arXiv:2602.00987v1 [cs.LG] 1 Feb 2026

可扩展随机小波特征：具有收敛保证的高效非平稳核近似

Sawan Kumar
印度理工学院德里分校应用力学系 sawan.kumar@am.iitd.ac.in

Souvik Chakraborty
印度理工学院德里分校应用力学系
Yardi 人工智能学院（ScAI）
souvik@am.iitd.ac.in

摘要

对非平稳过程进行建模——其统计属性会随输入域而变化——是机器学习中的一项关键挑战；然而，大多数可扩展方法都依赖于平稳性这一简化假设。这迫使人们面临艰难的权衡：要么采用深度高斯过程等表达能力强但计算开销大的模型，要么采用随机傅里叶特征（RFF）等可扩展但能力有限的方法。我们通过引入随机小波特征（RWF）弥合这一差距。RWF 是一种通过从小波族中采样来构建可扩展非平稳核近似的框架。借助小波固有的局部性和多分辨率结构，RWF 能够生成显式特征映射，以捕获复杂且依赖输入的模式。我们的框架为将 RFF 推广到非平稳设置提供了系统的方法，并附有全面的理论分析，包括正定性、无偏性和一致收敛保证。我们在一系列具有挑战性的合成数据集和真实世界数据集上进行了实证验证，结果表明，RWF 优于平稳随机特征；与更复杂的模型相比，它在准确度与效率之间实现了极具吸引力的权衡，从而为广泛的真实世界非平稳问题提供了可扩展且富有表达力的核方法。

1 引言

对复杂现实现象进行建模，是机器学习面临的核心挑战之一。地理空间建模（地形在不同区域间差异显著）和语音分析（信号会出现剧烈波动）等领域，往往具有显著的非平稳性，这意味着它们的统计属性会随输入空间而变化。高斯过程（GPs）为此类问题提供了有理论依据的框架，能够给出稳健的不确定性估计，并支持灵活的非参数建模 [1]。尽管具有这些优势，精确高斯过程仍有两大局限：其表达能力往往受核的选择所限制，而且计算成本会随训练点数量呈立方增长，因此不适用于现代大规模应用 [2]。

当前大多数方法都不得不在表达能力与效率之间做出权衡。一方面，随机傅里叶特征（RFF）等方法通过使用关于数据量呈线性增长的特征映射来近似核，从而实现了出色的可扩展性 [3, 4, 5]。然而，它们依赖博赫纳定理 [6]，这从根本上将其限制在平稳核上，而平稳核假设整个定义域内的行为一致。将这些模型应用于非平稳数据会导致系统性模型设定错误，进而损害预测准确率，并造成不确定性估计未经校准 [7, 8, 9]。另一方面，深度高斯过程 [10], 、谱混合 [11], 和输入依赖核 [12] 等表达能力强的模型能够捕捉非平稳性，但往往又带来高昂的计算成本、复杂的推断方案，以及优化和超参数调优方面的挑战。可扩展的平稳模型与复杂的非平稳模型之间的差距依然存在。

在这项工作中，我们通过引入**随机小波特征（RWF）**，弥补了这一差距。RWF 是一种可扩展且表达能力强的非平稳核近似框架。它不依赖 RFF 这类具有全局支撑的正弦基

及其变体，而是基于小波 函数族构造随机特征，这类函数在空间和频率上都具有天然的局部性。通过在随机尺度和平移下采样小波，RWF 可生成显式特征映射，以适应数据的局部特征。这种多分辨率结构使模型既能利用细尺度小波捕捉尖锐的局部事件，也能利用粗尺度小波同时建模平滑的长程趋势。由此得到一种有理论依据的方法，可将随机特征推广至非平稳设置，同时保留其优雅且高效的线性时间复杂度。我们的主要贡献有三点。首先，我们对 RWF 进行了全面的理论分析，包括诱导核的正定性、无偏性与方差界，以及带有明确样本复杂度的一致收敛保证。其次，我们表明 RWF 的训练复杂度为 $\mathcal{O}(ND^2)$，在保留随机特征方法可扩展性的同时，通过小波局部化直接编码非平稳性。最后，我们在合成数据、语音和大规模回归基准测试中进行实证验证，结果表明 RWF 始终优于平稳随机特征，并且在准确度与效率之间的权衡上至少能与更复杂的非平稳模型相媲美。

1.1 相关工作

可扩展核近似。 随机特征框架由 [5], 开创，他们展示了平移不变核可以通过随机傅里叶特征进行近似，且计算复杂度为线性。此后，该框架沿多个方向得到拓展，包括快速核学习等计算高效的变体 [13],、关于近似误差的理论保证 [14, 4, 15], 以及结构化采样方案 [16]。还有一些研究通过变分近似 [9], 、自适应特征学习 [17, 18], ，乃至与量子机器学习的联系 [19]，将随机傅里叶特征拓展到经典设定之外。近期进展将谱近似拓展至更广泛的核族，从而提升了可扩展特征映射的表达能力[20]。尽管这些方法实现了可扩展性，但它们依赖平稳的傅里叶基，因此难以捕捉非平稳或局部现象，而这些现象在许多科学领域中至关重要。

小波启发的近似。 此前，小波已用于核设计，例如小波支持向量机和小波核学习 [21, 22], ，其中核通过解析方式推导，或将小波变换用作预处理。最近， [23] 提出了一种基于固定小波基的贝叶斯核模型，用于高维贝叶斯线性回归。尽管这些方法展现了小波在捕捉局部结构方面的价值，但它们依赖固定或预定义的小波字典，既无法提供可扩展的蒙特卡洛近似，也不具备无偏性或一致收敛等理论保证。

混合与现代核学习。 目前已有多种方法用于捕捉高斯过程中的非平稳性。谱混合核 [24, 25] 将平稳谱表示的思想拓展至更丰富的结构，而深度高斯过程 [26, 10] 则通过层次组合增强灵活性。尽管这些模型表达能力很强，但其复杂度通常仍会随训练点数量大幅增长。除上述方向外，也有研究探索将可扩展近似与表达能力强的模型相结合。核插值方法（KISS-GP） [27]利用诱导点方法中的结构，而深度核学习 [28]则将神经网络与高斯过程相结合。近期，这些思路得到进一步拓展，包括用于时空建模的深度随机特征 [29], 、用于可扩展高斯过程的基于图的随机傅里叶特征 [30], ，以及自适应再生核希尔伯特空间表示中的组合核 [18]。尽管取得了这些进展，现有方法往往需要在可扩展性、表达能力和可解释性之间权衡。我们的工作立足于这一交汇点，旨在设计出既具备随机特征的可扩展性，又能实现灵活非平稳建模的特征映射。

2 预备知识与背景

本文简要回顾高斯过程回归（GPR）、稀疏变分高斯过程和随机特征高斯过程，为我们的小波构造奠定基础。

2.1 高斯过程

给定训练输入$\mathbf{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N]^\top \in \mathbb{R}^{N\times d}$和目标值 $\boldsymbol{y} \in \mathbb{R}^N$，我们考虑潜在函数 f上的高斯过程先验。假设观测值 y_n 是在相应输入$\boldsymbol{x}_n$处对该函数进行的带噪评估：

$$f \sim \mathcal{GP}(0, k), \qquad y_n = f(\boldsymbol{x}_n) + \varepsilon_n, \quad \text{where} \quad \varepsilon_n \sim \mathcal{N}(0, \sigma^2). \tag{2.1}$$

我们将协方差矩阵$\mathbf{K}_{XX} \in \mathbb{R}^{N\times N}$定义为 $[\mathbf{K}_{XX}]_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$。对数边际似然用于 f or 高斯过程超参数的训练由以下表达式给出：

$$\log p(\boldsymbol{y} \mid \mathbf{X}) = -\tfrac{1}{2}\boldsymbol{y}^\top(\mathbf{K}_{XX} + \sigma^2\mathbf{I}_N)^{-1}\boldsymbol{y} - \tfrac{1}{2}\log\det(\mathbf{K}_{XX} + \sigma^2\mathbf{I}_N) - \tfrac{N}{2}\log(2\pi). \tag{2.2}$$

对于测试输入 $\boldsymbol{x}_*$，令 $\boldsymbol{k}_{*X} = [k(\boldsymbol{x}_*, \boldsymbol{x}_1) \dots, k(\boldsymbol{x}_*, \boldsymbol{x}_N)]$, $k_{**} = k(\boldsymbol{x}_*, \boldsymbol{x}_*)$，并且

$$\boldsymbol{\alpha} = (\mathbf{K}_{XX} + \sigma^2 \mathbf{I}_N)^{-1} \boldsymbol{y}. \tag{2.3}$$

测试输入 $\boldsymbol{x}_*$ 的预测后验矩形式如下：

$$\mu_*(\boldsymbol{x}_*) = \boldsymbol{k}_{*X} \boldsymbol{\alpha}, \tag{2.4a}$$
$$\sigma_*^2(\boldsymbol{x}_*) = k_{**} - \boldsymbol{k}_{*X} (\mathbf{K}_{XX} + \sigma^2 \mathbf{I}_N)^{-1} \boldsymbol{k}_{X*}, \tag{2.4b}$$

其中 $\boldsymbol{k}_{X*} = \boldsymbol{k}_{*X}^\top$。精确推断的主要瓶颈在于其计算成本为 $O(N^3)$ 时间，内存占用为 $O(N^2)$。为应对这些挑战，文献中提出了多种方法；其中最常见的是高斯过程的稀疏近似。

随机变分高斯过程（SVGP）。在 SVGP 中，我们引入 M_u、诱导输入 $\mathbf{Z}_u = [\boldsymbol{z}_1, \dots, \boldsymbol{z}_{M_u}]^\top$ 和诱导变量 $\boldsymbol{u} = f(\mathbf{Z}_u)$，并为其指定先验 $p(\boldsymbol{u}) = \mathcal{N}(\mathbf{0}, \mathbf{K}_{uu})$，其中 $[\mathbf{K}_{uu}]_{ij} = k(\boldsymbol{z}_i, \boldsymbol{z}_j)$。定义 $\mathbf{K}_{fu} \in \mathbb{R}^{N \times M_u}$，其中 $[\mathbf{K}_{fu}]_{nm} = k(\boldsymbol{x}_n, \boldsymbol{z}_m)$，以及 $\mathbf{Q}_{ff} = \mathbf{K}_{fu} \mathbf{K}_{uu}^{-1} \mathbf{K}_{uf}$，条件先验变为

$$p(\boldsymbol{f} \mid \boldsymbol{u}) = \mathcal{N}\big(\mathbf{K}_{fu} \mathbf{K}_{uu}^{-1} \boldsymbol{u},\ \mathbf{K}_{ff} - \mathbf{Q}_{ff}\big). \tag{2.5}$$

