学习物理所在位置：面向刚性偏微分方程的概率自适应采样

Akshay Govind Srinivasan，Balaji Srinivasan印度理工学院马德拉斯分校，金奈，泰米尔纳德邦 600107，印度
me22b102@smail.iitm.ac.in，sbalaji@dsai.iitm.ac.in

arXiv:2603.06287v1 [cs.CE] 6 Mar 2026

摘要

对具有陡峭梯度的刚性偏微分方程（PDE）进行建模，仍是科学机器学习面临的一项重大挑战。物理信息神经网络（PINNs）受谱偏差和训练缓慢的困扰，而物理信息极限学习机（PIELM）能够快速求得闭式线性解，却根本上受限于与物理无关的随机初始化。我们提出高斯混合模型自适应 PIELM（GMM-PIELM），这是一种概率框架，能够学习表示“物理位置”的概率密度函数，从而对 PIELM 的核函数进行自适应采样。通过采用加权 EM 算法，GMM-PIELM 能够自主地将径向基函数中心集中在数值误差较大的区域，例如激波前沿和边界层。这种方法无需进行计算开销巨大的基于梯度的优化（如 PINNs）或贝叶斯搜索，即可动态改善隐藏层的条件性。我们在扩散系数为$\nu = 10^{-4}$的一维奇异摄动对流扩散方程上评估了该方法。与基线 RBF-PIELM 相比，我们的方法实现了最高 7 个数量级的 L_2 误差降低，成功解析了指数级薄的边界层，同时保留了 ELM 架构在速度上的数量级优势。

1 引言

对具有陡峭梯度的刚性常微分方程/偏微分方程进行建模，面临严峻的分辨率挑战（Madhavi 等人，2025年）。尽管已有高保真经典方法，但由于稳定性约束和复杂的网格划分要求，这些方法的成本往往高得难以承受（Yagawa & Okuda，2011）。因此，人们开始开发无网格、基于模型的求解器，以高效解析刚性动力学。物理信息神经网络（PINNs）（Raissi 等人，2019年）提供了一种灵活的无网格框架，但存在训练缓慢和超参数敏感性问题。近期基准测试表明，要达到相当的精度，这类方法所需的实际运行时间往往远超经典求解器；对于刚性偏微分方程，这一差距尤其明显（McGreivy & Hakim，2024年；Karnakov 等人，2024年）。此外，即使在过参数化网络中，谱偏差和优化病态性也常常导致陡峭梯度分辨不足（Krishnapriyan 等人，2021年）

物理信息极限学习机（PIELM）（Dwivedi & Srinivasan, 2020）以闭式线性最小二乘解替代迭代反向传播，实现了数量级的加速。然而，PIELM 的有效性从根本上受其初始化限制：随机的、与物理无关的隐藏特征往往与底层刚性动力学存在谱失配，从而导致病态表示（Huang et al., 2006; Dwivedi 等人, 2025b）。尽管径向基函数变体（如 RBF-PIELM）（Dwivedi et al., 2025a; Srinivasan et al., 2025）引入了可解释的局部感受性，但它们通常依赖静态节点分配，缺乏跟踪演化不连续性的灵活性，并且需要人工设计针对具体问题的启发式方法。因此，仍然迫切需要有原则的、数据驱动的自适应机制，以便在保留线性求解器计算效率的同时，使基函数与奇异摄动问题的移动波前动态对齐。

为应对这些挑战，本文提出了**高斯混合模型自适应 PIELM（GMM-PIELM）**，这是一种概率训练框架，将偏微分方程残差场视为计算域上的误差密度。我们假定 log(1 + |残差|) 场可被解释为表示“物理位置”的未归一化概率密度函数。随后，我们使用基于期望最大化（E-M）的算法学习该分布。GMM-PIELM 将隐单元中心集中在数值误差较高的区域。这种自适应重分布改善了隐藏层系统的条件数，并恰好在解呈现刚性行为的位置增强表达能力，同时保留 ELM 架构的线性最小二乘求解（不同于 Tang 等（2023））。与依赖迭代贝叶斯优化进行中心布置的 KAPI-ELM（Dwivedi 等人, 2025b）相比，我们的方法依赖于对残差场概率密度的无监督学习。在一维奇异摄动对流–扩散方程上的实验表明，所提方法在捕捉复杂边界层的同时，取得了远低于 RBF-PIELM 的 L_2 误差。

