以斜体和较小字号显示。

固定自回归系数 $\gamma = 0.5$，固定分辨率 $K = 200$，变化的样本量 N

	布朗运动		积分布朗运动		Matérn $\nu = 0.001$		Matérn $\nu = 0.01$		Matérn $\nu = 0.1$		Matérn $\nu = 1$	
N	经验值	神经网络	经验值	NN	经验值	神经网络	Emp	神经网络	经验值	神经网络	经验值	神经网络
100	38.10	36.98	30.09	30.60	254。28	43。38	241.38	50.45	82。34	51。13	36.30	36.45
	1.58	*1.63*	*1.81*	*1.96*	*1.03*	*1.70*	*1.01*	*1.60*	*1.11*	*1.50*	*1.68*	*1.69*
200	26.77	25.72	20.77	21.32	184.76	33.88	175.20	39.27	59.26	38.01	25.30	25.17
	1.13	*1.03*	*1.28*	*1.35*	*0.80*	*1.08*	*0.79*	*1.00*	*0.76*	*0.96*	*1.18*	*1.14*
400	19.51	19.06	15.32	15.42	134。16	28。18	126.93	30.29	42。35	27。93	18。59	18。58
	0.90	*0.91*	*1.02*	*1.09*	*0.42*	*0.65*	*0.41*	*0.72*	*0.51*	*0.72*	*0.93*	*0.92*
800	13.51	13.30	10.40	10.43	99.81	24.47	94.03	24.43	30。14	20。98	12.90	12.96
	0.53	*0.54*	*0.64*	*0.64*	*0.19*	*0.51*	*0.18*	*0.39*	*0.28*	*0.38*	*0.54*	*0.56*
1600	10.43	10.37	8.30	8.60	76.44	23.61	71.52	21.19	21。99	17。05	10.03	10.24
	0.52	*0.53*	*0.63*	*0.64*	*0.12*	*0.49*	*0.11*	*0.37*	*0.26*	*0.32*	*0.54*	*0.56*

固定自回归系数 $\gamma = 0.5$、固定样本量 $N = 250$、变化的分辨率 K

	布朗运动		积分布朗运动	d	Matérn $\nu = 0.001$		Matérn $\nu = 0.01$		Matérn $\nu = 0.1$		Matérn $\nu = 1$	
K	经验值	神经网络	经验值	NN	经验值	神经网络	Emp	神经网络	经验值	神经网络	经验值	神经网络
20	25.48	24.90	21.01	20.41	212。72	72。74	199.33	106.54	56.63	42.03	24.18	24.29
	0.77	*0.75*	*0.83*	*0.89*	*1.48*	*3.00*	*1.40*	*2.92*	*0.67*	*0.79*	*0.81*	*0.86*
40	27.03	26.55	22.33	22.34	188.81	47.54	177.69	63.38	55。65	38。90	25.93	25.86
	1.30	*1.33*	*1.47*	*1.51*	*1.00*	*2.13*	*0.95*	*1.71*	*0.85*	*1.09*	*1.35*	*1.37*
80	24.03	23.24	18.66	18.67	173.65	34.39	163.93	42.15	53.05	34.86	22.82	23.07
	0.91	*0.97*	*1.03*	*1.03*	*0.70*	*1.03*	*0.68*	*1.16*	*0.56*	*0.82*	*0.92*	*0.92*
160	24.20	23.76	19.17	19.34	166.56	31.17	157.66	35.90	52.39	33.27	22。99	23。00
	0.93	*0.91*	*1.10*	*1.06*	*0.60*	*0.84*	*0.58*	*0.72*	*0.51*	*0.74*	*0.97*	*0.98*
320	24.91	24.24	19.64	19.50	164。26	30。85	155.89	34.59	53.30	34.40	23.84	23.95
	1.01	*1.02*	*1.17*	*1.17*	*0.55*	*0.91*	*0.54*	*1.00*	*0.67*	*0.98*	*1.05*	*1.07*
640	24.83	24.29	19.46	19.71	163.55	30.61	155.30	34.11	53.58	33.58	23.54	23.46
	1.49	*1.51*	*1.68*	*1.66*	*0.74*	*1.34*	*0.74*	*1.32*	*1.01*	*1.21*	*1.55*	*1.58*
1280	25.19	24.50	19.87	19.99	162.13	30.81	154.07	33.56	53。20	34。28	24。20	24。37
	0.94	*0.99*	*1.17*	*1.15*	*0.48*	*0.80*	*0.48*	*0.76*	*0.58*	*0.97*	*0.99*	*1.03*
2560	23.86	22.95	18.64	18.36	161。14	29。70	153.08	32.18	52。20	32。86	22.78	22.69
	0.99	*0.92*	*1.16*	*1.17*	*0.42*	*0.83*	*0.43*	*0.79*	*0.59*	*0.85*	*1.04*	*1.02*