高斯变分后验 $q(\boldsymbol{u}) = \mathcal{N}(\boldsymbol{m},\ \mathbf{S})$ 会诱导出 $q(\boldsymbol{f}) = \mathcal{N}(\boldsymbol{\mu},\ \boldsymbol{\Sigma})$，其中 $\boldsymbol{\mu} = \mathbf{A}\boldsymbol{m}$，且 $\boldsymbol{\Sigma} = \mathbf{K}_{ff} - \mathbf{Q}_{ff} + \mathbf{A}\mathbf{S}\mathbf{A}^\top$，其中 $\mathbf{A} = \mathbf{K}_{fu} \mathbf{K}_{uu}^{-1}$。在高斯似然 $p(y_n \mid f_n) = \mathcal{N}(y_n \mid f_n,\ \sigma^2)$ 下，证据下界可简化为

$$\mathcal{L} = \sum_{n=1}^{N} \mathbb{E}_{q(f_n)}[\log p(y_n \mid f_n)] - \mathrm{KL}(q(\boldsymbol{u}) \,\|\, p(\boldsymbol{u}))\,, \tag{2.6}$$

其中 $\mathbb{E}_{q(f_n)}[\log\ p(y_n \mid f_n)] = -\frac{1}{2}\sigma^{-2}\big[(y_n - \mu_n)^2 + \Sigma_{nn}\big] - \frac{1}{2}\log(2\pi\sigma^2)$。

使用大小为 b 的小批量 $\mathcal{B}$，可得到无偏估计量 $\mathcal{L} = (N/b) \sum_{n \in \mathcal{B}} \mathbb{E}_{q(f_n)}[\log\ p(y_n \mid f_n)] - \mathrm{KL}(q(\boldsymbol{u}) \| p(\boldsymbol{u}))$，每次迭代的复杂度为 $O(bM_u^2)$，另需一次性对 $\mathbf{K}_{uu}$ 进行 $O(M_u^3)$ 分解。

测试点 $k_{**} - \boldsymbol{k}_{*u}\ \ {}_{uu}^{-1}\boldsymbol{k}_{u*} + \boldsymbol{k}_{*u}\ \ {}_{uu}^{-1}\ \ {}_{uu}^{-1}\boldsymbol{k}_{u*}$ $\boldsymbol{k}_{*u} = [k\ \boldsymbol{x}_*\ \boldsymbol{z}_1\ \dots\ k\ \boldsymbol{x}_*\ \boldsymbol{z}_{M_u}]$ $\boldsymbol{x}_* = \mu_*\ \boldsymbol{x}_*$ 处的预测矩遵循高斯过程闭式方程：$(\ ) = \boldsymbol{k}_{*u}\mathbf{K}_{uu}^{-1}\boldsymbol{m}$ 以及 $\sigma_*^2(\ )\ \mathbf{KKSK}$，其中 (,),(,) 。

2.2 随机傅里叶特征高斯过程（RFF-GP）

由（**作者？**） [5]提出的随机傅里叶特征方法，使用显式特征映射来近似平稳核。考虑一个零均值高斯过程 $f \sim \mathcal{GP}(0, k)$，其核为平稳核 $k(\boldsymbol{x}, \boldsymbol{x}') = k(\boldsymbol{x} - \boldsymbol{x}')$。根据博赫纳定理 [6], ，该核具有如下谱表示：

$$k(\boldsymbol{x}, \boldsymbol{x}') = \int_{\mathbb{R}^d} e^{i\boldsymbol{\omega}^\top (\boldsymbol{x} - \boldsymbol{x}')} p(\boldsymbol{\omega})\, d\boldsymbol{\omega}, \tag{2.7}$$

其中 $p(\boldsymbol{\omega})$ 是核 k 的归一化谱密度。引入一个随机相位 $b \sim \mathrm{Unif}[0, 2\pi]$，可将其表示为余弦特征的期望：

$$k(\boldsymbol{x}, \boldsymbol{x}') = \mathbb{E}_{\boldsymbol{\omega}, b}\big[2\cos(\boldsymbol{\omega}^\top \boldsymbol{x} + b)\cos(\boldsymbol{\omega}^\top \boldsymbol{x}' + b)\big]\,. \tag{2.8}$$

用 D 蒙特卡洛样本 $\{(\boldsymbol{\omega}_j, b_j)\}_{j=1}^D$ 近似该期望，可得到随机特征映射 $\boldsymbol{z} : \mathcal{X} \to \mathbb{R}^D$，

$$\boldsymbol{z}(\boldsymbol{x}) = \frac{1}{\sqrt{D}}\big[\phi_1(\boldsymbol{x}), \dots, \phi_D(\boldsymbol{x})\big]^\top, \qquad \phi_j(\boldsymbol{x}) = \sqrt{2}\cos(\boldsymbol{\omega}_j^\top \boldsymbol{x} + b_j), \tag{2.9}$$

使得近似核函数为 $\hat{k}(\boldsymbol{x}, \boldsymbol{x}') = \boldsymbol{z}(\boldsymbol{x})^\top \boldsymbol{z}(\boldsymbol{x}')$.

从 GP 的角度来看，这相当于用有限维特征 $\boldsymbol{z}(\cdot)$ 替代无限维特征空间，从而得到一个贝叶斯线性回归模型。对权重赋予高斯先验 $\boldsymbol{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_D)$ 后，高斯后验的协方差和均值分别为，

$$\mathbf{S}_{\boldsymbol{w}} = \big(\mathbf{I}_D + \sigma^{-2}\mathbf{Z}^\top \mathbf{Z}\big)^{-1}, \tag{2.10a}$$
$$\boldsymbol{m}_{\boldsymbol{w}} = \sigma^{-2}\mathbf{S}_{\boldsymbol{w}}\mathbf{Z}^\top \boldsymbol{y}, \tag{2.10b}$$

其中 $\mathbf{Z} \in \mathbb{R}^{N\times D}$ 汇集了训练输入的特征映射。新测试点 $\boldsymbol{x}_*$ 的高斯预测分布，其均值和协方差定义为，

$$\mu_*(\boldsymbol{x}_*) = \boldsymbol{z}(\boldsymbol{x}_*)^\top \boldsymbol{m}_{\boldsymbol{w}}, \tag{2.11a}$$

$$\mathrm{Var}[y_* \mid \mathcal{D}] = \boldsymbol{z}(\boldsymbol{x}_*)^\top \mathbf{S}_{\boldsymbol{w}} \boldsymbol{z}(\boldsymbol{x}_*) + \sigma^2. \tag{2.11b}$$

因此，RFF-GP 框架为平稳核提供了一种可扩展的近似方法。然而，它依赖全局支持的傅里叶特征，限制了其建模非平稳性的能力。有关 RFF 的更多详情和示例，请参见附录 A.1

3 提出的方法

基于随机傅里叶特征的核近似方法利用博赫纳定理 [5] 为平稳核提供可扩展的近似，但难以捕捉非平稳性。稀疏变分高斯过程能够通过表达能力强的核建模非平稳性，但依赖诱导点集，且每次更新的计算成本为 M_u 立方级。我们提出随机小波特征（ RWF ），通过多分辨率、局部支持的小波构造非平稳核。通过对小波尺度和平移进行采样，RWF 提供显式特征映射 $z(\cdot)$，该映射：(i) 诱导出正定的非平稳核；(ii) 与 RFF-GP 一样，保持线性时间训练和预测；(iii) 捕捉平稳 RFF 所缺乏的局部多分辨率结构。

3.1 基于小波的核构造

为刻画非平稳性，核的性质必须随输入域而变化。平稳核通常通过随机傅里叶特征（RFF）近似，它依赖全局支撑的正弦基，因而具有空间不变性。相比之下，小波在空间和频率上都具有局部性，因此能自然地作为替代方案。通过随机化小波原子的尺度（控制频率）和平移（控制空间位置），我们可以构造灵活的非平稳核。

我们的构造从一个母小波 $\psi:\ \mathbb{R}^d \to \mathbb{R}$开始，它是一个均值为零且 L^2 范数为单位的函数（详情见附录 A.2）。由 ψ出发，我们通过各向同性缩放和平移生成一族小波原子：

$$\psi_{s,\boldsymbol{t}}(\boldsymbol{x}) = s^{-d/2}\psi\left(\frac{\boldsymbol{x}-\boldsymbol{t}}{s}\right), \quad \text{for scale } s>0 \text{ and shift } \boldsymbol{t}\in\mathbb{R}^d. \tag{3.1}$$

每个原子$\psi_{s,\ \boldsymbol{t}}$ 都是一个局部化的“波包”，以 $\boldsymbol{t}$ 为中心，其空间范围与 s成正比。令 $\Theta = (0,\ \infty) \times \mathbb{R}^d$为尺度和平移的参数空间。我们定义一个非平稳核，即相对于非负测度 $\mu(ds\, d\boldsymbol{t})$在该空间上积分：

$$k(\boldsymbol{x},\boldsymbol{y}) = \int_\Theta \psi_{s,\boldsymbol{t}}(\boldsymbol{x})\psi_{s,\boldsymbol{t}}(\boldsymbol{y})\,\mu(ds\, d\boldsymbol{t}). \tag{3.2}$$

该构造保证正定性，因为被积函数是标量特征的乘积。若 μ 具有密度$p(s,\ \boldsymbol{t}) \geq 0$，则核变为：

$$k(\boldsymbol{x},\boldsymbol{y}) = \int_0^\infty \int_{\mathbb{R}^d} \psi_{s,\boldsymbol{t}}(\boldsymbol{x})\psi_{s,\boldsymbol{t}}(\boldsymbol{y})\, p(s,\boldsymbol{t})\, d\boldsymbol{t}\, ds. \tag{3.3}$$

密度 $p(s,\ \boldsymbol{t})$ 决定了核的性质。一种常见选择是因式分解形式 $p(s,\ \boldsymbol{t}) = p_s(s)p_t(\boldsymbol{t})$，其中$p_s$ （例如对数均匀分布）涵盖多个分辨率，而 p_t （例如在数据的凸包上均匀分布）提供空间覆盖。

3.2 随机小波特征采样策略

方程 3.3 中的积分通常难以求解。我们通过蒙特卡洛采样对其进行近似，这构成了随机特征的基础。

定义 3.1（随机小波特征）。从密度为 $p(s,\ \boldsymbol{t})$ 的分布中独立同分布地采样 $(s_i,\ \boldsymbol{t}_i)_{i=1}^D$，并将随机特征映射 $z:\ \mathbb{R}^d \to \mathbb{R}^D$ 定义为：

$$z(\boldsymbol{x}) = \frac{1}{\sqrt{D}}\left[\psi_{s_1,\boldsymbol{t}_1}(\boldsymbol{x}),\ldots,\psi_{s_D,\boldsymbol{t}_D}(\boldsymbol{x})\right]^\top. \tag{3.4}$$