贡献 本文提出一种基于期望最大化的算法，用于为 PIELM 自适应采样核函数。具体而言：

- 我们提出了一种快速且易于解释的、基于残差场的期望最大化算法，用于自适应采样核函数，以通过 RBF-PIELM 求解刚性偏微分方程
- 我们以具有单边界层和双边界层的一维边界层问题为基准，评估该算法相较于 RBF-PIELM 的能力与效率。

2 数学公式

我们考虑定义在有界域 $\Omega \subset \mathbb{R}^d$ 上的定常偏微分方程，其边界为 $\partial\Omega$：

$$\mathcal{L}[u](x) = f(x), \quad x \in \Omega \quad \mathcal{B}[u](x) = g(x), \quad x \in \partial\Omega \tag{1}$$

其中，$\mathcal{L}$ 是线性微分算子，$\mathcal{B}$ 是边界算子，$f(x)$是源项，$g(x)$是边界数据。在 RBF-PIELM 中，我们使用包含 N 个隐含神经元的单隐层前馈网络（SLFN）来近似解 $u(x)$：

$$\hat{u}(x;\beta) = \sum_{j=1}^{N} \beta_j \phi_j(x) = \sum_{j=1}^{N} \beta_j \exp\left(-\frac{\|x - x_0^j\|^2}{2s_j^2}\right) \tag{2}$$

其中，β_j 是隐藏单元权重，$\phi_j(x)$是非线性激活函数。x_0^j 和 s_j^2 是 RBF 核的参数。对这些点进行随机采样会带来挑战，尤其是在刚性问题中（参见附录 A.1 节）。Dwivedi 等人（2025a）和 Srinivasan 等（2025）使用人工设计的启发式方法对这些点进行采样。但这需要对底层物理机制有先验知识，而且难以适应动态问题。

2.1 GMM-PIELM：基于高斯混合模型的核自适应

上一节中的近似 $\hat{u}$ 的精度很大程度上取决于基函数中心 x_0^j 的位置和宽度 s_j的选择，尤其是在刚性问题中。为此，我们提出一种概率框架，根据问题的局部物理复杂度按比例分配资源。

概率公式化 令 $\mathcal{R}(x;\theta) = \mathcal{L}[\hat{u}_\theta](x) - f(x)$ 表示当前近似 $\hat{u}_\theta$ 的偏微分方程残差场。标准物理信息方法会最小化该残差的全局 L_2 范数，实际上是在整个区域内对误差进行平均。然而，对于奇异摄动问题而言，这一全局度量会受到光滑区域的主导，从而常常掩盖集中在激波或边界层处的高频误差（Krishnapriyan 等人，2021年）。
为解决这一问题，我们假设 log(1 + |残差|) 场可作为未归一化概率密度函数（概率密度函数），表示近似误差在空间上的集中程度。我们在附录 A.5 节中说明选择对数变换的理由。我们将残差能量密度 $p_{\text{res}}(x)$ 定义为：

$$p_{\text{res}}(x) = \frac{\log(1 + |\mathcal{R}(x;\theta)|)}{Z}, \quad \text{where} \quad Z = \int_\Omega \log(1 + |\mathcal{R}(z;\theta)|)\, dz. \tag{3}$$

直观而言， $p_{\text{res}}(x)$ 映射出“物理所在的位置”，突出显示频谱带宽不足以捕捉底层动力学的区域（ Nabian & Meidani，2019年；Wu 等人，2023年）。

GMM-PIELM 在这一假设下，隐含神经元的最优分配对应于基函数中心 $\{\mathbf{c}_j\}_{j=1}^{N}$ 的一种分布，该分布使从该残差分布中采样的似然最大化。因此，我们将该分布建模为高斯混合

$$x_0^j \sim p(x;\Theta) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x \mid \mu_k, \Sigma_k), \tag{4}$$

其中 $\Theta = \{\pi_k, \mu_k, \Sigma_k\}_{k=1}^{K}$ 分别为混合系数、均值和协方差。为将该模型拟合到不断演化的解，我们采用加权 EM 框架以最大化残差加权对数似然。在**E 步**中，我们评估当前近似，并计算“责任度” q_{ik}——即第 i个配点 x_i （残差权重为 $w_i = \log(1 + |\mathcal{R}(x_i)|)$）属于第 k个高斯分量的后验概率：

$$q_{ik} = \frac{\pi_k \mathcal{N}(x_i \mid \mu_k, \Sigma_k)}{\sum_{\ell=1}^{K} \pi_\ell \mathcal{N}(x_i \mid \mu_\ell, \Sigma_\ell)}. \tag{5}$$

