arXiv:2301.05729v1 [stat.ML] 13 Jan 2023

GAR：面向多保真融合的广义自回归

Yuxin Wang *
北京航空航天大学数学科学学院，中国北京，100191。
WYXtt_2011@163.com

Zheng Xing *
图形工程部，瑞芯微电子有限公司，中国福州，350003
zheng.xing@rock-chips.com

Wei W.Xing †北京航空航天大学集成电路科学与工程学院，中国北京，100191。
wayne.xingle@gmail.com

摘要

在许多需要对复杂系统进行重复仿真的科学研究和工程应用中，人们通常采用代理模型来快速估计整个系统。为降低生成训练样本的高昂成本，结合低保真度（速度快但精度低）与高保真度（速度慢但精度高）仿真的结果，已成为一种颇具前景的方法。尽管多保真融合技术发展迅速，但现有的大多数方法都要求特定的数据结构，且难以扩展至高维输出。为解决这些问题，我们对经典自回归（AR）进行推广。经典自回归因其简单、稳健、准确且易于处理而被广泛使用。我们采用张量表述和潜在特征，提出了广义自回归（GAR）。GAR能够处理任意维度的输出和任意多保真度数据结构，以满足复杂问题的多保真融合需求；它具有完全可处理的似然和后验分布，无需近似推断，并且能够良好扩展至高维问题。此外，我们证明了多保真度情形下基于GAR的自动克里金定理，并开发了CIGAR，这是一种简化的GAR，在计算量减少至原来的 d^3倍的同时，仍能保持预测均值的精确度，其中 d 是输出的维数。实证评估涵盖了多个典型偏微分方程和真实科学实例，结果表明，所提出的方法始终显著优于最先进方法（均方根误差最高提升6倍），且仅需少量高保真训练样本。

1 引言

许多科学与工程系统的设计、优化和控制，都高度依赖微分方程的数值分析，而这通常需要大量计算。在这种情况下，可利用数据驱动代理模型，根据数值求解器的输入输出数据近似系统，并在需要重复仿真的场景中帮助提高收敛效率，例如贝叶斯优化 (BO) [1] 和不确定性分析 (UQ) [2]。

有了代理模型后，剩下的挑战是，执行高保真数值模拟来生成训练数据仍然可能代价高昂。为了进一步降低计算

*作者对本文的贡献相同。

†通讯作者。

预印本。审稿中。

负担，可以结合低保真度结果来进行高保真度预测 [3]。具体来说，低保真度求解器通常基于简化的偏微分方程（例如，减少物理细节层次）或简单的求解器设置（例如，使用粗网格、较大的时间步长、较低阶的近似基，以及较大的误差容限）。它们能够快速求解原始问题，但结果不够准确；高保真度求解器则准确但成本高昂。多保真度融合技术的工作方式类似于迁移学习，通过利用大量低保真观测，在仅使用少量高保真度样本时提高精度。通常，这一方法涉及为不同保真度构建代理模型，并建立跨保真度转换过程。由于效率较高，多保真度方法在贝叶斯优化 [4, 5], 、不确定性量化 [6, 7], 和代理建模 [8, 9]中日益受到关注。有关精彩综述，请参阅 [10, 11]。

尽管许多最先进方法取得了成功，但它们通常假设：（1）所有保真度的输出维度相同且彼此对齐，而这通常不适用于多保真度仿真，因为其输出是在节点处的量值，天然无法对齐；（2）高保真度样本对应的输入构成低保真度输入的子集；（3）输出维度较小。然而，在科学计算中，这并不现实，因为维度可能高达 100 万（对于一个 $100 \times 100 \times 100$ 时空场而言）。这些假设严重限制了它们在实际问题中的应用，例如磁共振成像和科学计算中的偏微分方程求解。

为应对这些挑战，先前的研究要么采用插值来对齐维度 [12, 9] ，要么依赖近似推断并进行粗暴简化 [8, 13], ，从而导致性能不佳。我们注意到，经典自回归（AR）因其简单、稳健、准确且易于处理而得到广泛应用；尽管它无法处理高维问题，但在文献报道中，它在不同数据集上始终表现稳健且出色。因此，我们不再提出另一个需要预处理和简化的临时模型（这会导致模型难以调优且泛化能力较差），而是利用张量代数和潜在特征对 AR 进行推广，提出广义自回归（GAR）。该方法能够处理任意高维问题，且不需要多保真度数据采用子集结构。GAR 是一种完全可处理的贝叶斯模型，可扩展至极高维输出，且无需进行任何近似推断。本工作的创新点如下：

1. 将 AR 推广到任意非结构化高维输出。借助张量代数和潜在特征，GAR 能够有效地进行知识迁移，并以闭式形式实现，且可扩展至极高维问题。2. 将 AR 推广到非子集多保真度数据。据我们所知，我们首次将子集数据的 AR 闭式解推广到非子集情形，并将其推广至所提出的 GAR。3. 我们首次揭示了 AR 结构中多保真融合的自克里金能力 [14] ，并据此推导出条件独立广义自回归（CIGAR），这是一种高效的 GAR 实现，能够精确得到后验均值预测。

2 背景

2.1 问题陈述

给定多保真度数据 $D^i = \{\mathbf{x}_n^i,\ \mathbf{y}_n^i\}_{n=1}^{N^i}$，对于 $i = 1, \ldots, \tau$，其中 $\mathbf{x} \in \mathbb{R}^l$ 表示系统输入（出现在方程组和/或仿真的初始边界条件中的参数向量）；$\mathbf{y}^i \in \mathbb{R}^{d^i}$ 表示相应的输出，其中 d^i 是 i 保真度的维度；τ 是保真度的总数。一般而言，高保真度数据更接近真实值，获取成本也更高。因此，高保真度数据的样本更少，即 $N^1 \geqslant N^2 \geqslant \cdots \geqslant N^\tau$。不同保真度间的维度不一定相同或对齐。在大多数工作中，例如 [15, 16, 11], ，高保真度的系统输入被选为低保真度系统输入的子集，即 $\mathbf{X}^\tau \subset \cdots \subset \mathbf{X}^2 \subset \mathbf{X}^1$。我们将其称为多保真度数据集的子集结构，与任意数据结构相对。我们将在第 3.3 节中通过闭式解处理任意数据结构，并将其扩展至经典自回归模型。我们的目标是根据不同保真度下的观测数据 $\{D^i\}_{i=1}^\tau$，估计函数$\mathbf{f}^\tau$： $\mathbb{R}^l \to \mathbb{R}^{d^\tau}$。

2.2 自回归

为清晰起见，我们考虑一个双保真度情形，上标 h 表示高保真度，上标l 表示低保真度。不过，该公式可以递归推广至具有多个保真度的问题。对于所有保真度均为简单标量输出的情况，自回归模型 [3] 假设

$$f^h(\mathbf{x}) = \rho f^l(\mathbf{x}) + f^r(\mathbf{x}), \tag{1}$$

其中， ρ 是以线性方式从低保真度传递知识的因子，而 $f^r(\mathbf{x})$ 则试图捕捉残差信息。假设 $f^l(\mathbf{x})$ 和 $f^r(\mathbf{x})$ 服从零均值高斯过程（GP）先验 [17]（简要介绍见附录 A），即 $f^l(\mathbf{x}) \sim \mathcal{N}(\mathbf{0}, k^l(\mathbf{x},\mathbf{x}'))$ 且 $f^r(\mathbf{x}) \sim \mathcal{N}(\mathbf{0}, k^r(\mathbf{x},\mathbf{x}'))$，则高保真函数也服从高斯过程。这使得联合观测 $\mathbf{y} = [\mathbf{y}^l; \mathbf{y}^h]^T$ 可以构成优雅的联合高斯过程，

$$\begin{pmatrix} \mathbf{y}^l \\ \mathbf{y}^h \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0},\ \begin{matrix} \mathbf{K}^l(\mathbf{X}^l,\mathbf{X}^l) & \rho\mathbf{K}^l(\mathbf{X}^l,\mathbf{X}^h) \\ \rho\mathbf{K}^l(\mathbf{X}^h,\mathbf{X}^l) & \rho^2\mathbf{K}^l(\mathbf{X}^h,\mathbf{X}^h) + \mathbf{K}^r(\mathbf{X}^h,\mathbf{X}^h) \end{matrix}\right) \tag{2}$$

其中，$\mathbf{y}^l \in \mathbb{R}^{N_l\times 1}$，$\mathbf{y}^h \in \mathbb{R}^{N_h\times 1}$ 是与输入 $\mathbf{X}$ 对应的低保真观测，$\mathbf{y}$ 是高保真观测；$[\mathbf{K}^l(\mathbf{X}^l,\mathbf{X}^l)]_{ij} = k^l(\mathbf{x}_i,\mathbf{x}_j)$ 是低保真输入 $\mathbf{x}_i,\mathbf{x}_j \in \mathbf{X}^l$ 的协方差矩阵；$[\mathbf{K}^r(\mathbf{X}^l,\mathbf{X}^l)]_{ij} = k^r(\mathbf{x}_i,\mathbf{x}_j)$ 是高保真输入 $\mathbf{x}_i,\mathbf{x}_j \in \mathbf{X}^h$ 的协方差矩阵；$[\mathbf{K}^l(\mathbf{X}^l,\mathbf{X}^h)]_{ij} = k^l(\mathbf{x}_i,\mathbf{x}_j)$ 是低保真输入 $\mathbf{x}_i \in \mathbf{X}^l$ 与高保真输入 $\mathbf{x}_j \in \mathbf{X}^h$ 之间的跨保真度协方差矩阵，且 $\mathbf{K}^l(\mathbf{X}^h,\mathbf{X}^l) = (\mathbf{K}^l(\mathbf{X}^l,\mathbf{X}^h))^T$。自回归的一个直接优势是，联合高斯形式不仅允许对所有低保真度和高保真度数据进行联合训练，还允许通过条件高斯分布对任意给定的新输入进行预测（因为后验分布是在标准高斯过程框架 [17] 中推导的）。此外，Le Gratiet [15] 推导出引理 1，通过子集数据结构将复杂度从 $O((N^l+N^h)^3)$ 降至 $O((N^l)^3+(N^h)^3)$。

引理 1. *[15]* 如果$\mathbf{X}^h \subset \mathbf{X}^l$，自回归模型的联合似然和预测后验可分解为两个独立部分，分别对应低保真度和高保真度数据。

3 广义自回归

现在考虑更一般的高维情形。一种简单的方法是将多维输出转换为标量值，并将维度索引附加到输入上。然而，自回归模型最终会得到一个联合高斯过程，其协方差矩阵的大小为 $(N^l d^l + N^h d^h)^2$,，这使得它对于中等规模的高维问题也难以实现。

3.1 带潜在特征的张量分解推广

为了解决可扩展性问题，我们将所有输出重新排列到一个多维空间（即张量空间）中，并引入潜在坐标特征来索引输出，以捕捉它们之间的相关性，如 HOGP [18] 所示。更具体地说，我们将低保真度输出组织为一个 M 模张量，$\mathbf{Z}^l \in \mathbb{R}^{d_1^l\times\cdots\times d_M^l}$，其中输出维度为 $d^l = \prod_{m=1}^M d_m^l$。元素 $\mathcal{Z}^l$ 根据其坐标 $\mathbf{c} = (c_1,\ldots,c_M)(1 \leqslant c_k \leqslant d_k$（对于 $k = 1,\ldots,M$）进行索引。如果原始数据确实具有多数组结构，我们可以使用具有实际含义的原始索引来索引这些坐标。例如，二维空间数据集可以使用其原始空间坐标来索引单个位置（像素）。为了提高模型灵活性，我们不必局限于使用原始索引，尤其是在原始数据不具有多数组结构，或多数组结构规模过大的情况下。在这种情况下，我们可以采用任意张量化，并为模 m 中的每个坐标 c_m 使用一个潜在特征向量 $\mathbf{v}^l_{c_m}$（其值在模型训练过程中推断得到）。这样，元素$\mathcal{Z}^l$就由向量（$\mathbf{v}^l_{c_1},\ldots,\mathbf{v}^l_{c_M}$）进行索引。根据式（1）的线性变换，我们首先引入张量-矩阵积 [19],

$$\mathbf{F}^h(\mathbf{x}) = \mathbf{F}^l(\mathbf{x}) \times_1 \mathbf{W}_1,\ldots,\times_M \mathbf{W}_M + \mathbf{F}^r(\mathbf{x}), \tag{3}$$

其中，$\mathbf{F}^h(\mathbf{x})$ 表示目标函数 $\mathbf{f}^h(\mathbf{x})$，其输出组织成多数组 $\mathbf{Z}^h$；对于 $\mathbf{F}^l(\mathbf{x})$ 和 $\mathbf{F}^r(\mathbf{x})$，同样适用这一概念；$\times_m$ 表示在模 m 上的张量-矩阵积。具体而言，对于任意张量 $\mathbf{Z}^l \in \mathbb{R}^{d_1^l\times\cdots\times d_M^l}$ 和矩阵 $\mathbf{W}_m \in \mathbb{R}^{s\times d_m}$，$\times_m$ 积计算如下：$[\mathbf{Z}^l \times_m \mathbf{W}_m]_{i_1,\ldots,i_{m-1},j,i_{m+1}\cdots,i_M} = \sum_{k=1}^{c_m} w_{jk}\mathbf{Z}_{i_1,\ldots,i_{m-1},k,i_{m+1}\cdots,i_M}$，所得结果为一个 $d_1^l\times\cdots\times d_{m-1}^l\times s\times d_{m+1}^l\times\cdots\times d_M^l$ 张量。我们还可以将这组

线性变换矩阵表示为 Tucker 张量 $\mathbf{W} = [\mathbf{W}_1, \ldots, \mathbf{W}_M]$，并使用 Tucker 算子 [19], $\mathbf{F}^l(\mathbf{x}) \times \mathbf{W}$ 将式 (3) 紧凑地表示为:

$$\operatorname{vec}\left(\mathbf{F}^h(\mathbf{x}) - \mathbf{F}^r(\mathbf{x})\right) = (\mathbf{W}_1 \otimes \ldots \otimes \mathbf{W}_M) \operatorname{vec}\left(\mathbf{F}^l(\mathbf{x})\right). \tag{4}$$

受式 (1) 中自回归的启发，我们为低保真张量函数 $\mathbf{F}^l(\mathbf{x})$ 和残差张量函数$\mathbf{F}^r(\mathbf{x})$ 设定张量变量高斯过程（TGP）先验 [20]：

$$\mathbf{Z}^l(\mathbf{x}, \mathbf{x}') \sim \mathcal{TGP}\left(\mathbf{0}, k^l(\mathbf{x}, \mathbf{x}'), \mathbf{S}_1^l, \ldots, \mathbf{S}_M^l\right), \mathbf{Z}^r(\mathbf{x}, \mathbf{x}') \sim \mathcal{TGP}\left(\mathbf{0}, k^r(\mathbf{x}, \mathbf{x}'), \mathbf{S}_1^r, \ldots, \mathbf{S}_M^r\right), \tag{5}$$

其中，$\mathbf{S}_m^i \in \mathbb{R}^{d_m \times d_m}$ 为输出相关矩阵，$[\mathbf{S}_m^i]_{jk} = \tilde{k}_m^i(\mathbf{v}_{c_i}^i, \mathbf{v}_{c_k}^i)$ 和 $\tilde{k}_m^i(\cdot,\cdot)$ 为核函数（其超参数未知）。TGP 是多变量高斯过程的一种推广，本质上表示联合高斯过程先验 $\operatorname{vec}(\mathbf{Y}^l) \sim \mathcal{N}\left(0, \mathbf{K}^l(\mathbf{X}^l, \mathbf{X}^l) \bigotimes_{m=1}^M \mathbf{S}_m\right)$。类似于式 (2) 的联合概率，我们可以根据式 (3) 的 Tucker 变换推导出 $\mathbf{y} = [\operatorname{vec}(\mathbf{Y}^l)^T, \operatorname{vec}(\mathbf{Y}^h)^T]^T$ 的联合概率；为便于理解，我们将证明放在附录中。

引理 2. 给定$\mathbf{Y}^l(\mathbf{x},\mathbf{x}')$ 和$\mathbf{Y}^r(\mathbf{x},\mathbf{x}')$ 的张量高斯过程先验以及式(3)的 *Tucker* 变换，$\mathbf{y} = [\operatorname{vec}(\mathbf{Y}^l)^T$、$\operatorname{vec}(\mathbf{Y}^h)^T]^T$ 的联合概率为$\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$，其中 $\boldsymbol{\Sigma} =$

$$\begin{pmatrix} \mathbf{K}^l(\mathbf{X}^l, \mathbf{X}^l) \otimes \left(\bigotimes_{m=1}^M \mathbf{S}_m\right) & \mathbf{K}^l(\mathbf{X}^l, \mathbf{X}^h) \otimes \left(\bigotimes_{m=1}^M \mathbf{S}_m \mathbf{W}_m^T\right) \\ \mathbf{K}^l(\mathbf{X}^h, \mathbf{X}^l) \otimes \left(\bigotimes_{m=1}^M \mathbf{W}_m \mathbf{S}_m\right) & \mathbf{K}^l(\mathbf{X}^h, \mathbf{X}^h) \otimes \left(\bigotimes_{m=1}^M \mathbf{W}_m \mathbf{S}_m \mathbf{W}_m^T\right) + \mathbf{K}^r(\mathbf{X}^h, \mathbf{X}^h) \otimes \left(\bigotimes_{m=1}^M \mathbf{S}_m^r\right) \end{pmatrix}.$$

引理 2 允许不同保真度下的任意输出（这些输出可以属于不同空间、具有不同维度和/或模，且彼此不对齐）。此外，它不要求数据集之间存在子集关系。

推论 3.0.1. 如果添加冗余索引，使所有输出具有相同的 M模数，则引理 *2* 可应用于每种保真度下模数不同的数据，即 $M^l \neq M^h$。

引理 2 定义了我们的 GAR 模型，即一种具有特殊张量结构的广义自回归模型。低保真度下的协方差为 $\operatorname{cov}(\mathrm{Z}_{\mathbf{c}}^l(\mathrm{x}), \mathrm{Z}_{\mathbf{c}}^l(\mathrm{x}')) = k^l(\mathrm{x}, \mathrm{x}') \prod_{m=1}^M \tilde{k}_m^l(\mathrm{v}_{c_m}^m, \mathrm{v}_{c_m'}^m)$，跨保真度协方差为 $\operatorname{cov}(\mathrm{Z}_{\mathbf{c}}^l(\mathrm{x}), \mathrm{Z}_{\mathbf{c}'}^h(\mathrm{x}')) = k^l(\mathrm{x}, \mathrm{x}') \prod_{m=1}^M \tilde{k}_m^l(\mathrm{v}_{c_m}^l, \mathrm{v}_{c_m'}^l) w_c^m$，$_{c'}$(其中 $w_c^m{}_{,c'}$ 是 W^m 的第 c, c' 个元素)，高保真度下的协方差为 $\operatorname{cov}(\mathrm{Z}^{\ h}\mathrm{c}(\mathrm{x}), \mathrm{Z}_{\mathbf{c}'}^h(\mathrm{x}')) = k^l(\mathrm{x}, \mathrm{x}') \prod_{m=1}^M \tilde{k}_m^l(\mathrm{v}_{c_m}^l, \mathrm{v}_{c_m'}^l)(w_c^m{}_{,c'})^2 + k^h(\mathrm{x}, \mathrm{x}') \prod_{m=1}^M \tilde{k}_m^r(\mathrm{u}_{c_m}^m, \mathrm{u}_{c_m'}^m)(w_c^m{}_{,c'})^2$。不同保真度之间复杂的输出相关性通过潜在特征 $\{\mathrm{V}^m, \mathrm{U}^m\}_{m=1}^M$ 和任意核函数 $\tilde{k}_m^i$ 捕获，而跨保真度输出相关性则以复合方式捕获。这种组合克服了以往工作所假设的简单线性相关性；以往工作只是将输出分解为降维预处理 [12]。当 Z^l 和 Z^h 的维度对齐，因而 $d_m^l = d_m^h$ 时，我们可以令 $\mathrm{v}_j^m = \mathrm{u}_j^m$，在使用不同核函数的同时，共享两个保真度下的潜在特征。这样，潜在特征就更能抵抗过拟合。对于具有显式索引但未对齐的数据，我们可以使用核插值 [21] 来实现相同目的。为进一步促进潜在特征的稀疏性，我们施加拉普拉斯先验，即 $\mathrm{v}_j^m \sim \mathrm{Laplace}(\lambda) \propto \exp(-\lambda||\mathrm{v}_j^m||_1)$。

3.2 子集数据结构的高效模型推断

模型定义完成后，我们现在可以训练模型以获得所有未知模型参数。为简洁起见，我们使用以下紧凑记号 $\mathrm{S}^l = \bigotimes_{m=1}^M \mathrm{S}_m^l$, $\mathrm{S}^h = \bigotimes_{m=1}^M \mathrm{S}_m^h$, $\mathrm{W} = \bigotimes_{m=1}^M \mathrm{W}_m$, $\mathrm{K}^l = \mathrm{K}^l(\mathrm{X}^l, \mathrm{X}^l)$, $\mathrm{K}^{lh} = \mathrm{K}^l(\mathrm{X}^l, \mathrm{X}^h)$, $\mathrm{K}^{hl} = \mathrm{K}^l(\mathrm{X}^h, \mathrm{X}^l)$, $\mathrm{K}^{lr} = \mathrm{K}^l(\mathrm{X}^h, \mathrm{X}^h)$，以及 $\mathrm{K}^r = \mathrm{K}^r(\mathrm{X}^h, \mathrm{X}^h)$（这里稍微滥用了记号）。

引理 3. 引理 *1* 的张量推广。若 $\mathbf{X}^h \subset \mathbf{X}^l$，则 $\mathbf{y} = [\operatorname{vec}(\mathbf{Y}^l)^T$、$\operatorname{vec}(\mathbf{Y}^h)^T]^T$ 的联合似然 $\mathcal{L}$ 具有两个相互独立的可分离似然 $\mathcal{L} = \mathcal{L}^l + \mathcal{L}^r$，其中

$$\mathcal{L}^l = -\frac{1}{2}\operatorname{vec}\left(\mathbf{Y}^l\right)^T (\mathbf{K}^l \otimes \mathbf{S}^l)^{-1} \operatorname{vec}\left(\mathbf{Y}^l\right) - \frac{1}{2}\log|\mathbf{K}^l \otimes \mathbf{S}^l| - \frac{N^l D^l}{2}\log(2\pi),$$

$$\mathcal{L}^r = -\frac{1}{2} vec\left(Y^h - Y^l \times \hat{}W\right)^T (K^r \otimes S^r)^{-1} vec\left(Y^h - Y^l \times \hat{}W\right) - \frac{1}{2} log\ |K^r \otimes S^r| - \frac{N^h D^h}{2} log(2\pi),$$

其中 $\hat{}W = [E, W1, \ldots, W_M]$ 是一个带有选择矩阵 $E^T X^l = X^h$ 的 *Tucker* 张量。

为清晰起见，我们将证明保留在附录中。注意， $\mathcal{L}^l$ 和 $\mathcal{L}^r$ 分别是$\mathbf{Y}^l$ 和残差 $\mathbf{Y}^h - \mathbf{Y}^l \times \hat{\mathbf{W}}$ 的 HOGP 似然。由于 $\mathcal{L}^l$ 和 $\mathcal{L}^r$ 的计算相互独立，因此可以高效地并行进行模型训练。

预测后验分布。同样，经过一些繁琐的线性代数运算后（见附录），我们可以通过对潜在函数积分消去，推导出高保真输出的具体预测后验分布。该分布同样服从高斯分布：$\text{vec}(\mathbf{Z}_*^h) \sim \mathcal{N}(\text{vec}(\bar{\mathbf{Z}}_*^h), \mathbf{S}_*^h)$，其中

$$\begin{aligned} \text{vec}(\bar{\mathbf{Z}}_*^h) &= \left(\mathbf{k}_*^l \left(\mathbf{K}^l\right)^{-1} \otimes \mathbf{W}\right) \text{vec}(\mathbf{Y}^l) + \left(\mathbf{k}_*^r \left(\mathbf{K}^r\right)^{-1} \otimes \mathbf{I}\right) \text{vec}(\mathbf{Y}^r), \\ \mathbf{S}_*^h &= \left(k_{**}^l - (\mathbf{k}_*^l)^T \left(\mathbf{K}^l\right)^{-1} \mathbf{k}_*^l\right) \otimes \mathbf{W}\mathbf{S}^l\mathbf{W}^T + \left(k_{**}^r - (\mathbf{k}_*^r)^T \left(\mathbf{K}^r\right)^{-1} \mathbf{k}_*^r\right) \otimes \mathbf{S}^r, \end{aligned} \tag{6}$$

$\mathbf{k}_*^l = \mathbf{k}^l(\mathbf{x}_*, \mathbf{X}^l)$ 是给定输入 $\mathbf{x}_*$ 与低保真观测输入 $\mathbf{X}^l$ 之间的协方差向量；类似地，我们有 $\mathbf{k}_{**}^l = \mathbf{k}^l(\mathbf{x}_*, \mathbf{x}_*)$、$\mathbf{k}_*^r = \mathbf{k}^r(\mathbf{x}_*, \mathbf{X}^h)$、$\mathbf{k}_{**}^r = \mathbf{k}^r(\mathbf{x}_*, \mathbf{x}_*)$。

3.3 非子集数据的泛化：高效的模型推断与预测

在实践中，有时很难要求多保真度数据保持子集结构，尤其是在多保真贝叶斯优化的情形下 [22, 23]。这给大多数最先进的多保真度模型带来了挑战，例如 NAR [16], ResGP [9], 随机配置 [24]。相比之下，自回归模型的优势在于，即使多保真度数据不具备子集数据结构，仍可依据式 (2) 的联合似然，利用所有可用数据训练模型。然而，由于需要对大型联合协方差矩阵求逆，这种方法缺乏可扩展性 $\boldsymbol{\Sigma}$。如果处理的是包含两种以上保真度的多保真度数据，情况会更糟。为解决这一问题，我们提出一种基于虚拟子集的快速推断方法。具体而言，

将缺失的低保真度数据视为潜变量 $\hat{\mathbf{Y}}^l$，联合似然函数 is

$$\begin{aligned} \log p(\mathbf{Y}^l, \mathbf{Y}^h) &= \log \int p(\mathbf{Y}^l, \mathbf{Y}^h, \hat{\mathbf{Y}}^l) d\hat{\mathbf{Y}}^l = \log \int \left(p(\mathbf{Y}^h|\hat{\mathbf{Y}}^l, \mathbf{Y}^l) p(\hat{\mathbf{Y}}^l|\mathbf{Y}^l) p(\mathbf{Y}^l)\right) d\hat{\mathbf{Y}}^l \\ &= \log \int p(\mathbf{Y}^h|\hat{\mathbf{Y}}^l, \mathbf{Y}^l) p(\hat{\mathbf{Y}}^l|\mathbf{Y}^l) d\hat{\mathbf{Y}}^l + \log p(\mathbf{Y}^l), \end{aligned} \tag{7}$$

其中， $p(\mathbf{Y}^h|\hat{\mathbf{Y}}^l, \mathbf{Y}^l)$ 是引理 3 中给定互补虚拟子集时的似然，而 $p(\hat{\mathbf{Y}}^l|\mathbf{Y}^l) \sim \mathcal{N}(\bar{\mathbf{Y}}^l, \hat{\mathbf{S}}^l \otimes \mathbf{S}^l)$ 是给定低保真观测 $\mathbf{Y}^l$ 时的虚拟后验分布。该积分可以使用高斯求积或其他采样方法计算，如 [8, 25], 所示，但速度较慢且精度较低。

引理 4. *GAR* 对非子集（以及未对齐）数据的联合似然可分解为两个相互独立的高斯过程的似然

$$\begin{aligned} \log p(\mathbf{Y}^l, \mathbf{Y}^h) &= \mathcal{L}^l - \frac{N^h d^h}{2} \log(2\pi) - \frac{1}{2} \log \left|\mathbf{K}^r \otimes \mathbf{S}^r + \hat{\mathbf{E}}\hat{\mathbf{S}}^l\hat{\mathbf{E}}^T \otimes \mathbf{W}^T\mathbf{S}^l\mathbf{W}\right| \\ &- \frac{1}{2} \left[\begin{pmatrix} \text{vec}(\check{\mathbf{Y}}^h) \\ \text{vec}(\hat{\mathbf{Y}}^h) \end{pmatrix}^T - \begin{pmatrix} \text{vec}(\check{\mathbf{Y}}^l) \\ \text{vec}(\bar{\mathbf{Y}}^l) \end{pmatrix}^T \tilde{\mathbf{W}}^T\right] \left(\mathbf{K}^r \otimes \mathbf{S}^r + \hat{\mathbf{E}}\hat{\mathbf{S}}^l\hat{\mathbf{E}}^T \otimes \mathbf{W}^T\mathbf{S}^l\mathbf{W}\right)^{-1} \left[\begin{pmatrix} \text{vec}(\check{\mathbf{Y}}^h) \\ \text{vec}(\hat{\mathbf{Y}}^h) \end{pmatrix} - \tilde{\mathbf{W}} \begin{pmatrix} \text{vec}(\check{\mathbf{Y}}^l) \\ \text{vec}(\bar{\mathbf{Y}}^l) \end{pmatrix}\right], \end{aligned} \tag{8}$$

其中， $\mathcal{L}^l$ 是低保真数据 Y^l 的似然； $\tilde{W} = I_{N^h} \otimes W$ $\hat{Y}^h$ 是与虚拟低保真度输出 $\hat{Y}^l$ 对应的高保真观测集合； $\check{Y}^h$ 是与低保真度输出 $\check{Y}^l$ 对应的补集（其选择矩阵为 $\check{X}^h = E^T X^l$），即 $Y^h = [\check{Y}^h, \hat{Y}^h]$； $\hat{X}^h = \hat{E}^T X^h$ 是 $\hat{Y}^l$ 的选择矩阵。

我们保留附录中的证明。注意，$\hat{\mathbf{E}}\hat{\mathbf{S}}^l\hat{\mathbf{E}}^T \otimes \mathbf{W}^T\mathbf{S}^l\mathbf{W} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{S}}^l \end{pmatrix} \otimes \mathbf{W}^T\mathbf{S}^l\mathbf{W}\hat{}$ 是缺失低保真观测 $\text{vec}(\mathbf{Y})$ 的预测方差的右下块；我们可以很容易地将似然的最后一部分理解为一个在相应缺失点上加入累积不确定性（方差）的 GP。当输出为标量时，引理 4 自然适用于 AR，其中 $\mathbf{W} = \rho$、$\mathbf{S}^l = 1$ 和 $\mathbf{S}^r = 1$。