$20 \times 20 \times 20$，使用图形处理器的谱神经网络所需时间不到经验估计量的六分之一。在分辨率为 $25 \times 25 \times 25$ 时，这一差距更加显著：使用图形处理器的谱神经网络所需时间不到经验估计量的十五分之一。在该分辨率下，经验估计量需要近 42 吉字节内存，远超普通计算机的内存容量。相比之下，谱神经网络估计量所需内存不到 1.5 吉字节。在分辨率为$30 \times 30 \times 30$时，经验估计量完全无法运行，需要超过 100 吉字节内存。相比之下，谱神经网络估计量仅需 3.2 吉字节内存，大多数普通计算机都能轻松提供。

表 6：固定样本量$N = 250$和固定分辨率 $K = 200$下，不同一维示例中经验谱密度估计量（经验值）与谱神经网络估计量（神经网络）的相对误差率（%），其中自回归系数 γ各不相同。数值为基于 25 次模拟运行的平均值。相应的标准误在下一行以斜体和较小字号显示。I

		ntegrated			Matern					
	布朗运动	布朗运动			$\nu = 0.001$		$\nu = 0.01$		$\nu = 0.1$		$\nu = 1$	
γ	经验值	经验值	神经网络	NN	经验值	神经网络	经验值	神经网络	经验值	神经网络	经验值	神经网络
0.1	23.84	23.22	18.31	18.39	166。03	30。26	157。32	34。45	52.80	33.11	22.77	22.76
	0.82	*0.85*	*1.02*	*1.00*	*0.51*	*0.71*	*0.51*	*0.68*	*0.50*	*0.68*	*0.84*	*0.85*
0.25	24.39	23.68	18.88	19.13	166.23	31.05	157.53	35.18	53。20	33。64	23。34	23。43
	1.00	*0.97*	*1.21*	*1.22*	*0.59*	*0.91*	*0.59*	*0.84*	*0.62*	*0.76*	*1.02*	*1.04*
0.5	25.80	25.12	20.36	20.66	166.82	33.23	158.14	37.73	54.21	34.69	24。76	24。72
	1.40	*1.42*	*1.64*	*1.64*	*0.81*	*1.23*	*0.80*	*1.18*	*0.88*	*1.14*	*1.43*	*1.53*
0.75	30.24	29.87	25.36	25.71	170.14	37.77	161。42	43。20	57。47	39。25	29.23	29.28
	1.87	*1.96*	*2.10*	*2.16*	*1.46*	*1.65*	*1.40*	*1.52*	*1.25*	*1.54*	*1.91*	*1.96*
0.9	49.60	48.91	42.33	42.54	193.18	57.66	183.83	61.94	74。75	57。74	48.72	48.72
	1.60	*1.64*	*1.48*	*1.44*	*4.74*	*1.63*	*4.48*	*1.62*	*1.47*	*1.40*	*1.62*	*1.66*

表 7：一维示例中经验谱密度估计量（经验值）和谱神经网络估计量（神经网络）的平均计算时间（秒）及最大内存使用量（MB）。对于神经网络，使用图形处理器时的计算时间以斜体显示在下一行。代码在一台配备 64 GiB 内存、AMD Ryzen 9 5900X（3.7 GHz）中央处理器、NVIDIA GeForce RTX 3090 GPU 和 Ubuntu 24.04.2 LTS（64 位）操作系统的计算机上运行。