在随后的**M 步**中，我们更新 GMM 参数，使各分量集中于高误差区域。更新后的均值 μ_k^* 和协方差 Σ_k^* 计算如下：

$$N_k = \sum_{i=1}^{N_{eval}} w_i q_{ik}, \quad \mu_k^* = \frac{1}{N_k} \sum_{i=1}^{N_{eval}} w_i q_{ik} x_i, \quad \Sigma_k^* = \frac{1}{N_k} \sum_{i=1}^{N_{eval}} w_i q_{ik} (x_i - \mu_k^*)(x_i - \mu_k^*)^\top \tag{6}$$

其中， N_{eval} 是密集网格上的评估点数量。参数更新的推导见附录 A.4 节。最后，我们在自适应采样中加入一组均匀分布的中心，以确保区域的全局覆盖，并防止低残差区域出现基函数耗尽。随后，我们根据 k近邻距离确定局部核宽度 s_j ，以确保即使在高度聚集的排列中也能保持一致的重叠。

3 数值实验

我们针对具有单边界层和双边界层的奇异摄动一维稳态对流–扩散方程，评估 GMM-PIELM 框架相对于 RBF-PIELM 的性能。这是用于验证针对刚性动力学设计的数值格式的经典基准问题（Roos 等人，2008年）。其对应的偏微分方程定义在区域 $x \in (0,\ 1)$上：

$$-\nu u_{xx} + u_x = 0 \tag{7}$$

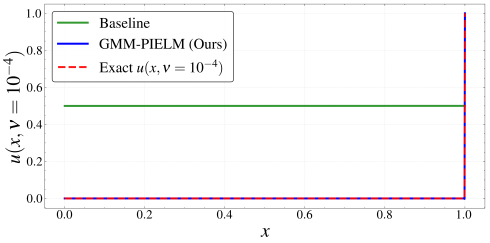
其中， $\nu > 0$表示扩散系数。精确解析解见 Dwivedi 等（2025b）。我们评估了其在 $\nu = 10^{-4}$刚性区域中的性能。该方程是流体力学中高 P´eclet 数输运现象的基本模型，可单独考察边界层形成的物理过程，其中平流力占主导，扩散作用较弱。

该问题场景下的主要计算挑战是出口附近（$x = 1$）形成宽度为 $\delta \sim \mathcal{O}(\nu)$的指数级薄边界层。要解析这一特征，均匀求解器通常需要极高的网格密度（$N \gg 1/\nu$），代价高得难以承受；而物理信息神经网络等基于学习的方法则会受到显著的谱偏差影响，在没有专门的损失加权时，往往无法捕捉高频跃迁（Krishnapriyan 等人, 2021年；Wang 等人, 2022年）。类似地，采用随机初始化的标准 PIELM 往往在这一狭窄区域缺乏足够的基函数支撑，导致病态线性系统并降低精度（参见附录 A.1 节）。实现细节请参见附录 A.6 节。

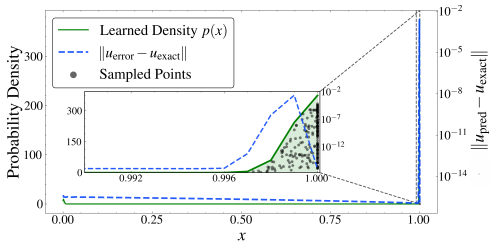
结果 从图 1a 和 1c 可以看出，与基线方法（RBF-PIELM）不同，我们的方法（GMM-PIELM）能够捕捉 $x = 1$ 处的尖锐边界层（双边界层情形下为 $x = 0,\ 1$）。与 RBF-PIELM 相比，我们的解精度最高可提升 7 个数量级（参见表 1）。此外，从图 1b 和 1d 可以看出，学得的 GMM分布与误差具有相似的结构。

边界条件	方法	平均 L_2 误差（RMSE）	时间（秒）
单次（$N = 300$）	基线 RBF-PIELM	5.00×10^{-1}	0.071
	我们的方法（GMM 自适应）	2.73×10^{-8}	0.696
双重（$N = 500$）	基线 RBF-PIELM	1.01×10^{-5}	0.312
	我们的方法（GMM 自适应）	1.04×10^{-9}	1.776