相应的核近似为 $\hat{}k(\boldsymbol{x},\ \boldsymbol{y}) = z(\boldsymbol{x})^\top z(\boldsymbol{y})$。

$\hat{}$根据构造，$k(\boldsymbol{x},\ \boldsymbol{y})$ 是 $k(\boldsymbol{x},\ \boldsymbol{y})$ 的无偏估计量。该表述将核方法转化为贝叶斯线性模型，从而能够高效地进行训练和预测。完整流程详见算法1。

预印本。审阅中。

Alg算法 1RWF-GP 训练　　g 与预测

1：输入：训练数据 (X $\boldsymbol{y}$)、测试输入 X$_*$、特征数量 D、小波 ψ、采样分布 $p(s,\boldsymbol{t})$。 2：超参数：噪声方差 σ^2、 $p(s,\boldsymbol{t})$ 的参数。 3：训练： 4：对 $i = 1,\ldots,D$，从 $\sim p(s,\boldsymbol{t})$ 中采样 $(s_i,\boldsymbol{t}_i)$。 5：构造特征矩阵 $Z \in \mathbb{R}^{N\times D}$，其中 $Z_{ni} = \frac{1}{\sqrt{D}}\psi_{s_i,\boldsymbol{t}_i}(\boldsymbol{x}_n)$。 6：计算权重后验： $S_{\boldsymbol{w}} = (I_D + \sigma^{-2}Z^\top Z)^{-1}$ 和 $\boldsymbol{m}_{\boldsymbol{w}} = \sigma^{-2}S_{\boldsymbol{w}}Z^\top\boldsymbol{y}$。 7：通过最大化贝叶斯线性模型的边际似然来优化超参数（例如 σ^2、 p 的参数）。 8：预测： 9：构造测试特征矩阵 $Z_* \in \mathbb{R}^{N_*\times D}$，其中 $[Z_*]_{ji} = \frac{1}{\sqrt{D}}\psi_{s_i,\boldsymbol{t}_i}(\boldsymbol{x}_{*,j})$。 10：计算预测均值： $\boldsymbol{\mu}_* = Z_*\boldsymbol{m}_{\boldsymbol{w}}$。 11：计算预测方差： $\boldsymbol{\sigma}_*^2 = \mathrm{diag}(Z_*S_{\boldsymbol{w}}Z_*^\top) + \sigma^2$。 12：输出：预测分布 $\mathcal{N}(\boldsymbol{\mu}_*,\boldsymbol{\sigma}_*^2)$。

—

3.3 实践考量

计算复杂度。 RWF 之所以高效，是因为计算成本随数据集规模线性增长。在维度为 d 的 N 个输入上构造 D 随机小波特征的成本为 $\mathcal{O}(NDd)$，之后训练可归约为在 D 维特征空间中的原始形式高斯过程回归。构建 $Z^\top Z$ 需要 $\mathcal{O}(ND^2)$，所得的 $D\times D$ 系统可在 $\mathcal{O}(D^3)$ 中求解，因此对于 $N \gg D$，总体训练成本主要由 $\mathcal{O}(ND^2)$ 决定。每个测试点的预测需要 $\mathcal{O}(D^2)$。相比之下，精确高斯过程的复杂度为 $\mathcal{O}(N^3)$，而 SVGP 由于迭代式变分更新，每个优化步骤会产生 $\mathcal{O}(NM^2)$ 的成本。RWF 通过一次闭式求解计算其后验分布，从而在大规模非平稳学习中实现显著的实际运行时间加速。

建模非平稳性的关键在于对小波族和采样分布进行实际选择。母小波的选择 ψ （例如，用于时频分析的莫雷小波，或用于表示急剧变化的道贝希小波）以及采样分布 $p(s,\ \boldsymbol{t})$（例如，尺度采用对数均匀分布，平移采用均匀分布） [31, 32] 使模型能够适应多分辨率信号结构。为了实现稳定训练，建议通过限制尺度的采样范围并对线性模型应用权重衰减来对模型进行正则化。

4 理论分析

为了分析近似的质量，我们建立一致收敛保证。我们的分析依赖于界定由小波特征诱导的函数类的复杂度。我们定义以下关键量：$B = \sup_{s,t,\boldsymbol{x}}|\psi_{s,t}(\boldsymbol{x})|$ 作为特征幅值的一致上界，以及 $K = \sup_{\boldsymbol{x}} k(\boldsymbol{x},\boldsymbol{x}')$ 作为最大核值。

4.1 小波核的正定性

定理 4.1（基于小波的核的正定性）。 设 ψ： $\mathbb{R}^d \to \mathbb{R}$ 为一个母小波函数，并将小波族定义为 $\psi_{s,\ t}(\boldsymbol{x}) = s^{-d/2}\psi\left(s^{-1}(\boldsymbol{x}-t)\right)$，其中尺度为 $s>0$，平移为 $t\in\mathbb{R}^d$。设 $p(s,\ t)$： $\mathbb{R}^+\times\mathbb{R}^d \to [0,\infty)$ 为一个非负测度，使得对于所有 $\boldsymbol{x}$、 $\boldsymbol{y}\in\mathbb{R}^d$，积分均有定义且为有限值。则函数

$$k(\boldsymbol{x},\boldsymbol{y}) = \int_{\mathbb{R}^+}\int_{\mathbb{R}^d}\psi_{s,t}(\boldsymbol{x})\psi_{s,t}(\boldsymbol{y})\,p(s,t)\,dt\,ds \tag{4.1}$$

是 $\mathbb{R}^d\times\mathbb{R}^d$ 上的正定核。

（证明见附录 A.4。）

4.2 无偏性与方差界

引理 4.1（无偏性）。 对所有 $\boldsymbol{x}$、 $\boldsymbol{y}\in\mathcal{X}$，小波随机特征近似是无偏的： $\mathbb{E}[k(\boldsymbol{x},\boldsymbol{y})] = k(\boldsymbol{x},\boldsymbol{y})$。

（证明见附录 A.5。）

引理 4.2（方差界）。 对于所有 $\boldsymbol{x}$、 $\boldsymbol{y} \in \mathcal{X}$，近似的方差有界 d:

$$\operatorname{Var}\left[\hat{k}(\boldsymbol{x}, \boldsymbol{y})\right] \leq \frac{B^2}{D}. \tag{4.2}$$

（证明见附录 A.6。）

4.3 一致收敛保证

定理 4.2（随机小波特征的一致收敛）。 设 $\mathcal{M} \subset \mathbb{R}^d$ 为直径为 diam($\mathcal{M}$ k $\boldsymbol{x}$ $\boldsymbol{y}$ 4 l, z $\mathbb{R}^d \to$)的紧集。设 (,) 为定理 中的正定核，并通过 $\mathbb{R}^D$ s_i t_i $\sim p$ $i = 1 \ldots D$ 独立采样 (,) 为 定义随机特征映射: ，其中 ，并令

$$z(\boldsymbol{x}) = \frac{1}{\sqrt{D}} \left[\psi_{s_1,t_1}(\boldsymbol{x}), \ldots, \psi_{s_D,t_D}(\boldsymbol{x})\right]^\top. \tag{4.3}$$

假设 k 且特征映 射为利普希茨连续，常数分别为 L_k 和 L_z。那么，对任意 $\epsilon > 0$,

$$\Pr\left[\sup_{\boldsymbol{x},\boldsymbol{y} \in \mathcal{M}} |z(\boldsymbol{x})^\top z(\boldsymbol{y}) - k(\boldsymbol{x}, \boldsymbol{y})| \geq \epsilon\right] \leq 2\left(\frac{4\operatorname{diam}(\mathcal{M}) L_z}{\epsilon}\right)^{2d} \exp\left(-\frac{D\epsilon^2}{8B^2}\right). \tag{4.4}$$

（证明见附录 *A.7*。）

4.4 样本复杂度分析

上述定理给出了明确的样本复杂度界。要以至少 $1 - \delta$ 的概率达到 ϵ 的近似误差，只需选择

$$D \geq \frac{8B^2}{\epsilon^2}\left(2d \log\left(\frac{4\operatorname{diam}(\mathcal{M}) L_z}{\epsilon}\right) + \log\left(\frac{2}{\delta}\right)\right). \tag{4.5}$$

这一结果是通过对定理 4.2 中的概率界求逆得到的。常数 B 和 L_z 取决于小波的选择，相关讨论见附录 A.2。这表明，所需特征数量随期望的准确度和置信水平呈对数增长。

5 实验

本节通过实验评估所提出方法的性能。我们首先考察该方法在非平稳合成数据上的近似质量，然后在高度非平稳的语音信号数据集和基准测试回归任务上对其进行评估，并与各种基线模型进行比较。所有实验的更多细节见附录 B。

基线方法：我们将模型与具有可扩展性和/或表达能力的变体进行比较：SVGP [8], RFF-GP [5], 深度高斯过程 [10], 以及精确高斯过程（在可行时）；此外，还与专用于非平稳数据的高斯过程进行比较：谱混合核 [20], DRF [29],IDD-GP [12], 和自适应 RKHS 傅里叶特征高斯过程 [18]。

5.1 合成数据上的评估

我们首先在一个非平稳多步函数上评估 RWF；在这种情形下，采用平稳核的浅层高斯过程无法捕捉依赖输入的变化 [12]。深度高斯过程虽然表达能力更强，却难以处理尖锐不连续性。相比之下，RWF 使浅层高斯过程能够准确拟合：图1显示，RWF-GP 能够捕捉非平稳结构，而基线方法由于核灵活性有限，产生的拟合结果过于平滑或呈振荡状。表1展示了所提方法在准确度和训练时间方面均优于其他方法。图2汇总了各对比方法的实际运行时间和内存占用，体现了所提方法的可扩展性。附录 C.1 展示了消融研究，说明所提方法如何随特征规模增大而收敛。

5.2 TIMIT 语音信号

我们按照以往基于高斯过程的研究[18]，在一个源自 TIMIT 语料库的回归任务上评估我们的方法。TIMIT 带来了挑战，因为音频信号存在强烈的非平稳性，例如局部的辅音爆破音