预测后验。 令人惊讶的是，后验分布也呈现为高斯分布，

$$
\begin{aligned}
p\left(\mathbf{Z}_*^h|\mathbf{Y}^l,\mathbf{Y}^h,\mathbf{x}_*\right) &= 2\pi^{-\frac{d^h}{2}} \times \left|\mathbf{S}_*^h + \mathbf{\Gamma}\left(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l\right)\mathbf{\Gamma}^T\right|^{-\frac{1}{2}} \\
&\times \exp\left[-\frac{1}{2}\left(\mathrm{vec}(\mathbf{Z}_*^h) - \mathrm{vec}(\bar{\mathbf{Z}})\right)^T \left(\mathbf{S}_*^h + \mathbf{\Gamma}\left(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l\right)\mathbf{\Gamma}^T\right)^{-1}\left(\mathrm{vec}(\mathbf{Z}_*^h) - \mathrm{vec}(\bar{\mathbf{Z}})\right)\right],
\end{aligned} \tag{9}
$$

其中， $\mathbf{\Gamma}$ 和预测后验的均值$\mathbf{Z}_*$ 如下所示，

$$
\begin{aligned}
\mathbf{\Gamma} &= \left([\mathbf{k}_*^r(\mathbf{K}^r)^{-1}\mathbf{E}_n^T - \mathbf{k}_*^l(\hat{\mathbf{K}}^l)^{-1}] \otimes \mathbf{W}\right)\mathbf{E}_m \otimes \mathbf{I}^l, \\
\mathrm{vec}(\bar{\mathbf{Z}}_*) &= \left(\mathbf{k}_*^l(\hat{\mathbf{K}}^l)^{-1} \otimes \mathbf{W}\right)\begin{pmatrix}\mathrm{vec}(\mathbf{Y}^l)\\ \mathrm{vec}(\bar{\mathbf{Y}}^l)\end{pmatrix} + \left(\mathbf{k}_*^r(\mathbf{K}^r)^{-1} \otimes \mathbf{I}\right)\left(\mathrm{vec}(\mathbf{Y}^h) - \hat{\mathbf{W}}\begin{pmatrix}\mathrm{vec}(\mathbf{Y}^l)\\ \mathrm{vec}(\bar{\mathbf{Y}}^l)\end{pmatrix}\right),
\end{aligned} \tag{10}
$$

其中，E_m 和 E_n 是选择矩阵，使得 $\hat{}\mathrm{X}^h = \mathrm{E}_m^T[\mathrm{X}^l、\hat{}\mathrm{X}^h]、\mathrm{X}^h = \mathrm{E}_n^T[\mathrm{X}^l、\mathrm{X}^h]、\hat{}\mathrm{W} = \mathrm{E}_n^T \otimes \hat{}$ W，并且 $\hat{}\mathrm{K}^l$ 是协方差矩阵，其中 $\hat{}\mathrm{K}^l = \mathrm{K}^l([\mathrm{X}^l, \hat{}\mathrm{X}^h], [\mathrm{X}^l, \hat{}\mathrm{X}^h])$。

3.4 自克里金能力、复杂度与进一步加速

对于子集结构化数据，GAR 的计算复杂度可分解为两个独立的 TGP，分别用于似然和预测后验。得益于张量代数（主要是 $(\mathbf{K}\otimes\mathbf{S})^{-1} = \mathbf{K}^{-1}\otimes\mathbf{S}^{-1}$），$i$保真度核矩阵求逆的复杂度从 $\mathcal{O}((N^i d^i)^3)$ 降低至 $\mathcal{O}(\sum_{m=1}^{M}(d_m^i)^3 + (N^i)^3)$。对于非子集情形，式（8）中的计算复杂度遗憾地为 $\mathcal{O}((N_m^i d^i)^3)$，其中 N_m 是虚拟低保真度点的数量。尽管如此，借助张量结构，我们仍然可以使用共轭梯度 [26] 高效地求解线性系统。

请注意，式（9）中的均值预测 $\bar{}\mathbf{Z}_*^h$ 不依赖于任何输出协方差矩阵$\{\mathbf{S}_m^h,\ \mathbf{S}_m^l\}_{m=1}^M$，这体现了基于 GAR 框架的自克里金能力（在无噪声情形下，均值预测不涉及知识迁移）。对于不关注预测变异的应用，我们可以引入条件独立输出相关性，即 $\mathbf{S}_m^h = \mathbf{I}$，$\mathbf{S}_m^l = \mathbf{I}$ 以及正交权重矩阵，即 $\mathbf{W}_m^T\mathbf{W}_m = \mathbf{I}$，将计算复杂度进一步降低至 $\mathcal{O}((N^i)^3)$（详细证明见附录）。我们将其称为 CIGAR，即条件独立 GAR 的简称。在我们的实证评估中，CIGAR 的表现略逊于 GAR，这是因为难以确保$\mathbf{W}_m^T\mathbf{W}_m = \mathbf{I}$，且存在数值噪声。

4 相关工作

用于高维输出的高斯过程是许多应用中的重要模型，例如空间数据建模和不确定性量化。关于优秀的综述，读者可参阅 [27]。共区域化线性模型（LMC） [28, 29] 可能是地统计学界为高维高斯过程开发的最通用框架。LMC 假设完整协方差矩阵可表示为常数矩阵与输入相关核的乘积之和。为降低模型复杂度，半参数潜因子模型（SLFM） [30] 通过假设矩阵为秩一矩阵，对 LMC 进行了简化。Higdon 等人 [31] 进一步利用输出集合的奇异值分解（SVD），固定秩一矩阵的基，从而简化 SLFM。为克服 LMC 的线性假设，可以通过流形学习（例如核主成分分析（KPCA） [32] 和 I soMap [33] ）或过程卷积 [34–36]以非线性方式构造（隐式）基。其他方法包括多任务高斯过程，它在类似于 LMC 的框架中将输出视为相互依赖的任务 [37–39] ；以及高斯过程回归网络（GPRN） [40, 41], ，该方法提出用高斯过程的乘积来建模非线性输出，但会导致难以处理的模型。尽管上述方法取得了成功，其复杂度至多为 $\mathcal{O}(N^3 + d^3)$，而有些方法的复杂度为 $\mathcal{O}(N^3 d^3)$，因此难以扩展到科学数据中的高维输出场景，其中 d 可高达一百万。引入张量代数 [42]构建 HOGP [18]，或采用可扩展模型推断（例如 GPRN 中的方法 [43]），可以很好地解决这一问题。

多保真度已成为一种很有前景的方法，可进一步减少构建代理模型 [13] 和贝叶斯优化所需的数据。Kennedy 提出的开创性自回归（AR）模型[3] 对单变量高保真度输出进行线性变换。Le Gratiet [15] 采用线性变换的确定性参数形式，对该模型进行了改进，以实现此前介绍的高效训练方案。

然而，目前尚不清楚 AR 如何处理高维输出。为克服 AR 的线性局限，Perdikaris 等人 [16] 提出了非线性自回归（NAR）。该方法忽略输出分布，直接将低保真度解用作高保真度高斯过程模型的输入，本质上是一种串联高斯过程结构，称为深度高斯过程 [44]。为在多保真度模型中传递不确定性，Cutajar 等人 [25] 采用计算开销高昂的近似推断，因此容易过拟合，且无法处理维度极高的问题。对于多保真度贝叶斯优化（MFBO），Poloczek 等人 [4]和 Kandasamy 等人 [45] 分别用高斯过程独立近似各个保真度；Zhang 等人 [46] 采用卷积核，类似于过程卷积 [34, 36]，以学习保真度之间的相关性；Song 等人 [5] 则将所有保真度的数据合并到单个高斯过程中，以降低不确定性。然而，大多数多保真度贝叶斯优化代理模型无法扩展到高维问题，因为它们是针对单个目标（或至多少数几个目标）设计的。

为处理高维输出（例如时空场），Xing 等人 [9] 通过假设一种简单的加性结构，并以可扩展的多输出高斯过程取代简单的高斯过程来扩展自回归，但代价是失去捕捉输出相关性的能力，导致性能较差且不确定性估计不准确；Xing 等人 [12] 提出深度共区域化，通过学习潜在过程 [30, 29] 来扩展 NAR。该潜在过程是通过所提出的残差主成分分析，将高维输出嵌入残差潜在空间后提取的；Wang 等人 [8] 进一步在深度高斯过程中引入基函数传播和潜在过程传播，以提高模型灵活性，但代价是模型参数数量大幅增加，且近似推断中需要进行一些简化。Parussini 等人 [6]将 NAR 推广至高维问题。然而，这些方法缺乏系统的联合模型训练方法，导致在小型数据集上不稳定且拟合效果不佳。Wu 等人 [47] 利用神经过程扩展高斯过程，从而有效地建模高维及非子集问题。在科学计算中，多保真融合已通过随机配置法（SC） [24] 应用于高维问题，该方法为多保真度问题提供闭式解和高效的实验设计。Xing 等人 [7] 指出，SC 本质上是自回归的一种特殊形式，并提出主动学习来选择高保真度实验的最佳子集。

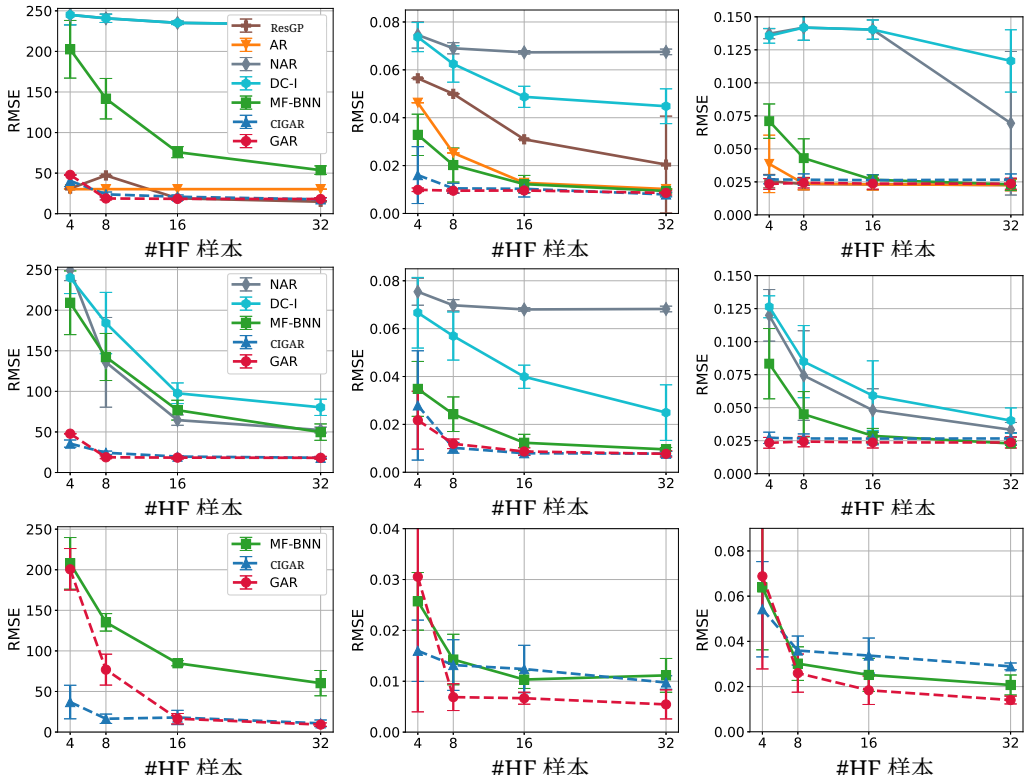
为了借助深度学习神经网络（NN）的优势，并兼容任意多保真度数据（即非子集结构），Perrone 等人 [22] 提出了一种基于神经网络的多任务方法，该方法可自然扩展至多保真度贝叶斯优化（MFBO）。Li 等人 [23] 进一步将其扩展为用于多保真度贝叶斯优化（MFBO）的贝叶斯神经网络（BNN）。Meng 和 Karniadakis [48] 添加了一个物理正则化层，该层需要问题偏微分方程的显式形式，以提高预测精度。为适用于各保真度下维度任意的高维问题，Li 等人 [13] 提出了一种贝叶斯网络方法，并采用主动学习技术来提高效率，实现多保真融合。

除多保真融合外，自回归（AR）还可用于对多变量问题建模 [49, 50]，，GAR 也可应用于此类问题。GAR 是一种基于高斯过程（GP）的高维输出多保真融合通用框架。具体而言，采用可分离核时，令$\mathbf{W} = \rho\mathbf{I}$，AR 即为 GAR 的特例；令$\mathbf{W} = \mathbf{I}$ 且 $\mathbf{S} = \mathbf{I}$，ResGP 即为 GAR 的特例；采用可分离核并以正态先验对$\mathbf{W}$ 积分消去时，NAR 即为 GAR 的特例；若 GAR 仅使用一个潜在过程，并像 NAR 一样以可分离核对$\mathbf{W}$ 积分消去，则 DC 即为 GAR 的特例；若仅使用一个隐藏层，MF-BNN 即为 GAR 的有限形式。有关最先进方法的比较，请参见附录 C。

5 实验结果

为评估 GAR 和 CIGAR，我们将其与用于高维输出的最先进多保真融合方法进行比较，包括：(1) AR [3], (2) NAR [16], (3) ResGP [9], (4) DC[3] [12], 和 (5) MF-BNN[4] [13]。为确保公平比较，所有高斯过程均采用径向基函数核。由于 ARD 核具有可分离性，AR 和 NAR 采用与 GAR 相同的克罗内克积结构进行加速，以实现可行的计算。采用残差主成分分析的原始 DC 无法处理未对齐输出，但使用独立主成分分析后即可处理，我们将其称为 DC-I。两种 DC 都保留 99% 的能量用于降维。MF-BNN 使用其默认设置进行实验。GAR、CIGAR、AR、NAR 和 ResGP 均使用 PyTorch[5] 实现。所有实验均在配备 AMD 5950x CPU 和 32 GB 内存的工作站上运行。

[3]https://github.com/wayXing/DC [4]https://github.com/shib0li/DNN-MFBO [5]https://pytorch.org /



(a) 泊松方程　(b) 伯格斯　(c) 热方程

图 1：对于泊松方程、伯格斯方程和热方程，随着高保真训练样本数量增加，均方根误差的变化。上排为对齐输出，中排为非对齐输出，下排为非子集数据。

5.1 典型偏微分方程的多保真融合

我们首先在经典偏微分方程仿真基准上评估 GAR，这些基准生成高维空间/时空场作为模型输出。具体而言，我们测试了文献中常用的伯格斯方程、泊松方程和热方程 [12, 51–53]。高保真度结果通过在 32×32 网格上使用有限差分求解这些方程得到，而低保真度结果则通过在 8×8粗网格上求解得到。记录这些网格点上的解，并将其向量化作为输出。由于网格不同，维度也各不相同。为了与只能处理对齐输出的标准多保真度方法进行比较，我们采用插值将低保真度结果上采样，并在高保真度网格节点处记录。相应的输入为偏微分方程参数以及参数化的初始条件或边界条件。详细的实验设置见附录 E.1

我们均匀生成 128 个样本用于测试，32 个用于训练。我们将高保真训练样本的数量增加到与低保真训练样本相同的 32 个。通过打乱样本进行五次比较。图 1 报告了均方根误差（RMSE）的统计结果（均值和标准差）。GAR 和 CIGAR 显著优于其他方法，RMSE 最高降低了 6 倍；同时，只需最多 8 个高保真度样本便可达到最优性能，这表明低保真度与高保真度的融合效果良好。CIGAR 的表现略逊于 GAR，这可能是因为实现时未对其权重矩阵的正交性施加硬约束。正如我们在文献中发现的那样，自回归方法始终表现良好。GAR 采用灵活的线性变换，在保留自回归方法稳健性的同时超越了它，从而取得最佳性能。对于未对齐输出，MF-BNN 的表现略逊于对齐情形，凸显了未对齐输出带来的挑战。相比之下，GAR 和 CIGAR 在两种情形下的表现几乎相同。尽管如此，与其他方法相比，MF-BNN 也表现良好，这与 [13] 中的发现一致。有趣的是，对于非子集数据，表现出色的方法比在子集情形下表现更好。GAR 和 CIGAR 仍以明显优势超越其他方法。

为了近似评估主动学习过程中的性能，我们改为按照 Sobol序列 [54]生成训练样本。结果见附录 E.2，其中 GAR 和 CIGAR 同样以显著优势优于其他方法。

5.2 面向现实世界应用的多保真融合

最优拓扑结构是指在给定某些设计规格（例如外力和角度）的情况下，对材料（如合金和混凝土）进行优化布局。这种拓扑优化是机械设计中的一项关键技术，但也以计算成本高昂而闻名，因此需要采用多保真融合。我们考虑悬臂梁的拓扑优化，并将点载荷的位置、点载荷的角度和滤波半径 [55] 作为系统输入。低保真度模型使用 16×16 规则网格进行有限元求解，而高保真度模型 64×64。详细设置请参见附录 E.3。

与上一个实验一样，图 2 展示了均方根误差（RMSE）统计数据随高保真训练样本数量增加的变化。显然，GAR 始终以显著优势胜过其他方法。当训练样本数量较大时，CIGAR 的表现可与 GAR 相当。

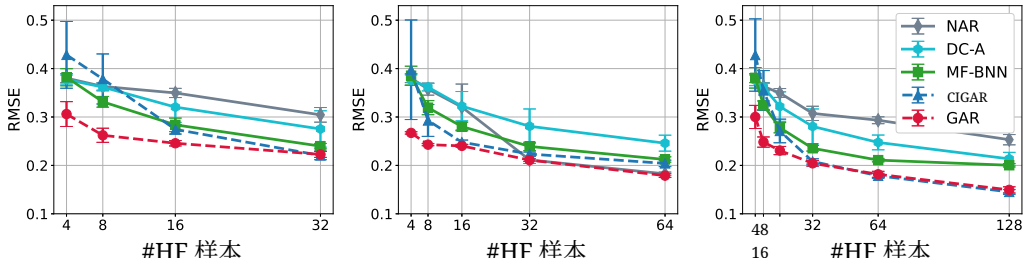


图 2：低保真训练样本数固定为 {32,64,128} 时的均方根误差。

稳态三维固体氧化物燃料电池是现代燃料电池优化的关键模型，可同时求解包括欧姆定律、纳维-斯托克斯方程、Brinkman 方程、Maxwell–Stefan 扩散和对流在内的复杂耦合偏微分方程。该模型使用 COMSOL 中的有限元方法求解。输入包括电极孔隙率、通道中的电池电压、温度和压力。低保真实验使用 3164 个单元和 0.1 的相对容差，而高保真实验使用 37064 个单元和 0.001 的相对容差。输出为位于通道中心的$x-z$平面上的电解质电流密度（ECD）和离子电势（IP）耦合场。

均方根误差统计结果如图 5a 所示，再次凸显了所提出方法的优越性，仅需四个高保真数据训练样本即可。为了进一步评估模型处理非结构化输出的能力，我们在低保真训练数据中仅保留 ECD（图 5b）和 IP（图 5c），以增加预测高保真 ECD+IP 场的难度。可以看出，移除部分低保真信息确实增加了预测难度，尤其是移除 ECD 时，在训练数据较少的情况下，MF-BNN 的表现优于 GAR 和 CIGAR。随着训练样本数增加，GAR 和 CIGAR 很快再次表现出优势。

等离激元纳米粒子阵列是一项复杂的物理模拟，使用耦合偶极近似（CDA）计算散射体数量各不相同的等离激元系统的消光效率和散射效率 Q_{ext} 和 Q_{sc}。耦合偶极近似是一种模拟由尺寸远小于光波长（此处为 25 nm）的相似非磁性金属纳米颗粒组成的阵列光学响应的方法。Q_{ext} 和 Q_{sc} 被定义为本研究中的感兴趣量。详细的实验设置请参见附录 E.5。我们使用打乱后的样本进行了 5 次实验，并将低保真度训练样本数量固定为 32、64 和 128，同时将高保真数据从 4 个样本逐步增加到 32、64 和 128 个样本。从图 3 可以看出，GAR 明显优于其他方法，尤其是在高保真数据仅包含 4 个样本时。当训练样本数据集较大时，CIGAR 也能达到与 GAR 同样出色的表现。

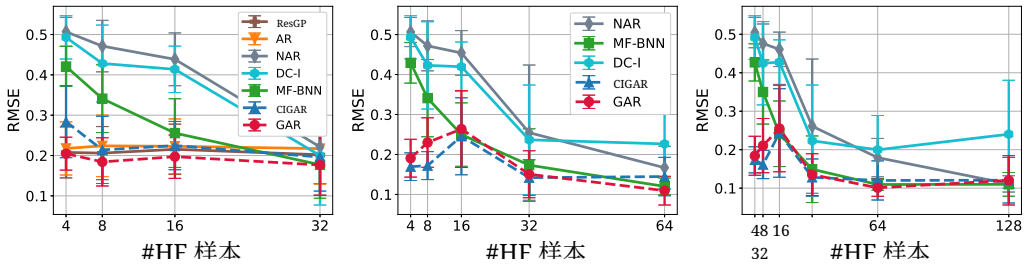
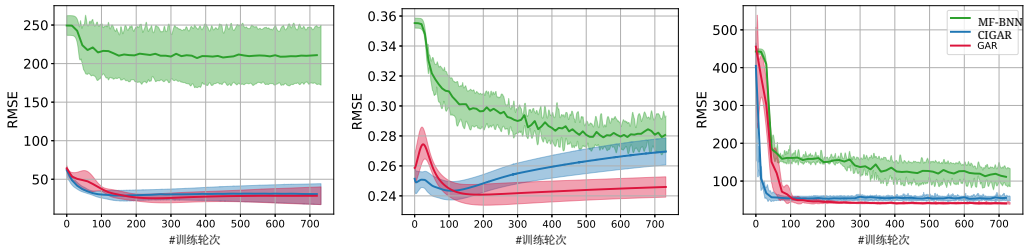


图 3：在低保真度训练样本数量固定为{32,64,128}的情况下，均方根误差随高保真度训练样本数量增加而变化，用于等离激元纳米粒子阵列模拟。

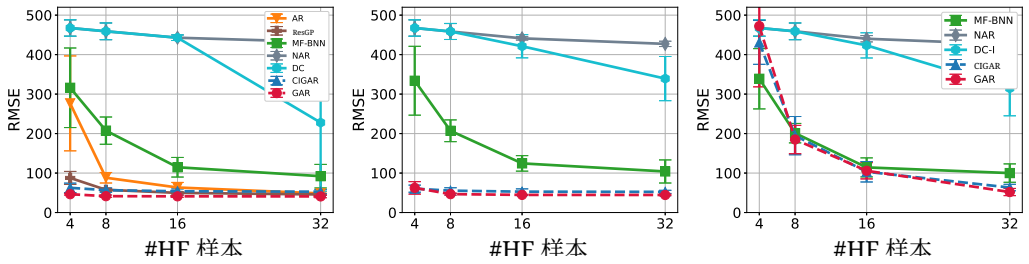


(a) 泊松　　(b) 拓扑优化　　(c) 固体氧化物燃料电池

图 4：固体氧化物燃料电池、拓扑优化和泊松数据集的测试均方根误差随训练轮次增加的变化。

5.3 稳定性测试

作为非参数模型，与基于神经网络的方法相比，我们预期 GAR 和 CIGAR 能够稳定地抵御过拟合。本节展示了 GAR、CIGAR 和 MF-BNN 在前述泊松方程、固体氧化物燃料电池（SOFC）和拓扑优化任务中的测试均方根误差随训练轮次的变化情况。为确保公平，实验重复进行了五次。结果如图 4 所示。显然，在几乎所有情况下，GAR 和 CIGAR 都比 MF-BNN 更稳定。最值得注意的是 GAR 和 CIGAR 的收敛速度：在拓扑优化和 SOFC 案例中，其速度快了 10 倍以上。对于泊松方程，无论训练轮次数有多大，MF-BNN 都不太可能达到 GAR 和 CIGAR 的性能。对于 SOFC、MF-BNN 和拓扑优化任务，MF-BNN 或许能在训练轮次极多时达到 GAR 的表现，但这会消耗昂贵的计算资源，因此并非明智之选。



(a) ECD+离子电势　　(b) ECD　　(c) 离子电势

图 5：均方根误差　固体氧化物燃料电池的低保真度训练样本数　量固定为 32。

6 结论

我们提出 GAR，这是首个能够推广至任意输出和非子集多保真度数据、且具有闭式解的自回归方法；并提出 CIGAR，通过揭示自回归中的自克里金能力，实现高效计算。本工作的局限在于样本规模方面的可扩展性不足、缺乏主动学习 [13], , 以及尚未将基于自回归的方法应用于时间序列和迁移学习[49, 50]等更广泛的问题。

参考文献

[1] Bobak Shahriari、Kevin Swersky、Ziyu Wang、Ryan P. Adams 和 Nando Freitas。《将人类排除在循环之外：贝叶斯优化综述》。104(1):148–175。ISSN 0018-9219, 1558-2256。doi: 10.1109/JPROC.2015.2494218。URL https://ieeexplore.ieee.org/document/7352306/.[2] Apostolos F. Psaros、Xuhui Meng、Zongren Zou、Ling Guo 和 George Em Karniadakis。《科学机器学习中的不确定性量化：方法、度量与比较》。[3] M. Kennedy。《当快速近似方法可用时，如何预测复杂计算机代码的输出》。87(1):1–13。ISSN 0006-3444, 1464-3510。doi: 10.1093/biomet/87.1.1。[4] Matthias Poloczek、Jialei Wang 和 Peter Frazier。《多信息源优化》。《神经信息处理系统进展》，第30卷，2017年。[5] Jialin Song、Yuxin Chen 和 Yisong Yue。《基于高斯过程的多保真贝叶斯优化通用框架》。载于《第22届人工智能与统计国际会议》，第3158–3167页。URL http://proceedings.mlr.press/v89/song19b.html.[6] L. Parussini、D. Venturi、P. Perdikaris 和 G.E. Karniadakis。《用于随机场预测的多保真高斯过程回归》。336(C):36–50。ISSN 0021-9991。doi: 10.1016/j.jcp.2017.01.047。[7] W. Xing、M. Razi、R. M. Kirby、K. Sun 和 A. A. Shah。《针对不同尺度多保真计算机模型的贪心非线性自回归》。1:100012。ISSN 2666-5468。doi: 10.1016/j.egyai.2020.100012。URL https://www.sciencedirect.com/science/article/pii/S2666546820300124.[8] Zheng Wang、Wei Xing、Robert Kirby 和 Shandian Zhe。《用于物理模拟的多保真高阶高斯过程》。载于《人工智能与统计国际会议》，第847–855页。PMLR。URL http://proceedings.mlr.press/v130/wang21c.html.[9] W. W. Xing、A. A. Shah、P. Wang、S. Zhe、Q. Fu 和 R. M. Kirby。《残差高斯过程：一种用于多保真仿真的易处理非参数贝叶斯模拟器》。97:36–56。ISSN 0307-904X。doi: 10.1016/j.apm.2021.03.041。URL https://www.sciencedirect.com/science/article/pii/ S0307904X21001724.[10] M. Giselle Fernández-Godino、Chanyoung Park、Nam-Ho Kim 和 Raphael T. Haftka。《多保真模型综述》。[11] B. Peherstorfer、K. Willcox 和 M. Gunzburger。《不确定性传播、推断与优化中的多保真方法综述》。60(3):550–591。ISSN 0036-1445。doi: 10.1137/16M1082469。[12] Wei W. Xing、Robert M. Kirby 和 Shandian Zhe。《用于模拟的时空场建模的深度共区域化》。428:109984。ISSN 0021-9991。doi: 10.1016/j.jcp.2020.109984。URL https://linkinghub.elsevier.com/retrieve/pii/S0021999120307580.[13] Shibo Li、Robert M Kirby 和 Shandian Zhe。《高维输出的深度多保真主动学习》。arXiv预印本 arXiv:2012.00901, 2020年。[14] Mauricio A. Alvarez、Lorenzo Rosasco 和 Neil D. Lawrence。《向量值函数的核：综述》。[15] Loic Le Gratiet。《分层多保真代码的贝叶斯分析》。1(1):244–269。ISSN 2166-2525。doi: 10.1137/120884122。[16] Paris Perdikaris、M Raissi、Andreas Damianou、N D. Lawrence 和 George Karniadakis。《用于数据高效多保真建模的非线性信息融合算法》，第473卷。英国皇家学会。doi: 10.1098/rspa.2016.0751。[17] Carl Edward Rasmussen 和 Christopher K I Williams。《机器学习中的高斯过程》。第266页。[18] Shandian Zhe、Wei Xing 和 Robert M. Kirby。《可扩展高阶高斯过程回归》。载于《第22届人工智能与统计国际会议》，第2611–2620页。PMLR。URL http://proceedings.mlr.press/v89/zhe19a.html.[19] Tamara Gibson Kolda。《高阶分解的多线性算子》。URL http://www.osti. gov/servlets/purl/923081-u0xXJa/.[20] Zenglin Xu、Feng Yan、Yuan 和 Qi。《无限Tucker分解：用于多路数据分析的非参数贝叶斯模型》。