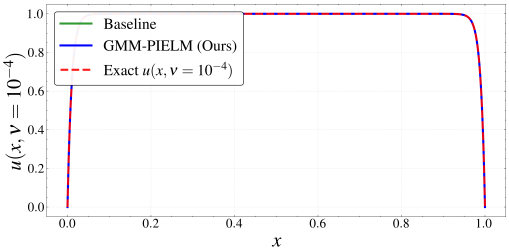
表 1：单边界层和双边界层问题（$\nu = 10^{-4}$）中基线 RBF-PIELM 与 GMM 自适应 RBF-PIELM 的比较，展示了精度的显著提升。



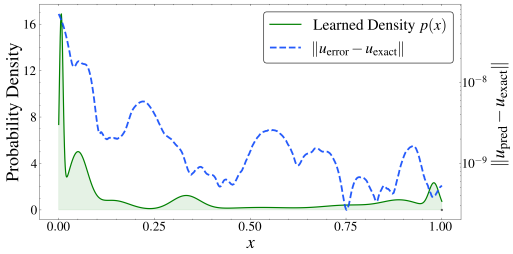
(a) 单边界层的解剖面



(b) 单边界层的误差密度和学习分布边界层



(c) 双边界层的解剖面



(d) 双边界层的误差密度和学习分布边界层

图 1：GMM-PIELM 在具有单边界层和双边界层的一维对流扩散方程上的性能分析。

4 结论

本文提出了高斯混合模型自适应 PIELM（GMM-PIELM），这是一种将 PDE残差视为代表“物理位置”的概率密度的框架。通过将 $\log(1+|\text{残差}|)$ 场解释为未归一化概率密度函数，我们使用期望最大化算法在数值误差较大的区域动态集中隐单元中心和宽度。这种方法解决了标准物理信息极限学习机在初始化方面的根本局限；该算法无需依赖物理无关随机采样或启发式节点布置，而是自动追踪解中的陡峭梯度。对一维奇异摄动对流–扩散方程进行的数值实验表明，GMM-PIELM 的误差比基线 RBF-PIELM 低最多 7 个数量级，并成功解析了标准方法无法捕捉的薄边界层（$\nu = 10^{-4}$）。尽管自适应过程带来了适度的计算开销，但与基于梯度的训练相比，它仍保留了 ELM 架构的根本速度优势。因此，对于传统无网格方法通常难以求解的刚性多尺度物理系统，GMM-PIELM 提供了一种高效且稳健的替代方案。

未来工作 未来工作将着重于把这一概率框架扩展到含时偏微分方程，使高斯混合模型质心能够实时追踪移动波前。此外，我们还计划研究该方法在高维问题和复杂几何中的可扩展性与稳定性，以便将其应用于多种场景。

参考文献

Vikas Dwivedi 和 Balaji Srinivasan。物理信息极限学习机（PIELM）——一种快速求解偏微分方程数值解的方法。*Neurocomputing*，391：96–118，2020年。doi：10.1016/j.neucom.2019.12.099。

Vikas Dwivedi、Bruno Sixou 和 Monica Sigovan。用于流体流动模拟的课程学习驱动的 PIEL M。*Neurocomputing*，616:130924，2025年a。doi：10.1016/j.neucom.2024.130924。

Vikas Dwivedi、Balaji Srinivasan 等。用于具有陡峭梯度的偏微分方程中正问题和逆问题的核自适应 PI-ELM。*arXiv* 预印本 *arXiv:2507.10241*，2025b。

Guang-Bin Huang、Qin-Yu Zhu 和 Chee-Kheong Siew。极限学习机：理论与应用。*Neurocomputing*，70:489–501，2006。doi：10.1016/j.neucom.2005.12.126。

Petr Karnakov、Sergey Litvinov 和 Petros Koumoutsakos。通过优化离散损失求解物理学中的逆问题：无需神经网络即可实现快速而准确的学习。*PNASNexus*，3（1）：pgae005，2024。doi：10.1093/pnasnexus/pgae005。

Aditi Krishnapriyan、Amir Gholami、Shandian Zhe、Robert Kirby 和 Michael W. Mahoney。刻画物理信息神经网络中可能的失效模式。载于神经信息处理系统进展，第34卷，第26548–26560页，2021年。