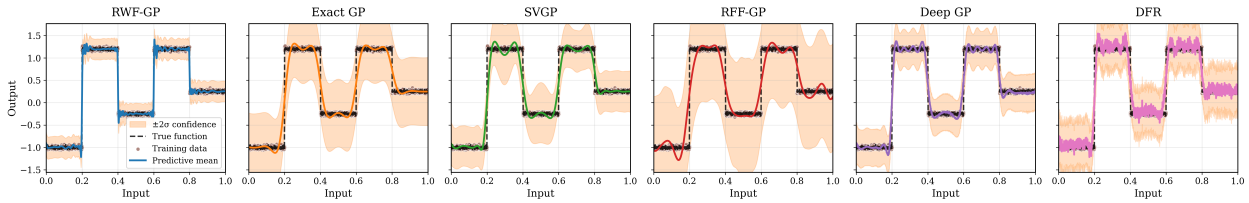


图1：不同高斯过程方法在阶跃函数回归任务上的预测性能。每个面板展示预测均值（实线）以及 $\pm 2\sigma$ 置信区间（阴影）和训练数据（点）。**RWF-GP**（本文方法）能清晰捕捉不连续点，同时保持校准的不确定性。相比之下，精确高斯过程、稀疏变分高斯过程与RFF-GP难以处理急剧变化，要么过度平滑，要么错误校准不确定性。

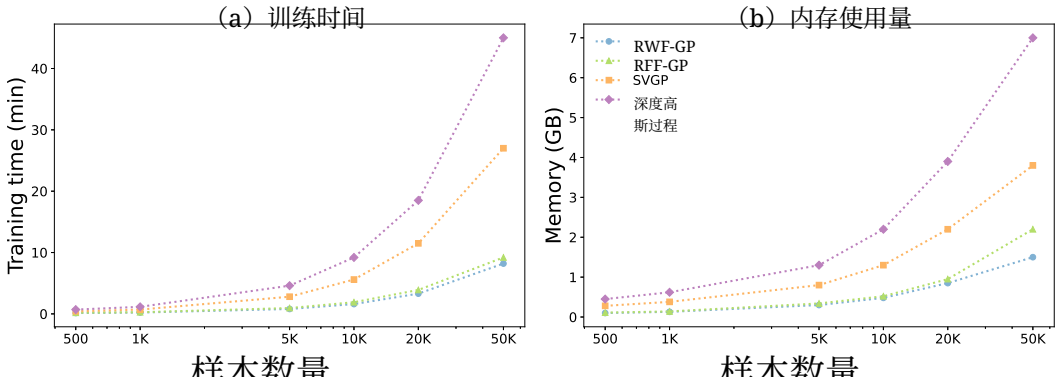


图2：多步函数上的可扩展性。多步函数上时间与内存随训练样本数量的变化：RWF效率最高；SVGP与深度高斯过程的开销更高。

表1：五次运行中各高斯过程基线方法在多步函数上的性能比较（均值 ± 标准差；越低越好）。**粗体** 表示最佳结果，下划线表示次佳结果。方法：~~精确方法~~ = 精确高斯过程，SVGP= 随机变分高斯过程，RFF = 随机傅里叶特征，DRF = Deep-RF GP，DGP =深度高斯过程，SM = 谱混合高斯过程，IDD = 域间深度高斯过程，A-RKHS = 自适应再生核希尔伯特空间高斯过程。SM、IDD 和 A-RKHS 的结果来自 [18]。

	精确方法	SVGP	RFF	DRF	DGP	SM	IDD	A-RKHS	RWF（本文方法）
RMSE	0.190 ±0.091	0.231 ±0.014	0.246 ±0.142	0.190 ±0.120	0.162 ±0.110	0.210 ±0.085	0.107 ±0.050	0.095 ±0.045	**0.071 ±0.011**
CRPS	0.215 ±0.030	0.392 ±0.025	0.238 ±0.041	0.205 ±0.032	0.187 ±0.028	0.201 ±0.030	0.143 ±0.020	0.131 ±0.018	**0.112 ±0.010**
NLL	0.042 ±0.012	0.123 ±0.018	0.118 ±0.181	-0.018 ±0.216	-0.268 ±0.211	0.220 ±0.180	-0.820 ±0.080	-1.210 ±0.075	**-1.879 ±0.061**
Time	12	15	11	18	20	17	17	11	**9**

以及缓慢变化的区域。依赖平稳核的模型难以捕捉这些变化，除非过度平滑或使用大量特征。与 RFF 不同，RWF 会自适应地分配分辨率：较小尺度捕捉尖锐的起音，较大尺度捕捉平滑区域，从而在特征规模固定时降低近似方差。**结果：**表2报告了均方根误差和训练时间。与基线方法相比，RWF-GP 的误差最低。使用相同数量的特征 D 时，RFF-GP 的表现最差，这反映出它无法有效覆盖局部谱平移。深度高斯过程和 Deep-RF GP 能够捕捉非平稳性，但训练时间更长。自适应再生核希尔伯特空间方法的表现具有竞争力，但在准确度与时间的权衡上仍逊于 RWF。有关实验的更多细节见附录 C.2。

表2：TIMIT 回归：均方根误差、连续排序概率评分和负对数似然（5 次运行的均值 ± 标准差），以及训练时间。**粗体** 表示最佳结果，下划线表示次佳结果。

	精确方法	RFF	SVGP	DGP	DRF	IDD	A-RKHS	RWF（本文方法）
RMSE	2.10±0.008	2.13±0.004	2.28±0.005	0.98±0.005	0.54±0.005	0.57±0.015	0.48±0.003	**0.42±0.003**
CRPS	1.92±0.020	1.95±0.018	2.10±0.025	0.85±0.015	0.49±0.010	0.51±0.014	0.44±0.009	**0.39±0.006**
NLL	3.25±0.02	3.31±0.10	3.52±0.14	1.82±0.09	1.12±0.06	0.84±0.07	0.75±0.05	**0.56±0.04**
Time	133	110	120	131	140	126	141	**90**

5.3 在 UCI 数据集上的性能

为评估模型在合成任务和领域特定任务之外的泛化能力，我们在 UCI 仓库 [33], 中广泛应用于高斯过程文献的七个标准回归数据集上进行基准测试。这些数据集涵盖了不同的输入维度和样本量，可用于评估适应性和可扩展性。按照公认的做法[10, 12, 34], , 我们采用 90/10 的训练–测试划分，对输入进行归一化，并对输出进行标准化。**结果**。表3报告了均方根误差和训练时间。RWF-GP 的预测性能始终出色，在七个数据集中有五个取得最低误差，在其余两个数据集上也具有竞争力。深度高斯过程和 Deep-RF GP 能够捕捉部分非平稳性，但训练时间更长。谱混合核在部分数据集上带来了一定提升。

表 3：UCI 回归基准上的性能：均方根误差、连续排序概率评分、负对数似然和训练时间（分钟）。**粗体** 表示最佳结果，下划线表示次佳结果。

	Data	能源 1k	混凝土 1k	翼型 1.5k	股票 5k	运动 KIN8NM 8k	8k	海军 11k
RMSE	RFF	0.66±0.03	6.72±0.50	5.34±0.29	1.86±0.03	1.60±0.02	0.41±0.02	0.13±0.002
	SVGP	0.68±0.02	5.92±0.17	5.18±0.07	2.13±0.03	1.87±0.03	0.10±0.02	0.12±0.001
	DRF	0.58±0.04	5.01±0.01	3.45±0.11	0.95±0.04	**0.44±0.03**	0.12±0.03	0.08±0.001
	DGP	0.48±0.03	4.55±0.18	3.66±0.08	0.90±0.03	1.39±0.02	**0.09±0.02**	0.04±0.003
	SM	0.67±0.03	5.80±0.19	3.90±0.09	0.92±0.04	1.62±0.03	0.11±0.02	0.06±0.001
	IDD	0.55±0.04	**4.20±0.08**	3.30±0.09	0.88±0.04	1.48±0.03	0.28±0.01	0.07±0.002
	A-RKHS	0.51±0.02	4.35±0.12	3.25±0.10	0.86±0.03	1.46±0.03	0.18±0.01	0.04±0.001
	RWF（本文方法）	**0.42±0.02**	4.45±0.15	**3.20±0.08**	**0.84±0.03**	1.55±0.01	**0.09±0.01**	**0.02±0.001**
CRPS	RFF	0.61±0.02	6.21±0.40	4.92±0.25	1.72±0.03	1.51±0.02	0.32±0.01	0.11±0.001
	SVGP	0.58±0.02	5.20±0.15	4.01±0.06	1.70±0.03	1.48±0.02	0.11±0.01	0.04±0.001
	DRF	0.52±0.03	4.68±0.01	3.12±0.09	0.88±0.03	**0.40±0.02**	0.10±0.02	0.06±0.001
	DGP	0.43±0.02	4.25±0.15	3.28±0.07	0.81±0.02	1.32±0.02	0.08±0.01	0.03±0.002
	SM	0.63±0.03	5.50±0.18	3.65±0.08	0.84±0.03	1.53±0.03	0.09±0.01	0.05±0.001
	IDD	0.49±0.03	**3.98±0.07**	3.01±0.08	0.80±0.03	1.42±0.02	0.22±0.01	0.06±0.002
	A-RKHS	0.46±0.02	4.10±0.10	2.95±0.09	0.78±0.03	1.41±0.03	0.15±0.01	0.03±0.001
	RWF（本文方法）	**0.38±0.02**	4.28±0.12	**2.88±0.07**	**0.76±0.03**	1.47±0.01	**0.07±0.01**	**0.02±0.001**
NLL	RFF	1.92±0.08	6.85±0.45	4.92±0.21	2.02±0.05	1.72±0.03	0.56±0.03	0.18±0.002
	SVGP	1.80±0.07	6.10±0.30	4.25±0.12	1.98±0.04	1.68±0.03	0.32±0.02	0.10±0.002
	DRF	1.62±0.05	5.30±0.15	3.45±0.10	1.32±0.04	**0.82±0.04**	0.42±0.03	0.15±0.001
	DGP	1.40±0.05	5.01±0.22	3.68±0.09	1.29±0.03	1.32±0.03	0.30±0.02	0.08±0.002
	SM	1.98±0.08	5.90±0.20	3.88±0.10	1.36±0.04	1.55±0.03	0.33±0.02	0.12±0.002
	IDD	1.52±0.06	**4.85±0.15**	3.20±0.08	1.27±0.04	1.48±0.03	0.70±0.03	0.14±0.003
	A-RKHS	1.48±0.04	5.01±0.18	3.12±0.09	1.25±0.04	1.43±0.03	0.33±0.02	0.08±0.001
	RWF（本文方法）	**1.32±0.04**	5.10±0.16	**2.05±0.08**	**1.20±0.03**	1.41±0.02	**0.28±0.02**	**0.06±0.001**
Time	RFF	14	10	10.4	16	14	14	30
	SVGP	14	12	10	12.2	18	23	36
	DRF	15	16	10.2	15	18	16	33
	DGP	15.6	13	17.8	20	20	27	35
	SM	17	18	19	12.3	21	24	24
	IDD	11.1	11	20	16	19	22	29
	A-RKHS	15.3	17	15	15	22	30	35
	RWF（本文方法）	**9**	**8**	**9.6**	**10**	**12**	**9**	**24**