[21] Andrew Gordon Wilson 和 Hannes Nickisch。用于可扩展结构化高斯过程的核插值（KISS-GP）。载于第32届国际机器学习会议论文集——第37卷，ICML'15，第1775–1784页。JMLR.org。[22] Valerio Perrone、Rodolphe Jenatton、Matthias W Seeger 和 Cédric Archambeau。可扩展的超参数迁移学习。《神经信息处理系统进展》，31，2018年。[23] Shibo Li、Wei Xing、Robert Kirby 和 Shandian Zhe。通过深度神经网络实现多保真贝叶斯优化。33：8521–8531，。网址：https://papers.nips.cc/paper/2020/hash/60e1deb043af37db5ea4ce9ae8d2c9ea-Abstract.html。[24] Akil Narayan、Claude Gittelson 和 Dongbin Xiu。多保真模型随机配置算法。36(2)：A495–A521。ISSN 1064-8275。doi：10.1137/130929461。[25] Kurt Cutajar、Mark Pullin、Andreas Damianou、Neil Lawrence 和 Javier González。用于多保真建模的深度高斯过程。[26] Andrew Wilson 和 Ryan Adams。用于模式发现和外推的高斯过程核。载于国际机器学习会议论文集，第1067–1075页。PMLR，2013年。[27] Mauricio A. Álvarez、Lorenzo Rosasco 和 Neil D. Lawrence。向量值函数的核：综述。4(3)：195–266。ISSN 1935-8237，1935-8245。doi：10.1561/2200000036。网址：http://www.nowpublishers.com/article/Details/MAL-036。[28] Michel Goulard 和 Marc Voltz。线性共区域化模型：交叉变异函数矩阵的估计与选择工具。《数学地质学》，24(3)：269–286，1992年。[29] Pierre Goovaerts 等。自然资源评价地统计学。牛津大学出版社，1997年。[30] Yee Whye Teh、Matthias Seeger 和 Michael I Jordan。半参数潜因子模型。载于人工智能与统计国际研讨会论文集，第333–340页。PMLR，2005年。[31] Dave Higdon、James Gattiker、Brian Williams 和 Maria Rightley。使用高维输出进行计算机模型校准。103(482)：570–583。ISSN 0162-1459，1537-274X。doi：10.1198/ 016214507000000888。[32] W.W. Xing、V. Triantafyllidis、A.A. Shah、P.B. Nair 和 N. Zabaras。通过流形学习模拟计算模型中的空间场。326：666–690，。ISSN 0021-9991。doi：10.1016/j.jcp.2016.07.040。网址：https://linkinghub.elsevier.com/retrieve/pii/S0021999116303722。[33] Wei Xing、Akeel A. Shah 和 Prasanth B. Nair。基于 IsoMap 的参数化偏微分方程降维高斯过程仿真器。471 5021，1471 2174 20140697，1364 2946 10 1098 2014 0697 ( )：。ISSN - - 。doi：。/rspa。..[34] Mauricio A Álvarez、Wil Ward 和 Cristian Guarnizo。用于多输出高斯过程的非线性过程卷积。载于第22届人工智能与统计国际会议论文集，第1969–1977页。PMLR，2019年。[35] Phillip Boyle 和 Marcus Frean。相关高斯过程。《神经信息处理系统进展》，17，2004年。[36] Dave Higdon。使用过程卷积进行空间及时空建模。载于 Clive W. Ander- son、Vic Barnett、Philip C. Chatwin 和 Abdel H. El-Shaarawi 编，《当前环境问题的定量方法》，第37–56页。Springer London。ISBN 978-1-4471-1171-9 978-1-4471- 0657-9。doi：10.1007/978-1-4471-0657-9_2。[37] Edwin V Bonilla、Felix V Agakov 和 Christopher KI Williams。利用任务特定特征进行核多任务学习。载于人工智能与统计论文集，第43–50页。PMLR，2007年。[38] Barbara Rakitsch、Christoph Lippert、Karsten Borgwardt 和 Oliver Stegle。噪声决定一切：使用结构化残差进行高效的多任务高斯过程推断。《神经信息处理系统进展》，26，2013年。[39] Ping Li 和 Songcan Chen。用于多任务学习的分层高斯过程模型。74：134–144。ISSN 0031-3203。doi：10.1016/j.patcog.2017.09.021。网址：https://linkinghub.elsevier.com/retrieve/pii/S0031320317303746。

[40] Andrew Gordon Wilson、David A. Knowles 和 Zoubin Ghahramani。高斯过程回归网络。载于第29届国际机器学习会议论文集，ICML'12，第1139–1146页，美国威斯康星州麦迪逊，2012年。Omnipress。ISBN 9781450312851。[41] Trung Nguyen 和 Edwin Bonilla。高斯过程回归网络的高效变分推断。载于人工智能与统计论文集，第472–480页。PMLR，2013年。[42] Tamara G. Kolda 和 Brett W. Bader。张量分解及其应用。51(3)：455–500。ISSN 0036-1445，1095-7200。doi：10.1137/07070111X。[43] Shibo Li、Wei Xing、Robert M. Kirby 和 Shandian Zhe。可扩展的高斯过程回归网络。第3卷，第2456–2462页，。doi：10.24963/ijcai.2020/340。网址：https://www.ijcai.org/ proceedings/2020/340。[44] Andreas Damianou。深度高斯过程与不确定性的变分传播。[45] Kirthevasan Kandasamy、Gautam Dasarathy、Junier B Oliva、Jeff Schneider 和 Barnabás Póczos。具有多保真度评估的高斯过程老虎机优化。《神经信息处理系统进展》，29，2016年。[46] Yehong Zhang、Trong Nghia Hoang、Bryan Kian Hsiang Low 和 Mohan Kankanhalli。基于信息的多保真度贝叶斯优化。载于神经信息处理系统贝叶斯优化研讨会，2017年。[47] Dongxia Wu、Matteo Chinazzi、Alessandro Vespignani、Yi-An Ma 和 Rose Yu。多保真度分层神经过程。arXiv预印本 arXiv:2206.04872，2022年。[48] Xuhui Meng 和 George Em Karniadakis。一种从多保真度数据中学习的复合神经网络：函数逼近与偏微分方程逆问题中的应用。401：109020。ISSN 0021-9991。doi：10.1016/j.jcp.2019.109020。网址：http://www.sciencedirect.com/science/article/pii/ S0021999119307260。[49] James Requeima、William Tebbutt、Wessel Bruinsma 和 Richard E Turner。高斯过程自回归回归模型（gpar）。载于第22届人工智能与统计国际会议论文集，第1860–1869页。PMLR，2019年。[50] Rui Xia、Wessel Bruinsma、William Tebbutt 和 Richard E Turner。高斯过程潜在自回归模型。载于第三届近似贝叶斯推断进展研讨会，2020年。[51] Rui Tuo、C. F. Jeff Wu 和 Dan Yu。不同网格密度下计算机实验的代理建模。56(3)：372–380。ISSN 0040-1706，1537-2723。doi：10.1080/00401706.2013.842935。[52] Mehmet Onder Efe 和 Hitay Ozbay。用于降阶建模的正交适当分解：二维热流。载于2003年IEEE控制应用会议论文集，2003年。CCA 2003.，第2卷，第1273–1277页。IEEE。[53] Maziar Raissi 和 George Em Karniadakis。使用高斯过程学习线性微分方程。348：683–693。ISSN 0021-9991。doi：10/gbzfgr。[54] I.M Sobol'。立方体中的点分布与积分的近似计算。7(4)：86–112。ISSN 0041-5553。doi：10.1016/0041-5553(67)90144-9。网址：https://linkinghub.elsevier.com/ retrieve/pii/0041555367901449。[55] Tyler E. Bruns 和 Daniel A. Tortorelli。非线性弹性结构与柔顺机构的拓扑优化。《计算力学与工程方法》，190 3459，2001 26 3443 0045 7825 10 1016 0045 7825 00 00278 4 ( )：–。ISSN -。doi：https://doi.org/ . /S - ( ) -。网址：http://www.sciencedirect.com/science/article/pii/S0045782500002784。[56] TJ Chung。计算流体力学。剑桥大学出版社。[57] N Sugimoto。具有分数阶导数的伯格斯方程；对非线性声波的遗传效应。225：631–653。[58] Kai Nagel。粒子跳跃模型与交通流理论。53(5)：4655。[59] S Kutluay、AR Bahadir 和 A Özdeş。一维伯格斯方程的数值解：显式与精确显式有限差分方法。103(2)：251–261。[60] A. A. Shah、W. W. Xing 和 V. Triantafyllidis。参数依赖的线性与非线性动态偏微分方程模型的降阶建模。473 5021，1471 2200 20160809 1364 2946 10 1098 2016 0809 ( )：。ISSN - -。doi：。/rspa。..

[61] Maziar Raissi、Paris Perdikaris 和 George Em Karniadakis。物理信息深度学习（第一部分）：非线性偏微分方程的数据驱动解。[62] Steven C Chapra、Raymond P Canale 等。工程师数值方法。波士顿：McGraw-Hill Higher Education，。[63] S Persides。施瓦西时空中的拉普拉斯方程和泊松方程。43(3):571–578。[64] I. E. Lagaris、A. Likas 和 D. I. Fotiadis。用于求解常微分方程和偏微分方程的人工神经网络。9(5):987–1000，1998年9月。ISSN 1045-9227。doi: 10.1109/72.712178。[65] Frank Spitzer。静电容量、热流和布朗运动。3(2):110–121。[66] Krzysztof Burdzy、Zhen-Qing Chen、John Sylvester 等。时变区域中的热方程与反射布朗运动。32(1B):775–804。[67] Fischer Black 和 Myron Scholes。期权和公司债务的定价。81(3):637–654。[68] Wei Xing、Shireen Y. Elhabian、Vahid Keshavarzzadeh 和 Robert M. Kirby。共享高斯过程：学习跨数据空间的可解释共享隐藏结构，以进行设计空间分析和探索。142(8)，。ISSN 1050-0472、1528-9001。doi: 10.1115/1.4046074。[69] Erik Andreassen、Anders Clausen、Mattias Schevenels、Boyan S. Lazarov 和 Ole Sigmund。使用88行代码在 MATLAB 中高效进行拓扑优化。Structural and Multidisciplinary Optimization，43 (1):1–16，2011年1月。ISSN 1615-1488。doi: 10.1007/s00158-010-0594-7。URL https://doi.org/10. 1007/s00158-010-0594-7。[70] Martin Philip Bendsoe 和 Ole Sigmund。拓扑优化：理论、方法与应用。Springer，2004。[71] Charles-Antoine Guérin、Pierre Mallet 和 Anne Sentenac。有限尺寸聚集体的有效介质理论。JOSA A，23(2):349–358，2006。[72] Mani Razi、Ren Wang、Yanyan He、Robert M. Kirby 和 Luca Dal Negro。等离激元纳米粒子大型沃格尔螺旋阵列的优化。14(1):253–261。ISSN 1557-1955、1557-1963。doi: 10.1007/s11468-018-0799-y。[73] Aristi Christofi、Felipe A Pinheiro 和 Luca Dal Negro。利用格林矩阵谱方法探测沃格尔螺旋的散射共振。Optics Letters，41(9):1933–1936，2016年。[74] Dongxia Wu、Liyao Gao、Xinyue Xiong、Matteo Chinazzi、Alessandro Vespignani、Yi-An Ma 和 Rose Yu。量化深度时空预测中的不确定性。arXiv 预印本 arXiv:2105.11982，2021年。

核对清单

核对清单位于参考文献之后。请仔细阅读核对清单指南，了解如何回答这些问题。对于每个问题，请将默认选项 **[待办]** 改为 [是] , [否] , 或 [不适用] 。我们强烈建议您为答案提供**论证说明**，可以引用论文中的相应章节，也可以简要地直接说明。例如：

- 您是否注明了代码和数据集的许可证？ [是] 参见章节**??**。
- 您是否注明了代码和数据集的许可证？ [否] 代码和数据均为专有内容。
- 您是否注明了代码和数据集的许可证？ [不适用]

请勿修改问题，回答时仅使用提供的宏。请注意，核对清单章节不计入页数限制。在论文中，请删除此说明文本块，仅保留上方的核对清单标题以及下方的问题和答案。

1. 对所有作者.....

   （a）摘要和引言中提出的主要论断是否准确反映了论文的贡献和范围？ [是] 请参见贡献、摘要和引言。（b）您是否描述了工作的局限性？ [是]请参见结论和复杂度分析章节。（c）您是否讨论了工作可能产生的负面社会影响？ [不适用] 我们认为这项工作没有明显的负面社会影响，因为它属于基础性且相当理论性的研究。

（d）您是否阅读了伦理审查指南，并确保论文符合相关要求？ [是]

2. 如果您包含理论结果.....

(a) 您是否陈述了所有理论结果的完整假设集合？ [是](b) 您是否提供了所有理论结果的完整证明？ [是] 请参阅附录

3. 如果您进行了实验.....

(a) 您是否提供了复现主要实验结果所需的代码、数据和说明（可置于补充材料中或通过 URL 提供）？ [是] 请参阅补充材料 (b) 您是否说明了所有训练细节（例如，数据划分、超参数及其选择方式）？ [是] 请参阅附录 (c) 您是否报告了误差条（例如，多次运行实验后针对随机种子给出的误差条）？ [是]请参阅实验章节 (d) 您是否列出了计算资源总量及所用资源的类型（例如，图形处理器类型、内部集群或云服务提供商）？ [是] 请参阅实验章节

4. 如果您使用了现有资产（例如，代码、数据、模型），或整理/发布了新资产.....

(a) 如果您的工作使用了现有资产，是否注明了创作者？ [是] 请参见实验章节 (b) 是否提及了这些资产的许可证？ [是](c) 是否在补充材料中或通过 URL 提供了任何新资产？ [是] (d) 是否讨论了您如何获得所使用或整理的数据相关人员的同意？ [是] 请参见实验章节和附录 (e) 是否讨论了您使用或整理的数据是否包含个人身份信息或冒犯性内容？ [不适用] 我们不使用此类数据

5. 如果您使用了众包或开展了涉及人类受试者的研究.....

(a) 您是否提供了向参与者提供的完整说明文本，以及适用时的截图? [不适用](b) 您是否说明了参与者可能面临的风险，并在适用时附上机构审查委员会（IRB）的批准链接？ [不适用](c) 您是否提供了支付给参与者的预计时薪，以及参与者报酬的总支出？ [不适用]

附录

一个高斯过程

高斯过程（GP）是代理模型的典型选择，因为它能够表示复杂的黑箱函数，并进行不确定性量化。暂且考虑一个简化情形：我们有含噪观测 $\{y_i = g(\mathbf{x}_i) + \epsilon_i\}_{i=1}^N$。在 GP 模型中，我们对由$\mathbf{x}$索引的 $f(\mathbf{x})$ 赋予一个先验分布：

$$\eta(\mathbf{x})|\boldsymbol{\theta} \sim \mathcal{GP}\left(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'|\boldsymbol{\theta})\right), \tag{A.1}$$

其均值函数和协方差函数为：

$$\begin{array}{ll} m_0(\mathbf{x}) & = \mathbb{E}[f(\mathbf{x})], \\ k(\mathbf{x}, \mathbf{x}'|\boldsymbol{\theta}) & = \mathbb{E}[(f(\mathbf{x}) - m_0(\mathbf{x}))(f(\mathbf{x}') - m_0(\mathbf{x}'))], \end{array} \tag{A.2}$$

其中，$\mathbb{E}[\cdot]$ 是期望，$\boldsymbol{\theta}$ 是控制核函数的超参数。通过对数据进行中心化，可以假设均值函数为一个常数，$m_0(\mathbf{x}) \equiv m_0$。也可以采用其他选项，例如 $\mathbf{x}$ 的线性函数，但除非已掌握函数形状的先验知识，否则很少使用这些选项。协方差函数可以采用多种形式，其中自动相关性判定（ARD）核最为常用。

$$k(\mathbf{x}, \mathbf{x}'|\boldsymbol{\theta}) = \theta_0 \exp\left(-(\mathbf{x} - \mathbf{x}')^T \text{diag}(\theta_1^{-2}, \ldots, \theta_l^{-2})(\mathbf{x} - \mathbf{x}')\right). \tag{A.3}$$

从现在起，我们省略 $k(x, x')$ 对 $\boldsymbol{\theta}$ 的显式记号。在此情况下，超参数 $\theta_1, \ldots, \theta_l$ 被称为长度尺度。对于常数参数 $\mathbf{x}$，$f(\mathbf{x})$ 是其随机变量。相较之下，一组值 $f(\mathbf{x}_i)$，$i = 1,\ldots, N$，是该高斯过程的部分实现。高斯过程的实现是 x 的确定性函数。高斯过程的主要特征是，$eta(xi)$ $i = 1, hdots, N$ 的联合分布是多元高斯分布。

假设模型缺陷 $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ 同样服从高斯分布，我们可以利用先验 (A.1) 和现有数据推导模型似然。

$$\begin{aligned} \mathcal{L} \triangleq p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) &= \int (f(\mathbf{x}) + \varepsilon) df = \mathcal{N}(\mathbf{y}|m_0\mathbf{1}, \mathbf{K} + \sigma^2\mathbf{I}) \\ &= -\frac{1}{2}(\mathbf{y} - m_0\mathbf{1})^T (\mathbf{K} + \sigma^2\mathbf{I})^{-1} (\mathbf{y} - m_0\mathbf{1}) \\ &\quad - \frac{1}{2}\ln|\mathbf{K} + \sigma^2\mathbf{I}| - \frac{N}{2}\log(2\pi), \end{aligned} \tag{A.4}$$

其中 $\mathbf{K} = [K_{ij}]$ 是协方差矩阵，其中 $K_{ij} = k(\mathbf{x}_i, \mathbf{x}j)$，$i, j = 1,\ldots, N$。超参数 $\boldsymbol{\theta}$ 通常通过对 $\boldsymbol{\theta}$ 求式 (A.4) 的最大似然 (MLE) 进行点估计得到。$\mathbf{y}$ 与 $f(\mathbf{x})$ 的联合分布也是一个联合高斯分布，其均值为 $m_0\mathbf{1}$，协方差矩阵为.....

$$\mathbf{K}' = \left[\begin{array}{c|c} \mathbf{K} + \sigma^2\mathbf{I} & \mathbf{k}(\mathbf{x}) \\ \hline \mathbf{k}^T(\mathbf{x}) & k(\mathbf{x}, \mathbf{x}) + \sigma^2 \end{array}\right], \tag{A.5}$$

其中 $\mathbf{k}(\mathbf{x}) = (k(\mathbf{x}_1, \mathbf{x}) \ldots, k(\mathbf{x}_N, \mathbf{x}))^T$。给定 $\mathbf{y}$ 后，可得到 $\mathbf{x}$ 处的条件预测分布。

$$\begin{aligned} \hat{f}(\mathbf{x})|\mathbf{y} &\sim \mathcal{N}\left(\mu(\mathbf{x}), v(\mathbf{x}, \mathbf{x}')\right), \\ \mu(\mathbf{x}) &= m_0\mathbf{1} + \mathbf{k}(\mathbf{x})^T \left(\mathbf{K} + \sigma^2\mathbf{I}\right)^{-1} (\mathbf{y} - m_0\mathbf{1}), \\ v(\mathbf{x}) &= \sigma^2 + k(\mathbf{x}, \mathbf{x}) - \mathbf{k}^T(\mathbf{x}) \left(\mathbf{K} + \sigma^2\mathbf{I}\right)^{-1} \mathbf{k}(\mathbf{x}). \end{aligned} \tag{A.6}$$

期望值 $\mathbb{E}[f(\mathbf{x})]$ 由 $\mu(\mathbf{x})$ 给出，预测方差为 $v(\mathbf{x})$。从式 (A.5) 推导至式 (A.6) 至关重要，因为该唤醒过程的预测后验分布基于一个相应的分块协方差矩阵。

B 定理证明

引理 1. *[16]* 若 $\mathbf{X}^h \subset \mathbf{X}^l$，则自回归的联合似然可分解为低保真度和高保真度对应的两个独立似然。

该引理由 [15] 证明。然而，记号和推导不易理解。为奠定 GAR 的基础，我们采用更清晰的方式和易懂的记号来证明它。

证明。 根据公式 (2)，协方差矩阵的求逆为

$$\mathbf{\Sigma}^{-1} = \begin{pmatrix} (\mathbf{K}^l)^{-1} + \begin{pmatrix} 0 & 0 \\ 0 & \rho^2(\mathbf{K}^r)^{-1} \end{pmatrix} & -\begin{pmatrix} 0 \\ \rho(\mathbf{K}^r)^{-1} \end{pmatrix} \\ -\begin{pmatrix} 0 & \rho(\mathbf{K}^r)^{-1} \end{pmatrix} & (\mathbf{K}^r)^{-1} \end{pmatrix}.$$

我们可以将所有低保真度和高保真度观测的对数似然写为:

$$\begin{aligned}
&\log p(\mathbf{Y}^l, \mathbf{Y}^h) \\
=&-\frac{N^h+N^l}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{\Sigma}| - \frac{1}{2}(\mathbf{Y}^l, \rho\mathbf{E}^T\mathbf{Y}^l + \mathbf{Y}^r)^T\mathbf{\Sigma}^{-1}\begin{pmatrix} \mathbf{Y}^l \\ \rho\mathbf{E}^T\mathbf{Y}^l + \mathbf{Y}^r \end{pmatrix} \\
=&-\frac{1}{2}\log|\mathbf{\Sigma}| - \frac{N^h+N^l}{2}\log(2\pi) - \frac{1}{2}[(\mathbf{Y}^l)^T(\mathbf{K}^l)^{-1}\mathbf{Y}^l + (\mathbf{Y}^l)^T\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \rho^2(\mathbf{K}^r)^{-1} \end{pmatrix}\mathbf{Y}^l \\
&-\rho(\mathbf{Y}^l)^T\mathbf{E}\big(0,\ \rho(\mathbf{K}^r)^{-1}\big)\mathbf{Y}^l - \Big(0,\ (\mathbf{Y}^r)^T\rho(\mathbf{K}^r)^{-1}\Big)\mathbf{Y}^l - (\mathbf{Y}^l)^T\mathbf{E}\rho{\mathbf{K}_r}^{-1}(\rho\mathbf{E}^T\mathbf{Y}^l + \mathbf{Y}^r) \\
&+\rho\mathbf{Y}^l\mathbf{E}(\mathbf{K}^r)^{-1}(\rho\mathbf{E}^T\mathbf{Y}^l + \mathbf{Y}^r) + \mathbf{Y}^r(\mathbf{K}^r)^{-1}(\rho\mathbf{E}^T\mathbf{Y}^l + \mathbf{Y}^r)] \\
=&-\frac{1}{2}\log|\mathbf{\Sigma}| - \frac{N^h+N^l}{2}\log(2\pi) - \frac{1}{2}[(\mathbf{Y}^l)^T(\mathbf{K}^l)^{-1}\mathbf{Y}^l - \Big(0,\ (\mathbf{Y}^r)^T\rho(\mathbf{K}^r)^{-1}\Big)\mathbf{Y}^l \\
&+\mathbf{Y}^r(\mathbf{K}^r)^{-1}\rho\mathbf{E}^T\mathbf{Y}^l + \mathbf{Y}^r(\mathbf{K}^r)^{-1}\mathbf{Y}^r] \\
=&-\frac{1}{2}\log|\mathbf{K}^l| - \frac{1}{2}\log|\mathbf{K}^r| - \frac{N^l+N^h}{2}log(2\pi) - \frac{1}{2}(\mathbf{Y}^l)^T(\mathbf{K}^l)^{-1}\mathbf{Y}^l - \frac{1}{2}\mathbf{Y}^r(\mathbf{K}^r)^{-1}\mathbf{Y}^r \\
=&\underbrace{-\frac{N^L}{2}log(2\pi) - \frac{1}{2}\log|\mathbf{K}^l| - \frac{1}{2}(\mathbf{Y}^l)^T(\mathbf{K}^l)^{-1}\mathbf{Y}^l}_{\mathcal{L}^l}\underbrace{-\frac{N^h}{2}log(2\pi) - \frac{1}{2}\log|\mathbf{K}^r| - \frac{1}{2}(\mathbf{Y}^r)^T(\mathbf{K}^r)^{-1}\mathbf{Y}^r}_{\mathcal{L}^r}
\end{aligned} \tag{A.7}$$

□

其中 $\mathbf{Y}^r = \mathbf{Y}^h - \rho\mathbf{E}^T\mathbf{Y}^l$， $\mathcal{L}^l$ 是使用低保真核的低保真数据的对数似然，而 $\mathcal{L}^r$ 是使用残差核的残差数据的对数似然； $\mathcal{L}^l$ 和 $\mathcal{L}^r$ 相互独立，因此可以并行训练。

基于联合概率式 (2)，我们可以同样通过标准的高斯过程后验推导，得出高保真度的预测后验分布。给定$\mathbf{Y}^h$和$\mathbf{Y}^l$后，对于新的输入$\mathbf{x}_*$，高保真度的预测后验同样服从高斯分布 $\mathcal{N}(\mu_*^h, \sigma_*^h)$:

$$\begin{aligned}
\mu_*^h =& \Big(\rho\mathbf{k}^l(\mathbf{x}_*, \mathbf{X}^l),\ \rho^2\mathbf{k}^l(\mathbf{x}_*, \mathbf{X}^h) + \mathbf{k}^r(\mathbf{x}_*, \mathbf{X}^h)\Big)\mathbf{K}^{-1}\begin{pmatrix} \mathbf{Y}^l \\ \rho\mathbf{E}^T\mathbf{Y}^l + \mathbf{Y}^r \end{pmatrix} \\
=&\rho\mathbf{k}^l(\mathbf{x}_*, \mathbf{X}^l)\mathbf{K}^l(\mathbf{X}^l, \mathbf{X}^l)^{-1}\mathbf{Y}^l \\
&+\rho^3\mathbf{k}^l(\mathbf{x}_*, \mathbf{X}^l)\mathbf{E}(\mathbf{K}^r)^{-1}\mathbf{E}^T\mathbf{Y}^l - \rho^3\mathbf{k}^l(\mathbf{x}_*, \mathbf{X}^h)(\mathbf{K}^r)^{-1}\mathbf{E}^T\mathbf{Y}^l \\
&-\rho\mathbf{k}^r(\mathbf{x}_*, \mathbf{X}^h)(\mathbf{K}^r)^{-1}\mathbf{E}^T\mathbf{Y}^l - \rho^2\mathbf{k}^l(\mathbf{x}_*, \mathbf{X}^l)\mathbf{E}(\mathbf{K}^r)^{-1}\left[\rho\mathbf{E}^T\mathbf{Y}^l + \mathbf{Y}^r\right] \\
&+\rho^2\mathbf{k}^l(\mathbf{x}_*, \mathbf{X}^h)(\mathbf{K}^r)^{-1}\left[\rho\mathbf{E}^T\mathbf{Y}^l + \mathbf{Y}^r\right] + \mathbf{k}^r(\mathbf{x}_*, \mathbf{X}^h)(\mathbf{K}^r)^{-1}\left[\rho\mathbf{E}^T\mathbf{Y}^l + \mathbf{Y}^r\right] \\
=&\left[\rho\mathbf{k}^l(\mathbf{x}_*, \mathbf{X}^l)(\mathbf{K}^l)^{-1}\right]\mathbf{Y}^l + \mathbf{k}^r(\mathbf{x}_*, \mathbf{X}^h)(\mathbf{K}^r)^{-1}\mathbf{Y}^r \\
&-\left[\rho\mathbf{k}^r(\mathbf{x}_*, \mathbf{X}^h)(\mathbf{K}^r)^{-1}\right]\mathbf{E}\mathbf{Y}^l + \left[\rho\mathbf{k}^r(\mathbf{x}_*, \mathbf{X}^h)(\mathbf{K}^r)^{-1}\right]\mathbf{E}\mathbf{Y}^l \\
=&\left[\rho\mathbf{k}^l(\mathbf{x}_*, \mathbf{X}^l)(\mathbf{K}^l)^{-1}\right]\mathbf{Y}^l + \mathbf{k}^r(\mathbf{x}_*, \mathbf{X}^h)(\mathbf{K}^r)^{-1}\mathbf{Y}^r
\end{aligned} \tag{A.8}$$

and

$$\begin{aligned}
\sigma_*^h =& \Big(\rho^2\mathbf{k}^l(\mathbf{x}_*, \mathbf{x}_*) + \mathbf{k}^r(\mathbf{x}_*, \mathbf{x}_*)\Big) - (\rho\mathbf{k}_*^l, \rho^2\mathbf{k}_*^l(\mathbf{X}^h) + \mathbf{k}_*^r)^T\mathbf{K}^{-1}(\rho\mathbf{k}_*^l, \rho^2\mathbf{k}_*^l(\mathbf{X}^h) + \mathbf{k}_*^r) \\
=&\Big(\rho^2\mathbf{k}^l(\mathbf{x}_*, x_*) + \mathbf{k}^r(\mathbf{x}_*, \mathbf{x}_*)\Big) - \Big(\rho(\mathbf{k}_*^l)^T(\mathbf{K}^l)^{-1}\rho\mathbf{k}_*^l\Big) + \Big(\mathbf{0}, \rho(\mathbf{k}_*^r)^T(\mathbf{K}^r)^{-1}\Big)\rho\mathbf{k}_*^l \\
&-(\mathbf{k}_*^r)^T(\mathbf{K}^r)^{-1}\rho^2\mathbf{k}_*^l(\mathbf{X}^h) - (\mathbf{k}_*^r)^T(\mathbf{K}^r)^{-1}\mathbf{k}_*^r \\
=&\rho^2\left(\mathbf{k}^l(\mathbf{x}_*, \mathbf{x}_*) - (\mathbf{k}_*^l)^T(\mathbf{K}^l)^{-1}\mathbf{k}_*^l\right) + \left(\mathbf{k}^r(\mathbf{x}_*, \mathbf{x}_*) - (\mathbf{k}_*^r)^T(\mathbf{K}^r)^{-1}\mathbf{k}_*^r\right)
\end{aligned} \tag{A.9}$$

其中，$\mathbf{k}_*^l(\mathbf{X}^h) = \mathbf{k}^l(\mathbf{x}_*,\mathbf{X}^h)$ 是新输入$\mathbf{x}_*$ 与$\mathbf{X}^h$之间的协方差向量。注意，预测后验分布也分解为与低保真高斯过程和残差高斯过程相关的两个独立部分，这有利于并行计算并节省计算资源。

Lemma 2. 给定$\mathbf{Y}^l(\mathbf{x},\mathbf{x}')$ 和$\mathbf{Y}^r(\mathbf{x},\mathbf{x}')$ 的张量高斯过程先验以及式(3)中的 *Tucker* 变换，　　*e*
联合 概率对于$\mathbf{y} = [\mathrm{vec}(\mathbf{Y}^l)^T,\ \mathrm{vec}(\mathbf{Y}^h)^T]^T$为$\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}$，其中 $\mathbf{\Sigma} =$
$$\begin{pmatrix} \mathbf{K}^l(\mathbf{X}^l,\mathbf{X}^l) \otimes \left(\bigotimes_{m=1}^M \mathbf{S}_m^l\right) & \mathbf{K}^l(\mathbf{X}^l,\mathbf{X}^h) \otimes \left(\bigotimes_{m=1}^M \mathbf{S}_m^l \mathbf{W}_m^T\right) \\ \mathbf{K}^l(\mathbf{X}^h,\mathbf{X}^l) \otimes \left(\bigotimes_{m=1}^M \mathbf{W}_m \mathbf{S}_m^l\right) & \mathbf{K}^l(\mathbf{X}^h,\mathbf{X}^h) \otimes \left(\bigotimes_{m=1}^M \mathbf{W}_m \mathbf{S}_m^l \mathbf{W}_m\right) + \mathbf{K}^r(\mathbf{X}^h,\mathbf{X}^h) \otimes \left(\bigotimes_{m=1}^M \mathbf{S}_m^r\right) \end{pmatrix}$$