M. V. D. N. S. Madhavi、Mir Sohail Ali、Satyawan L. Dhondge、Vijay Narayan Deshmukh、Gajendra R. Gandhe 和 P. V. S. Sairam。求解流体力学中非线性偏微分方程的高级数值方法：在航空航天工程中的应用。*InternationalJournal of AppliedMathematics*，38(3s)，2025年。网址：https://ijamjournal.org/ijam/publication/index.php/ijam/article/view/135。

Nick McGreivy 和 Ammar Hakim。薄弱的基准和报告偏差导致机器学习应用于流体相关偏微分方程时过度乐观。*Nature MachineIntelligence*，6(10):1256–1269，2024年。网址：https://www.nature.com/articles/s42256-024-00897-5。

Mohammad Amin Nabian和Rael Meidani。通过自适应重要性采样高效训练物理信息神经网络。*Computer-Aided Civil and Infrastructure Engineering*，34(12)：1026–1037，2019年。

Maziar Raissi、Paris Perdikaris和George E. Karniadakis。物理信息神经网络：用于求解涉及非线性偏微分方程的正问题与反问题的深度学习框架。计算物理杂志，378：686–707，2019年。doi：10.1016/j. jcp.2018.10.045。

Hans-G¨org Roos、Martin Stynes和Lutz Tobiska。奇异摄动微分方程的稳健数值方法：对流-扩散-反应和流动问题，第24卷，属于计算数学丛书。施普林格出版社，柏林、海德堡，第2版，2008年。doi：10.1007/978-3-540-34467-4。

Akshay Govind Srinivasan、Vikas Dwivedi 与 Balaji Srinivasan。深度与浅层：在双调和方程上对物理信息神经架构进行基准测试。*arXiv*预印本 *arXiv:2510.04490*，2025年。

Kejun Tang、Xiaoliang Wan 与 Chao Yang。DAS-PINNs：一种用于求解高维偏微分方程的深度自适应采样方法。计算物理杂志，476: 111868，2023年。ISSN 0021-9991。doi: https://doi.org/10.1016/j.jcp.2022.111868。URL https: //www.sciencedirect.com/science/article/pii/S0021999122009317。

Sifan Wang、Xinling Yu 与 Paris Perdikaris。物理信息神经网络何时以及为何训练失败：神经正切核视角。计算物理杂志，449:110768，2022年。

Chuizheng Wu、Min Zhu、QiquanTan、Yogesh Ktha 和 E.Lu。《物理信息神经网络综合综述：关于失效模式和求解策略的观点》。*arXiv* 预印本 *arXiv:2303.08823*，2023年。

Genki Yagawa 和 Hiroshi Okuda。无网格方法应用于连续介质力学问题的计算性能。*Interaction and Multiscale Mechanics*，4(1):1–28，2011年。doi：10.12989/imm.2011.4.1.001。

A 附录

A.1 条件数问题

为了确定 $\beta = [\beta_1, \ldots, \beta_L]^T$，我们将试探解 $\hat{u}$ 代入偏微分方程和边界条件，并在一组配点上对其进行求值。设 $\{x_f^{(i)}\}_{i=1}^{N_f}$ 为内部点，$\{x_b^{(k)}\}_{k=1}^{N_b}$ 为边界点。

对于线性算子 $\mathcal{L}$，残差约束关于 β 是线性的：

$$\sum_{j=1}^{L} \beta_j \mathcal{L}[\phi_j](x_f^{(i)}) = f(x_f^{(i)}) \tag{8}$$

$$\sum_{j=1}^{L} \beta_j \mathcal{B}[\phi_j](x_b^{(k)}) = g(x_b^{(k)}) \tag{9}$$

该系统可以写成矩阵形式 $\mathbf{H}\beta = \mathbf{T}$，其中：

$$\mathbf{H} = \begin{bmatrix} \mathcal{L}[\phi_1](x_f^{(1)}) & \ldots & \mathcal{L}[\phi_L](x_f^{(1)}) \\ \vdots & \ddots & \vdots \\ \mathcal{L}[\phi_1](x_f^{(N_f)}) & \ldots & \mathcal{L}[\phi_L](x_f^{(N_f)}) \\ \lambda\mathcal{B}[\phi_1](x_b^{(1)}) & \ldots & \lambda\mathcal{B}[\phi_L](x_b^{(1)}) \\ \vdots & \ddots & \vdots \end{bmatrix}, \quad \mathbf{T} = \begin{bmatrix} f(x_f^{(1)}) \\ \vdots \\ f(x_f^{(N_f)}) \\ \lambda g(x_b^{(1)}) \\ \vdots \end{bmatrix} \tag{10}$$