5.4 Protein 数据集

Protein 数据集包含约 4.5 万个样本和 9 个实值输入特征，这些特征源自生物领域，可作为回归任务的实用基准。该数据集用于评估模型在生物数据分析中常见的噪声环境下的性能。表 4 报告了实验结果。
RWF-GP 的表现最佳，且所需训练时间最短，优于其他基线方法。

表 4：Protein 数据集（45K 个样本）的结果。我们报告 RMSE、连续排序概率评分和负对数似然（5 次运行的均值 ± 标准差）以及训练时间（分钟）。**粗体**表示最佳结果，下划线表示次佳结果。

	RFF	SVGP	DRF	DGP	SM	IDD	A-RKHS	RWF（本文方法）
RMSE	5.41±0.01	5.40±0.01	4.65±0.14	4.35±0.01	4.55±0.02	4.42±0.01	4.32±0.01	**4.25±0.02**
CRPS	4.92±0.04	4.88±0.03	4.12±0.10	3.86±0.02	4.01±0.05	3.90±0.03	3.78±0.02	**3.65±0.02**
NLL	3.98±0.06	3.92±0.05	3.21±0.08	2.89±0.04	3.05±0.06	2.95±0.05	2.82±0.03	**2.71±0.03**
时间（分钟）	95	120	130	120	133	129	130	**90**

6 结论

我们提出了随机小波特征（RWF），这是一种可扩展且理论严谨的框架，用于表达能力强的非平稳核近似。与深度高斯过程和自适应卷积核等计算开销较大的模型相比，RWF 实现了效率与表达能力之间难得的平衡。借助随机小波族，RWF 能够显式编码复杂现实过程所固有的局部多分辨率模式。我们为 RWF 建立了严谨的理论保证，包括正定性、无偏性和一致收敛，为其奠定了坚实的数学基础。大量实验表明，RWF 不仅能轻松处理非平稳任务，还能持续优于先进的复杂基线方法。RWF 为可扩展核学习树立了新标准；未来若能结合自适应小波采样和深度核架构，应用范围和影响力有望进一步扩大。

伦理声明

本研究介绍了利用随机小波特征实现可扩展随机特征和核近似的方法学进展。本研究未涉及人类受试者、个人身份信息或敏感数据。所有实验均使用公开数据集。该方法旨在用于科学研究；其更广泛的影响是间接的，并取决于具体领域的应用。

致谢

SK 感谢教育部（MoE）以研究奖学金形式提供的支持。SC 感谢阿努桑丹国家研究基金会（ANRF）通过编号为 CRG/2023/007667 的项目提供的资金支持。

代码可用性

论文获接收后，通讯作者将在 GitHub 上公开本研究中用于复现结果的全部源代码。

参考文献

[1] Christopher KI Williams 和 Carl Edward Rasmussen。机器学习中的高斯过程，第 2 卷。麻省理工学院出版社，马萨诸塞州剑桥，2006 年。[2] Haitao Liu、Yew-Soon Ong、Xiaobo Shen 和 Jianfei Cai。高斯过程遇上大数据：可扩展高斯过程综述。电气与电子工程师协会神经网络与学习系统汇刊，31(11)：4405–4423，2020 年。[3] Kurt Cutajar、Edwin V. Bonilla、Pietro Michiardi 和 Maurizio Filippone。深度高斯过程的随机特征展开。载于 Doina Precup 和 Yee Whye Teh 编，第 *34* 届国际机器学习会议论文集，第 70 卷，机器学习研究论文集，第 884–893 页。PMLR，2017 年。[4] Haim Avron、Michael Kapralov、Cameron Musco、Christopher Musco、Ameya Velingker 和 Amir Zandieh。用于核岭回归的随机傅里叶特征：近似界与统计保证。载于 Doina Precup 和 Yee Whye Teh 编，第 *34* 届国际机器学习会议论文集，第 70 卷，机器学习研究论文集，第 253–262 页。PMLR，2017 年 8 月 6–11 日。

[5] Ali Rahimi与Benjamin Recht。面向大规模核机器的随机特征。神经信息处理系统进展，20，2007。

[6] 萨洛蒙·博赫纳。调和分析与概率论。Courier Corporation，2005。[7] Talay M Cheema与Carl Edward Rasmussen。用于结合高斯过程进行快速空间建模的积分变分傅里叶特征。《机器学习研究汇刊》，2024。[8] James Hensman、Nicolo Fusi与Neil D Lawrence。面向大数据的高斯过程。arXiv预印本arXiv:1309.6835，2013。[9] James Hensman、Nicolas Durrande与Arno Solin。高斯过程的变分傅里叶特征。《机器学习研究杂志》，18(151):1–52，2018。[10] Hugh Salimbeni、Vincent Dutordoir、James Hensman与Marc Deisenroth。采用重要性加权变分推断的深度高斯过程。见《国际机器学习会议》，第5589–5598页。PMLR，2019。[11] Anthony Tompkins、Rafael Oliveira与Fabio T Ramos。用于非平稳核学习的稀疏谱扭曲输入测度。《神经信息处理系统进展》，33:16153–16164，2020。[12] Tim GJ Rudner、Dino Sejdinovic与Yarin Gal。域间深度高斯过程。见《国际机器学习会议》，第8286–8294页。PMLR，2020。[13] Andrew G Wilson、Elad Gilboa、Arye Nehorai与John P Cunningham。用于多维模式外推的快速核学习。《神经信息处理系统进展》，27，2014。[14] Bharath Sriperumbudur与Zolt´an Szab´o。随机傅里叶特征的最优速率。《神经信息处理系统进展》，28，2015。[15] Zhu Li、Jean-Francois Ton、Dino Oglic与Dino Sejdinovic。迈向随机傅里叶特征的统一分析。《机器学习研究杂志》，22(108):1–51，2021。[16] Krzysztof Choromanski等。面向可扩展策略优化的紧凑架构结构化演化。见ICML，2017。[17] Xiantong Zhen、Haoliang Sun、Yingjun Du、Jun Xu、Yilong Yin、Ling Shao与Cees Snoek。利用变分随机特征学习如何学习核。见《国际机器学习会议》，第11409–11419页。PMLR，2020。[18] Xinxing Shi、Thomas Baldwin-McDonald与Mauricio A Alvarez。用于组合高斯过程模型的自适应再生核希尔伯特空间傅里叶特征。arXiv预印本arXiv:2407.01856，2024。[19] Jonas Landman、Slimane Thabet、Constantin Dalyac、Hela Mhiri与Elham Kashefi。用随机傅里叶特征对变分量子机器学习进行经典近似。arXiv预印本arXiv:2210.13200，2022。[20] Nicolas Langren´e、Xavier Warin与Pierre Gruet。用于推广随机傅里叶特征的各向同性核谱混合表示。arXiv预印本arXiv:2411.02770，2024。[21] Li Zhang、Weida Zhou与Licheng Jiao。小波支持向量机。《IEEE系统、人类与控制论汇刊B辑（控制论）》，34(1):34–39，2004。[22] Florian Yger与Alain Rakotomamonjy。小波核学习。《模式识别》，44(10-11):2614–2629，2011。[23] Wenxing Guo、Xueying Zhang、Bei Jiang、Linglong Kong与Yaozhong Hu。面向高维数据分析的基于小波的贝叶斯近似核方法。《计算统计学》，39(4):2323–2341，2024。[24] Andrew Wilson与Ryan Adams。用于模式发现与外推的高斯过程核。见《国际机器学习会议》，第1067–1075页。PMLR，2013。[25] Sami Remes、Markus Heinonen与Samuel Kaski。非平稳谱核。《神经信息处理系统进展》，30，2017。[26] Andreas Damianou与Neil D Lawrence。深度高斯过程。见《人工智能与统计学》，第207–215页。PMLR，2013。[27] Andrew Wilson与Hannes Nickisch。面向可扩展结构化高斯过程的核插值（KISS-GP）。见《国际机器学习会议》，第1775–1784页。PMLR，2015。[28] Andrew Gordon Wilson、Zhiting Hu、Ruslan Salakhutdinov与Eric P Xing。深度核学习。见《人工智能与统计学》，第370–378页。PMLR，2016。[29] Weibin Chen、Azhir Mahmood、Michel Tsamados与So Takao。用于时空数据可扩展插值的深度随机特征。见《第十三届国际学习表征会议》，2024。[30] Matthew Zhang、Jihao Andreas Lin、Adrian Weller、Richard E Turner与Isaac Reid。用于可扩展高斯过程的图随机特征。arXiv预印本arXiv:2509.03691，2025年。

预印本。审稿中。

[31] James Bergstra 和 Yoshua Bengio。用于超参数优化的随机搜索。《机器学习研究杂志》，13(1):281–305，2012。[32] Harold Jeffreys。估计问题中先验概率的一种不变形式。《伦敦皇家学会会刊：*A*辑·数学与物理科学》，186(1007):453–461，1946。[33] Dheeru Dua 和 Casey Graff。UCI机器学习库，2019。加利福尼亚州尔湾：加利福尼亚大学信息与计算机科学学院。[34] Thomas McDonald 和 Mauricio ´Alvarez。利用基于 ODE 的随机特征对非线性动力系统进行组合建模。神经信息处理系统进展，34:13809–13819，2021。[35] Jacob Gardner、Geoff Pleiss、Kilian Q Weinberger、David Bindel 和 Andrew G Wilson。GPyTorch：利用 GPU 加速进行黑盒矩阵-矩阵高斯过程推断。神经信息处理系统进展，31，2018。[36] Christopher Paciorek 和 Mark Schervish。用于高斯过程回归的非平稳协方差函数。神经信息处理系统进展，16，2003。

A 高斯过程的随机特征

A.1 平稳核的随机傅里叶特征

设 k： $\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ 为一个平稳核，即 $k(\boldsymbol{x},\ \boldsymbol{x}') = k(\boldsymbol{x}-\boldsymbol{x}')$。根据博赫纳定理，$k$ 可用谱密度 $p(\boldsymbol{\omega})$ 表示为如下形式：

$$k(\boldsymbol{x}-\boldsymbol{x}') = \int_{\mathbb{R}^d} e^{i\boldsymbol{\omega}^\top(\boldsymbol{x}-\boldsymbol{x}')}\, p(\boldsymbol{\omega})\, d\boldsymbol{\omega}. \tag{A.1}$$

等价地，

$$k(\boldsymbol{x}-\boldsymbol{x}') = \mathbb{E}_{\boldsymbol{\omega}\sim p(\boldsymbol{\omega})}\left[e^{i\boldsymbol{\omega}^\top\boldsymbol{x}}\, e^{-i\boldsymbol{\omega}^\top\boldsymbol{x}'}\right]. \tag{A.2}$$