固定分辨率 $K = 200$，样本量变化 N。

N	100		200		400		800		1600	
	Emp	NN	Emp	NN	Emp	NN	Emp	NN	Emp	NN
Fit	0.12	113.92	0.26	164.85	0.53	217.73	1.03	341.88	2.07	2798.25
		245.14		*269.01*		*315.34*		*406.07*		*623.84*
Eval	37.81	101.50	36.20	100.57	36.04	100.90	35.84	101.38	37.40	100.35
		1.72		*1.72*		*1.72*		*1.72*		*1.75*
总计	37.93	215.42	36.46	265.42	36.57	318.63	36.87	443.26	39.47	2898.61
		246.86		*270.73*		*317.06*		*407.79*		*625.59*
内存	121	689	114	694	121	695	120	706	121	699

固定样本量 $N = 250$，分辨率变化 K。

K	160		320		640		1280		2560	
	Emp	NN	Emp	NN	Emp	NN	Emp	NN	Emp	NN
Fit	0.26	172.19	0.34	188.46	0.40	221.23	0.72	265.94	7.96	403.10
		279.26		*277.88*		*277.73*		*277.94*		*280.23*
Eval	34.28	97.75	40.08	97.80	38.08	99.63	41.39	99.05	55.02	98.31
		1.72		*1.73*		*1.73*		*1.73*		*1.73*
总计	34.54	269.95	40.42	286.26	38.48	320.86	42.11	364.99	62.98	501.41
		280.98		*279.61*		*279.46*		*279.67*		*281.96*
内存	120	695	140	688	190	689	396	691	1235	698

表 8：不同三维示例中经验谱密度估计量（经验值）和谱神经网络估计量（神经网络）的相对误差率（%）。数值为基于 25 次模拟运行计算的平均值。相应的标准误以斜体和较小字号显示在下一行。破折号（—）表示程序因内存不足而运行失败。

固定自回归系数 $\gamma = 0.5$，固定分辨率 $K = 15$，变化的样本量 N

	积分				Matérn							
	布朗片	布朗片			$\nu = 0.001$		$\nu = 0.01$		$\nu = 0.1$		$\nu = 1$	
N	经验值	神经网络经验值		NN	Emp	NN	Emp	神经网络经验值	神经网络经验值		神经网络	
100	56.05	46.77	33.95	36.86	15348。49	202。79	13121.01	162.75	535.52	98.74	50.49	48.80
	0.87	*1.07*	*1.59*	*1.65*	*33.24*	*10.13*	*21.91*	*5.45*	*1.46*	*2.75*	*0.83*	*1.11*
200	43.19	37.20	28.12	27.31	11046.12	216.70	9495.02	157.97	387.94	81.10	38.57	37.27
	0.78	*0.89*	*1.34*	*1.26*	*15.06*	*7.11*	*8.57*	*5.62*	*1.00*	*1.97*	*1.25*	*1.30*
400	31.57	29.24	22.35	21.27	7902。86	361。22	6794.62	177.41	280。35	63。53	28.67	28.01
	0.65	*0.92*	*0.54*	*0.86*	*8.48*	*9.30*	*5.22*	*4.23*	*0.81*	*1.70*	*0.54*	*0.56*
800	22.48	20.64	18.33	18.37	5676.39	307.82	4874.80	168.72	203.10	52.58	19.98	20.10
	0.26	*0.37*	*0.34*	*1.59*	*4.59*	*9.53*	*3.04*	*2.46*	*0.36*	*0.84*	*0.41*	*0.50*
1600	17.93	17.18	17.66	17.17	4090。39	265。60	3519.45	163.90	150。16	46。29	13.95	14.41
	0.28	*0.37*	*0.32*	*1.66*	*1.91*	*6.19*	*2.33*	*3.37*	*0.27*	*0.44*	*0.28*	*0.27*