这里，λ 是边界条件的惩罚权重。由于通常 $N_f + N_b > L$，该系统是超定的。最小二乘意义下的最优权重可通过 Moore-Penrose伪逆求得：

$$\beta^* = \mathbf{H}^\dagger \mathbf{T} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{T} \tag{11}$$

最小二乘解的精度很大程度上取决于条件数 $\kappa(\mathbf{H})$。对于刚性偏微分方程，如果配点均匀分布，那么$\mathbf{H}$ 中对应光滑区域的行表现良好，但对应边界层（梯度极大处）的行却很稀疏。如果落入边界层的点很少，矩阵$\mathbf{H}$ 就无法捕捉高频动态，实质上会将边界层视为噪声。这会导致近似效果不佳β^*，并将激波平滑掉。为解决这一问题，我们必须提高激波区域的行密度，从而明确约束求解器遵循该区域的物理规律。

A.2 算法实现

这一概率反馈回路在算法1中得到了形式化描述。该过程在快速线性求解（ELM步骤）与对基函数进行统计适应（EM步骤）之间交替进行，直至残差能量趋于稳定。

算法1GMM-PIELM 自适应求解

要求： 算子 $\mathcal{L}$，$\mathcal{B}$；迭代次数 T；混合比 α
1: **初始化：** 中心 $x_0^{(0)}$（均匀），宽度 $s_j^{(0)}$（常数）
2: **对于** $t = 0$ 至 T **执行**
3: // 步骤 *1*：线性求解（*PIELM*）
4: 构造矩阵 $H^{\text{in}}, H^{\text{bd}}$ 使用当前的 $\{x_0^j, s_j\}$
5: 求解 $\mathbf{w}^{(t)} = \arg\min_{\mathbf{w}} \|H\mathbf{w} - \mathbf{b}\|_2^2$
6: 构造近似：$\widehat{u}^{(t)}(x) = \sum w_j \phi(x; x_0^j, s_j)$
7: // 步骤 *2*：评估
8: 在网格上评估残差 $\mathcal{R}(x;\theta) = \mathcal{L}[\widehat{u}^{(t)}] - f$
9: 计算密度 $p_{\text{res}}(x) \propto \log(1 + |\mathcal{R}(x;\theta)|)$
10: **如果** $t < T$ **则**
11: // 步骤 *3*：自适应（期望最大化与混合采样）
12: 将 GMM 参数拟合到来自 $p_{\text{res}}(x)$ 的样本 Θ
13: 采样 $x_{\text{gmm}} \sim p(x;\Theta)$ 和 $x_{\text{uni}} \sim \mathcal{U}(\Omega)$
14: 更新中心：$x_0^{(t+1)} \leftarrow \alpha x_{\text{gmm}} \cup (1-\alpha) x_{\text{uni}}$
15: 更新宽度：$s_j^{(t+1)} \leftarrow \beta \cdot \text{分布}_k(x_0^j) + \epsilon$
16: **结束 if**
17: **结束 for**
18: **返回** 最终近似 $\widehat{u}^{(T)}$

A.3 计算规范

实验中，我们仅在一台CPU上运行该实验，即一台配备16GB RAM的Ryzen 7 5800H。所有进程均以单线程方式运行。

A.4 EM步骤的推导

GMM-PIELM框架的目标是自适应地分布径向基函数的中心，使其匹配PDE误差的空间分布。我们假设平方PDE残差场表示一个未归一化的概率密度函数（PDF），用于指示"物理位置"。

令**残差能量密度**，$p_{res}(x)$，定义为：

$$p_{\text{res}}(x) = \frac{\log(1 + |\mathcal{R}(x;\theta)|)}{Z}, \quad \text{where} \quad Z = \int_\Omega \log(1 + |\mathcal{R}(z;\theta)|)\, dz. \tag{12}$$

我们使用高斯混合模型（GMM）$p(x;\Theta)$来近似该目标分布，其参数化形式为 $\Theta = \{\pi_k, \mu_k, \Sigma_k\}_{k=1}^K$：

$$p(x;\Theta) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k, \Sigma_k) \tag{13}$$

我们寻求参数 Θ，使模型分布 $p(x;\Theta)$能够最佳近似目标残差分布 $p_{res}(x)$。衡量两个概率分布之间差异的信息论度量是**Kullback-Leibler（KL）散度**。