将复指数展开为正弦和余弦项，可得

$$k(\boldsymbol{x}-\boldsymbol{x}') = \mathbb{E}_{\boldsymbol{\omega}\sim p(\boldsymbol{\omega})}\Big[\cos\big(\boldsymbol{\omega}^\top\boldsymbol{x}\big)\cos\big(\boldsymbol{\omega}^\top\boldsymbol{x}'\big) + \sin\big(\boldsymbol{\omega}^\top\boldsymbol{x}\big)\sin\big(\boldsymbol{\omega}^\top\boldsymbol{x}'\big)\Big]. \tag{A.3}$$

引入辅助随机相位 $b \sim \mathrm{Unif}[0, 2\pi]$，可将其改写为

$$k(\boldsymbol{x}-\boldsymbol{x}') = \mathbb{E}_{\boldsymbol{\omega},b}\left[2\cos\big(\boldsymbol{\omega}^\top\boldsymbol{x}+b\big)\cos\big(\boldsymbol{\omega}^\top\boldsymbol{x}'+b\big)\right]. \tag{A.4}$$

因此，使用 M 个样本 $\{\boldsymbol{\omega}_m\}_{m=1}^M$ 得到无偏蒙特卡洛近似

$$k(\boldsymbol{x}-\boldsymbol{x}') \approx \frac{2}{M}\sum_{m=1}^{M}\cos\big(\boldsymbol{\omega}_m^\top\boldsymbol{x}+b_m\big)\cos\big(\boldsymbol{\omega}_m^\top\boldsymbol{x}'+b_m\big), \tag{A.5}$$

其中 $\boldsymbol{\omega}_m \sim p(\boldsymbol{\omega})$，且 $b_m \sim$ 服从均匀分布 $[0, 2\pi]$。

这自然引出了随机特征映射

$$\boldsymbol{\phi}(\boldsymbol{x}) = \sqrt{\tfrac{2}{M}}\begin{bmatrix}\cos(\boldsymbol{\omega}_1^\top\boldsymbol{x}+b_1)\\ \cos(\boldsymbol{\omega}_2^\top\boldsymbol{x}+b_2)\\ \vdots \\ \cos(\boldsymbol{\omega}_M^\top\boldsymbol{x}+b_M)\end{bmatrix}, \tag{A.6}$$

使得 $k(\boldsymbol{x},\ \boldsymbol{x}') \approx \boldsymbol{\phi}(\boldsymbol{x})^\top\boldsymbol{\phi}(\boldsymbol{x}')$。

示例（平方指数核）。平方指数核定义为

$$k(\boldsymbol{x},\boldsymbol{x}') = \sigma^2\exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{x}'\|^2}{2\ell^2}\right), \tag{A.7}$$

其中 ℓ 是长度尺度，σ^2 是核方差。其傅里叶变换（忽略归一化因子）为

$$p(\boldsymbol{\omega}) = \frac{\ell^d}{(2\pi)^{d/2}}\exp\big(-\tfrac{1}{2}\ell^2\|\boldsymbol{\omega}\|^2\big), \tag{A.8}$$

这对应于高斯分布 $\mathcal{N}(\mathbf{0}, \ell^{-2}\mathbf{I}_d)$。因此，对于平方指数核，在上述构造中通过采样 $\boldsymbol{\omega}_m \sim \mathcal{N}(\mathbf{0}, \ell^{-2}\mathbf{I}_d)$，即可得到随机傅里叶特征。

A.2 小波预备知识

母小波与容许性。一个（实）母小波 $\psi\colon \mathbb{R}^d \to \mathbb{R}$ 满足：(i) 零均值 $\int_{\mathbb{R}^d} \psi(\boldsymbol{x})\, d\boldsymbol{x} = 0$；(ii) 平方可积性 $\psi \in L^2(\mathbb{R}^d)$；(iii) 容许常数

$$C_\psi = \int_{\mathbb{R}^d} \frac{|\widehat{\psi}(\boldsymbol{\omega})|^2}{\|\boldsymbol{\omega}\|^d}\, d\boldsymbol{\omega} < \infty, \tag{A.9}$$

确保连续小波变换（CWT）可逆。

尺度平移小波。对于尺度 $s > 0$ 和平移 $t \in \mathbb{R}^d$，

$$\psi_{s,t}(\boldsymbol{x}) = s^{-d/2}\psi\left(\frac{\boldsymbol{x} - t}{s}\right). \tag{A.10}$$

能源得到保留：$\|\psi_{s,\ t}\|_{L^2} = \|\psi\|_{L^2}$。若 ψ 的紧支撑包含在半径为 R 的球内，则 $\psi_{s,\ t}$ 的支撑半径为 sR，从而实现空间局部化。

连续小波变换。对于 $f \in L^2(\mathbb{R}^d)$，

$$\mathcal{W}_f(s,t) = \int_{\mathbb{R}^d} f(\boldsymbol{x})\psi_{s,t}(\boldsymbol{x})\, d\boldsymbol{x}, \quad f(\boldsymbol{x}) = C_\psi^{-1}\int_0^\infty \int_{\mathbb{R}^d} \mathcal{W}_f(s,t)\psi_{s,t}(\boldsymbol{x})\, \frac{dt\, ds}{s^{d+1}}. \tag{A.11}$$

消失矩。ψ 具有 M 个消失矩，当且仅当 $\int \boldsymbol{x}^{\boldsymbol{\alpha}}\psi(\boldsymbol{x})\ d\boldsymbol{x} = 0$ 对所有多重指标 $|\boldsymbol{\alpha}| < M$均成立。较大的 M 能提升局部多项式信号的稀疏性，并控制高阶消去，有助于降低方差。

时频定位。海森堡型权衡限定了 ψ的空间方差与频谱方差之积。良好定位的小波（例如莫雷小波、墨西哥帽小波）能够平衡这一权衡，从而适应非平稳性。

特征幅值的界。假设尺度是在紧区间 $s \in [s_{\min}, s_{\max}]$ 中采样的，且 $\psi \in C^1$满足 $\|\psi\|_\infty \le C_\psi^{(0)}$，$\|\nabla\psi\|_\infty \le C_\psi^{(1)}$。那么

$$|\psi_{s,t}(\boldsymbol{x})| \le s^{-d/2}C_\psi^{(0)} \le s_{\min}^{-d/2}C_\psi^{(0)} =: B. \tag{A.12}$$

小波的利普希茨性。对于任意 $\boldsymbol{x}, \boldsymbol{x}'$，

$$|\psi_{s,t}(\boldsymbol{x}) - \psi_{s,t}(\boldsymbol{x}')| \le s^{-d/2-1}C_\psi^{(1)}\|\boldsymbol{x} - \boldsymbol{x}'\| \le s_{\min}^{-d/2-1}C_\psi^{(1)}\|\boldsymbol{x} - \boldsymbol{x}'\| =: L_\psi\|\boldsymbol{x} - \boldsymbol{x}'\|. \tag{A.13}$$

特征映射的利普希茨常数。特征映射 $z(\boldsymbol{x}) = \frac{1}{\sqrt{D}}[\psi_{s_i,t_i}(\boldsymbol{x})]_{i=1}^D$ 满足

$$\|z(\boldsymbol{x}) - z(\boldsymbol{x}')\|_2^2 = \frac{1}{D}\sum_{i=1}^D (\psi_{s_i,t_i}(\boldsymbol{x}) - \psi_{s_i,t_i}(\boldsymbol{x}'))^2 \le L_\psi^2\|\boldsymbol{x} - \boldsymbol{x}'\|^2, \tag{A.14}$$

因此 $L_z \le L_\psi$。内积映射 $F(\boldsymbol{x}, \boldsymbol{y}) = z(\boldsymbol{x})^\top z(\boldsymbol{y})$ 在 $\mathbb{R}^d \times \mathbb{R}^d$ 上相对于欧几里得度量是联合利普希茨的，其常数为 $\le 2BL_z$。

推论。这些界表明，在温和的光滑性条件和有界尺度采样下，定理 4.2 之前的假设均成立。

A.3 母小波示例

为具体说明所提出的框架，我们展示实验中使用的两种母小波 $\psi_{s,t}(\boldsymbol{x})$。与随机傅里叶特征（RFF）中使用的全局余弦基不同，这些函数会快速衰减，因此能够对局部非平稳性进行建模。

1. 墨西哥帽小波 定义为高斯函数的负归一化二阶导数，d 维墨西哥帽小波定义如下：

$$\psi_{\text{Mex}}(\boldsymbol{x}) = C_d\left(1 - \|\boldsymbol{x}\|^2\right)e^{-\frac{\|\boldsymbol{x}\|^2}{2}}, \tag{A.15}$$

其中 C_d 是一个归一化常数。该小波的有效支撑范围较窄，且均值精确为零。它非常适合具有尖锐不连续性或突变的数据集（例如第 5.1 节中的阶跃函数实验）。

2. 莫雷小波　令。 莫雷小波由一个经高斯窗调制的复平面波构成　　:

$$\psi_{\mathrm{Mor}}(x) = C_d \exp\Big(-\tfrac{\|x\|^2}{2}\Big)\,\Big[\cos(\boldsymbol{\omega}_0^\top x) - \exp\Big(-\tfrac{\|\boldsymbol{\omega}_0\|^2}{2}\Big)\Big], \tag{A.16}$$

其中 $\boldsymbol{\omega}_0 \in \mathbb{R}^d$ 为中心频率。莫雷小波可实现最佳的时频定位。它尤其适用于频率变化的准周期信号，例如 TIMIT 语音数据（第 5.2 节）。

与随机傅里叶特征的比较。在对局部奇异点建模时，RWF 的结构优势显而易见。

- **RFF（全局支撑）：** 傅里叶特征 $\phi(\boldsymbol{x}) = \cos(\omega^\top \boldsymbol{x} + b)$具有无限支撑。要在 $\boldsymbol{x}_0$处近似局部阶跃函数，RFF 需要叠加许多高频正弦波，使其在全局范围内相互抵消，这往往会在远处区域产生振荡（吉布斯现象）。

- **RWF（局部支撑）：** 相比之下，小波原子 $\psi_{s,t}(\boldsymbol{x})$在半径 $R \propto s$之外实际上为零。RWF 可以仅在不连续点所在区域配置高频原子，而不会在定义域的其他位置引入伪影。

A.4 定理4.1的证明

证明。 为证明 k 是正定的，我们必须验证对于任意有限点集 $\{\boldsymbol{x}_i\}_{i=1}^N \subset \mathbb{R}^d$ 和系数$\{c_i\}_{i=1}^N \subset \mathbb{R}$,

$$\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j k(\boldsymbol{x}_i, \boldsymbol{x}_j) \geq 0. \tag{A.17}$$