证明。 由于 $\mathbf{\Sigma}$ 是$\mathbf{y}$的协方差矩阵，因此可将其表示为分块形式：

$$\mathbf{\Sigma} = \begin{pmatrix} \mathrm{cov}(\mathrm{vec}(\mathbf{Y}^l), \mathrm{vec}(\mathbf{Y}^l)) & \mathrm{cov}(\mathrm{vec}(\mathbf{Y}^l), \mathrm{vec}(\mathbf{Y}^h)) \\ \mathrm{cov}(\mathrm{vec}(\mathbf{Y}^h), \mathrm{vec}(\mathbf{Y}^l)) & \mathrm{cov}(\mathrm{vec}(\mathbf{Y}^h), \mathrm{vec}(\mathbf{Y}^h)) \end{pmatrix},$$

其中 cov(vec($\mathbf{Y}^l$),vec($\mathbf{Y}^h$)) $= \mathrm{cov}(\mathrm{vec}(\mathbf{Y}^l), \mathrm{vec}(\mathbf{Y}^h))^T$是$\mathbf{Y}^l$ 与$\mathbf{Y}^h$之间的交叉协方差。假设$\mathbf{Y}^h \in \mathbb{R}^{N^h \times d_1^h \times \ldots \times d_M^h}$和$\mathbf{Y}^l \in \mathbb{R}^{N^l \times d_1^l \times \ldots \times d_M^l}$，结合式 (3) 中 Tucker 算子的性质，高保真数据和低保真数据具有如下变换,

$$\begin{aligned} \mathbf{Y}^h &= \mathbf{Y}^l \times_1 \mathbf{E} \times_2 \mathbf{W}_1 \times_3 \ldots \times_M \mathbf{W}_{M-1} \times_{M+1} \mathbf{W}_M \\ \mathrm{vec}(\mathbf{Y}^h) &= \left[\mathbf{E} \otimes \left(\bigotimes_{m=1}^M \mathbf{W}_m\right)\right] \mathrm{vec}(\mathbf{Y}^l) + \mathrm{vec}(\mathbf{Y}^r), \end{aligned} \tag{A.10}$$

其中 $\forall i = 1, 2, ..., M, \mathbf{W}_i \in \mathbb{R}^{d_i^h \times d_i^l}$，且$\mathbf{E}^T = (\mathbf{0}, \mathbf{I}_{N^h}) \in \mathbb{R}^{N^h \times N^l}$是选择矩阵，使得$\mathbf{X}^h = \mathbf{E}^T \mathbf{X}^l$。根据我们的高斯过程先验定义，低保真数据的联合概率为:

$$\mathrm{vec}(\mathbf{Y}^l) \sim \mathcal{N}\left(0, \mathbf{K}^l(\mathbf{X}^l,\mathbf{X}^l) \otimes \left(\bigotimes_{m=1}^M \mathbf{S}_m^l\right)\right)$$

因此，低保真数据的协方差矩阵为 cov(vec($\mathbf{Y}^l$)，vec($\mathbf{Y}^l$)) $= \mathbf{K}^l(\mathbf{X}^l,\ \mathbf{X}^l) \otimes \left(\bigotimes_{m=1}^M \mathbf{S}_m^l\right)$。随后，我们可以推导出的另一部分。首先，假设残差信息 vec($\mathbf{Y}^r$) 与 vec($\mathbf{Y}^l$) 相互独立，则 vec($\mathbf{Y}^h$) 与 vec($\mathbf{Y}^l$) 之间的协方差为

$$\begin{aligned} \mathrm{cov}(\mathrm{vec}(\mathbf{Y}^l), \mathrm{vec}(\mathbf{Y}^h)) =& \mathrm{cov}\left(\mathrm{vec}(\mathbf{Y}^l), \left[\mathbf{E} \otimes \left(\bigotimes_{m=1}^M \mathbf{W}_m\right)\right] \mathrm{vec}(\mathbf{Y}^l) + \mathrm{vec}(\mathbf{Y}^r)\right) \\ =& \mathrm{cov}\left(\mathrm{vec}(\mathbf{Y}^l), \mathrm{vec}(\mathbf{Y}^r)\right) + \mathrm{cov}\left(\mathrm{vec}(\mathbf{Y}^l), \left[\mathbf{E} \otimes \left(\bigotimes_{m=1}^M \mathbf{W}_m\right)\right] \mathrm{vec}(\mathbf{Y}^l)\right) \\ =& \mathrm{cov}(\mathrm{vec}(\mathbf{Y}^l), \mathrm{vec}(\mathbf{Y}^l)) \left[\mathbf{E} \otimes \left(\bigotimes_{m=1}^M \mathbf{W}_m\right)\right]^T \\ =& \left[\mathbf{K}^l(\mathbf{X}^l,\mathbf{X}^l) \otimes \left(\bigotimes_{m=1}^M \mathbf{S}_m^l\right)\right] \left[\mathbf{E}^T \otimes \left(\bigotimes_{m=1}^M \mathbf{W}_m^T\right)\right] \\ =& \mathbf{K}^l(\mathbf{X}^l,\mathbf{X}^h) \otimes \left(\bigotimes_{m=1}^M \mathbf{S}_m^l \mathbf{W}_m^T\right). \end{aligned} \tag{A.11}$$

由于 cov(vec($\mathbf{Y}^l$),vec($\mathbf{Y}^h$)) 是 cov(vec($\mathbf{Y}^h$),vec($\mathbf{Y}^l$)) 的转置，因此 $\mathbf{\Sigma}$ 的右上部分为

$$\mathrm{cov}(\mathrm{vec}(\mathbf{Y}^h), \mathrm{vec}(\mathbf{Y}^l)) = \mathrm{cov}(\mathrm{vec}(\mathbf{Y}^l), \mathrm{vec}(\mathbf{Y}^h))^T = \mathbf{K}^l(\mathbf{X}^h,\mathbf{X}^l) \otimes \left(\bigotimes_{m=1}^M \mathbf{W}_m \mathbf{S}_m^l\right). \tag{A.12}$$

对于 $\mathbf{\Sigma}$ 的下部和右部，cov(vec($\mathbf{Y}^h$),vec($\mathbf{Y}^h$)) 之间的协方差为

$$
\begin{aligned}
&\operatorname{cov}(\operatorname{vec}(\mathbf{Y}^h), \operatorname{vec}(\mathbf{Y}^h)) \\
=&\operatorname{cov}\left(\left[\mathbf{E}\otimes\left(\bigotimes_{m=1}^{M}\mathbf{W}_m\right)\right]\operatorname{vec}(\mathbf{Y}^l)+\operatorname{vec}(\mathbf{Y}^r),\left[\mathbf{E}\otimes\left(\bigotimes_{m=1}^{M}\mathbf{W}_m\right)\right]\operatorname{vec}(\mathbf{Y}^l)+\operatorname{vec}(\mathbf{Y}^r)\right) \\
=&\operatorname{cov}\left(\left[\mathbf{E}\otimes\left(\bigotimes_{m=1}^{M}\mathbf{W}_m\right)\right]\operatorname{vec}(\mathbf{Y}^l),\left[\mathbf{E}\otimes\left(\bigotimes_{m=1}^{M}\mathbf{W}_m\right)\right]\operatorname{vec}(\mathbf{Y}^l)\right)+\operatorname{cov}\left(\operatorname{vec}(\mathbf{Y}^r),\operatorname{vec}(\mathbf{Y}^r)\right) \\
&+\operatorname{cov}\left(\left[\mathbf{E}\otimes\left(\bigotimes_{m=1}^{M}\mathbf{W}_m\right)\right]\operatorname{vec}(\mathbf{Y}^l),\operatorname{vec}(\mathbf{Y}^r)\right)+\operatorname{cov}\left(\operatorname{vec}(\mathbf{Y}^r),\left[\mathbf{E}\otimes\left(\bigotimes_{m=1}^{M}\mathbf{W}_m\right)\right]\operatorname{vec}(\mathbf{Y}^l)\right) \\
=&\left[\mathbf{E}\otimes\left(\bigotimes_{m=1}^{M}\mathbf{W}_m\right)\right]\left(\operatorname{cov}(\operatorname{vec}(\mathbf{Y}^l),\operatorname{vec}(\mathbf{Y}^l))\right)\left[\mathbf{E}\otimes\left(\bigotimes_{m=1}^{M}\mathbf{W}_m\right)\right]^T+\operatorname{cov}\left(\operatorname{vec}(\mathbf{Y}^r),\operatorname{vec}(\mathbf{Y}^r)\right) \\
=&\mathbf{K}^l(\mathbf{X}^h,\mathbf{X}^h)\otimes\left(\bigotimes_{m=1}^{M}\mathbf{W}_m\mathbf{S}_m^l\mathbf{W}_m^T\right)+\mathbf{K}^r(\mathbf{X}^h,\mathbf{X}^h)\left(\bigotimes_{m=1}^{M}\mathbf{S}_m^r\right).
\end{aligned}
\tag{A.13}
$$

将各部分组合起来，可得联合协方差矩阵 $\mathbf{\Sigma}$:

$$
\begin{pmatrix}
\mathbf{K}^l(\mathbf{X}^l,\mathbf{X}^l)\otimes\left(\bigotimes_{m=1}^{M}\mathbf{S}_m^l\right) & \mathbf{K}^l(\mathbf{X}^l,\mathbf{X}^h)\otimes\left(\bigotimes_{m=1}^{M}\mathbf{S}_m^l\mathbf{W}_m^T\right) \\
\mathbf{K}^l(\mathbf{X}^h,\mathbf{X}^l)\otimes\left(\bigotimes_{m=1}^{M}\mathbf{W}_m\mathbf{S}_m^l\right) & \mathbf{K}^l(\mathbf{X}^h,\mathbf{X}^h)\otimes\left(\bigotimes_{m=1}^{M}\mathbf{W}_m\mathbf{S}_m^l\mathbf{W}_m\right)+\mathbf{K}^r(\mathbf{X}^h,\mathbf{X}^h)\otimes\left(\bigotimes_{m=1}^{M}\mathbf{S}_m^r\right)
\end{pmatrix}
\tag{A.14}
$$

□

继续下一个证明之前，我们先介绍矩阵求逆性质，稍后会用到。**性质 1.** 对于任意可逆分块矩阵 $\begin{pmatrix}\mathbf{A}&\mathbf{B}\\\mathbf{B}^T&\mathbf{C}\end{pmatrix}$，若其子矩阵也可逆，则有 $(\mathbf{B}^T,\ \mathbf{C})\begin{pmatrix}\mathbf{A}&\mathbf{B}\\\mathbf{B}^T&\mathbf{C}\end{pmatrix}^{-1}=(\mathbf{0},\ \mathbf{I})$ 且 $\begin{pmatrix}\mathbf{A}&\mathbf{B}\\\mathbf{B}^T&\mathbf{C}\end{pmatrix}^{-1}\begin{pmatrix}\mathbf{B}\\\mathbf{C}\end{pmatrix}=\begin{pmatrix}\mathbf{0}\\\mathbf{I}\end{pmatrix}$。

证明。根据 Sherman-Morrison公式，分块矩阵的逆（若其可逆）为

$$
\begin{pmatrix}\mathbf{A}&\mathbf{B}\\\mathbf{B}^T&\mathbf{C}\end{pmatrix}^{-1}=\begin{pmatrix}\mathbf{P^{-1}}&-\mathbf{P^{-1}BC^{-1}}\\-\mathbf{C^{-1}B}^T\mathbf{P^{-1}}&\mathbf{C}^{-1}+\mathbf{C^{-1}B}^T\mathbf{P^{-1}BC^{-1}}\end{pmatrix},a \tag{A.15}
$$

其中 $\mathbf{P}=\mathbf{A}-\mathbf{BC}^{-1}\mathbf{B}^T$，然后我们可以根据分块矩阵乘法的规则推导出性质 1 中的乘法:

$$
\begin{aligned}
&(\mathbf{B}^T,\mathbf{C})\begin{pmatrix}\mathbf{A}&\mathbf{B}\\\mathbf{B}^T&\mathbf{C}\end{pmatrix}^{-1} \\
=&(\mathbf{B}^T,\mathbf{C})\begin{pmatrix}\mathbf{P^{-1}}&-\mathbf{P^{-1}BC^{-1}}\\-\mathbf{C^{-1}B}^T\mathbf{P^{-1}}&\mathbf{C}^{-1}+\mathbf{C^{-1}B}^T\mathbf{P^{-1}BC^{-1}}\end{pmatrix} \\
=&\left(\mathbf{B}^T\mathbf{P^{-1}}-\mathbf{C}(\mathbf{C^{-1}B^TP^{-1}}),-\mathbf{B^TP^{-1}BC^{-1}}+\mathbf{CC^{-1}}+\mathbf{C}(\mathbf{C^{-1}B^TP^{-1}BC^{-1}})\right) \\
=&\left(\mathbf{B}^T\mathbf{P^{-1}}-\mathbf{B^TP^{-1}},-\mathbf{B^TP^{-1}BC^{-1}}+\mathbf{I}+\mathbf{B^TP^{-1}BC^{-1}}\right) \\
=&(\mathbf{0},\mathbf{I}).
\end{aligned}
\tag{A.16}
$$

同样，结论的另一部分也可以推导出来

$$
\begin{aligned}
&\begin{pmatrix}\mathbf{A}&\mathbf{B}\\\mathbf{B}^T&\mathbf{C}\end{pmatrix}^{-1}\begin{pmatrix}\mathbf{B}\\\mathbf{C}\end{pmatrix} \\
=&\begin{pmatrix}\mathbf{P^{-1}}&-\mathbf{P^{-1}BC^{-1}}\\-\mathbf{C^{-1}B}^T\mathbf{P^{-1}}&\mathbf{C}^{-1}+\mathbf{C^{-1}B}^T\mathbf{P^{-1}BC^{-1}}\end{pmatrix}\begin{pmatrix}\mathbf{B}\\\mathbf{C}\end{pmatrix} \\
=&\begin{pmatrix}\mathbf{P^{-1}B}-\mathbf{P^{-1}BC^{-1}C}\\-\mathbf{C^{-1}B}^T\mathbf{P^{-1}B}+(\mathbf{C}^{-1}+\mathbf{C^{-1}B^TP^{-1}BC^{-1}})\mathbf{C}\end{pmatrix} \\
=&\begin{pmatrix}\mathbf{P^{-1}B}-\mathbf{P^{-1}B}\\-\mathbf{C^{-1}B}^T\mathbf{P^{-1}B}+\mathbf{I}+\mathbf{C^{-1}B^TP^{-1}B}\end{pmatrix} \\
=&\begin{pmatrix}\mathbf{0}\\\mathbf{I}\end{pmatrix},
\end{aligned}
\tag{A.17}
$$

□

如果假设矩阵是对称的，这似乎相当显而易见且直观。

引理 3. *GAR 中引理 1 的推广。若* $\mathbf{X}^h \subset \mathbf{X}^l$，*则关于* $\mathbf{y}[\quad^{l\ T}\quad\quad^{h\ T}]^T$ $\mathcal{L} = \mathcal{L}^l + \mathcal{L}^r$ =
vec(**Y**),vec(**Y**) *的联合似然* $\mathcal{L}$ *可分解为两个相互独立的可分离似然，其中*

$$\mathcal{L}^l = -\frac{1}{2}\mathrm{vec}\left(\mathbf{Y}^l\right)^T(\mathbf{K}^l \otimes \mathbf{S}^l)^{-1}\mathrm{vec}\left(\mathbf{Y}^l\right) - \frac{1}{2}\log|\mathbf{K}^l \otimes \mathbf{S}^l| - \frac{N^l D^l}{2}\log(2\pi),$$

$$\mathcal{L}^r = -\frac{1}{2}\mathrm{vec}\left(\mathbf{Y}^h - \mathbf{Y}^l \times \hat{\mathbf{W}}\right)^T(\mathbf{K}^r \otimes \mathbf{S}^r)^{-1}\mathrm{vec}\left(\mathbf{Y}^h - \mathbf{Y}^l \times \hat{\mathbf{W}}\right) - \frac{1}{2}\log|\mathbf{K}^r \otimes \mathbf{S}^r| - \frac{N^h D^h}{2}\log(2\pi),$$

其中 $\hat{}\mathbf{W} = [\mathbf{E}, \mathbf{W}]$ 是将原始权重张量与选择矩阵 $\mathbf{X}^h = \mathbf{E}^T\mathbf{X}^l$拼接得到的。

证明。 将核矩阵划分为四个块。我们再次使用性质 1 中 Sherman-Morrison公式的矩阵求逆形式，并作如下略微修改:

$$\begin{pmatrix} \mathbf{T} & \mathbf{U} \\ \mathbf{V} & \mathbf{M} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{T}^{-1} + \mathbf{T}^{-1}\mathbf{U}\mathbf{Q}^{-1}\mathbf{V}\mathbf{T}^{-1} & -\mathbf{T}^{-1}\mathbf{U}\mathbf{Q}^{-1} \\ -\mathbf{Q}^{-1}\mathbf{V}\mathbf{T}^{-1} & \mathbf{Q}^{-1} \end{pmatrix}$$

其中

$$\mathbf{Q} = \mathbf{M} - \mathbf{V}\mathbf{T}^{-1}\mathbf{U}.$$

我们从性质 1 中的矩阵$\mathbf{V}\mathbf{T}^{-1}\mathbf{U}$ 开始证明，由此可得:

$$\mathbf{K}^l(\mathbf{X}^h, \mathbf{X}^l)\mathbf{K}^l(\mathbf{X}^l, \mathbf{X}^l)^{-1}\mathbf{K}^l(\mathbf{X}^l, \mathbf{X}^h) = (\mathbf{0}, \mathbf{I})\,\mathbf{K}^l(\mathbf{X}^l, \mathbf{X}^h) = \mathbf{K}^l(\mathbf{X}^h, \mathbf{X}^h).$$

因此，矩阵 $\mathbf{Q}$ 的最后一部分$\mathbf{V}\mathbf{T}^{-1}\mathbf{U}$ 为

$$\begin{aligned}
&\mathbf{V}\mathbf{T}^{-1}\mathbf{U} \\
=&\left[\mathbf{K}^l(\mathbf{X}^h, \mathbf{X}^l) \otimes \mathbf{W}\mathbf{S}^l\right]\left[\mathbf{K}^l(\mathbf{X}^l, \mathbf{X}^l) \otimes \mathbf{S}^l\right]^{-1}\left[\mathbf{K}^l(\mathbf{X}^l, \mathbf{X}^h) \otimes \mathbf{S}^l\mathbf{W}^T\right] \\
=&\left[\mathbf{K}^l(\mathbf{X}^h, \mathbf{X}^l) \otimes \mathbf{W}\mathbf{S}^l\right]\left[\mathbf{K}^l(\mathbf{X}^l, \mathbf{X}^l)^{-1} \otimes (\mathbf{S}^l)^{-1}\right]\left[\mathbf{K}^l(\mathbf{X}^l, \mathbf{X}^h) \otimes \mathbf{S}^l\mathbf{W}^T\right] \\
=&\left[\mathbf{K}^l(\mathbf{X}^h, \mathbf{X}^l)\mathbf{K}^l(\mathbf{X}^l, \mathbf{X}^l)^{-1}\mathbf{K}^l(\mathbf{X}^l, \mathbf{X}^h)\right] \otimes \left[(\bigotimes_{m=1}^{M}\mathbf{W}_m\mathbf{S}_m^l)(\bigotimes_{m=1}^{M}(\mathbf{S}_m^l)^{-1})(\bigotimes_{m=1}^{M}\mathbf{S}_m^l\mathbf{W}_m^T)\right] \\
=&\left[\mathbf{K}^l(\mathbf{X}^h, \mathbf{X}^l)\mathbf{K}^l(\mathbf{X}^l, \mathbf{X}^l)^{-1}\mathbf{K}^l(\mathbf{X}^l, \mathbf{X}^h)\right] \otimes \left[\bigotimes_{m=1}^{M}(\mathbf{W}_m\mathbf{S}_m^l(\mathbf{S}_m^l)^{-1}\mathbf{S}_m^l\mathbf{W}_m^T)\right] \\
=&\mathbf{K}^l(\mathbf{X}^h, \mathbf{X}^h) \otimes \left[\bigotimes_{m=1}^{M}(\mathbf{W}_m\mathbf{S}_m^l\mathbf{W}_m^T)\right].
\end{aligned} \tag{A.18}$$

将公式（A.18）代回矩阵求逆中，我们可以推导出矩阵 $\mathbf{Q}^{-1}$, $-\mathbf{T}^{-1}$**不确定性量化**$^{-1}$、$\mathbf{Q}^{-1}\mathbf{V}\mathbf{T}^{-1}$，以及 $\mathbf{T}^{-1} + \mathbf{T}^{-1}$**不确定性量化**$^{-1}\mathbf{V}\mathbf{T}^{-1}$，其表达式为

$$\begin{aligned}
&\mathbf{Q}^{-1} \\
=&(\mathbf{M} - \mathbf{V}\mathbf{T}^{-1}\mathbf{U})^{-1} \\
=&\left(\mathbf{K}^l(\mathbf{X}^h, \mathbf{X}^h) \otimes \mathbf{W}\mathbf{S}^l\mathbf{W}^T + \mathbf{K}^r(\mathbf{X}^h, \mathbf{X}^h) \otimes \mathbf{S}^r - \mathbf{K}^l(\mathbf{X}^h, \mathbf{X}^h) \otimes \mathbf{W}\mathbf{S}^l\mathbf{W}^T\right)^{-1} \\
=&\left(\mathbf{K}^r(\mathbf{X}^h, \mathbf{X}^h) \otimes \mathbf{S}^r\right)^{-1},
\end{aligned} \tag{A.19}$$

$$\begin{aligned}
&-\mathbf{T}^{-1}\mathbf{U}\mathbf{Q}^{-1} \\
=&-\left[\mathbf{K}^l(\mathbf{X}^l, \mathbf{X}^l)^{-1} \otimes (\mathbf{S}^l)^{-1}\right]\left[\mathbf{K}^l(\mathbf{X}^l, \mathbf{X}^h) \otimes \mathbf{S}^l\mathbf{W}^T\right]\left[\mathbf{K}^r(\mathbf{X}^h, \mathbf{X}^h)^{-1} \otimes (\mathbf{S}^r)^{-1}\right] \\
=&-\left[\mathbf{K}^l(\mathbf{X}^l, \mathbf{X}^l)^{-1}\mathbf{K}^l(\mathbf{X}^l, \mathbf{X}^h)\mathbf{K}^r(\mathbf{X}^h, \mathbf{X}^h)^{-1}\right] \otimes \left[(\mathbf{S}^l)^{-1}\mathbf{S}^l\mathbf{W}^T(\mathbf{S}^r)^{-1}\right] \\
=&-\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{I} \end{pmatrix}\mathbf{K}^r(\mathbf{X}^h, \mathbf{X}^h)^{-1}\right]\left[\mathbf{W}^T(\mathbf{S}^r)^{-1}\right] \\
=&-\begin{pmatrix} \mathbf{0} \\ \mathbf{K}^r(\mathbf{X}^h, \mathbf{X}^h)^{-1} \otimes \mathbf{W}^T(\mathbf{S}^r)^{-1} \end{pmatrix},
\end{aligned} \tag{A.20}$$

$$
\begin{aligned}
&\mathbf{Q}^{-1}\mathbf{V}\mathbf{T}^{-1} \\
=&-\left[\mathbf{K}^r(\mathbf{X}^h,\mathbf{X}^h)^{-1}(\mathbf{S}^r)^{-1}\right]\left[\mathbf{K}^l(\mathbf{X}^h,\mathbf{X}^l)\otimes\mathbf{W}\mathbf{S}^l\right]\left[\mathbf{K}^l(\mathbf{X}^l,\mathbf{X}^l)^{-1}\otimes(\mathbf{S}^l)^{-1}\right] \\
=&-\left[\mathbf{K}^r(\mathbf{X}^h,\mathbf{X}^h)^{-1}\mathbf{K}^l(\mathbf{X}^h,\mathbf{X}^l)\mathbf{K}^l(\mathbf{X}^l,\mathbf{X}^l)^{-1}\right]\otimes\left[(\mathbf{S}^r)^{-1}\mathbf{W}\mathbf{S}^l(\mathbf{S}^l)^{-1}\right] \\
=&-\left(\mathbf{0},\quad \mathbf{K}^r(\mathbf{X}^h,\mathbf{X}^h)^{-1}\otimes(\mathbf{S}^r)^{-1}\mathbf{W}\right),
\end{aligned}
\tag{A.21}
$$

and

$$
\begin{aligned}
&\mathbf{T}^{-1}+\mathbf{T}^{-1}\mathbf{U}\mathbf{Q}^{-1}\mathbf{V}\mathbf{T}^{-1} \\
=&\left[\mathbf{K}^l(\mathbf{X}^l,\mathbf{X}^l)^{-1}\otimes(\mathbf{S}^l)^{-1}\right]+\left[\mathbf{K}^l(\mathbf{X}^l,\mathbf{X}^l)^{-1}\otimes(\mathbf{S}^l)^{-1}\right]\left[\mathbf{K}^l(\mathbf{X}^l,\mathbf{X}^h)\otimes\mathbf{S}^l\mathbf{W}^T\right] \\
&\times\left[\mathbf{K}^r(\mathbf{X}^h,\mathbf{X}^h)^{-1}\otimes(\mathbf{S}^r)^{-1}\right]\left[\mathbf{K}^l(\mathbf{X}^h,\mathbf{X}^l)\otimes\mathbf{W}\mathbf{S}^l\right]\left[\mathbf{K}^l(\mathbf{X}^l,\mathbf{X}^l)^{-1}\otimes(\mathbf{S}^l)^{-1}\right] \\
=&\left[\mathbf{K}^l(\mathbf{X}^l,\mathbf{X}^l)^{-1}\otimes(\mathbf{S}^l)^{-1}\right]+\left[\mathbf{K}^l(\mathbf{X}^l,\mathbf{X}^l)^{-1}\mathbf{K}^l(\mathbf{X}^l,\mathbf{X}^h)\mathbf{K}^r(\mathbf{X}^h,\mathbf{X}^h)^{-1}\mathbf{K}^l(\mathbf{X}^h,\mathbf{X}^l)\mathbf{K}^l(\mathbf{X}^l,\mathbf{X}^l)^{-1}\right] \\
&\otimes\left[(\mathbf{S}^l)^{-1}\mathbf{S}^l(\mathbf{S}^r)^{-1}\mathbf{S}^l(\mathbf{S}^l)^{-1}\right] \\
=&\mathbf{K}^l(\mathbf{X}^l,\mathbf{X}^l)^{-1}\otimes(\mathbf{S}^l)^{-1}+\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{K}^r(\mathbf{X}^h,\mathbf{X}^h)^{-1}\otimes\mathbf{W}^T\mathbf{S}^r\mathbf{W}\end{pmatrix}.
\end{aligned}
\tag{A.22}
$$

将所有这些元素组合起来，我们就得到联合核矩阵的求逆 $\mathbf{\Sigma}^{-1}=$

$$
\begin{bmatrix}
(\mathbf{K}^l)^{-1}\otimes(\mathbf{S}^l)^{-1}+\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & (\mathbf{K}^r)^{-1}\otimes\mathbf{W}^T(\mathbf{S}^r)^{-1}\mathbf{W}\end{pmatrix} & -\begin{pmatrix}\mathbf{0}\\ (\mathbf{K}^r)^{-1}\otimes\mathbf{W}^T(\mathbf{S}^r)^{-1}\end{pmatrix} \\
-\left(\mathbf{0},(\mathbf{K}^r)^{-1}\otimes(\mathbf{S}^r)^{-1}\mathbf{W}\right) & (\mathbf{K}^r)^{-1}\otimes(\mathbf{S}^r)^{-1}
\end{bmatrix},
\tag{A.23}
$$

其中 $\mathbf{S}^l=\bigotimes_{m=1}^M\mathbf{S}_m^l$、$\mathbf{S}^r=\bigotimes_{m=1}^M\mathbf{S}_m^r$、$\mathbf{W}=\bigotimes_{m=1}^M\mathbf{W}_m$、$\mathbf{K}^l=\mathbf{K}^l(\mathbf{X}^l,\mathbf{X}^l)$，以及$\mathbf{K}^r=\mathbf{K}^r(\mathbf{X},\mathbf{X})$，其定义见主论文。利用式 (A.10) 的性质，并定义 $\mathbf{y}=[\mathrm{vec}(\mathbf{Y}^l)^T,\mathrm{vec}(\mathbf{Y}^h)^T]^T$，我们可以将式 (A.23) 代入联合似然，推导出联合似然的数据拟合部分

$$
\begin{aligned}
&\mathbf{y}^T\mathbf{\Sigma}^{-1}\mathbf{y} \\
=&\left(\mathrm{vec}(\mathbf{Y}^l)^T,\mathrm{vec}(\mathbf{Y}^l)^T(\mathbf{E}\otimes\mathbf{W}^T)+\mathrm{vec}(\mathbf{Y}^r)^T\right)\mathbf{\Sigma}^{-1}\begin{pmatrix}\mathrm{vec}(\mathbf{Y}^l)\\ (\mathbf{E}^T\otimes\mathbf{W})\mathrm{vec}(\mathbf{Y}^l)+\mathrm{vec}(\mathbf{Y}^r)\end{pmatrix} \\
=&\mathrm{vec}(\mathbf{Y}^l)^T\left(\mathbf{K}^l\otimes\mathbf{S}^l\right)^{-1}\mathrm{vec}(\mathbf{Y}^l)+\mathrm{vec}(\mathbf{Y}^l)^T\mathbf{E}\left((\mathbf{K}^r)^{-1}\otimes\mathbf{W}^T(\mathbf{S}^r)^{-1}\mathbf{W}\right)\mathbf{E}^T\mathrm{vec}(\mathbf{Y}^l) \\
&-\mathrm{vec}(\mathbf{Y}^l)^T\left(\mathbf{E}\otimes\mathbf{W}^T\right)\left((\mathbf{K}^r)^{-1}\mathbf{E}^T\otimes(\mathbf{S}^r)^{-1}\mathbf{W}\right)\mathrm{vec}(\mathbf{Y}^l) \\
&-\mathrm{vec}(\mathbf{Y}^r)^T\left((\mathbf{K}^r)^{-1}\mathbf{E}^T\otimes(\mathbf{S}^r)^{-1}\mathbf{W}\right)\mathrm{vec}(\mathbf{Y}^l) \\
&-\mathrm{vec}(\mathbf{Y}^l)^T\left(\mathbf{E}(\mathbf{K}^r)^{-1}\otimes\mathbf{W}^T(\mathbf{S}^r)^{-1}\right)\left[\left(\mathbf{E}^T\otimes\mathbf{W}\right)\mathrm{vec}(\mathbf{Y}^l)+\mathrm{vec}(\mathbf{Y}^r)\right] \\
&+\mathrm{vec}(\mathbf{Y}^l)^T\left(\mathbf{E}\otimes\mathbf{W}^T\right)(\mathbf{K}^r\otimes\mathbf{S}^r)^{-1}\left[\left(\mathbf{E}^T\otimes\mathbf{W}\right)\mathrm{vec}(\mathbf{Y}^l)+\mathrm{vec}(\mathbf{Y}^r)\right] \\
&+\mathrm{vec}(\mathbf{Y}^r)^T(\mathbf{K}^r\otimes\mathbf{S}^r)^{-1}\left[\left(\mathbf{E}^T\otimes\mathbf{W}\right)\mathrm{vec}(\mathbf{Y}^l)+\mathrm{vec}(\mathbf{Y}^r)\right] \\
=&\mathrm{vec}(\mathbf{Y}^l)^T\left(\mathbf{K}^l\otimes\mathbf{S}^l\right)^{-1}\mathrm{vec}(\mathbf{Y}^l)+\mathrm{vec}(\mathbf{Y}^r)^T(\mathbf{K}^r\otimes\mathbf{S}^r)^{-1}\mathrm{vec}(\mathbf{Y}^r) \\
&-\mathrm{vec}(\mathbf{Y}^r)^T\left((\mathbf{K}^r)^{-1}\mathbf{E}^T\otimes(\mathbf{S}^r)^{-1}\mathbf{W}\right)\mathrm{vec}(\mathbf{Y}^l)+\mathrm{vec}(\mathbf{Y}^r)^T(\mathbf{K}^r\otimes\mathbf{S}^r)^{-1}\left(\mathbf{E}^T\otimes\mathbf{W}\right)\mathrm{vec}(\mathbf{Y}^l) \\
=&\mathrm{vec}(\mathbf{Y}^l)^T\left(\mathbf{K}^l\otimes\mathbf{S}^l\right)^{-1}\mathrm{vec}(\mathbf{Y}^l)+\mathrm{vec}(\mathbf{Y}^r)^T(\mathbf{K}^r\otimes\mathbf{S}^r)^{-1}\mathrm{vec}(\mathbf{Y}^r).
\end{aligned}
\tag{A.24}
$$