我们最小化真实分布 p_{res} 与模型 p_Θ 之间的 KL 散度：

$$\Theta^* = \arg\min_\Theta D_{KL}(p_{res}||p_\Theta) \tag{14}$$

$$= \arg\min_\Theta \int_\Omega p_{res}(x) \ln\left(\frac{p_{res}(x)}{p(x;\Theta)}\right) dx \tag{15}$$

$$= \arg\min_\Theta \left[\int_\Omega p_{res}(x) \ln p_{res}(x) dx - \int_\Omega p_{res}(x) \ln p(x;\Theta) dx\right] \tag{16}$$

第一项表示残差场的熵。由于残差 $\mathcal{R}(x)$ 在自适应步骤中是固定的（它仅取决于上一次迭代的权重），因此该项相对于 Θ 是常数。因此，最小化 KL 散度等价于最大化第二项（交叉熵）：

$$\Theta^* = \arg\max_{\Theta} \int_{\Omega} p_{res}(x) \ln p(x;\Theta) dx \tag{17}$$

我们通过黎曼和，利用一组密集的配点 $\{x_i\}_{i=1}^N$ 近似该积分。代入 $p_{res}(x_i) \propto \log(1+|\mathcal{R}(x_i)|)$：

$$\Theta^* \approx \arg\max_{\Theta} \sum_{i=1}^{N} \log\left(1+|\mathcal{R}(x_i)|\right) \ln p(x_i;\Theta) \tag{18}$$

定义权重 $w_i = \log(1+|\mathcal{R}(x_i)|)$，得到**加权对数似然**目标函数：

$$\mathcal{L}(\Theta) = \sum_{i=1}^{N} w_i \ln\left(\sum_{k=1}^{K} \pi_k \mathcal{N}(x_i|\mu_k,\Sigma_k)\right) \tag{19}$$

由于对数内部存在求和项，直接最大化 $\mathcal{L}(\Theta)$ 是难以处理的。我们采用 EM 算法。引入潜变量 $z_i \in \{1,\dots,K\}$，表示点 x_i 的分量分配。

加权数据集的期望完整数据对数似然（Q 函数）为：

$$Q(\Theta,\Theta^{(t)}) = \sum_{i=1}^{N} w_i \sum_{k=1}^{K} q_{ik} \left[\ln \pi_k + \ln \mathcal{N}(x_i|\mu_k,\Sigma_k)\right] \tag{20}$$

其中， $q_{ik} = P(z_i=k|x_i,\Theta^{(t)})$ 是在 E 步中计算的责任度（论文中的式 5 r).

我们通过将梯度设为零，关于 μ_k 和 Σ_k 最大化 Q 。

考虑 Q 中依赖于 μ_k的项：

$$J(\mu_k) = \sum_{i=1}^{N} w_i q_{ik} \left(-\frac{1}{2}(x_i-\mu_k)^T \Sigma_k^{-1}(x_i-\mu_k)\right) \tag{21}$$

对 μ_k求导：

$$\frac{\partial J}{\partial \mu_k} = \sum_{i=1}^{N} w_i q_{ik} \Sigma_k^{-1}(x_i-\mu_k) \tag{22}$$

令$\frac{\partial J}{\partial \mu_k} = 0$并乘以 Σ_k：

$$\sum_{i=1}^{N} w_i q_{ik}(x_i-\mu_k) = 0 \tag{23}$$

$$\sum_{i=1}^{N} w_i q_{ik} x_i = \mu_k \sum_{i=1}^{N} w_i q_{ik} \tag{24}$$

求解 μ_k：

$$\boxed{\mu_k^* = \frac{\sum_{i=1}^{N} w_i q_{ik} x_i}{\sum_{i=1}^{N} w_i q_{ik}}} \tag{25}$$

考虑依赖于 Σ_k的项：

$$J(\Sigma_k) = \sum_{i=1}^{N} w_i q_{ik} \left(-\frac{1}{2}\ln|\Sigma_k| - \frac{1}{2}(x_i-\mu_k)^T \Sigma_k^{-1}(x_i-\mu_k)\right) \tag{26}$$

使用矩阵导数恒等式 $\frac{\partial \ln|\Sigma|}{\partial \Sigma} = \Sigma^{-1}$ 和 $\frac{\partial a^T \Sigma^{-1} a}{\partial \Sigma} = -\Sigma^{-1} a a^T \Sigma^{-1}$:

$$\frac{\partial J}{\partial \Sigma_k} = \sum_{i=1}^{N} w_i q_{ik} \left(-\frac{1}{2}\Sigma_k^{-1} + \frac{1}{2}\Sigma_k^{-1}(x_i - \mu_k)(x_i - \mu_k)^T \Sigma_k^{-1} \right) \tag{27}$$