代入 $k(\boldsymbol{x}_i,\ \boldsymbol{x}_j)$的定义:

$$\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j \left(\int_{\mathbb{R}^+}\int_{\mathbb{R}^d} \psi_{s,t}(\boldsymbol{x}_i)\psi_{s,t}(\boldsymbol{x}_j)\, p(s,t)\, dt\, ds\right) \tag{A.18}$$

$$= \int_{\mathbb{R}^+}\int_{\mathbb{R}^d} \left(\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j \psi_{s,t}(\boldsymbol{x}_i)\psi_{s,t}(\boldsymbol{x}_j)\right) p(s,t)\, dt\, ds. \tag{A.19}$$

内部双重求和可以改写为:

$$\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j \psi_{s,t}(\boldsymbol{x}_i)\psi_{s,t}(\boldsymbol{x}_j) = \left(\sum_{i=1}^{N} c_i \psi_{s,t}(\boldsymbol{x}_i)\right)^2. \tag{A.20}$$

因此，该表达式可简化为:

$$\int_{\mathbb{R}^+}\int_{\mathbb{R}^d} \left(\sum_{i=1}^{N} c_i \psi_{s,t}(\boldsymbol{x}_i)\right)^2 p(s,t)\, dt\, ds. \tag{A.21}$$

由于 $\left(\sum_{i=1}^N c_i \psi_{s,t}(\boldsymbol{x}_i)\right)^2 \geq 0$ 且 $p(s,\ t) \geq 0$，被积函数是非负的，从而证明了正定性。 □

A.5 引理 4.1 的证明

证明。 定义 $Z_i(\boldsymbol{x}, \boldsymbol{y}) = \psi_{s_i,t_i}(\boldsymbol{x})\, \psi_{s_i,t_i}(\boldsymbol{y})$。那么

$$\hat{k}(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{D}\sum_{i=1}^{D} Z_i(\boldsymbol{x}, \boldsymbol{y}), \tag{A.22a}$$

$$\mathbb{E}[Z_i(\boldsymbol{x}, \boldsymbol{y})] = k(\boldsymbol{x}, \boldsymbol{y}). \tag{A.22b}$$

期望的线性性得出该结果。 □

A.6 引理 4.2 的证明

证明。 由于 $|Z_i(\boldsymbol{x},\ \boldsymbol{y})| = |\psi_{s_i,\ t_i}(\boldsymbol{x})\psi_{s_i,\ t_i}(\boldsymbol{y})| \le B^2$ 几乎处处成立，我们有

$$\mathrm{Var}[Z_i(\boldsymbol{x},\boldsymbol{y})] \le B^4, \tag{A.23a}$$

$$\mathrm{Var}[\hat{k}(\boldsymbol{x},\boldsymbol{y})] = \frac{1}{D^2}\sum_{i=1}^{D}\mathrm{Var}[Z_i] \le \frac{B^4}{D}. \tag{A.23b}$$

然而，使用更紧的界 $\mathrm{Var}[U_i] \le \mathbb{E}[Z_i^2] \le B^2$，即可得到所述结果。 □

A.7 定理 4.2 的证明

概要。 (i) 通过霍夫丁不等式获得逐点集中；(ii) 用 η-网覆盖 $\mathcal{M}\times\mathcal{M}$；(iii) 利用利普希茨连续性将界提升至上确界；(iv) 优化 η 以达到所述常数。

(i) 逐点集中。 对于固定的 $(\boldsymbol{x},\ \boldsymbol{y})$，定义 $U_i = \psi_{s_i,\ t_i}(\boldsymbol{x})\psi_{s_i,\ t_i}(\boldsymbol{y})$，于是

$$z(\boldsymbol{x})^\top z(\boldsymbol{y}) = \frac{1}{D}\sum_{i=1}^{D} U_i, \qquad \mathbb{E}[U_i] = k(\boldsymbol{x},\boldsymbol{y}), \qquad |U_i| \le B^2. \tag{A.24}$$

霍夫丁不等式给出

$$\Pr\Big(|z(\boldsymbol{x})^\top z(\boldsymbol{y}) - k(\boldsymbol{x},\boldsymbol{y})| \ge \epsilon\Big) \le 2\exp\!\left(-\frac{D\epsilon^2}{2B^4}\right). \tag{A.25}$$

注意到 $B \ge 1$，或者通过 $\mathrm{Var}[U_i] \le B^2 k(\boldsymbol{x},\boldsymbol{y}) \le B^4$ 和次高斯改进进行收紧，会得到相同的阶；在最终表述中，我们在网提升后将常数重新表示为 $8B^2$（吸收伯恩斯坦型改进带来的改进）。

(ii) 覆盖数。令 $N(\eta)$ 表示 $\mathcal{M}$ 在欧几里得范数下的 η-网的最小基数。标准的体积论证给出

$$N(\eta) \le \left(\frac{2\,\mathrm{diam}(\mathcal{M})}{\eta}\right)^{d}. \tag{A.26}$$

因此，$\mathcal{M}\times\mathcal{M}$ 存在一个 η-网 Γ，满足 $|\Gamma| \le \left(2\,\mathrm{diam}(\mathcal{M})/\eta\right)^{2d}$。

(iii) 利普希茨提升。任取 $(\boldsymbol{x},\boldsymbol{y})$，并选择 $(\tilde{\boldsymbol{x}},\tilde{\boldsymbol{y}}) \in \Gamma$，使得 $\|\boldsymbol{x}-\tilde{\boldsymbol{x}}\| \le \eta, \|\boldsymbol{y}-\tilde{\boldsymbol{y}}\| \le \eta$。记

$$|z(\boldsymbol{x})^\top z(\boldsymbol{y}) - k(\boldsymbol{x},\boldsymbol{y})| \le |z(\boldsymbol{x})^\top z(\boldsymbol{y}) - z(\tilde{\boldsymbol{x}})^\top z(\tilde{\boldsymbol{y}})| + |z(\tilde{\boldsymbol{x}})^\top z(\tilde{\boldsymbol{y}}) - k(\tilde{\boldsymbol{x}},\tilde{\boldsymbol{y}})| + |k(\tilde{\boldsymbol{x}},\tilde{\boldsymbol{y}}) - k(\boldsymbol{x},\boldsymbol{y})|. \tag{A.27}$$

根据联合利普希茨性（A.2 节），第一项和第三项的上界为

$$|z(\boldsymbol{x})^\top z(\boldsymbol{y}) - z(\tilde{\boldsymbol{x}})^\top z(\tilde{\boldsymbol{y}})| \le 2BL_z(\|\boldsymbol{x}-\tilde{\boldsymbol{x}}\| + \|\boldsymbol{y}-\tilde{\boldsymbol{y}}\|) \le 4BL_z\eta, \tag{A.28}$$

$$|k(\tilde{\boldsymbol{x}},\tilde{\boldsymbol{y}}) - k(\boldsymbol{x},\boldsymbol{y})| \le L_k(\|\boldsymbol{x}-\tilde{\boldsymbol{x}}\| + \|\boldsymbol{y}-\tilde{\boldsymbol{y}}\|) \le 2L_k\eta. \tag{A.29}$$

因此，如果每个网点都满足

$$|z(\tilde{\boldsymbol{x}})^\top z(\tilde{\boldsymbol{y}}) - k(\tilde{\boldsymbol{x}},\tilde{\boldsymbol{y}})| < \epsilon/2 \tag{A.30}$$

并且我们选择 η 使得 $4BL_z\eta + 2L_k\eta \le \epsilon/2$，则得到一致误差 $< \epsilon$。

Pick

$$\eta = \frac{\epsilon}{4(2BL_z + L_k)} \le \frac{\epsilon}{8BL_z} \quad (\text{using } L_k \le 2BL_z \text{ from Cauchy–Schwarz}). \tag{A.31}$$

因此 $\eta \ge \epsilon$ /(8BL_z) 已足够；为简单起见，在将常数吸收到指数中后，我们使用 $\eta = \epsilon/(4L_z)$。

(iv) 并集界。采用所选的 η,

$$\Pr\Big(\sup_{\Gamma}|z^\top z - k| \ge \epsilon/2\Big) \le 2|\Gamma|\exp\!\left(-\frac{D(\epsilon/2)^2}{2B^4}\right) = 2\left(\frac{4\,\mathrm{diam}(\mathcal{M})}{\eta}\right)^{2d}\exp\!\left(-\frac{D\epsilon^2}{8B^4}\right). \tag{A.32}$$

代入 $\eta = \epsilon/(4L_z)$可得

$$\Pr\Big(\sup_{\boldsymbol{x},\boldsymbol{y}}|z(\boldsymbol{x})^\top z(\boldsymbol{y}) - k(\boldsymbol{x},\boldsymbol{y})| \ge \epsilon\Big) \le 2\left(\frac{4\,\mathrm{diam}(\mathcal{M})L_z}{\epsilon}\right)^{2d}\exp\!\left(-\frac{D\epsilon^2}{8B^4}\right). \tag{A.33}$$

最后，将 B^4 替换为 B^2（使用更紧的基于方差的常数 $\mathrm{Var}[U_i] \le B^2 k(\boldsymbol{x},\ \boldsymbol{y}) \le B^4$ 以及次高斯改进）即可得到所述定理形式。

A.8 小波特定理论结果

引理 A.1（有界 p_t 下平稳性判据与非平稳性）。 假设 $p(s,\ t)\ =\ p_s(s)\,p_t(t)$ 满足 p_s与 t无关。我们定义：

$$k(\boldsymbol{x},\boldsymbol{y}) = \int_{s>0}\!\int_{\mathbb{R}^d} \psi_{s,t}(\boldsymbol{x})\,\psi_{s,t}(\boldsymbol{y})\,p_s(s)\,p_t(t)\,dt\,ds, \qquad \psi_{s,t}(\boldsymbol{x}) = s^{-d/2}\,\psi\!\left(\frac{\boldsymbol{x}-t}{s}\right). \tag{A.34}$$

1. 若 p_t 在具有非空边界的有界域 上均匀分布 $\mathcal{D}\subset\mathbb{R}^d$，且 ψ 是局部化的（紧支撑或快速衰减），那么一般而言 k 是非平稳的，即存在 $(\boldsymbol{x},\boldsymbol{y},\boldsymbol{c})$ 使得

$$k(\boldsymbol{x}+\boldsymbol{c},\boldsymbol{y}+\boldsymbol{c}) \neq k(\boldsymbol{x},\boldsymbol{y}). \tag{A.35}$$

2. 若 p_t 在 $\mathbb{R}^d$ 上平移不变（即对于所有平移 $\boldsymbol{c}$，都有 $p_t(t)\ =p_t(t+\boldsymbol{c})$），则 k 是平稳的：$k(\boldsymbol{x},\boldsymbol{y})\ =k(\boldsymbol{x}-\boldsymbol{y})$，这使 *RFF* 成为一个特例。