利用分块矩阵的行列式公式，我们还可以推导出联合核矩阵的行列式 ，

$$
|\mathbf{\Sigma}|=\left|\mathbf{K}^l\otimes\mathbf{S}^l\right|\times|\mathbf{Q}|=\left|\mathbf{K}^l\otimes\mathbf{S}^l\right|\times|\mathbf{K}^r\otimes\mathbf{S}^r|\,.
\tag{A.25}
$$

在此处，我们不再进一步分解它们，以便为低保真度和高保真度分别构建独立的高斯过程。根据式 (A.24) 的结论，完整的联合对数似然为

$$
\begin{aligned}
&\log p(\mathbf{Y}^l, \mathbf{Y}^h) \\
=&-\frac{1}{2}\mathbf{y}^T\boldsymbol{\Sigma}^{-1}\mathbf{y}-\frac{1}{2}\log|\boldsymbol{\Sigma}|-\frac{d^lN^l+d^hN^h}{2}\log(2\pi) \\
=&\underbrace{-\frac{1}{2}\mathrm{vec}(\mathbf{Y}^l)^T\left(\mathbf{K}^l\otimes\mathbf{S}^l\right)^{-1}\mathrm{vec}(\mathbf{Y}^l)-\frac{1}{2}\log\left|\mathbf{K}^l\otimes\mathbf{S}^l\right|-\frac{N^ld^l}{2}\log(2\pi)}_{TGP\ for\ low-fidelity\ data} \\
&\underbrace{-\frac{1}{2}\mathrm{vec}(\mathbf{Y}^r)^T\left(\mathbf{K}^r\otimes\mathbf{S}^r\right)^{-1}\mathrm{vec}(\mathbf{Y}^r)-\frac{1}{2}\log|\mathbf{K}^r\otimes\mathbf{S}^r|-\frac{N^hd^h}{2}\log(2\pi)}_{TGP\ for\ residual\ information} \\
=&\log p(\mathbf{Y}^l)+\log p(\mathbf{Y}^h|\mathbf{Y}^l)
\end{aligned}
\tag{A.26}
$$

$\mathbf{S}^l$、$\mathbf{S}^r$、$\mathbf{W}$和$\mathbf{K}^r$ 的含义与公式 A.23 中的定义相同。

B.1 后验分布

对于后验分布，我们在假设低保真度和高保真度数据满足严格的子集要求$\mathbf{X}^h \subseteq \mathbf{X}^l$的情况下，计算均值函数和协方差矩阵。根据引理 2 的结论和分块矩阵乘法规则，均值函数和协方差矩阵可表示为:

$$
\begin{aligned}
&\mathrm{vec}(\mathbf{Z}_*^h) \\
&=\left(\mathbf{k}_*^l\otimes\mathbf{W}\mathbf{S}^l, \mathbf{k}_*^l(\mathbf{X}^h)\otimes\mathbf{W}\mathbf{S}^l\mathbf{W}^T+\mathbf{k}_*^r\otimes\mathbf{S}^r\right)\boldsymbol{\Sigma}^{-1}\begin{pmatrix}\mathrm{vec}(\mathbf{Y}^l)\\ \mathrm{vec}(\mathbf{Y}^h)\end{pmatrix} \\
&=\left(\mathbf{k}_*^l\otimes\mathbf{W}\mathbf{S}^l\right)\left(\mathbf{K}^l\otimes\mathbf{S}^l\right)^{-1}\mathrm{vec}(\mathbf{Y}^l)+\left(\mathbf{k}_*^l\otimes\mathbf{W}\mathbf{S}^l\right)\left(\mathbf{E}(\mathbf{K}^r)^{-1}\mathbf{E}^T\otimes\mathbf{W}^T(\mathbf{S}^r)^{-1}\mathbf{W}\right)\mathrm{vec}(\mathbf{Y}^l) \\
&\quad-\left(\mathbf{k}_*^l(\mathbf{X}^h)\otimes\mathbf{W}\mathbf{S}^l\mathbf{W}^T\right)\left(\mathbf{K}^r\mathbf{E}^T\otimes(\mathbf{S}^r)^{-1}\mathbf{W}\right)\mathrm{vec}(\mathbf{Y}^l) \\
&\quad-\left(\mathbf{k}_*^r\otimes\mathbf{S}^r\right)\left(\mathbf{K}^r\mathbf{E}^T\otimes(\mathbf{S}^r)^{-1}\mathbf{W}\right)\mathrm{vec}(\mathbf{Y}^l) \\
&\quad-\left(\mathbf{k}_*^l\otimes\mathbf{W}\mathbf{S}^l\right)\left(\mathbf{E}(\mathbf{K}^r)^{-1}\otimes\mathbf{W}^T(\mathbf{S}^r)^{-1}\right)\mathrm{vec}(\mathbf{Y}^h) \\
&\quad+\left(\mathbf{k}_*^l(\mathbf{X}^h)\otimes\mathbf{W}\mathbf{S}^l\mathbf{W}^T\right)\left(\mathbf{K}^r\otimes(\mathbf{S}^r)^{-1}\right)\mathrm{vec}(\mathbf{Y}^h)+\left(\mathbf{k}_*^r\otimes\mathbf{S}^r\right)\left(\mathbf{K}^r\otimes(\mathbf{S}^r)^{-1}\right)\mathrm{vec}(\mathbf{Y}^h) \\
&=\left(\mathbf{k}_*^l\otimes\mathbf{W}\mathbf{S}^l\right)\left(\mathbf{K}^l\otimes\mathbf{S}^l\right)^{-1}\mathrm{vec}(\mathbf{Y}^l)-\left(\mathbf{k}_*^r\otimes\mathbf{S}^r\right)\left(\mathbf{K}^r\mathbf{E}^T\otimes(\mathbf{S}^r)^{-1}\mathbf{W}\right)\mathrm{vec}(\mathbf{Y}^l) \\
&\quad+\left(\mathbf{k}_*^r\otimes\mathbf{S}^r\right)\left(\mathbf{K}^r\otimes(\mathbf{S}^r)^{-1}\right)\left(\mathbf{E}^T\otimes\mathbf{W}\right)\mathrm{vec}(\mathbf{Y}^l)+\left(\mathbf{k}_*^r\otimes\mathbf{S}^r\right)\left(\mathbf{K}^r\otimes(\mathbf{S}^r)^{-1}\right)\mathrm{vec}(\mathbf{Y}^r) \\
&=\left(\mathbf{k}_*^l(\mathbf{K}^l)^{-1}\otimes\mathbf{W}\right)\mathrm{vec}(\mathbf{Y}^l)+\left(\mathbf{k}_*^r(\mathbf{K}^r)^{-1}\otimes\mathbf{I}_r\right)\mathrm{vec}(\mathbf{Y}^r),
\end{aligned}
\tag{A.27}
$$

$$
\begin{aligned}
\mathbf{S}_*^h =& \left(\mathbf{k}^l(\mathbf{x}_*,\mathbf{x}_*)\otimes\mathbf{W}\mathbf{S}^l\mathbf{W}^T+\mathbf{k}^r(\mathbf{x}_*,\mathbf{x}_*)\otimes\mathbf{S}^r\right)- \\
&\left(\mathbf{k}_*^l\otimes\mathbf{W}\mathbf{S}^l,\mathbf{k}_*^l(\mathbf{X}^h)\otimes\mathbf{W}\mathbf{S}^l\mathbf{W}^T+\mathbf{k}_*^r\otimes\mathbf{S}^r\right)\boldsymbol{\Sigma}^{-1}\begin{pmatrix}(\mathbf{k}_*^l)^T\otimes\mathbf{S}^l\mathbf{W}^T\\ \mathbf{k}_*^l(\mathbf{X}^h)^T\otimes\mathbf{W}^T\mathbf{S}^l\mathbf{W}+(\mathbf{k}_*^r)^T\otimes\mathbf{S}^r\end{pmatrix} \\
=&\left(\mathbf{k}^l(\mathbf{x}_*,\mathbf{x}_*)\otimes\mathbf{W}\mathbf{S}^l\mathbf{W}^T+\mathbf{k}^r(\mathbf{x}_*,\mathbf{x}_*)\otimes\mathbf{S}^r\right) \\
&-\left(\mathbf{k}_*^l\otimes\mathbf{W}\mathbf{S}^l\right)\left(\mathbf{K}^l\otimes\mathbf{S}^l\right)^{-1}\left((\mathbf{k}_*^l)^T\otimes\mathbf{S}^l\mathbf{W}^T\right) \\
&-\left(\mathbf{k}_*^l\otimes\mathbf{W}\mathbf{S}^l\right)\left(\mathbf{E}(\mathbf{K}^l)^{-1}\mathbf{E}^T\otimes(\mathbf{S}^l)^{-1}\right)\left((\mathbf{k}_*^l)^T\otimes\mathbf{S}^l\mathbf{W}^T\right) \\
&+\left(\mathbf{k}_*^l(\mathbf{X}^h)\otimes\mathbf{W}\mathbf{S}^l\mathbf{W}^T\right)\left(\mathbf{K}^r\mathbf{E}^T\otimes(\mathbf{S}^r)^{-1}\mathbf{W}\right)\left((\mathbf{k}_*^l)^T\otimes\mathbf{S}^l\mathbf{W}^T\right) \\
&+(\mathbf{k}_*^r\otimes\mathbf{S}^r)\left(\mathbf{K}^r\mathbf{E}^T\otimes(\mathbf{S}^r)^{-1}\mathbf{W}\right)\left((\mathbf{k}_*^l)^T\otimes\mathbf{S}^l\mathbf{W}^T\right) \\
&+\left(\mathbf{k}_*^l\otimes\mathbf{W}\mathbf{S}^l\right)\left(\mathbf{E}(\mathbf{K}^r)^{-1}\otimes\mathbf{W}^T(\mathbf{S}^r)^{-1}\right)\left(\mathbf{k}_*^l(\mathbf{X}^h)^T\otimes\mathbf{W}^T\mathbf{S}^l\mathbf{W}+(\mathbf{k}_*^r)^T\otimes\mathbf{S}^r\right) \\
&-\left(\mathbf{k}_*^l(\mathbf{X}^h)\otimes\mathbf{W}\mathbf{S}^l\mathbf{W}^T\right)(\mathbf{K}^r\otimes(\mathbf{S}^r)^{-1})\left(\mathbf{k}_*^l(\mathbf{X}^h)^T\otimes\mathbf{W}^T\mathbf{S}^l\mathbf{W}+(\mathbf{k}_*^r)^T\otimes\mathbf{S}^r\right) \\
&-(\mathbf{k}_*^r\otimes\mathbf{S}^r)(\mathbf{K}^r\otimes(\mathbf{S}^r)^{-1})\left(\mathbf{k}_*^l(\mathbf{X}^h)^T\otimes\mathbf{W}^T\mathbf{S}^l\mathbf{W}+(\mathbf{k}_*^r)^T\otimes\mathbf{S}^r\right) \\
=&\left(\mathbf{k}^l(\mathbf{x}_*,\mathbf{x}_*)\otimes\mathbf{W}\mathbf{S}^l\mathbf{W}^T+\mathbf{k}^r(\mathbf{x}_*,\mathbf{x}_*)\otimes\mathbf{S}^r\right) \\
&-\left(\mathbf{k}_*^l\otimes\mathbf{W}\mathbf{S}^l\right)\left(\mathbf{K}^l\otimes\mathbf{S}^l\right)^{-1}\left((\mathbf{k}_*^l)^T\otimes\mathbf{S}^l\mathbf{W}^T\right) \\
&+(\mathbf{k}_*^r\otimes\mathbf{S}^r)\left(\mathbf{K}^r\mathbf{E}^T\otimes(\mathbf{S}^r)^{-1}\mathbf{W}\right)\left((\mathbf{k}_*^l)^T\otimes\mathbf{S}^l\mathbf{W}^T\right) \\
&-(\mathbf{k}_*^r\otimes\mathbf{S}^r)(\mathbf{K}^r\otimes(\mathbf{S}^r)^{-1})\left(\mathbf{k}_*^l(\mathbf{X}^h)^T\otimes\mathbf{W}^T\mathbf{S}^l\mathbf{W}\right) \\
&-(\mathbf{k}_*^r\otimes\mathbf{S}^r)(\mathbf{K}^r\otimes(\mathbf{S}^r)^{-1})\left((\mathbf{k}_*^r)^T\otimes\mathbf{S}^r\right) \\
=&\left(k_{**}^l-(\mathbf{k}_*^l)^T(\mathbf{K}^l)^{-1}\mathbf{k}_*^l\right)\otimes\mathbf{W}\mathbf{S}^l\mathbf{W}^T+\left(k_{**}^r-(\mathbf{k}_*^r)^T(\mathbf{K}^r)^{-1}\mathbf{k}_*^r\right)\otimes\mathbf{S}^r,
\end{aligned}
\tag{A.28}
$$

其中，$\mathbf{W}$、$\mathbf{S}^l$ 和$\mathbf{S}^r$ 的含义与主论文中相同。

B.2 非子集多保真数据的联合似然

I在主论文和子集章节中，我们将联合似然 log p(**Y**^{h}，**Y**^{l}) 分解为两部分 a s

$$
\log p(\mathbf{Y}^l,\mathbf{Y}^h)=\log p(\mathbf{Y}^l)+\log\int p(\mathbf{Y}^h|\hat{\mathbf{Y}}^l,\mathbf{Y}^l)p(\hat{\mathbf{Y}}^l|\mathbf{Y}^l)d\hat{\mathbf{Y}}^l
$$

其中，p(**Y**^{h}|ˆ**Y**^{l}，**Y**^{l}) 是在高保真数据为低保真数据子集时推导出的预测后验概率。

$$
\begin{aligned}
&p(\mathbf{Y}^h|\hat{\mathbf{Y}}^l,\mathbf{Y}^l)=2\pi^{-\frac{N^hd^h}{2}}\times|\mathbf{K}^r\otimes\mathbf{S}^r|^{-\frac{1}{2}} \\
&\times\exp\left[-\frac{1}{2}\left[\begin{pmatrix}\mathrm{vec}(\check{\mathbf{Y}}^h)\\ \mathrm{vec}(\hat{\mathbf{Y}}^h)\end{pmatrix}-\hat{\mathbf{W}}\begin{pmatrix}\mathrm{vec}(\mathbf{Y}^l)\\ \mathrm{vec}(\hat{\mathbf{Y}}^l)\end{pmatrix}\right]^T(\mathbf{K}^r\otimes\mathbf{S}^r)^{-1}\left[\begin{pmatrix}\mathrm{vec}(\check{\mathbf{Y}}^h)\\ \mathrm{vec}(\hat{\mathbf{Y}}^h)\end{pmatrix}-\hat{\mathbf{W}}\begin{pmatrix}\mathrm{vec}(\mathbf{Y}^l)\\ \mathrm{vec}(\hat{\mathbf{Y}}^l)\end{pmatrix}\right]\right]
\end{aligned}
\tag{A.29}
$$

其中，我们定义 ˆ**W** $=\mathbf{E}\otimes\bigotimes_{m=1}^{M}\mathbf{W}_m$。根据低保真训练数据，我们还有 p(ˆ**Y**^{l}|**Y**^{l}) $\sim\mathcal{N}$(¯**Y**^{l}，ˆ**S** $^l\otimes\mathbf{S}^l$) 服从高斯分布。

$$
\begin{aligned}
&p(\hat{\mathbf{Y}}^l|\mathbf{Y}^l) \\
=&2\pi^{-\frac{N^md^l}{2}}\times\left|\hat{\mathbf{S}}^l\otimes\mathbf{S}^l\right|^{-\frac{1}{2}}\times\exp\left[-\frac{1}{2}\left(\mathrm{vec}(\hat{\mathbf{Y}}^l)-\mathrm{vec}(\bar{\mathbf{Y}}^l)\right)^T(\hat{\mathbf{S}}^l\otimes\mathbf{S}^l)^{-1}\left(\mathrm{vec}(\hat{\mathbf{Y}}^l)-\mathrm{vec}(\bar{\mathbf{Y}}^l)\right)\right],
\end{aligned}
\tag{A.30}
$$

其中，$\hat{}\mathbf{S}^l \otimes \mathbf{S}^l$ 是 $\hat{}\mathbf{Y}^l$ 的后验协方差矩阵。我们可以结合式 (A.29) 和式 (A.30)，推导出联合似然的积分部分

$$
\begin{aligned}
&\log \int p(\mathbf{Y}^h|\hat{\mathbf{Y}}^l, \mathbf{Y}^l) p(\hat{\mathbf{Y}}^l|\mathbf{Y}^l) d\hat{\mathbf{Y}}^l \\
=& -\frac{N^h d^h + N^m d^l}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{K}^r \otimes \mathbf{S}^r| - \frac{1}{2} \log \left| \hat{\mathbf{S}}^l \otimes \mathbf{S}^l \right| \\
&+ \log \int \exp \left\{ -\frac{1}{2} \left[\begin{pmatrix} \mathrm{vec}(\check{\mathbf{Y}}^h) \\ \mathrm{vec}(\hat{\mathbf{Y}}^h) \end{pmatrix} - \left(\mathbf{E}_n^T \otimes \mathbf{W} \right) \begin{pmatrix} \mathrm{vec}(\check{\mathbf{Y}}^l) \\ \mathrm{vec}(\hat{\mathbf{Y}}^l) \end{pmatrix} \right]^T (\mathbf{K}^r \otimes \mathbf{S}^r)^{-1} \right. \\
&\quad \left[\begin{pmatrix} \mathrm{vec}(\check{\mathbf{Y}}^h) \\ \mathrm{vec}(\hat{\mathbf{Y}}^h) \end{pmatrix} - \left(\mathbf{E}_n^T \otimes \mathbf{W} \right) \begin{pmatrix} \mathrm{vec}(\check{\mathbf{Y}}^l) \\ \mathrm{vec}(\hat{\mathbf{Y}}^l) \end{pmatrix} \right] \\
&\quad \left. -\frac{1}{2} \left(\mathrm{vec}(\hat{\mathbf{Y}}^l)^T - \mathrm{vec}(\bar{\mathbf{Y}}^l)^T \right) (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)^{-1} \left(\mathrm{vec}(\hat{\mathbf{Y}}^l) - \mathrm{vec}(\bar{\mathbf{Y}}^l) \right) \right\} d\mathrm{vec}(\hat{\mathbf{Y}}^l),
\end{aligned}
\tag{A.31}
$$

其中 $\mathbf{X}^h = \mathbf{E}_n^T [\mathbf{X}^l, \hat{}\mathbf{X}^h]$。由于我们知道 $\hat{}\mathbf{X}^h = \hat{}\mathbf{E}^T \mathbf{X}^h$，并且假设 $\check{}\mathbf{X}^h = \check{}\mathbf{E}^T \mathbf{X}^h$，因此可以推导出

$$
\begin{aligned}
&\begin{pmatrix} \mathrm{vec}(\check{\mathbf{Y}}^h) \\ \mathrm{vec}(\hat{\mathbf{Y}}^h) \end{pmatrix} - (\mathbf{E}_n^T \otimes \mathbf{W}) \begin{pmatrix} \mathrm{vec}(\mathbf{Y}^l) \\ \mathrm{vec}(\hat{\mathbf{Y}}^l) \end{pmatrix} \\
=& \begin{pmatrix} \mathrm{vec}(\check{\mathbf{Y}}^h) \\ \mathbf{0} \end{pmatrix} - \tilde{\mathbf{W}} \begin{pmatrix} \mathrm{vec}(\check{\mathbf{Y}}^l) \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ \mathrm{vec}(\hat{\mathbf{Y}}^h) \end{pmatrix} - \hat{\mathbf{W}} \begin{pmatrix} \mathbf{0} \\ \mathrm{vec}(\hat{\mathbf{Y}}^l) \end{pmatrix} \\
=& \left(\check{\mathbf{E}} \otimes \mathbf{I}_h \right) \mathrm{vec}(\check{\mathbf{Y}}^h) - \left(\check{\mathbf{E}} \otimes \mathbf{W} \right) \mathrm{vec}(\check{\mathbf{Y}}^l) + \left(\hat{\mathbf{E}} \otimes \mathbf{I}_h \right) \mathrm{vec}(\check{\mathbf{Y}}^h) - \left(\hat{\mathbf{E}} \otimes \mathbf{W} \right) \mathrm{vec}(\check{\mathbf{Y}}^l).
\end{aligned}
\tag{A.32}
$$

其中，$\tilde{}\mathbf{W} = \mathbf{I}_{N^h} \otimes \mathbf{W}$，而 $\check{}\mathbf{Y}^l$ 表示 $\check{}\mathbf{Y}^h$ 中对应的部分。为方便起见，我们选择首先计算式 (A.31) 中的指数部分。我们尝试将其分解为子集部分，即 ($\check{}\mathbf{Y}^h$ 和 $\check{}\mathbf{Y}^l$)，以及非子集部分，即 ($\hat{}\mathbf{Y}^h$ 和 $\hat{}\mathbf{Y}^l$)。式 (A.32) 将有助于后续推导。我们先通过将式 (A.32) 代入式 (A.31)，来考察数据拟合部分，

$$
\begin{aligned}
&-\frac{1}{2} \left[\begin{pmatrix} \mathrm{vec}(\check{\mathbf{Y}}^h) \\ \mathrm{vec}(\hat{\mathbf{Y}}^h) \end{pmatrix} - \hat{\mathbf{W}} \begin{pmatrix} \mathrm{vec}(\check{\mathbf{Y}}^l) \\ \mathrm{vec}(\hat{\mathbf{Y}}^l) \end{pmatrix} \right]^T (\mathbf{K}^r \otimes \mathbf{S}^r)^{-1} \left[\begin{pmatrix} \mathrm{vec}(\check{\mathbf{Y}}^h) \\ \mathrm{vec}(\hat{\mathbf{Y}}^h) \end{pmatrix} - \hat{\mathbf{W}} \begin{pmatrix} \mathrm{vec}(\check{\mathbf{Y}}^l) \\ \mathrm{vec}(\hat{\mathbf{Y}}^l) \end{pmatrix} \right] \\
=& -\frac{1}{2} \left[\mathrm{vec}(\check{\mathbf{Y}}^h)^T \left(\check{\mathbf{E}}^T \otimes \mathbf{I}_h \right) - \mathrm{vec}(\check{\mathbf{Y}}^l)^T \left(\check{\mathbf{E}}^T \otimes \mathbf{W}^T \right) \right] (\mathbf{K}^r \otimes \mathbf{S}^r)^{-1} \left[\left(\check{\mathbf{E}} \otimes \mathbf{I}_h \right) \mathrm{vec}(\check{\mathbf{Y}}^h) - \left(\check{\mathbf{E}} \otimes \mathbf{W} \right) \mathrm{vec}(\check{\mathbf{Y}}^h) \right] \\
&- \left[\mathrm{vec}(\check{\mathbf{Y}}^h)^T \left(\check{\mathbf{E}}^T \otimes \mathbf{I}_h \right) - \mathrm{vec}(\check{\mathbf{Y}}^l)^T \left(\check{\mathbf{E}}^T \otimes \mathbf{W}^T \right) \right] (\mathbf{K}^r \otimes \mathbf{S}^r)^{-1} \left[\left(\hat{\mathbf{E}} \otimes \mathbf{I}_h \right) \mathrm{vec}(\hat{\mathbf{Y}}^h) - \left(\hat{\mathbf{E}} \otimes \mathbf{W} \right) \mathrm{vec}(\hat{\mathbf{Y}}^l) \right] \\
&- \frac{1}{2} \left[\mathrm{vec}(\hat{\mathbf{Y}}^h)^T \left(\hat{\mathbf{E}}^T \otimes \mathbf{I}_h \right) - \mathrm{vec}(\hat{\mathbf{Y}}^l)^T \left(\hat{\mathbf{E}}^T \otimes \mathbf{W} \right) \right] (\mathbf{K}^r \otimes \mathbf{S}^r)^{-1} \left[\left(\hat{\mathbf{E}} \otimes \mathbf{I}_h \right) \mathrm{vec}(\hat{\mathbf{Y}}^h) - \left(\hat{\mathbf{E}} \otimes \mathbf{W} \right) \mathrm{vec}(\hat{\mathbf{Y}}^l) \right],
\end{aligned}
\tag{A.33}
$$