将其置零，并在两边乘以 Σ_k：

$$\sum_{i=1}^{N} w_i q_{ik} I = \sum_{i=1}^{N} w_i q_{ik}(x_i - \mu_k)(x_i - \mu_k)^T \Sigma_k^{-1} \tag{28}$$

$$\Sigma_k \left(\sum_{i=1}^{N} w_i q_{ik} \right) = \sum_{i=1}^{N} w_i q_{ik}(x_i - \mu_k)(x_i - \mu_k)^T \tag{29}$$

求解 Σ_k:

$$\boxed{\Sigma_k^* = \frac{\sum_{i=1}^{N} w_i q_{ik}(x_i - \mu_k^*)(x_i - \mu_k^*)^T}{\sum_{i=1}^{N} w_i q_{ik}}} \tag{30}$$

这些方程与 GMM-PIELM 论文公式中的公式 6 完全一致。

A.5 关于对数变换的论证

$p_{res}(x) \propto \log(1 + |\mathcal{R}(x)|)$ 是动态范围压缩的关键机制，尤其适用于我们框架所针对的刚性区域。使用原始残差平方场进行的实证测试显示，所有中心都集中在边界处，因为在奇异摄动问题中，边界层与区域其余部分之间的残差相差数个数量级。这导致基函数分布发生“塌缩”，全局区域信息丢失，并且线性系统在较平滑的区域中变得病态。通过应用对数变换，我们成功抑制了这些极端变化，使高斯混合模型能够解析误差分布的“肩部”，而不仅仅是奇异性。这确保了在优先处理高梯度区域的同时，整个区域仍能保留足够的信息和基函数支撑，从而获得稳定且准确的解。

A.6 实现细节

实验设置 所有实验均使用 scikit-learn 进行高斯混合模型估计，并使用 scipy 执行线性代数运算，同时将随机种子固定为 42。对于单边界层和双边界层实验，我们从区域 $\Omega = (0, 1)$ 中均匀采样固定数量的 $N_{\text{eval}} = 1500$ 配点，以构建线性系统。通过将边界坐标附加到配点集合中，利用软约束施加边界条件。

基线 RBF-PIELM 基线模型使用 N 个神经元初始化隐藏层（$N = 300$个用于单边界层，$N = 500$ 个用于双边界层）。中心均匀采样如下:

$$x_0^j \sim \text{Uniform}(0, 1), \quad j = 1, \ldots, N. \tag{31}$$

宽度设置为常数值 $s_j = |\Omega_N| \times 2.5$，与 Dwivedi 等人（2025a）提出的重叠类比一致。

提出的方法：自适应 GMM-PIELM 该自适应方法在 T 次迭代中逐步优化网络架构。在每次迭代中，我们在密集网格上计算残差场 $\mathcal{R}(x;\ \theta)$，并根据方程 12 的定义构造残差能量密度 $p_{\text{res}}(x)$。随后，我们对从 $p_{\text{res}}(x)$ 中采样的数据拟合高斯混合模型，并使用混合采样策略生成一组新的中心 $\{x_0^j\}_{j=1}^N$：

- **残差引导（100α%）：** 从拟合的GMM分量中采样，以针对高误差区域（边界层）。
- **全局均匀采样（$100(1-\alpha)$%）：** 从 Ω 均匀采样，作为防止光滑区域基函数耗尽的安全网。

为适应中心的非均匀聚类，局部宽度 s_j 采用 k-最近邻居（k-NN）进行自适应调整：

$$s_j = \beta \cdot \text{dist}_k(x_0^j) + \epsilon, \tag{32}$$

where dist_k 是到第 k个邻居（$k = 2$）的欧几里得距离， β 是重叠缩放因子 r.

超参数表2详细列出了单边界层（BL）和双边界层实验所使用的具体参数。

表2：实现的超参数设置。

参数	**单边界层**	**双边界层**
扩散率（ν）	10^{-4}	10^{-4}
神经元数量（N）	300	500
GMM分量（K）	8	16
混合比例（α）	0.7	0.7
适配迭代次数（T）	3	3
西格玛缩放 (β)	1.1	1.5
随机种子	42	42