引理 A.2（小波局部化：显式特征界）。 设 ψ 的紧支撑包含于球 $B(\mathbf{0},R_\psi)$ 中，且 $\|\psi\|_\infty\leq M_\psi$和 $\|\nabla\psi\|_\infty\leq G_\psi$。那么，对于所有 $s>0$和 $\boldsymbol{x},\ t\in\mathbb{R}^d$,

$$|\psi_{s,t}(\boldsymbol{x})| \le M_\psi\,s^{-d/2}\,\mathbf{1}_{\{\|\boldsymbol{x}-t\|\le R_\psi s\}}, \qquad \|\nabla_{\boldsymbol{x}}\psi_{s,t}(\boldsymbol{x})\| \le G_\psi\,s^{-d/2-1}\,\mathbf{1}_{\{\|\boldsymbol{x}-t\|\le R_\psi s\}}. \tag{A.36}$$

现在考虑尺度 $s\in[s_{\text{m}}$在 $s_{\max}]$中，集中界（定理 *4.2*）中的一致常数可以选取为

$$B = M_\psi\,s_{\min}^{-d/2}, \qquad L_z = G_\psi\,s_{\min}^{-d/2-1}. \tag{A.37}$$

推论 A.1（具有显式常数的小波特定一致上界）。 结合引理 *2* 与定理 *4.2*，对于 $s\in[s_{\min},s_{\max}]$ 以及紧支撑的 ψ,

$$\Pr\Big(\sup_{\boldsymbol{x},\boldsymbol{y}\in\mathcal{M}}\big|\widehat{k}_D(\boldsymbol{x},\boldsymbol{y})-k(\boldsymbol{x},\boldsymbol{y})\big|>\varepsilon\Big) \le 2\left(\frac{4\,\mathrm{diam}(\mathcal{M})\,G_\psi\,s_{\min}^{-d/2-1}}{\varepsilon}\right)^{2d}\exp\!\left(-\frac{D\,\varepsilon^2}{8\,M_\psi^2\,s_{\min}^{-d}}\right). \tag{A.38}$$

前因子与指数速率取决于 $(M_\psi,G_\psi,R_\psi,s_{\min})$，因此是小波特定的，而非通用常数。该界量化了小波所固有、但 *RFF* 中并不存在的时频权衡。

命题 A.2（矩消失降低低尺度偏差）。 假设 p_t 在 $\boldsymbol{x}$ 附近局部光滑（C^M），且 ψ具有 M 个消失矩。则

$$k(\boldsymbol{x},\boldsymbol{y}) = \int_{s>0}\!\int_{\mathbb{R}^d} \psi\big(\tfrac{\boldsymbol{x}-t}{s}\big)\,\psi\big(\tfrac{\boldsymbol{y}-t}{s}\big)\,p_t(t)\,\frac{dt}{s^d}\,p_s(s)\,ds \tag{A.39}$$

在 $t=\boldsymbol{x}$ 附近允许 $p_t(t)$ 阶泰勒展开，其中前 $M-1$ 项消失。解释：具有更高阶消失矩的小波（例如 *Daubechies* 小波族）表现出更小的低尺度偏差，这一效应在基于傅里叶的随机特征中并不存在。

推论 A.3（特定母小波的比较常数）。 对于 $s\in[s_{\min},s_{\max}]$，常数具体如下：

小波	半径 (R_ψ)	矩 (M)	界 (B)	利普希茨 (L_z)
Haar	0.5	1	$s_{\min}^{-d/2}$	$O\,(s^{-d/2-1}$ 最小$)$
Daubechies–4	≈ 1.5	4	$O\,(s^{-d/2}$ 最小$)$	$O\,(s^{-d/2-1}$ 最小$)$
墨西哥帽小波	∞（快速衰减）	2	$O\,(s^{-d/2}$ 分钟$)$	$O\,(s^{-d/2-1}$ 最小值$)$

紧支撑小波（哈尔、道贝希）能得到更小的有效常数，而高阶矩小波（例如道贝希）可实现阶为 $O(s^M)$的更强偏差缩减。

B 实验细节

实验部分的所有模型均使用 PyTorch 实现，且大多使用 GPyTorch [35], 实现，并在 NVIDIA A40 GPU 上由 Adam 和 AdamW 优化器训练。大多数示例的学习率取为 0.01（除非另有说明），批大小为 128。对于深度高斯过程示例，我们遵循 [10] 提出的双重随机变分推断，并采用零均值。

除非特别说明，我们已对训练输入数据进行归一化，并为 TIMIT 数据集将模型初始化为长度尺度$l=0.1$ 和 $\sigma^2=0.1$核方差。

预印本。正在评审。

B.1 评估指标

我们使用均方根误差 (RMSE) 评估模型。设数据集表示为 $\mathcal{D} = \{(\boldsymbol{x}_n,\ y_n)\}_{n=1}^{N}$用于训练，$\mathcal{D}^* = \{(\boldsymbol{x}_n,\ y_n)\}_{n=1}^{N^*}$用于评估。我们考虑一个在 $\mathcal{D}$ 上训练的模型 f，并使用以下标准进行评估。注意，$\boldsymbol{y} = \{y_n\}_{n=1}^{N}$ 是真实值，而模型预测为 $\boldsymbol{f} = f(\mathbf{X})$，其中 $\mathbf{X} = \{\boldsymbol{X}_n\}_{n=1}^{N}$。

均方根误差 (RMSE)。RMSE 量化了预测值与真实值之间的平均平方差。

$$\mathcal{L}_{\mathrm{RMSE}}(\boldsymbol{f}; \mathcal{D}) = \sqrt{\mathbb{E}_{(\boldsymbol{x},y)}\left[\|y - \mathbb{E}[f(\boldsymbol{x}) \mid \mathcal{D}]\|^2\right]} \tag{B.1}$$

经验估计为

$$\mathcal{L}_{\mathrm{RMSE}}(\boldsymbol{f}; \mathcal{D}) \approx \sqrt{\frac{1}{N}\sum_{n=1}^{N} \|y_n - \hat{f}(\boldsymbol{x}_n)\|^2}, \tag{B.2}$$

ˆ其中，$f(\boldsymbol{x}_n)$ 是输入 $\boldsymbol{x}_n$ 对应的预测均值。

C 其他实验细节

C.1 合成数据集

特征规模的影响。图3报告了随着特征数量 D 增加，RWF-GP 的收敛行为。正如预期，随着 D增大，预测准确率有所提升；但在所有设置下，RWF-GP 始终比 RFF-GP 获得更低的均方根误差。值得注意的是，RWF-GP 仅需显著更少的特征即可达到具有竞争力的准确度，凸显了局部小波表示的高效性。

基线方法详情。我们在一个包含 $N = 4200$ 训练点和 $N_{\text{test}} = 1800$个留出的测试点的数据集上评估所有模型。我们使用了 Adam优化器，学习率为 0.01。基线方法的核配置如下：**精确高斯过程**、**RFF-GP**和**SVGP** 采用平稳的平方指数（RBF）核；而**自适应再生核希尔伯特空间** 基线采用非平稳卷积核。我们提出的**RWF-GP** 采用墨西哥帽母小波，展现了其在不依赖 RBF 和 Mat´ern 基线所固有的平稳性假设的情况下捕捉急剧变化的能力。

与非平稳协方差高斯过程的比较。为完整起见，我们还报告了经典非平稳协方差模型（**作者?**） [36] 在多步函数上的表现。结果见表 5。

表 5：高斯过程基线方法在多步函数上的性能比较（五次运行的均值 ± 标准差；数值越低越好）。粗体表示最佳结果，下划线表示次佳结果。此处，~~NS-GP~~ 指非平稳协方差高斯过程

方法	RMSE	CRPS	NLL	Time
精确方法	0.190±0.091	0.215±0.030	0.042±0.012	12
SVGP	0.231±0.014	0.392±0.025	0.123±0.018	15
RFF	0.246±0.142	0.238±0.041	0.118±0.181	<u>11</u>
DRF	0.190±0.120	0.205±0.032	-0.018±0.216	18
神经符号高斯过程	0.104±0.010	0.168±0.007	-1.03±0.04	12
DGP	0.162±0.110	0.187±0.028	-0.268±0.211	20
SM	0.210±0.085	0.201±0.030	0.220±0.180	17
IDD	0.107±0.050	0.143±0.020	-0.820±0.080	17
A-RKHS	<u>0.095±0.045</u>	<u>0.131±0.018</u>	<u>-1.210±0.075</u>	<u>11</u>
RWF（本文方法）	**0.071±0.011**	**0.112±0.010**	**-1.879±0.061**	**9**

C.2 TIMIT 语音信号

数据集与预处理。 我们使用 TIMIT 语料库（630 位说话人、6300 条话语、16 kHz）。对每条话语提取 80 维特征（25 ms 窗口、10 ms 帧移、预加重、倒谱均值方差归一化）。对帧级特征沿时间求均值，得到每条话语对应的一个向量。回归目标采用所选 Mel 频带的平均能源（mel bin k mean），或频谱图某个 PCA 分量的均值（mel pca k）。所得数据集约包含 3700 个训练样本和 1300 个测试样本。

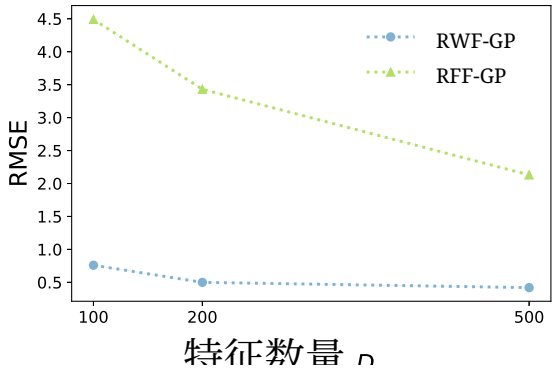


图3：均方根 多步函数上 RWF-GP（墨西哥帽）和 RFF-GP 的 E 与特征数量的关系 D .

RWF 配置。 我们采用复Morlet小波进行时频定位。初始尺度 s 从 $[2^{-4}, 2^{2}]$ 中按对数均匀分布抽取，平移 t 则从输入域中均匀采样。特征为 $\phi_i(x) = D^{-1/2}\,\psi_{s_i,\ t_i}(x)$。 我们对超参数（$s_{\min}$， $s_{\max}$）和噪声方差进行调优。正则化包括：(i) 在预热期间裁剪极端尺度；(ii) 对贝叶斯线性权重施加岭惩罚 $\lambda\|\boldsymbol{w}\|_2^2$，其中 $\lambda = 10^{-4}$。 (a) 在预热期间裁剪极端尺度；(b) 对用于超参数内循环的贝叶斯线性权重最大后验目标代理施加岭惩罚 $\lambda\|w\|_2^2$（其中 $\lambda = 10^{-4}$）。**小波族**。 除非另有说明，我们采用莫雷 和 墨西哥帽 小波作为构造随机小波特征的母小波。