由此得到的分解包括子集部分、非子集部分，以及它们之间的交互部分。现在，我们可以将式 (A.33) 代入式 (A.31) 中的积分部分，

$$
\begin{aligned}
&\log \int \exp[-\frac{1}{2}\left[\mathrm{vec}(\check{\mathbf{Y}}^h)^T\left(\check{\mathbf{E}}^T\otimes\mathbf{I}_h\right)-\mathrm{vec}(\check{\mathbf{Y}}^l)^T\left(\check{\mathbf{E}}^T\otimes\mathbf{W}^T\right)\right](\mathbf{K}^r\otimes\mathbf{S}^r)^{-1}\left[\left(\check{\mathbf{E}}\otimes\mathbf{I}_h\right)\mathrm{vec}(\check{\mathbf{Y}}^h)-\left(\check{\mathbf{E}}\otimes\mathbf{W}\right)\right.^{\text{向量化}(\check{\mathbf{Y}}^h)}\\
&\quad-\left[\mathrm{vec}(\check{\mathbf{Y}}^h)^T\left(\check{\mathbf{E}}^T\otimes\mathbf{I}_h\right)-\mathrm{vec}(\check{\mathbf{Y}}^l)^T\left(\check{\mathbf{E}}^T\otimes\mathbf{W}^T\right)\right](\mathbf{K}^r\otimes\mathbf{S}^r)^{-1}\left[\left(\hat{\mathbf{E}}\otimes\mathbf{I}_h\right)\mathrm{vec}(\hat{\mathbf{Y}}^h)-\left(\hat{\mathbf{E}}\otimes\mathbf{W}\right)\mathrm{vec}(\hat{\mathbf{Y}}^l)\right]\\
&\quad-\frac{1}{2}\left[\mathrm{vec}(\hat{\mathbf{Y}}^h)^T\left(\hat{\mathbf{E}}^T\otimes\mathbf{I}_h\right)-\mathrm{vec}(\hat{\mathbf{Y}}^l)^T\left(\hat{\mathbf{E}}^T\otimes\mathbf{W}\right)\right](\mathbf{K}^r\otimes\mathbf{S}^r)^{-1}\left[\left(\hat{\mathbf{E}}\otimes\mathbf{I}_h\right)\mathrm{vec}(\hat{\mathbf{Y}}^h)-\left(\hat{\mathbf{E}}\otimes\mathbf{W}\right)\mathrm{vec}(\hat{\mathbf{Y}}^l)\right]\\
&\quad-\frac{1}{2}\mathrm{vec}(\hat{\mathbf{Y}}^l)^T(\hat{\mathbf{S}}^l\otimes\mathbf{S}^l)^{-1}\mathrm{vec}(\hat{\mathbf{Y}}^l)-\frac{1}{2}\mathrm{vec}(\bar{\mathbf{Y}}^l)^T(\hat{\mathbf{S}}^l\otimes\mathbf{S}^l)^{-1}\mathrm{vec}(\bar{\mathbf{Y}}^l)+\mathrm{vec}(\hat{\mathbf{Y}}^l)^T(\hat{\mathbf{S}}^l\otimes\mathbf{S}^l)^{-1}\mathrm{vec}(\bar{\mathbf{Y}}^l)]d\mathrm{vec}(\hat{\mathbf{Y}}^l)\\
=&-\frac{1}{2}\left[\mathrm{vec}(\check{\mathbf{Y}}^h)^T\left(\check{\mathbf{E}}^T\otimes\mathbf{I}_h\right)-\mathrm{vec}(\check{\mathbf{Y}}^l)^T\left(\check{\mathbf{E}}^T\otimes\mathbf{W}^T\right)\right](\mathbf{K}^r\otimes\mathbf{S}^r)^{-1}\left[\left(\check{\mathbf{E}}\otimes\mathbf{I}_h\right)\mathrm{vec}(\check{\mathbf{Y}}^h)-\left(\check{\mathbf{E}}\otimes\mathbf{W}\right)\mathrm{vec}(\check{\mathbf{Y}}^h)\right]\\
&\quad-\left[\mathrm{vec}(\check{\mathbf{Y}}^h)^T\left(\check{\mathbf{E}}^T\otimes\mathbf{I}_h\right)-\mathrm{vec}(\check{\mathbf{Y}}^l)^T\left(\check{\mathbf{E}}^T\otimes\mathbf{W}^T\right)\right](\mathbf{K}^r\otimes\mathbf{S}^r)^{-1}\left(\hat{\mathbf{E}}\otimes\mathbf{I}_h\right)\mathrm{vec}(\hat{\mathbf{Y}}^h)\\
&\quad-\frac{1}{2}\mathrm{vec}(\hat{\mathbf{Y}}^h)^T\left(\hat{\mathbf{E}}^T\otimes\mathbf{I}_h\right)(\mathbf{K}^r\otimes\mathbf{S}^r)^{-1}\left(\hat{\mathbf{E}}\otimes\mathbf{I}_h\right)\mathrm{vec}(\hat{\mathbf{Y}}^h)-\frac{1}{2}\mathrm{vec}(\bar{\mathbf{Y}}^l)^T(\hat{\mathbf{S}}^l\otimes\mathbf{S}^l)^{-1}\mathrm{vec}(\bar{\mathbf{Y}}^l)\\
&\quad+\log\int\exp[\left[\mathrm{vec}(\check{\mathbf{Y}}^h)^T\left(\check{\mathbf{E}}^T\otimes\mathbf{I}_h\right)-\mathrm{vec}(\check{\mathbf{Y}}^l)^T\left(\check{\mathbf{E}}^T\otimes\mathbf{W}^T\right)\right](\mathbf{K}^r\otimes\mathbf{S}^r)^{-1}\left(\hat{\mathbf{E}}\otimes\mathbf{W}\right)\mathrm{vec}(\hat{\mathbf{Y}}^l)\\
&\quad+\mathrm{vec}(\hat{\mathbf{Y}}^h)^T\left(\hat{\mathbf{E}}^T\otimes\mathbf{I}_h\right)(\mathbf{K}^r\otimes\mathbf{S}^r)^{-1}\left(\hat{\mathbf{E}}\otimes\mathbf{W}\right)\mathrm{vec}(\hat{\mathbf{Y}}^l)\\
&\quad-\frac{1}{2}\mathrm{vec}(\hat{\mathbf{Y}}^l)^T\left(\hat{\mathbf{E}}^T\otimes\mathbf{W}^T\right)(\mathbf{K}^r\otimes\mathbf{S}^r)^{-1}\left(\hat{\mathbf{E}}\otimes\mathbf{W}\right)\mathrm{vec}(\hat{\mathbf{Y}}^l)\\
&\quad-\frac{1}{2}\mathrm{vec}(\hat{\mathbf{Y}}^l)^T(\hat{\mathbf{S}}^l\otimes\mathbf{S}^l)^{-1}\mathrm{vec}(\hat{\mathbf{Y}}^l)+\mathrm{vec}(\bar{\mathbf{Y}}^l)^T(\hat{\mathbf{S}}^l\otimes\mathbf{S}^l)^{-1}\mathrm{vec}(\hat{\mathbf{Y}}^l)]d\mathrm{vec}(\hat{\mathbf{Y}}^l)\\
=&-\frac{1}{2}\boldsymbol{\phi}^T(\mathbf{K}^r\otimes\mathbf{S}^r)^{-1}\boldsymbol{\phi}-\frac{1}{2}\mathrm{vec}(\bar{\mathbf{Y}}^l)^T(\hat{\mathbf{S}}^l\otimes\mathbf{S}^l)^{-1}\mathrm{vec}(\bar{\mathbf{Y}}^l)\\
&\quad+\frac{1}{2}\left(\boldsymbol{\Psi}^T(\mathbf{K}^r\otimes\mathbf{S}^r)^{-1}\boldsymbol{\phi}+(\hat{\mathbf{S}}^l\otimes\mathbf{S}^l)^{-1}\mathrm{vec}(\bar{\mathbf{Y}}^l)\right)^T\left(\boldsymbol{\Psi}^T(\mathbf{K}^r\otimes\mathbf{S}^r)^{-1}\boldsymbol{\phi}+(\hat{\mathbf{S}}^l\otimes\mathbf{S}^l)^{-1}\right)^{-1}\\
&\quad\left(\boldsymbol{\Psi}^T(\mathbf{K}^r\otimes\mathbf{S}^r)^{-1}\boldsymbol{\phi}+(\hat{\mathbf{S}}^l\otimes\mathbf{S}^l)^{-1}\mathrm{vec}(\bar{\mathbf{Y}}^l)\right)\\
&\quad+\frac{N^md^l}{2}\log 2\pi+\frac{1}{2}\log\left|\boldsymbol{\Psi}^T(\mathbf{K}^r\otimes\mathbf{S}^r)^{-1}\boldsymbol{\Psi}+(\hat{\mathbf{S}}^l\otimes\mathbf{S}^l)^{-1}\right|\\
=&\frac{N^md^l}{2}\log 2\pi+\frac{1}{2}\log\left|\boldsymbol{\Psi}^T(\mathbf{K}^r\otimes\mathbf{S}^r)^{-1}\boldsymbol{\Psi}+(\hat{\mathbf{S}}^l\otimes\mathbf{S}^l)^{-1}\right|\\
&\quad-\frac{1}{2}\boldsymbol{\phi}^T\underbrace{\left[(\mathbf{K}^r\otimes\mathbf{S}^r)^{-1}-(\mathbf{K}^r\otimes\mathbf{S}^r)^{-1}\boldsymbol{\Psi}\left(\boldsymbol{\Psi}^T(\mathbf{K}^r\otimes\mathbf{S}^r)^{-1}\boldsymbol{\Psi}+(\hat{\mathbf{S}}^l\otimes\mathbf{S}^l)^{-1}\right)^{-1}\boldsymbol{\Psi}^T(\mathbf{K}^r\otimes\mathbf{S}^r)^{-1}\right]}_{\text{part 1}}\boldsymbol{\phi}\\
&\quad-\frac{1}{2}\mathrm{vec}(\bar{\mathbf{Y}}^l)^T\underbrace{\left[(\hat{\mathbf{S}}^l\otimes\mathbf{S}^l)^{-1}-(\hat{\mathbf{S}}^l\otimes\mathbf{S}^l)^{-1}\left(\boldsymbol{\Psi}^T(\mathbf{K}^r\otimes\mathbf{S}^r)^{-1}\boldsymbol{\Psi}+(\hat{\mathbf{S}}^l\otimes\mathbf{S}^l)^{-1}\right)^{-1}(\hat{\mathbf{S}}^l\otimes\mathbf{S}^l)^{-1}\right]}_{\text{part 2}}\mathrm{vec}(\bar{\mathbf{Y}}^l)\\
&\quad+\boldsymbol{\phi}^T\underbrace{(\mathbf{K}^r\otimes\mathbf{S}^r)^{-1}\boldsymbol{\Psi}\left(\boldsymbol{\Psi}^T(\mathbf{K}^r\otimes\mathbf{S}^r)^{-1}\boldsymbol{\Psi}+(\hat{\mathbf{S}}^l\otimes\mathbf{S}^l)^{-1}\right)^{-1}(\hat{\mathbf{S}}^l\otimes\mathbf{S}^l)^{-1}}_{\text{part 3}}\mathrm{vec}(\bar{\mathbf{Y}}^l)
\end{aligned}
\tag{A.34}
$$

其中 $\boldsymbol{\phi}$ 由式 (A.35) 定义，

$$
\begin{aligned}
\boldsymbol{\phi}&=\left((\mathrm{vec}(\check{\mathbf{Y}}^h)^T,\mathrm{vec}(\hat{\mathbf{Y}}^h)^T)-(\mathrm{vec}(\check{\mathbf{Y}}^l)^T,\mathbf{0})\tilde{\mathbf{W}}^T\right)\\
\boldsymbol{\Psi}&=\hat{\mathbf{E}}\otimes\mathbf{W}.
\end{aligned}
\tag{A.35}
$$

利用Sherman-Morrison公式，我们可以进一步将式（A.34）中的第1部分简化为

$$
\begin{aligned}
&(\mathbf{K}^r\otimes\mathbf{S}^r)^{-1}-(\mathbf{K}^r\otimes\mathbf{S}^r)^{-1}\boldsymbol{\Psi}\left(\boldsymbol{\Psi}^T(\mathbf{K}^r\otimes\mathbf{S}^r)^{-1}\boldsymbol{\Psi}+(\hat{\mathbf{S}}^l\otimes\mathbf{S}^l)^{-1}\right)^{-1}\boldsymbol{\Psi}^T(\mathbf{K}^r\otimes\mathbf{S}^r)^{-1}\\
=&\left(\mathbf{K}^r\otimes\mathbf{S}^r+\boldsymbol{\Psi}(\hat{\mathbf{S}}^l\otimes\mathbf{S}^l)\boldsymbol{\Psi}^T\right)^{-1},
\end{aligned}
\tag{A.36}
$$

式（A.34）中的第2部分为

$$
\begin{aligned}
&(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)^{-1} - (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)^{-1} \left(\boldsymbol{\Psi}^T (\mathbf{K}^r \otimes \mathbf{S}^r)^{-1} \boldsymbol{\Psi} + (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)^{-1}\right)^{-1} (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)^{-1} \\
=&(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)^{-1} - \left((\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)\boldsymbol{\Psi}^T (\mathbf{K}^r \otimes \mathbf{S}^r)^{-1} \boldsymbol{\Psi} (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l) + (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)\right)^{-1} \\
=&(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)^{-1} - (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)^{-1} + \boldsymbol{\Psi}^T (\mathbf{K}^r \otimes \mathbf{S}^r + \boldsymbol{\Psi}^T (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l) \boldsymbol{\Psi}^T)^{-1} \boldsymbol{\Psi} \\
=&\boldsymbol{\Psi}^T (\mathbf{K}^r \otimes \mathbf{S}^r + \boldsymbol{\Psi} (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l) \boldsymbol{\Psi}^T)^{-1} \boldsymbol{\Psi},
\end{aligned}
\tag{A.37}
$$

以及式 (A.34) 中的第 3 部分可表示为

$$
\begin{aligned}
&(\mathbf{K}^r \otimes \mathbf{S}^r)^{-1} \boldsymbol{\Psi} \left(\boldsymbol{\Psi}^T (\mathbf{K}^r \otimes \mathbf{S}^r)^{-1} \boldsymbol{\Psi} + (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)^{-1}\right)^{-1} (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)^{-1} \\
=&(\mathbf{K}^r \otimes \mathbf{S}^r)^{-1} \boldsymbol{\Psi} \left(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l - (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l) \boldsymbol{\Psi}^T (\mathbf{K}^r \otimes \mathbf{S}^r + \boldsymbol{\Psi} (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l) \boldsymbol{\Psi}^T)^{-1} \boldsymbol{\Psi} (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)\right) (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)^{-1} \\
=&(\mathbf{K}^r \otimes \mathbf{S}^r)^{-1} \boldsymbol{\Psi} - (\mathbf{K}^r \otimes \mathbf{S}^r)^{-1} \boldsymbol{\Psi} (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l) \boldsymbol{\Psi}^T (\mathbf{K}^r \otimes \mathbf{S}^r + \boldsymbol{\Psi} (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l) \boldsymbol{\Psi}^T)^{-1} \boldsymbol{\Psi} \\
=&(\mathbf{K}^r \otimes \mathbf{S}^r)^{-1} \boldsymbol{\Psi} - (\mathbf{K}^r \otimes \mathbf{S}^r)^{-1} \left((\mathbf{K}^r \otimes \mathbf{S}^r) + \boldsymbol{\Psi} (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l) \boldsymbol{\Psi}^T - (\mathbf{K}^r \otimes \mathbf{S}^r)\right) \\
&(\mathbf{K}^r \otimes \mathbf{S}^r + \boldsymbol{\Psi} (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l) \boldsymbol{\Psi}^T)^{-1} \boldsymbol{\Psi} \\
=&(\mathbf{K}^r \otimes \mathbf{S}^r)^{-1} \boldsymbol{\Psi} - (\mathbf{K}^r \otimes \mathbf{S}^r)^{-1} \left(\mathbf{I} - (\mathbf{K}^r \otimes \mathbf{S}^r)(\mathbf{K}^r \otimes \mathbf{S}^r + \boldsymbol{\Psi} (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l) \boldsymbol{\Psi}^T)^{-1}\right) \boldsymbol{\Psi} \\
=&(\mathbf{K}^r \otimes \mathbf{S}^r)^{-1} \boldsymbol{\Psi} - (\mathbf{K}^r \otimes \mathbf{S}^r)^{-1} \boldsymbol{\Psi} + \left(\mathbf{K}^r \otimes \mathbf{S}^r + \boldsymbol{\Psi} (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l) \boldsymbol{\Psi}^T\right)^{-1} \boldsymbol{\Psi} \\
=&\left(\mathbf{K}^r \otimes \mathbf{S}^r + \boldsymbol{\Psi} (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l) \boldsymbol{\Psi}^T\right)^{-1} \boldsymbol{\Psi}.
\end{aligned}
\tag{A.38}
$$

利用对第 1、2 和 3 部分所作的简化，可得式 (A.36)、式 (A.37) 和式 (A.38)。将式 (A.36)、式 (A.37) 和式 (A.38) 代回式 (A.34)，积分部分会变得更加简洁，结果为

$$
\begin{aligned}
&\frac{N^m d^l}{2} \log 2\pi + \frac{1}{2} \log \left|\boldsymbol{\Psi}^T (\mathbf{K}^r \otimes \mathbf{S}^r)^{-1} \boldsymbol{\Psi} + (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)^{-1}\right| - \frac{1}{2} \boldsymbol{\phi}^T \left(\mathbf{K}^r \otimes \mathbf{S}^r + \boldsymbol{\Psi} (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l) \boldsymbol{\Psi}^T\right)^{-1} \boldsymbol{\phi} \\
&- \frac{1}{2} \mathrm{vec}(\bar{\mathbf{Y}}^l)^T \boldsymbol{\Psi}^T (\mathbf{K}^r \otimes \mathbf{S}^r + \boldsymbol{\Psi} (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l) \boldsymbol{\Psi}^T)^{-1} \boldsymbol{\Psi} \mathrm{vec}(\bar{\mathbf{Y}}^l) \\
&+ \boldsymbol{\phi}^T \left(\mathbf{K}^r \otimes \mathbf{S}^r + \boldsymbol{\Psi} (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l) \boldsymbol{\Psi}^T\right)^{-1} \boldsymbol{\Psi} \mathrm{vec}(\bar{\mathbf{Y}}^l) \\
=&\frac{N^m d^l}{2} \log 2\pi + \frac{1}{2} \log \left|\boldsymbol{\Psi}^T (\mathbf{K}^r \otimes \mathbf{S}^r)^{-1} \boldsymbol{\Psi} + (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)^{-1}\right| \\
&- \frac{1}{2} (\boldsymbol{\phi} - \boldsymbol{\Psi} \mathrm{vec}(\bar{\mathbf{Y}}^l))^T \left(\mathbf{K}^r \otimes \mathbf{S}^r + \boldsymbol{\Psi} (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l) \boldsymbol{\Psi}^T\right)^{-1} (\boldsymbol{\phi} - \boldsymbol{\Psi} \mathrm{vec}(\bar{\mathbf{Y}}^l)).
\end{aligned}
\tag{A.39}
$$

矩阵的式 (A.31) 中的行列式部分也可以按以下方式分解：

$$
\begin{aligned}
&-\frac{1}{2} \log |\mathbf{K}^r \otimes \mathbf{S}^r| - \frac{1}{2} \log \left|\hat{\mathbf{S}}^l \otimes \mathbf{S}^l\right| + \frac{1}{2} \log \left|\boldsymbol{\Psi}^T (\mathbf{K}^r \otimes \mathbf{S}^r)^{-1} \boldsymbol{\Psi} + (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)^{-1}\right| \\
=&-\frac{1}{2} \log |\mathbf{K}^r \otimes \mathbf{S}^r| - \frac{1}{2} \log \left|\hat{\mathbf{S}}^l \otimes \mathbf{S}^l\right| \\
&+\frac{1}{2} \log |\mathbf{K}^r \otimes \mathbf{S}^r| + \frac{1}{2} \log \left|\hat{\mathbf{S}}^l \otimes \mathbf{S}^l\right| - \frac{1}{2} \log \left|\mathbf{K}^r \otimes \mathbf{S}^r + \boldsymbol{\Psi} (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l) \boldsymbol{\Psi}^T\right| \\
=&-\frac{1}{2} \log \left|\mathbf{K}^r \otimes \mathbf{S}^r + \boldsymbol{\Psi} (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l) \boldsymbol{\Psi}^T\right|.
\end{aligned}
\tag{A.40}
$$

综合目前推导出的所有结果，非子集数据的联合似然为：

$$
\begin{aligned}
&\log p(\mathbf{Y}^l, \mathbf{Y}^h) \\
=&\log p(\mathbf{Y}^l) - \frac{N^h d^h}{2} \log(2\pi) - \frac{1}{2} \log \left|\mathbf{K}^r \otimes \mathbf{S}^r + \boldsymbol{\Psi} (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l) \boldsymbol{\Psi}^T\right| \\
&-\frac{1}{2} (\boldsymbol{\phi} - \boldsymbol{\Psi} \mathrm{vec}(\bar{\mathbf{Y}}^l))^T \left(\mathbf{K}^r \otimes \mathbf{S}^r + \boldsymbol{\Psi} (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l) \boldsymbol{\Psi}^T\right)^{-1} (\boldsymbol{\phi} - \boldsymbol{\Psi} \mathrm{vec}(\bar{\mathbf{Y}}^l))
\end{aligned}
\tag{A.41}
$$

其中 $\boldsymbol{\phi} - \boldsymbol{\Psi}$向量化($\bar{\mathrm{Y}}^l$) $= \begin{pmatrix} \text{向量化}(\bar{\mathrm{Y}}^h) \\ \text{向量化}(\hat{\mathrm{Y}}^h) \end{pmatrix} - \mathrm{W} \begin{pmatrix} \text{向量化}(\hat{\mathrm{Y}}^l) \\ \text{向量化}(\bar{\mathrm{Y}}^l) \end{pmatrix}$，且

$\mathrm{K}^r \otimes \mathrm{S}^r + \boldsymbol{\Psi}(\hat{\mathrm{S}}^l \otimes \mathrm{S}^l)\boldsymbol{\Psi}^T = \mathrm{K}^r \otimes \mathrm{S}^r + \hat{\mathrm{E}}\hat{\mathrm{S}}^l \hat{\mathrm{E}}^T \otimes \mathrm{W}^T \mathrm{S}^l \mathrm{W}$。

B.3 非子集多保真数据的后验分布

接下来，我们探讨这种非子集数据结构的后验分布。首先，我们利用积分来表示后验分布，

$$\begin{aligned} p\left(\mathbf{Y}^*|\mathbf{Y}^l,\mathbf{Y}^h\right) &= \int p(\mathbf{Y}^*,\hat{\mathbf{Y}}^l|\mathbf{Y}^h,\mathbf{Y}^l)d\hat{\mathbf{Y}}^l \\ &= \int p(\mathbf{Y}^*|\mathbf{Y}^h,\mathbf{Y}^l,\hat{\mathbf{Y}}^l)p(\hat{\mathbf{Y}}^l)d\hat{\mathbf{Y}}^l \end{aligned} \tag{A.42}$$

我们尝试将积分表示为不同部分。一旦确定了 $\hat{\mathbf{Y}}^l$，预测后验 $p(\mathbf{Y}^*|\mathbf{Y}^h,\mathbf{Y}^l,\hat{\mathbf{Y}}^l)$ 就可以用标准子集方式表示，即 $p(\mathbf{Y}^*|\mathbf{Y}^h,\mathbf{Y}^l,\hat{\mathbf{Y}}^l) \sim \mathcal{N}\left(\text{vec}(\bar{\mathbf{Z}}_*^h),\mathbf{S}_*^h\right)$，其中均值函数和协方差矩阵为

$$\begin{aligned} \text{vec}(\bar{\mathbf{Z}}_*^h) &= \left(\mathbf{k}_*^l(\mathbf{K}^l)^{-1}\otimes\mathbf{W}\right)\begin{pmatrix}\text{vec}(\mathbf{Y}^l)\\ \text{vec}(\hat{\mathbf{Y}}^l)\end{pmatrix} + \left(\mathbf{k}_*^r(\mathbf{K}^r)^{-1}\otimes\mathbf{I}^h\right)\text{vec}(\mathbf{Y}^r), \\ \mathbf{S}_*^h &= \left(k_{**}^l-(\mathbf{k}_*^l)^T(\mathbf{K}^l)^{-1}\mathbf{k}_*^l\right)\otimes\mathbf{W}\mathbf{S}^l\mathbf{W}^T + \left(k_{**}^r-(\mathbf{k}_*^r)^T(\mathbf{K}^r)^{-1}\mathbf{k}_*^r\right)\otimes\mathbf{S}^r. \end{aligned} \tag{A.43}$$

我们进一步简化这一情形，并引入如下定义：

$$\begin{aligned} \mathbf{K}_*^l &= k_{**}^l-(\mathbf{k}_*^l)^T(\mathbf{K}^l)^{-1}\mathbf{k}_*^l, \\ \mathbf{K}_*^r &= k_{**}^r-(\mathbf{k}_*^r)^T(\mathbf{K}^r)^{-1}\mathbf{k}_*^r \end{aligned}$$

由此可将式 (A.43) 简化为

$$\mathbf{S}_*^h = \mathbf{K}_*^l\otimes\mathbf{W}\mathbf{S}^l\mathbf{W}^T + \mathbf{K}_*^r\otimes\mathbf{S}^r.$$

同时，由于 $\hat{\mathbf{Y}}^l$ 是从$\mathbf{Y}^l$中抽取的样本，因此它也服从子集方式下的后验分布，这意味着 $\hat{\mathbf{Y}}^l \sim \mathcal{N}\left(\text{vec}(\bar{\mathbf{Y}}^l),\hat{\mathbf{S}}^l\otimes\mathbf{S}^l\right)$，其中 $\text{vec}(\bar{\mathbf{Y}}^l)$ 和 $\mathbf{S}_*^l\otimes\mathbf{S}^l$为

$$\begin{aligned} \text{vec}(\bar{\mathbf{Y}}^l) &= \left(\mathbf{k}_*^l(\mathbf{K}^l)^{-1}\otimes\mathbf{I}^l\right)\text{vec}(\mathbf{Y}^l), \\ \hat{\mathbf{S}}^l\otimes\mathbf{S}^l &= \left(k_{**}^l-(\mathbf{k}_*^l)^T(\mathbf{K}^l)^{-1}\mathbf{k}_*^l\right)\otimes\mathbf{S}^l. \end{aligned}$$

因此，非子集数据结构的后验分布为

$$
\begin{aligned}
&p\left(\mathbf{Y}^{*}|\mathbf{Y}^{l},\mathbf{Y}^{h}\right)\\
=&\int p(\mathbf{Y}^{*},\hat{\mathbf{Y}}^{l}|\mathbf{Y}^{h},\mathbf{Y}^{l})d\hat{\mathbf{Y}}^{l}\\
=&\int p(\mathbf{Y}^{*}|\mathbf{Y}^{h},\mathbf{Y}^{l},\hat{\mathbf{Y}}^{l})p(\hat{\mathbf{Y}}^{l})d\hat{\mathbf{Y}}^{l}\\
=&\int 2\pi^{-\frac{N^{p}d^{h}}{2}}\times\left|\mathbf{S}_{*}^{h}\right|^{-\frac{1}{2}}\\
&\times\exp\left[-\frac{1}{2}\left(\mathrm{vec}(\mathbf{Y}^{*})-\left(\mathbf{k}_{*}^{l}(\hat{\mathbf{K}}^{l})^{-1}\otimes\mathbf{W}\right)\begin{pmatrix}\mathrm{vec}(\mathbf{Y}^{l})\\ \mathrm{vec}(\hat{\mathbf{Y}}^{l})\end{pmatrix}-\left(\mathbf{k}_{*}^{r}(\mathbf{K}^{r})^{-1}\otimes\mathbf{I}^{h}\right)\mathrm{vec}(\mathbf{Y}^{r})\right)^{T}\right.\\
&\left.(\mathbf{S}_{*}^{h})^{-1}\left(\mathrm{vec}(\mathbf{Y}^{*})-\left(\mathbf{k}_{*}^{l}(\hat{\mathbf{K}}^{l})^{-1}\otimes\mathbf{W}\right)\begin{pmatrix}\mathrm{vec}(\mathbf{Y}^{l})\\ \mathrm{vec}(\hat{\mathbf{Y}}^{l})\end{pmatrix}-\left(\mathbf{k}_{*}^{r}(\mathbf{K}^{r})^{-1}\otimes\mathbf{I}^{h}\right)\mathrm{vec}(\mathbf{Y}^{r})\right)\right]\\
&\times 2\pi^{-\frac{N^{m}d^{l}}{2}}\times\left|\hat{\mathbf{S}}^{l}\otimes\mathbf{S}^{l}\right|^{-\frac{1}{2}}\times\exp[-\frac{1}{2}(\mathrm{vec}(\hat{\mathbf{Y}}^{l})-\mathrm{vec}(\bar{\mathbf{Y}}^{l}))^{T}(\hat{\mathbf{S}}^{l}\otimes\mathbf{S}^{l})^{-1}(\mathrm{vec}(\hat{\mathbf{Y}}^{l})-\mathrm{vec}(\bar{\mathbf{Y}}^{l}))]d\hat{\mathbf{Y}}^{l}\\
=&2\pi^{-\frac{N^{p}d^{h}+N^{m}d^{l}}{2}}\times\left|\mathbf{S}_{*}^{h}\right|^{-\frac{1}{2}}\times\left|\hat{\mathbf{S}}^{l}\otimes\mathbf{S}^{l}\right|^{-\frac{1}{2}}\times\exp\left[-\frac{1}{2}\tilde{\mathbf{Y}}^{T}(\mathbf{S}_{*}^{h})^{-1}\tilde{\mathbf{Y}}-\frac{1}{2}\mathrm{vec}(\bar{\mathbf{Y}}^{l})^{T}(\hat{\mathbf{S}}^{l}\otimes\mathbf{S}^{l})^{-1}\mathrm{vec}(\bar{\mathbf{Y}}^{l})\right]\\
&\times\int\exp[+\tilde{\mathbf{Y}}^{T}(\mathbf{S}_{*}^{h})^{-1}\mathbf{\Gamma}\mathrm{vec}(\hat{\mathbf{Y}}^{l})+\mathrm{vec}(\bar{\mathbf{Y}}^{l})^{T}(\hat{\mathbf{S}}^{l}\otimes\mathbf{S}^{l})^{-1}\mathrm{vec}(\hat{\mathbf{Y}}^{l})\\
&-\frac{1}{2}\mathrm{vec}(\hat{\mathbf{Y}}^{l})^{T}\mathbf{\Gamma}^{T}(\mathbf{S}_{*}^{h})^{-1}\mathbf{\Gamma}\mathrm{vec}(\hat{\mathbf{Y}}^{l})-\frac{1}{2}\mathrm{vec}(\hat{\mathbf{Y}}^{l})^{T}(\hat{\mathbf{S}}^{l}\otimes\mathbf{S}^{l})^{-1}\mathrm{vec}(\hat{\mathbf{Y}}^{l})]d\hat{\mathbf{Y}}^{l}\\
=&2\pi^{-\frac{N^{p}d^{h}+N^{m}d^{l}}{2}}\times\left|\mathbf{S}_{*}^{h}\right|^{-\frac{1}{2}}\times\left|\hat{\mathbf{S}}^{l}\otimes\mathbf{S}^{l}\right|^{-\frac{1}{2}}\times\exp\left[-\frac{1}{2}\tilde{\mathbf{Y}}^{T}(\mathbf{S}_{*}^{h})^{-1}\tilde{\mathbf{Y}}-\frac{1}{2}\mathrm{vec}(\bar{\mathbf{Y}}^{l})^{T}(\hat{\mathbf{S}}^{l}\otimes\mathbf{S}^{l})^{-1}\mathrm{vec}(\bar{\mathbf{Y}}^{l})\right]\\
&\times 2\pi^{\frac{N^{m}d^{l}}{2}}\times\left|\mathbf{\Gamma}^{T}(\mathbf{S}_{*}^{h})^{-1}\mathbf{\Gamma}+(\hat{\mathbf{S}}^{l}\otimes\mathbf{S}^{l})^{-1}\right|^{-\frac{1}{2}}\times\exp\left[\frac{1}{2}\left(\tilde{\mathbf{Y}}^{T}(\mathbf{S}_{*}^{h})^{-1}\mathbf{\Gamma}+\mathrm{vec}(\bar{\mathbf{Y}}^{l})^{T}(\hat{\mathbf{S}}^{l}\otimes\mathbf{S}^{l})^{-1}\right)\right.\\
&\left.\left(\mathbf{\Gamma}^{T}(\mathbf{S}_{*}^{h})^{-1}\mathbf{\Gamma}+(\hat{\mathbf{S}}^{l}\otimes\mathbf{S}^{l})^{-1}\right)^{-1}\left(\tilde{\mathbf{Y}}^{T}(\mathbf{S}_{*}^{h})^{-1}\mathbf{\Gamma}+\mathrm{vec}(\bar{\mathbf{Y}}^{l})^{T}(\hat{\mathbf{S}}^{l}\otimes\mathbf{S}^{l})^{-1}\right)^{T}\right]\\
=&2\pi^{-\frac{N^{p}d^{h}}{2}}\times\underbrace{\left|\mathbf{S}_{*}^{h}\right|^{-\frac{1}{2}}\times\left|\hat{\mathbf{S}}^{l}\otimes\mathbf{S}^{l}\right|^{-\frac{1}{2}}\times\left|\mathbf{\Gamma}^{T}(\mathbf{S}_{*}^{h})^{-1}\mathbf{\Gamma}+(\hat{\mathbf{S}}^{l}\otimes\mathbf{S}^{l})^{-1}\right|^{-\frac{1}{2}}}_{\text{part d}}\\
&\times\exp\left[-\frac{1}{2}\tilde{\mathbf{Y}}^{T}\underbrace{\left((\mathbf{S}_{*}^{h})^{-1}-(\mathbf{S}_{*}^{h})^{-1}\mathbf{\Gamma}\left(\mathbf{\Gamma}^{T}(\mathbf{S}_{*}^{h})^{-1}\mathbf{\Gamma}+(\hat{\mathbf{S}}^{l}\otimes\mathbf{S}^{l})^{-1}\right)^{-1}\mathbf{\Gamma}^{T}(\mathbf{S}_{*}^{h})^{-1}\right)}_{\text{part a}}\tilde{\mathbf{Y}}\right.\\
&-\frac{1}{2}\mathrm{vec}(\bar{\mathbf{Y}}^{l})^{T}\underbrace{\left((\hat{\mathbf{S}}^{l}\otimes\mathbf{S}^{l})^{-1}-(\hat{\mathbf{S}}^{l}\otimes\mathbf{S}^{l})^{-1}\left(\mathbf{\Gamma}^{T}(\mathbf{S}_{*}^{h})^{-1}\mathbf{\Gamma}+(\hat{\mathbf{S}}^{l}\otimes\mathbf{S}^{l})^{-1}\right)^{-1}(\hat{\mathbf{S}}^{l}\otimes\mathbf{S}^{l})^{-1}\right)}_{\text{part b}}\mathrm{vec}(\bar{\mathbf{Y}}^{l})\\
&\left.+\tilde{\mathbf{Y}}^{T}\underbrace{(\mathbf{S}_{*}^{h})^{-1}\mathbf{\Gamma}\left(\mathbf{\Gamma}^{T}(\mathbf{S}_{*}^{h})^{-1}\mathbf{\Gamma}+(\hat{\mathbf{S}}^{l}\otimes\mathbf{S}^{l})^{-1}\right)^{-1}(\hat{\mathbf{S}}^{l}\otimes\mathbf{S}^{l})^{-1}}_{\text{part c}}\mathrm{vec}(\bar{\mathbf{Y}}^{l})\right]
\end{aligned}
\tag{A.44}
$$