固定自回归系数 $\gamma = 0.5$、固定样本量 $N = 250$、变化的分辨率 K

	积分				Matérn							
	布朗片	布朗片			$\nu = 0.001$		$\nu = 0.01$		$\nu = 0.1$		$\nu = 1$	
K	经验值	神经网络经验值		NN	Emp	NN	Emp	神经网络经验值	神经网络经验值		神经网络	
10	40.14	35.35	31.96	41.57	10015.70	362.65	8631。20	287。94	354.54	96.48	34。84	33。37
	0.71	*0.96*	*0.93*	*3.89*	*11.61*	*10.03*	*12.17*	*6.60*	*1.07*	*2.13*	*0.63*	*0.68*
15	37.66	32.18	24.62	30.96	9937.27	234.32	8517.69	176.77	348.75	71.70	34.39	32.66
	0.62	*0.73*	*0.78*	*1.76*	*10.27*	*6.99*	*12.25*	*5.20*	*0.96*	*1.88*	*0.74*	*0.73*
20	38.37	32.83	24.11	37.88	9899.69	168.16	8492。14	118。38	346。48	65。13	33.92	32.45
	0.85	*0.88*	*1.51*	*3.32*	*6.84*	*7.29*	*8.58*	*2.30*	*0.73*	*1.35*	*0.80*	*0.80*
25	38.21	33.09	21.49	33.19	9878.38	153.95	8503.70	95.31	345.88	60.11	33.50	32.02
	0.51	*0.65*	*0.91*	*3.09*	*9.79*	*10.05*	*8.41*	*3.10*	*0.71*	*0.97*	*0.55*	*0.60*
30	—	33.53	—	33.94	—	143.89	—	80.50	—	57。34	—	32.18
		0.73		*2.75*		*7.92*		*2.52*		*0.85*		*0.65*

表 9：不同三维示例中经验谱密度估计量（经验值）和谱神经网络估计量（神经网络）的平均计算时间（秒）和最大内存使用量（MB）。对于神经网络，使用图形处理器时的计算时间以斜体显示在下一行。代码运行于一台配备 64 GiB 内存、AMD Ryzen 9 5900X（3.7 GHz）中央处理器、NVIDIA GeForce RTX 3090 GPU，且安装 Ubuntu 24.04.2 LTS（64 位）操作系统的计算机上。破折号（—）表示程序因内存不足而运行失败。

固定分辨率 $K = 15$，变化的样本量 N。

N	100		200		400		800		1600	
	Emp	NN	Emp	NN	Emp	NN	Emp	NN	Emp	NN
Fit	33.72	338.60	55.45	432.08	141.82	568.61	290.24	902.78	596。87	2362。70
		247.69		*271.48*		*316.58*		*405.20*		*627.75*
Eval	578。62	100。09	578.23	98.39	573.87	99.27	579.75	101.93	580.84	103.20
		1.97		*1.97*		*1.97*		*1.96*		*2.00*
总计	612.34	438.69	633.68	530.46	715。69	667。88	869.99	1004.71	1177.71	2465.90
		249.66		*273.45*		*318.55*		*407.16*		*629.75*
内存	2068	674	2069	699	2071	696	2074	698	2080	754

固定样本量 $N = 250$，分辨率变化 K。

K	10		15		20		25		30	
	Emp	NN	Emp	NN	Emp	NN	Emp	NN	Emp	NN
Fit	0.50	224.34	85.94	465.16	578。62	873。62	2292。85	1522。48	—	2782.41
		280.72		*279.68*		*286.78*		*292.13*		*317.63*
Eval	190.34	97.01	578.56	96.68	1402.81	99.14	2236.63	98.02	—	102.75
		1.98		*1.96*		*1.97*		*1.97*		*1.97*
总计	190.84	321.35	664.50	561.84	1981.44	972.76	4529.48	1620.50	—	2885.16
		282.70		*281.64*		*288.75*		*294.10*		*319.60*
内存	294	699	2070	698	11090	850	41961	1433	—	3162