˜其中 $\mathbf{Y}$ 和 $\mathbf{\Gamma}$由下式定义，

$$\begin{aligned}\tilde{\mathbf{Y}} &= \left(\mathrm{vec}(\mathbf{Y}^*) - \left(\mathbf{k}_*^l(\hat{\mathbf{K}}^l)^{-1} \otimes \mathbf{W}\right)\begin{pmatrix}\mathrm{vec}(\mathbf{Y}^l)\\ \mathbf{0}\end{pmatrix} - \left(\mathbf{k}_*^r(\mathbf{K}^r)^{-1} \otimes \mathbf{I}^h\right)\left(\mathrm{vec}(\mathbf{Y}^h) - \begin{pmatrix}\mathrm{vec}(\check{\mathbf{Y}}^l)\\ \mathbf{0}\end{pmatrix}\right)\right),\\ \mathbf{\Gamma} &= \left([\mathbf{k}_*^r(\mathbf{K}^r)^{-1}\mathbf{E}_n^T - \mathbf{k}_*^l(\hat{\mathbf{K}}^l)^{-1}] \otimes \mathbf{W}\right)\mathbf{E}_m \otimes \mathbf{I}^l.\end{aligned} \tag{A.45}$$

然后，我们利用 Sherman-Morrison公式将式 (A.44) 中的 a、b 和 c 部分简化如下。对于式 (A.44) 中的 a 部分，

$$\begin{aligned}&(\mathbf{S}_*^h)^{-1} - (\mathbf{S}_*^h)^{-1}\mathbf{\Gamma}\left(\mathbf{\Gamma}^T(\mathbf{S}_*^h)^{-1}\mathbf{\Gamma} + (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)^{-1}\right)^{-1}\mathbf{\Gamma}^T(\mathbf{S}_*^h)^{-1}\\ =&\left(\mathbf{S}_*^h + \mathbf{\Gamma}(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)\mathbf{\Gamma}^T\right)^{-1},\end{aligned} \tag{A.46}$$

对于式 (A. 中的 b 部分)

$$\begin{aligned}&(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)^{-1} - (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)^{-1}\left(\mathbf{\Gamma}^T(\mathbf{S}_*^h)^{-1}\mathbf{\Gamma} + (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)^{-1}\right)^{-1}(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)^{-1}\\ =&(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)^{-1} - \left((\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)\mathbf{\Gamma}^T(\mathbf{S}_*^h)^{-1}\mathbf{\Gamma}(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l) + (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)\right)^{-1}\\ =&(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)^{-1} - \left((\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)^{-1} - \mathbf{\Gamma}^T(\mathbf{S}_*^h + \mathbf{\Gamma}(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)\mathbf{\Gamma}^T)^{-1}\mathbf{\Gamma}\right)\\ =&\mathbf{\Gamma}^T(\mathbf{S}_*^h + \mathbf{\Gamma}(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)\mathbf{\Gamma}^T)^{-1}\mathbf{\Gamma},\end{aligned} \tag{A.47}$$

对于式 (A44) 中的.c 部分，

$$\begin{aligned}&(\mathbf{S}_*^h)^{-1}\mathbf{\Gamma}\left(\mathbf{\Gamma}^T(\mathbf{S}_*^h)^{-1}\mathbf{\Gamma} + (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)^{-1}\right)^{-1}(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)^{-1}\\ =&(\mathbf{S}_*^h)^{-1}\mathbf{\Gamma}\left((\hat{\mathbf{S}}^l \otimes \mathbf{S}^l) - (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)\mathbf{\Gamma}^T(\mathbf{S}_*^h + \mathbf{\Gamma}(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)\mathbf{\Gamma}^T)^{-1}\mathbf{\Gamma}(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)\right)(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)^{-1}\\ =&(\mathbf{S}_*^h)^{-1}\mathbf{\Gamma} - (\mathbf{S}_*^h)^{-1}\mathbf{\Gamma}(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)\mathbf{\Gamma}^T(\mathbf{S}_*^h + \mathbf{\Gamma}(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)\mathbf{\Gamma}^T)^{-1}\mathbf{\Gamma}\\ =&(\mathbf{S}_*^h)^{-1}\mathbf{\Gamma} - (\mathbf{S}_*^h)^{-1}\left(\mathbf{S}_*^h + \mathbf{\Gamma}(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)\mathbf{\Gamma}^T - \mathbf{S}_*^h\right)(\mathbf{S}_*^h + \mathbf{\Gamma}(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)\mathbf{\Gamma}^T)^{-1}\mathbf{\Gamma}\\ =&(\mathbf{S}_*^h)^{-1}\mathbf{\Gamma} - (\mathbf{S}_*^h)^{-1}\left(\mathbf{I} - \mathbf{S}_*^h(\mathbf{S}_*^h + \mathbf{\Gamma}(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)\mathbf{\Gamma}^T)^{-1}\right)\mathbf{\Gamma}\\ =&(\mathbf{S}_*^h)^{-1}\mathbf{\Gamma} - (\mathbf{S}_*^h)^{-1}\mathbf{\Gamma} - (\mathbf{S}_*^h + \mathbf{\Gamma}(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)\mathbf{\Gamma}^T)^{-1}\mathbf{\Gamma}\\ =&-(\mathbf{S}_*^h + \mathbf{\Gamma}(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)\mathbf{\Gamma}^T)^{-1}\mathbf{\Gamma}.\end{aligned} \tag{A.48}$$

行列式（式 (A.44) 中的 d 部分）也可以利用 Sherman-Morrison公式推导出更简洁的形式，

$$\begin{aligned}&\left|\mathbf{S}_*^h\right|^{-\frac{1}{2}} \times \left|\hat{\mathbf{S}}^l \otimes \mathbf{S}^l\right|^{-\frac{1}{2}} \times \left|\mathbf{\Gamma}^T(\mathbf{S}_*^h)^{-1}\mathbf{\Gamma} + (\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)^{-1}\right|^{-\frac{1}{2}}\\ =&\left|\mathbf{S}_*^h\right|^{-\frac{1}{2}} \times \left|\hat{\mathbf{S}}^l \otimes \mathbf{S}^l\right|^{-\frac{1}{2}} \times \left|(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)^{-1}\right|^{-\frac{1}{2}} \times \left|(\mathbf{S}_*^h)^{-1}\right|^{-\frac{1}{2}} \times \left|(\mathbf{S}_*^h)^{-1} + \mathbf{\Gamma}(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)^{-1}\mathbf{\Gamma}^T\right|^{-\frac{1}{2}}\\ =&\left|(\mathbf{S}_*^h)^{-1} + \mathbf{\Gamma}(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)^{-1}\mathbf{\Gamma}^T\right|^{-\frac{1}{2}}\end{aligned} \tag{A.49}$$

将 a、b、c 和 d 部分代回式 (A 44)，可得简洁形式

$$\begin{aligned}&p\left(\mathbf{Y}^*|\mathbf{Y}^l, \mathbf{Y}^h\right)\\ =&2\pi^{-\frac{Np_d h}{2}} \times \left|(\mathbf{S}_*^h)^{-1} + \mathbf{\Gamma}(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)^{-1}\mathbf{\Gamma}^T\right|^{-\frac{1}{2}} \times \exp[-\frac{1}{2}\tilde{\mathbf{Y}}^T\left(\mathbf{S}_*^h + \mathbf{\Gamma}(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)\mathbf{\Gamma}^T\right)^{-1}\tilde{\mathbf{Y}}\\ &-\frac{1}{2}\mathrm{vec}(\bar{\mathbf{Y}}^l)^T\mathbf{\Gamma}^T\left(\mathbf{S}_*^h + \mathbf{\Gamma}(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)\mathbf{\Gamma}^T\right)^{-1}\mathbf{\Gamma}\mathrm{vec}(\bar{\mathbf{Y}}^l) - \tilde{\mathbf{Y}}^T\left(\mathbf{S}_*^h + \mathbf{\Gamma}(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)\mathbf{\Gamma}^T\right)^{-1}\mathbf{\Gamma}\mathrm{vec}(\bar{\mathbf{Y}}^l)]\\ =&2\pi^{-\frac{Np_d h}{2}} \times \left|(\mathbf{S}_*^h)^{-1} + \mathbf{\Gamma}(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)^{-1}\mathbf{\Gamma}^T\right|^{-\frac{1}{2}}\\ &\times \exp[-\frac{1}{2}\left(\tilde{\mathbf{Y}} - \mathbf{\Gamma}\mathrm{vec}(\bar{\mathbf{Y}}^l)\right)^T\left(\mathbf{S}_*^h + \mathbf{\Gamma}(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l)\mathbf{\Gamma}^T\right)^{-1}\left(\tilde{\mathbf{Y}} - \mathbf{\Gamma}\mathrm{vec}(\bar{\mathbf{Y}}^l)\right)]\\ =&2\pi^{-\frac{d^h}{2}} \times \left|\mathbf{S}_*^h + \mathbf{\Gamma}\left(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l\right)\mathbf{\Gamma}^T\right|^{-\frac{1}{2}}\\ &\times \exp\left[-\frac{1}{2}\left(\mathrm{vec}(\mathbf{Z}_*^h) - \mathrm{vec}(\bar{\mathbf{Z}})\right)^T\left(\mathbf{S}_*^h + \mathbf{\Gamma}\left(\hat{\mathbf{S}}^l \otimes \mathbf{S}^l\right)\mathbf{\Gamma}^T\right)^{-1}\left(\mathrm{vec}(\mathbf{Z}_*^h) - \mathrm{vec}(\bar{\mathbf{Z}})\right)\right].\end{aligned} \tag{A.50}$$

We 可以看到，联合似然最终呈现出一种关于低保真度TGP和残差TG的优雅形式 P.

B.4 CIGAR

正如我们在章节 3.4 中提到的，我们假设输出协方差矩阵$\mathbf{S}_m^h$和$\mathbf{S}_m^l$是相同的矩阵，且权重矩阵$\mathbf{W}_m^T$ $\mathbf{W}_m$ = 相互正交，即$\mathbf{I}$。将这些假设代入 (A.41)，可得到简化后的协方差矩阵，

$$\begin{aligned}&\mathbf{K}^r\otimes\mathbf{I}^r+\hat{\mathbf{E}}\hat{\mathbf{S}}^l\hat{\mathbf{E}}^T\otimes\mathbf{W}^T\mathbf{I}^l\mathbf{W}\\=&\mathbf{K}^r\otimes\mathbf{I}^r+\hat{\mathbf{E}}\hat{\mathbf{S}}^l\hat{\mathbf{E}}^T\otimes\mathbf{W}^T\mathbf{W}\\=&\mathbf{K}^r\otimes\mathbf{I}^r+\hat{\mathbf{E}}\hat{\mathbf{S}}^l\hat{\mathbf{E}}^T\otimes\mathbf{I}^h\\=&(\mathbf{K}^r+\hat{\mathbf{E}}\hat{\mathbf{S}}^l\hat{\mathbf{E}}^T)\otimes\mathbf{I}^h\end{aligned}\tag{A.51}$$

其中$\mathbf{I}^r$是大小为 $d^r\times d^r$的单位矩阵；$\mathbf{I}^r$同样适用此规则；并且$\mathbf{I}^r=\mathbf{I}^h$。非子集数据的联合似然变为

$$\begin{aligned}\log p(\mathbf{Y}^l,\mathbf{Y}^h)=&\log p(\mathbf{Y}^l)-\frac{N^hd^h}{2}\log(2\pi)-\frac{1}{2}\log\left|(\mathbf{K}^r+\hat{\mathbf{E}}\hat{\mathbf{S}}^l\hat{\mathbf{E}}^T)\otimes\mathbf{I}^h\right|\\&-\frac{1}{2}(\boldsymbol{\phi}-\boldsymbol{\Psi}\mathrm{vec}(\bar{\mathbf{Y}}^l))^T\left((\mathbf{K}^r+\hat{\mathbf{E}}\hat{\mathbf{S}}^l\hat{\mathbf{E}}^T)\otimes\mathbf{I}^h\right)^{-1}(\boldsymbol{\phi}-\boldsymbol{\Psi}\mathrm{vec}(\bar{\mathbf{Y}}^l))\end{aligned}\tag{A.52}$$

where $\left(\boldsymbol{\phi}-\boldsymbol{\Psi}\mathrm{vec}(\bar{\mathbf{Y}}^l)\right)=\begin{pmatrix}\mathrm{vec}(\check{\mathbf{Y}}^h)\\\mathrm{vec}(\hat{\mathbf{Y}}^h)\end{pmatrix}-\tilde{\mathbf{W}}\begin{pmatrix}\mathrm{vec}(\bar{\mathbf{Y}}^l)\\\mathrm{vec}(\bar{\mathbf{Y}}^l)\end{pmatrix}$.

我们可以看到，核矩阵求逆的复杂度降低至 $\mathcal{O}((N^h)^3)$。

B.5 τ-保真度自回归模型

正如我们在章节 2.1 中提到的，我们可以将自回归应用于更多保真度层级，GAR 也是如此。在本节中，我们尝试将 GAR 扩展到更多保真度层级。假设$\mathbf{F}^\tau(\mathbf{x})=\mathbf{F}^{\tau-1}(\mathbf{x})\times_1\mathbf{W}_1^{\tau-1}\times_2\cdots\times_M\mathbf{W}_M^{\tau-1}+\mathbf{F}_\tau^r(\mathbf{x})$，我们可以推导出联合协方差矩阵，

$$\boldsymbol{\Sigma}^\tau=\begin{pmatrix}\mathbf{K}^{\tau-1}(\mathbf{X}^{\tau-1},\mathbf{X}^{\tau-1})\otimes\mathbf{S}^{\tau-1} & \mathbf{K}^{\tau-1}(\mathbf{X}^{\tau-1},\mathbf{X}^{\tau})\otimes\mathbf{S}^{\tau-1}(\mathbf{W}^{\tau-1})^T\\\mathbf{K}^{\tau-1}(\mathbf{X}^{\tau},\mathbf{X}^{\tau-1})\otimes\mathbf{W}^{\tau-1}\mathbf{S}^{\tau-1} & \mathbf{K}^{\tau-1}(\mathbf{X}^{\tau},\mathbf{X}^{\tau})\otimes\mathbf{W}^{\tau-1}\mathbf{S}^{\tau-1}(\mathbf{W}^{\tau-1})^T+\mathbf{K}_\tau^r(\mathbf{X}^{\tau},\mathbf{X}^{\tau})\otimes\mathbf{S}_\tau^r\end{pmatrix},$$

其中$\mathbf{S}^{\tau-1}=\bigotimes_{m=1}^M\mathbf{S}_m^{\tau-1}$且$\mathbf{W}^{\tau-1}=\bigotimes_{m=1}^M\mathbf{W}_m^{\tau-1}$。

与 GAR 的证明相同，我们可以推导出联合协方差矩阵的求逆，$(\boldsymbol{\Sigma}^\tau)^{-1}=$

$$\begin{bmatrix}(\mathbf{K}^{\tau-1})^{-1}\otimes(\mathbf{S}^{\tau-1})^{-1}+\begin{pmatrix}\mathbf{0}&\mathbf{0}\\\mathbf{0}&(\mathbf{K}_\tau^r)^{-1}\otimes\mathbf{W}^{\tau-1T}(\mathbf{S}_\tau^r)^{-1}\mathbf{W}^{\tau-1}\end{pmatrix} & -\begin{pmatrix}\mathbf{0}\\(\mathbf{K}_\tau^r)^{-1}\otimes\mathbf{W}^{\tau-1T}(\mathbf{S}_\tau^r)^{-1}\end{pmatrix}\\-\begin{pmatrix}\mathbf{0},(\mathbf{K}_\tau^r)^{-1}\otimes(\mathbf{S}_\tau^r)^{-1}\mathbf{W}^{\tau-1}\end{pmatrix} & (\mathbf{K}_\tau^r)^{-1}\otimes(\mathbf{S}_\tau^r)^{-1}\end{bmatrix}$$

Ther因此，我们在此展示了将序列拆分为多个子序列的过程，对于将其拆分为相互独立的 d TGP . We pres后验分布的均值函数和协方差矩阵，

$$\begin{aligned}\mathrm{vec}(\mathbf{Z}_*^\tau)=&\left(\mathbf{k}_*^{\tau-1}(\mathbf{K}^{\tau-1})^{-1}\otimes\mathbf{W}^{\tau-1}\right)\mathrm{vec}(\mathbf{Y}^{\tau-1})+\left((\mathbf{k}_\tau^r)_*(\mathbf{K}_\tau^r)^{-1}\otimes\mathbf{I}_r\right)\mathrm{vec}(\mathbf{Y}_\tau^r)\\\mathbf{S}_*^\tau=&\left(k_{**}^{\tau-1}-(\mathbf{k}_*^{\tau-1})^T(\mathbf{K}^{\tau-1})^{-1}\mathbf{k}_*^{\tau-1}\right)\otimes\mathbf{W}^{\tau-1}\mathbf{S}^{\tau-1}(\mathbf{W}^{\tau-1})^T+\left((k_\tau^r)_{**}-(\mathbf{k}_\tau^r)_*^T(\mathbf{K}_\tau^r)^{-1}(\mathbf{k}_\tau^r)_*\right)\otimes\mathbf{S}_\tau^r.\end{aligned}$$

(A.53)

C 最先进方法总结

我们在表1中比较并总结了最先进方法、GAR 和 CIGAR 的能力与复杂度。

D 实现与复杂度

现在，我们使用张量代数给出 GAR 和 CIGAR 的训练与预测算法，从而无需构造或显式计算完整协方差矩阵，以提高计算效率。以正态 TGP 为例，给定数据集 $(\mathbf{X},\mathbf{Y})$，$\mathrm{vec}(\mathbf{Y})\sim\mathcal{N}\left(\mathbf{0},\mathbf{K}(\mathbf{X},\mathbf{X})\otimes\left(\bigotimes_{m=1}^M\mathbf{S}_m\right)\right)$。推断需要估计所有协方差矩阵 $\bigotimes_{m=1}^M\mathbf{S}_m$ 和 $\mathbf{K}(\mathbf{X},\mathbf{X})$。为简洁起见，我们用 $\mathbf{S}$ 和 $\mathbf{K}$ 分别表示 $\bigotimes_{m=1}^M\mathbf{S}_m$ 和 $\mathbf{K}(\mathbf{X},\mathbf{X})$，并用 $\boldsymbol{\Sigma}=\mathbf{K}\otimes\mathbf{S}+\epsilon^{-1}\mathbf{I}$表示。我们通过最小化模型的负对数似然来估计参数，

$$\mathcal{L}=\frac{Nd}{2}\log(2\pi)+\frac{1}{2}\log|\boldsymbol{\Sigma}|+\frac{1}{2}\mathrm{vec}(\mathbf{Y})^T\boldsymbol{\Sigma}^{-1}\mathrm{vec}(\mathbf{Y}).$$

表1：高维问题中最先进的多保真融合方法比较

模型	任意输出?	非子集数据?	复杂度
NAR [16]	Yes	No	$\mathcal{O}(\sum_i (N^i)^3)$
ResGP[9]	No	No	$\mathcal{O}(\sum_i (N^i)^3)$
MF-BNN [13]	Yes	Yes	$\mathcal{O}(\sum_i (N^i)(A_i^2+\omega))$*
直流 [12]	Yes	No	$\mathcal{O}(\sum_i (N^i)^3)$
自回归 [3]	No	No	$\mathcal{O}(\sum_i (N^i d^i)^3)$
GAR	Yes	Yes	$\mathcal{O}(\sum_i \sum_{m=1}^{M} (d_m^i)^3 + (N^i)^3)$
CIGAR	Yes	Yes	$\mathcal{O}(\sum_i (N^i)^3)$

*A_i 是第 i 个保真度下神经网络的总权重规模， ω 是所有参数的数量

然而，由于 S 是一个大小为 $Nd \times Nd$ 的矩阵，当输出的数量较大时，将无法计算 K $\otimes$S 的求逆。因此，对于 TGP，我们利用 K $\otimes$S 中的克罗内克积来高效地计算负对数似然。首先，我们通过特征值分解将联合核矩阵表示为 K $=$U^{T}diag(λ)U，且 S$_m$ $=$U$_m^T$diag(λ_m)U$_m$。然后，我们用 $\mathbf{\Sigma}$ 表示联合核矩阵： $\mathbf{\Sigma}$ $=$K $\otimes$S $+\epsilon^{-1}$I $=$ $\left(\text{U}^T\text{diag}(\lambda)\text{U}\right) \otimes \left(\text{U}_1^T\text{diag}(\lambda_1)\text{U}_1\right) \otimes \cdots \otimes \left(\text{U}_M^T\text{diag}(\lambda_M)\text{U}_M\right) + \epsilon^{-1}$I。利用克罗内克积的性质，可得

$$\mathbf{\Sigma} = \mathbf{P}^T \Lambda \mathbf{P} + \epsilon^{-1}\mathbf{I} \tag{A.54}$$

其中 $\mathbf{P} = \mathbf{U} \otimes \mathbf{U}_1 \otimes \cdots \otimes \mathbf{U}_M$ 且 $\Lambda =$diag($\lambda \otimes \lambda_1 \otimes \cdots \otimes \lambda_M$)，由于 $\mathbf{U}$ 和 $\mathbf{U}_m$ 是特征向量且相互正交，因此 $\mathbf{P}^T\mathbf{P} = \mathbf{P}\mathbf{P}^T = \mathbf{I}$。因此，我们可以得到

$$\log|\mathbf{\Sigma}| = \log\left|\mathbf{P}^T\Lambda\mathbf{P} + \epsilon^{-1}\mathbf{I}\right| = \log\left|\mathbf{P}^T(\Lambda + \epsilon^{-1}\mathbf{I})\mathbf{P}\right| = \log\left|\Lambda + \epsilon^{-1}\mathbf{I}\right|. \tag{A.55}$$

因此，我们只需计算 Nd 对角元素，即可计算负对数似然的一部分 .

之后，我们计算负对数似然中的 vec(Y)$^T\mathbf{\Sigma}^{-1}$vec(Y)部分。首先，我们有 A
$\lambda \circ \lambda_1 \circ \cdots \circ \lambda_M + \epsilon^{-1}$1 1 $\circ$ =，其中 是一个全 1 张量， 是克鲁斯卡尔算子。接下来有

$$\begin{aligned}
\text{vec}(\mathbf{Y})^T\mathbf{\Sigma}^{-1}\text{vec}(\mathbf{Y}) &= \text{vec}(\mathbf{Y})^T\mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{\Sigma}^{-\frac{1}{2}}\text{vec}(\mathbf{Y}) \\
&= \text{vec}(\mathbf{Y})^T\mathbf{P}^T(\Lambda + \epsilon^{-1}\mathbf{I})^{-\frac{1}{2}}\mathbf{P}\mathbf{P}(\Lambda + \epsilon^{-1}\mathbf{I})^{-\frac{1}{2}}\mathbf{P}^T\text{vec}(\mathbf{Y}) \\
&= \eta^T\eta,
\end{aligned} \tag{A.56}$$

其中 $\eta =$P($\Lambda + \epsilon^{-1}$I)$^{-\frac{1}{2}}$P^{T}vec(Y)。由于 $\mathbf{P}$ 是克罗内克积矩阵，我们可以利用 Tucker 算子的性质 [19] 来计算 $\mathbf{b}$。

$$\begin{aligned}
&\mathbf{T}_1 = \mathbf{Y} \times_1 \mathbf{U}^T \times_2 \mathbf{U}_1^T \times_3 \cdots \times_{M+1} \mathbf{U}_M^T \\
&\mathbf{T}_2 = \mathbf{T}_1 \odot \mathbf{A}^{\cdot -\frac{1}{2}} \\
&\mathbf{T}_3 = \mathbf{T}_2 \times_1 \mathbf{U} \times_2 \mathbf{U}_1 \times_3 \cdots \times_{M+1} \mathbf{U}_M \\
&\eta = \text{vec}(\mathbf{T}_3)
\end{aligned} \tag{A.57}$$

whe其中 $\odot$ 表示逐元素乘积，而 $(\cdot)^{\cdot -\frac{1}{2}}$ 表示对 $-\frac{1}{2}$ 的各个元素分别取幂。因此， e
com负对数似然的复杂度为 $\mathcal{O}(\sum_{m=1}^{M}(d_m)^3 + (N)^3)$。
基于上述结论，我们还可以更高效地计算 GAR。根据引理 3，联合似然可分解为两个可分离似然 $\mathcal{L}^l$ 和 $\mathcal{L}^r$。对于这两者中的每一个，我们都可以利用这些技巧将复杂度降至 $\mathcal{O}(\sum_{m=1}^{M}(d_m^l)^3 + (N^l)^3) + \mathcal{O}(\sum_{m=1}^{M}(d_m^r)^3 + (N^r)^3)$。由于，

$$\begin{aligned}
&\log\left|\mathbf{K}^l \otimes \mathbf{S}^l\right| = \log\left|\Lambda^l + \epsilon^{-1}\mathbf{I}^l\right|, && \log|\mathbf{K}^r \otimes \mathbf{S}^r| = \log\left|\Lambda^r + \epsilon^{-1}\mathbf{I}^r\right|, \\
&\text{vec}(\mathbf{Y}^l)^T(\mathbf{K}^l \otimes \mathbf{S}^l)^{-1}\text{vec}(\mathbf{Y}^l) = (\eta^l)^T\eta^l; && \text{vec}(\mathbf{Y}^r)^T(\mathbf{K}^r \otimes \mathbf{S}^r)^{-1}\text{vec}(\mathbf{Y}^r) = (\eta^r)^T\eta^r,
\end{aligned} \tag{A.58}$$

其中 η^h、 η^l 和 Λ^h、 Λ^l 分别是低保真数据、残差及其对应的向量和特征值。因此，联合对数似然为:

$$\begin{aligned}
\mathcal{L} &= \mathcal{L}^l + \mathcal{L}^r \\
&= \text{const} - \frac{1}{2}\log\left|\Lambda^l + \epsilon^{-1}\mathbf{I}^l\right| - \frac{1}{2}(\eta^l)^T\eta^l - \frac{1}{2}\log\left|\Lambda^r + \epsilon^{-1}\mathbf{I}^r\right| - \frac{1}{2}(\eta^r)^T\eta^r
\end{aligned} \tag{A.59}$$

给定新的输入 $\mathbf{x}_*$，输出张量化为 vec(Z$_*^h$) 后的预测服从条件高斯分布 $\bar{}$vec(Z$_*$) $\sim \mathcal{N}$(vec(Z$_*$)，S$_*$)，其中

$$\begin{aligned}
\text{vec}(\bar{\mathbf{Z}}_*) &= \left(\mathbf{k}_* \left(\mathbf{K}\right)^{-1} \otimes \mathbf{I}\right)\text{vec}(\mathbf{Y}) \\
\mathbf{S}_* &= \left(k_{**} - \left(\mathbf{k}_*\right)^T \left(\mathbf{K}\right)^{-1}\mathbf{k}_*\right) \otimes \mathbf{S}.
\end{aligned} \tag{A.60}$$

我们可以使用 Tucker 算子，以更高效的方式计算预测均值 vec($\bar{\mathbf{Z}}_*$) 和 $\mathbf{S}_*$ 。利用核矩阵的特征值分解，可以推导出 $\mathbf{S}_* = k_{**} \otimes \mathbf{S} - \mathbf{L}\mathbf{L}^T$，其中 $\mathbf{L} = ((\mathbf{k}_*)^T(\mathbf{K})^{-1}\mathbf{U} \otimes \mathbf{U}_1 \otimes \cdots \otimes \mathbf{U}_M)(\Lambda(\Lambda + \epsilon^{-1}\mathbf{I})^{-\frac{1}{2}})$。因此，$\mathrm{diag}(\mathbf{S}_*) = k_{**} \otimes \mathrm{diag}(\mathbf{S}) - \mathrm{diag}(\mathbf{L}\mathbf{L}^T)$。我们也可以使用张量代数来计算预测协方差矩阵

$$\mathrm{diag}(\mathbf{S}_*) = \mathrm{vec}(\mathbf{M}),$$

其中 $\mathbf{M} = k_{**} \circ_1 \circ \cdots \circ_M + \left(\lambda \circ \lambda_1 \circ \cdots \circ \lambda_M \odot \cdot^{-\frac{1}{2}}\right)^{\cdot 2} \times_1 * \cdot^{-1} \cdot^{2} \times_2 \cdot^{2} \times_3 (\mathrm{diag}(\mathbf{S}) \cdots \times_{M+1} \cdot^{2}$

$\mathrm{diag}(\mathbf{S}))$ ( ) $\mathbf{A}(\mathbf{k}\mathbf{K}\mathbf{U})$ (U) (U) 。因此，我们也可以在 GAR 中高效地计算预测协方差矩阵 $\mathbf{S}$。

$$\mathrm{diag}(\mathbf{S}_*^h) = \mathrm{vec}(\mathbf{M}^l) + \mathrm{vec}(\mathbf{M}^r) \tag{A.61}$$

其中 vec($\mathbf{M}^l$) 和 vec($\mathbf{M}^r$) 分别是低保真度数据和残差数据对应的向量。计算 vec($\mathbf{M}^l$) 时，我们需要注意，输出核矩阵应为 $\mathbf{W}\mathbf{S}^l\mathbf{W}^T$。

E 实验详情

E.1 典型偏微分方程

我们考虑三种典型偏微分方程：泊松方程、热方程和伯格斯方程。这些偏微分方程在科学技术应用中发挥着关键作用 [62, 56, 66]。它们提供了常见的仿真场景，例如高维时空场输出、非线性和不连续性，并且经常被用作代理模型的基准问题 [12, 51–53]。 x 和 y 表示空间坐标， t 指定时间坐标，这与主论文中的记号相矛盾。附录中的这些记号仅用于使信息更清晰；它们与主论文无关，也不存在关联。

伯格斯方程被视为一种标准的非线性双曲型偏微分方程；它常用于描述各种物理现象，包括流体动力学 [56], 非线性声学 [57], 和交通流 [58]。由于它能够基于常规守恒方程产生不连续性（激波），因此被用作多种数值求解器和代理模型 [59–61] 的基准测试案例。该方程的黏性形式为

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = v\frac{\partial^2 u}{\partial x^2},$$

其中， u 表示体积， x 表示空间位置， t 表示时间， v 表示黏度。我们设定 $x \in [0, 1]$、 $t \in [0, 3]$ 和 u（x, 0） $= \sin(x\pi/2)$，并采用齐次狄利克雷边界条件。我们将黏度 $v \in [0.001, 0.1]$ 均匀采样作为输入参数，以生成解场。

在空间域和时间域中，该问题分别使用带帽函数的有限元法和后向欧拉法求解。对于第一个（保真度最低的）解，时空域被离散为 16×16 规则矩形网格。更高保真度的求解器会将网格每个维度上的节点数加倍，例如，第二种保真度为 32×32，第三种保真度为 64×64。结果场（即输出）使用 $128by128$ 规则时空网格计算。

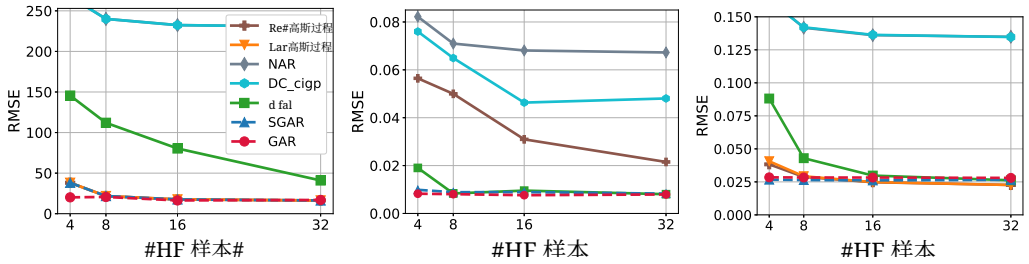
泊松方程 是机械工程和物理学中用于模拟势场（如引力场和静电场 [62]）的典型椭圆型偏微分方程。其表达式为

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0.$$

它是拉普拉斯方程 [63] 的推广。尽管形式简单，泊松方程在物理学中十分常见，也经常被用作代理模型的基础测试案例 [51, 64]。在我们的实验中，我们在二维空间域上施加狄利克雷边界条件，其中 $\mathbf{x} \in [0, 1] \times [0, 1]$。输入参数由矩形区域四条边界和中心的常数值组成，每个值的变化范围为 0.1 到 0.9。我们对输入参数进行均匀采样，以生成相应的势场作为输出。采用一阶中心差分格式和规则矩形网格，通过有限差分法求解该偏微分方程。对于最粗级别的解，我们使用了 8×8 网格。改进后的求解器采用更细的网格，每个维度的节点数是原来的两倍。最终的势场使用由 32×32 个单元组成的时空规则网格进行估计。

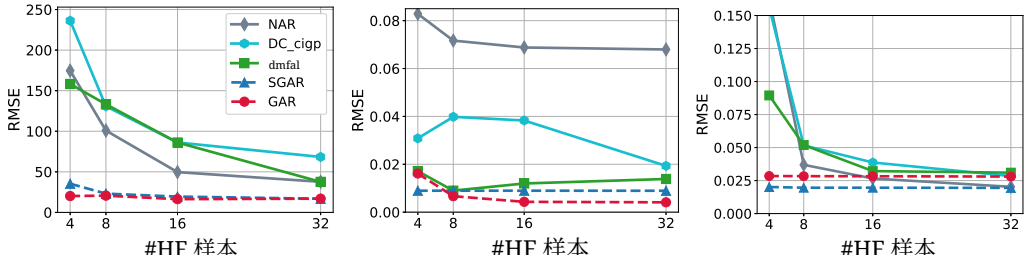
热方程是一种基础的偏微分方程，用于描述热通量随时间的演化。尽管热方程于 1822 年创立时仅用于描述热通量，但如今已广泛应用于概率论 [65, 66] 和金融数学 [67]等多个科学领域。因此，它常被用作替代模型。热方程如下：

$$\frac{\partial}{\partial x}\left(k\frac{\partial T}{\partial x}\right) + \frac{\partial}{\partial y}\left(k\frac{\partial T}{\partial y}\right) + \frac{\partial}{\partial z}\left(k\frac{\partial T}{\partial z}\right) + q_V = \rho c_p \frac{\partial T}{\partial t}$$



(a) 泊松 (b) 伯格斯 (c) 热方程

图 6：随着高保真训练样本数量增加，均方根误差的变化；训练样本通过 Sobol序列增加，且输出经过对齐（插值）。



(a) 泊松 (b) 伯格斯 (c) 热方程

图 7：随着高保真训练样本数量增加，均方根误差的变化；训练样本通过 Sobol序列增加，且输出未对齐。

其中， k 是材料导热率， q_V 是介质单位体积内能量的生成速率， ρ 是密度， c_p 是比热容。输入参数包括左侧边界在 $x = 0$ 处的通量率（范围为 0 至 1）、右侧边界在 $x = 1$ 处的通量率（范围为 -1 至 0），以及热导率（范围为 0.01 至 0.1）。

We 建立二维时空域 $x \in [0, 1], t \in [0, 5]$ ，并在x =处施加诺伊曼边界条件 0 and $x = 1$，以及 $u(x,\ 0) = H(x - 0.25) - H(x - 0.75)$，其中 $H(\cdot)$ 是赫维赛德阶跃函数。该方程在空间域中采用有限差分法求解，在时间域中采用后向欧拉法求解。时空域被离散为 16×16 规则矩形网格，用作第一个（最低）保真度求解器。精细化的求解器则使用 32×32 网格，实现第二种保真度。结果场在 100×100 时空网格上计算。

该方程在空间域中采用有限差分法求解，在时间域中采用反向欧拉法求解。时空域被离散为 8×8 规则矩形网格，用作第一个（精度最低）的解。改进求解器的网格所对应的第二种保真度为 32×32 网格。结果场在 100×100 时空网格上计算。

E.2 典型偏微分方程的多保真融合

在这些实验中，我们采用与第5.1节相同的实验设置，唯一的区别是使用 Sobol序列生成训练数据。我们生成了256个数据样本用于测试，32个样本用于训练。我们逐步将高保真度训练数据的数量从4增加到32，同时将高保真度训练数据固定为32。图6和图7展示了使用插值输出和原始未对齐输出时，对齐输出的均方根误差统计结果。与主论文中的情况一样，在高保真度训练数据稀缺时，GAR和CIGAR以明显优势胜过其他方法。同样，在处理非对齐输出时，GAR和CIGAR的优势更加显著：当高保真训练样本为4个和8个时，二者的均方根误差降低了5倍，大幅领先其他方法。

E.3 拓扑优化的多保真融合

我们在拓扑结构优化问题中使用 GAR：给定外力和角度等设计参数，输出材料布局（如合金和混凝土）的最优拓扑结构，即使刚度等力学指标最大化。拓扑结构优化是一种重要的方法

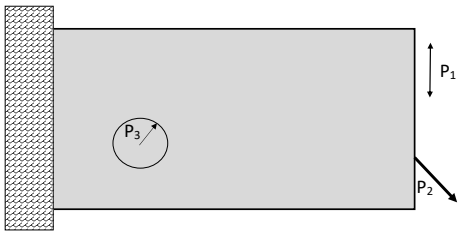
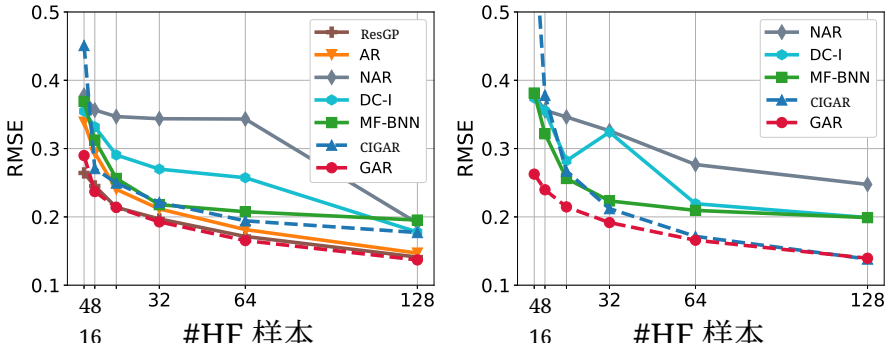


图8：悬臂梁的几何形状、边界条件和仿真参数

应用于翼型和板式桥梁等机械设计，尤其是在近期的 3D打印工艺中，材料以微量沉积。然而，众所周知，拓扑优化的计算开销很大，因为它涉及基于梯度的优化和力学特性的仿真。高保真度解需要庞大的离散网格，并带来巨大的时空计算开销，这使问题更加棘手。

利用数据驱动方法，在此过程中提供合适的结构 [68, 13]，这种做法日益普及。在这里，我们研究悬臂梁的拓扑优化（见附录）。我们采用快速实现 [69]，通过降低柔度 C 并满足体积限制 $V \leq \bar{V}$，来进行基于密度的拓扑优化。

我们使用 SIMP方法 [70]，将连续密度测量值转换为离散最优拓扑。我们将点载荷的位置 $P1$、点载荷的角度 $P2$和滤波半径 $P3$ [55] 设为系统输入。我们使用 40×80 规则网格求解低保真度问题，并使用 40×80 规则网格求解高保真度问题。本实验仅包含能够处理任意输出的技术。



（a）对齐输出　　（b）原始输出

图 9：使用 Sobol序列进行拓扑优化时，RMSE 随高保真训练样本数量增加的变化。

与早期实验一样，我们使用 Sobol序列生成 128 个测试样本和 64 个训练样本，以近似评估主动学习中的性能。结果如图 9 所示。可以看出，对于未经插值对齐的原始输出和对齐输出，所有可用方法的表现相似。尽管如此，GAR 始终以明显优势胜过其他方法。相比之下，CIGAR 在原始输出上的表现更好。

E.4 固体氧化物燃料电池的多保真融合

在此测试问题中，我们考虑一个稳态三维固体氧化物燃料电池模型。图 10 展示了其几何结构。该模型包含电子与离子电荷平衡（欧姆定律）、气体通道中的流量分布（纳维-斯托克斯方程）、多孔电极中的流动（Brinkman 方程），以及气体通道和多孔电极中的气相质量平衡（麦克斯韦-斯特凡扩散与对流）。假设阳极（$H_2 + O^{2-} \rightarrow H_2O + 2e^-$）和阴极（$O_2 + 4e^- \rightarrow 2O^{2-}$）中的反应遵循巴特勒-伏尔默电荷转移动力学。电池以恒电位仪方式运行（电池电压恒定）。我们使用采用有限元方法的 COMSOL Multiphysics[6]（应用程序 ID：514）求解该模型。

[6] https://www.comsol.com/model/current-density-distribution-in-a-solid-oxide-fuel-cell-514

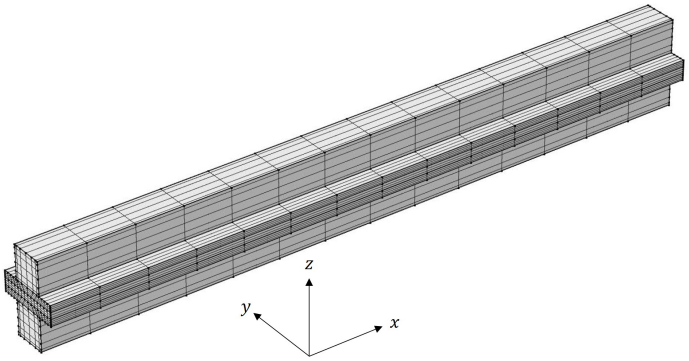


图10：在 SOFC 示例中，阴极位于计算域顶部，该计算域由气体通道、电极和电解质组成。从上至下，各层依次为通道、电极、电解质、电极和通道。通道的尺寸（x * y * z）为 1 cm * 0.5 mm * 0.5 mm，电极的尺寸为 1 cm * 1 mm * 0.1 mm，电解质的尺寸为 1 cm * 1 mm * 0.1 mm。阴极入口位于 $x = 1$ cm 处，阳极入口位于 $x = 0$cm 处。

假设的输入包括电极孔隙率 $\epsilon \in [0.4, 0.85]$、电池电压 $E_c \in [0.2, 0.85]$ V、温度 $T \in [973, 1273]$ K 和通道压力 $P \in [0.5, 2.5]$ atm。使用 Sobol序列在低保真度和高保真度仿真所指定的范围内选取 60 个输入。再从上述范围内随机选取 40 个高保真度测试点以完成测试。低保真度F1模型使用了 3164 个映射单元，相对容差为 0.1；高保真度模型则使用了 37064 个单元，相对容差为 0.001。此外，COMSOL 模型采用 V循环几何多重网格。感兴趣量为以通道为中心的 $x - z$ 平面内电解质电流密度（A m-2）和离子电势（V）的分布（图 10）。两种情况下均采集 $d = 100 \times 50 = 5000$ 个点，并将两种分布向量化，以生成训练和测试输出。

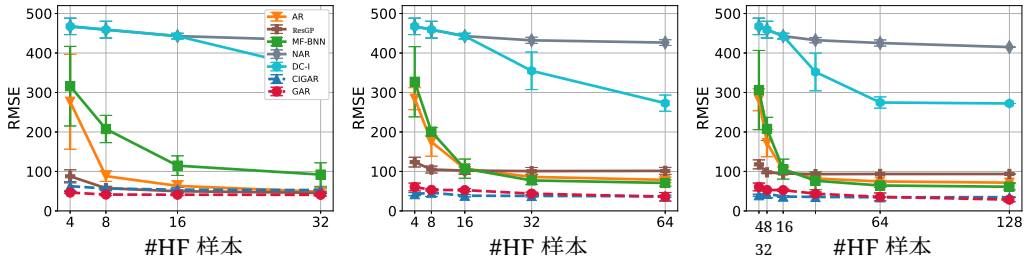


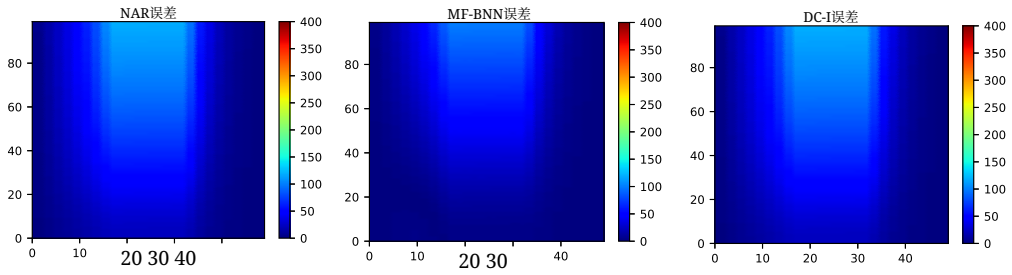
图11：对于固体氧化物燃料电池，当低保真度训练样本数固定为 {32,64,128} 时，RMSE 随高保真训练样本数增加的变化。

我们补充了经典实验，其中低保真训练样本数固定为 {32,64,128}，而高保真训练样本数则从 4 逐步增加至 {32,64,128}。通过插值对齐输出，实验结果如图11所示。可以看出，GAR 和 CIGAR 方法始终优于其他方法，尤其是在仅使用少量高保真度训练数据时。这与先前的实验结果一致。我们还可以看到，AR 的表现也不错，这表明这些数据并非高度非线性且复杂，因此相对容易求解。不过，AR 和 MF-BNN 都收敛到较高的误差，而 GAR 和 CIGAR 则收敛到较低的误差界。

为详细考察预测误差，我们将平均均方根误差场$\mathbf{Z}^{(\mathrm{AEF})}$ 定义为

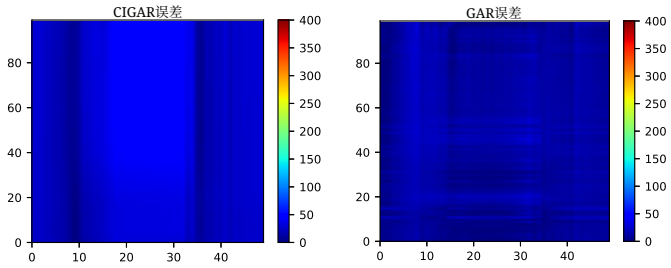
$$\mathbf{Z}^{(\mathrm{AEF})} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathbf{Z}_i - \tilde{\mathbf{Z}}_i)^2},$$

其中，$\tilde{\mathbf{Z}}_i$ 表示预测值，$\mathbf{Z}_i$ 表示真实值，平方根为逐元素运算。图 12 展示了 NAR、MF-BNN 和 DC 方法在 SOFC 数据中的 ECD 上的平均均方根误差场，使用了 32 个低保真度训练样本、16 个高保真训练样本和 128 个测试样本。为突出 GAR 和 CIGAR 的优势，图 13 展示了相同设置下的平均均方根误差场，但**仅使用 4 个高保真训练样本**。可以清楚地看到，即使只有 4 个高保真



（a）NAR　　（b）MF-BNN　　（c）DC

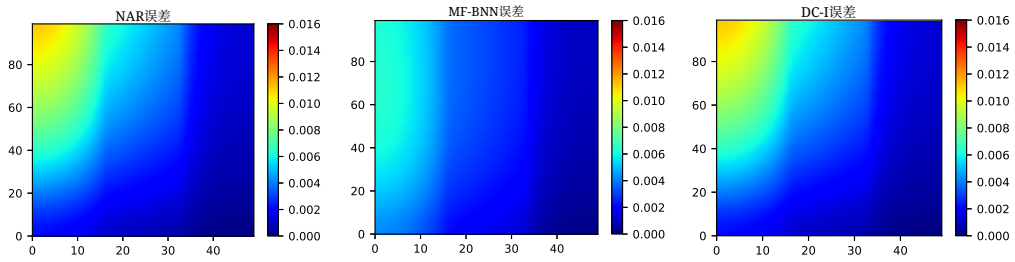
图 12：使用 32 个低保真度训练样本和 16 个高保真训练样本，对 128 个测试样本得到的 ECD 均方根误差场。



（a）CIGAR　　(b) GAR

图 13：使用 32 个低保真度和 **4** 个高保真训练样本，对 128 个测试样本的 ECD 均方根误差场。

与 NAR、MF-BNN 和 DC 相比，使用 16 个高保真训练样本时的训练样本。此外请注意，GAR 似乎存在一些棋盘格伪影，这可能是使用完整传递矩阵进行过度参数化所导致的。我们将留待后续工作解决这个问题。CIGAR 的棋盘格伪影较少，但均方根误差略有增加。



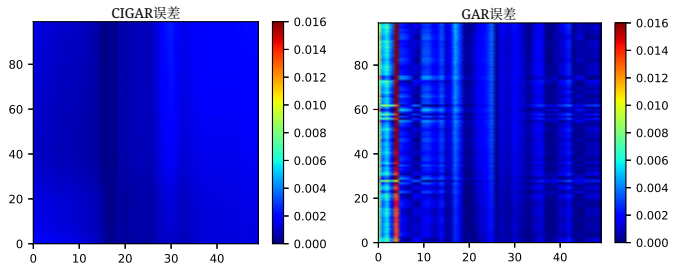
(a) NAR　　(b) MF-BNN　　(c) 直流

图 14：使用 32 个低保真度和 16 个高保真训练样本，对 128 个测试样本的离子电势均方根误差场。

在图 14 和图 15 中，与之前的实验设置类似，我们使用 SOFC数据集中的 IP 场，通过 128 个测试样本绘制平均均方根误差。NAR、MF-BNN 和 DC 使用 16 个高保真度样本进行训练，而 GAR 和 CIGAR 仅使用 4 个高保真度样本进行训练。可以看出，我们的方法明显优于其他方法。然而，在这种情况下，GAR 的棋盘格伪影更加严重，而 CIGAR 则成功减轻了此类伪影，同时保持较低的误差。

E.5 等离激元纳米粒子阵列仿真

在最后一个示例中，我们使用耦合偶极近似（CDA）方法，计算散射体数量各异的等离激元系统的消光效率和散射效率 Q_{ext} 和 Q_{sc}。CDA 是一种模拟尺寸



(a) CIGAR (b) GAR

图15：使用32个低保真度训练样本和**4** 个高保真训练样本时，128个测试样本的离子电势均方根误差场。

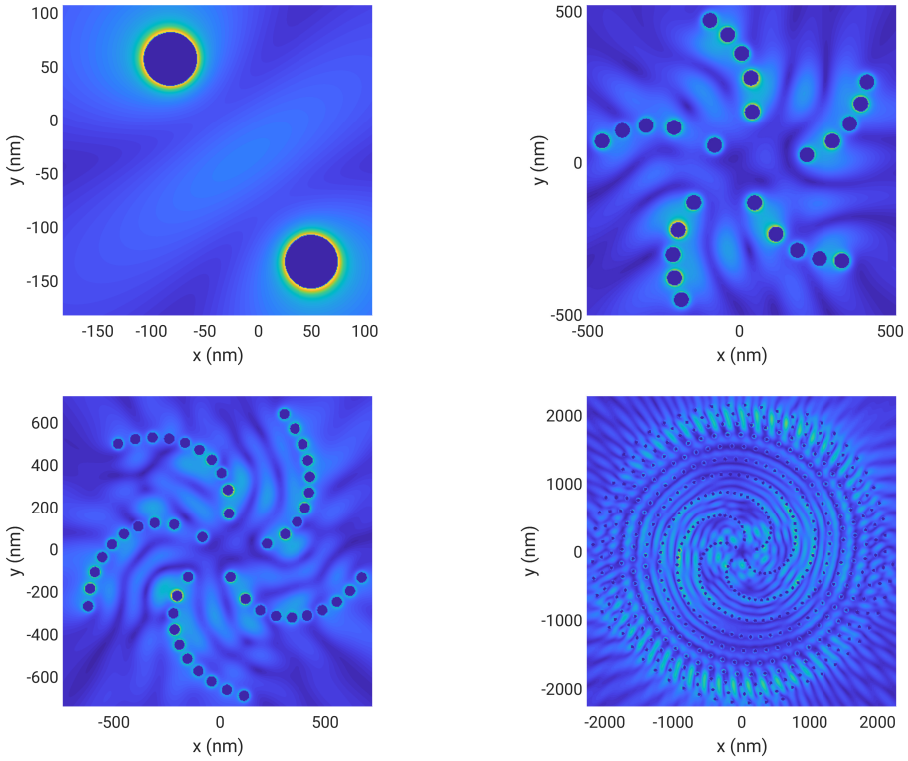


图16：包含 $\{2, 25, 50, 500\}$ 个粒子的沃格尔螺旋样本配置。

远小于光波长（此处为 25 nm）的相似非磁性金属纳米粒子阵列光学响应的方法。 Q_{ext} 和 Q_{sc} 被定义为本文档中的感兴趣量。我们使用所提出的方法，为效率构建最多包含三种保真度的代理模型。我们考察由沃格尔螺旋形成的粒子阵列。粒子发出的入射波之间的相互作用数量会影响磁场，因此，等离激元阵列中的纳米粒子数量会对其产生的局部消光场产生显著影响。图 16 展示了粒子数量属于集合 $\{2, 25, 50\}$、对应保真度 F1 至 F3 的沃格尔螺旋构型。 $\lambda \in [200, 800]$ nm、 $\alpha_{vs} \in [0, 2\pi]$ rad 和 $a_{vs} \in (1,\ 1500)$ 被确定为参数空间，分别对应入射波长、发散角和缩放因子。我们使用 Sobol序列选择输入。随着纳米粒子数量增加，执行 CDA 所需的计算时间呈指数增长。因此，所提出的采样方法显著降低了计算成本。

等离激元阵列对电磁辐射的响应可通过求解每个纳米球的局部电场 $\mathbf{E}_{loc}(\mathbf{r}_j)$ 来计算。考虑由相同体积定义的 N 金属粒子

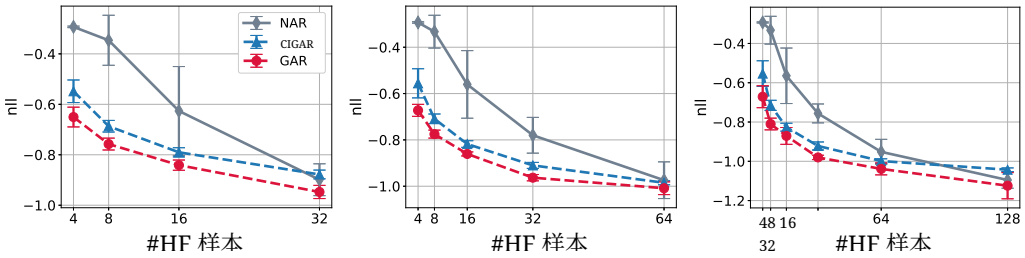
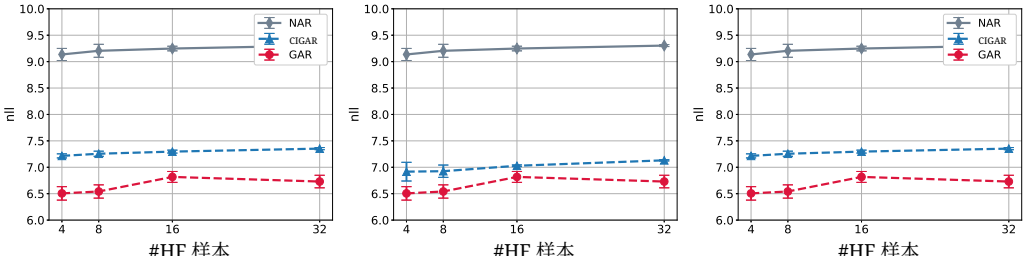


图 17：拓扑结构预测的负对数似然，其中低保真度训练样本数量固定为 {32,64,128}。



(a) ECD+离子电势　(b) ECD　(c) 离子电势

图 18：低保真度训练样本数固定为 32 时，固体氧化物燃料电池的负对数似然。

其极化率为 $\alpha(\omega)$，且位于向量坐标 $\mathbf{r}_i$ 处，通过求解 [71] 相应的线性方程，即可计算局部场 $\mathbf{E}_{loc}(\mathbf{r}_i)$。

$$\mathbf{E}_{loc}(\mathbf{r}_i) = \mathbf{E}_0(\mathbf{r}_i) - \frac{\alpha k^2}{\epsilon_0} \sum_{j=1, j\neq i}^{N} \tilde{\mathbf{G}}_{ij} \mathbf{E}_{loc}(\mathbf{r}_j) \tag{A.62}$$

其中，$\mathbf{E}_0(\mathbf{r}_i)$ 为入射场，k 为背景介质中的波数，ϵ_0 表示真空的介电常数（$\epsilon_0 = 1$ 在 CGS单位制中），而$\tilde{\mathbf{G}}^{ij}$ 由整体格林矩阵中与第 i个和第 j个粒子对应的 3×3 块构成。$\tilde{\mathbf{G}}^{ij}$ 在 $j = i$时为零矩阵，否则按如下方式计算：

$$\tilde{\mathbf{G}}_{ij} = \frac{\exp(ikr_{ij})}{r_{ij}} \left\{ \mathbf{I} - \hat{\mathbf{r}}_{ij}\hat{\mathbf{r}}_{ij}^T - \left[\frac{1}{ikr_{ij}} + \frac{1}{(kr_{ij})^2} (\mathbf{I} - 3\hat{\mathbf{r}}_{ij}\hat{\mathbf{r}}_{ij}^T) \right] \right\} \tag{A.63}$$

其中，$\hat{\mathbf{r}}^{ij}$ 表示从粒子 j 到 i 的单位位置向量，而 $r_{ij} = |\mathbf{r}_{ij}|$ 通过求解方程 A.62 和 A.63，可计算出总局部场$\mathbf{E}_{loc}(\mathbf{r}_i)$，进而计算出散射截面和消光截面。数值解的详细信息见 [72]。

Q_{ext} 和 Q_{sc} 是通过将散射截面和消光截面分别相对于阵列的总投影面积进行归一化得到的。我们考虑了沃格尔螺旋粒子阵列，其定义如下： [73]

$$\rho_n = \sqrt{n} a_{vs} \quad \text{and} \quad \theta_n = n\alpha_{vs}, \tag{A.64}$$

其中，ρ_n 和 θ_n 分别表示沃格尔螺旋阵列中第 n 个粒子的径向距离和极角。因此，沃格尔螺旋构型可由入射波长 λ、发散角 α_{vs}、缩放因子 a_{vs}以及粒子数 n唯一确定。

E.6 预测不确定性的度量指标

尽管均方根误差已被用作评估多保真度融合算法 [9, 12, 13, 16], 性能的标准指标，但考虑预测不确定性的指标也同样重要 [47], ，尤其是在下游应用高度依赖预测置信度的质量时，例如多保真度贝叶斯优化 [23]。为了更全面地评估所提出的方法，我们使用最常用的指标——负对数似然（nll）来评估预测后验的质量。

我们使用完全相同的实验设置，但采用负对数似然（nll）度量，重新绘制了图 2 和图 5，结果见图 17 和图 18。请注意，直流和 MF-BNN 的 nll 表现很差，可能是由于我们的实现所致，无法纳入图中，因此未予展示。另请注意，有些图中显示的 nll 为负值，这是因为我们在计算 nll 时省略了常数项。不过，

这种修改不会影响比较。可以看到，对于图 17 中的拓扑结构后验，结果与根据均方根误差得出的结论一致。由于 CIGAR 忽略了输出间相关性，因此会高估协方差行列式，导致其 nll 高于 GAR。训练数据较少时，NAR 的表现较差；随着训练数据增加，其表现持续改善，最终与 GAR 和 CIGAR 相近。同样，固体氧化物燃料电池的结果也与根据均方根误差得出的结论一致。然而，所展示的所有方法在增加训练数据后，性能都没有显著提升。这是由 nll 的计算方式以及数据本身造成的。具体而言，在 ECD 和离子电势场中，有少数空间位置记录的值几乎恒定（这是狄利克雷边界条件导致的）。在这种情况下，nll 会受到方差对数的主导，因此对于评估预测方差的质量而言，信息量较低。因此我们看到，无论使用多少训练点，图 18 中的 nll 都在相同数值附近波动。考虑到本工作的范围，我们将来会进一步研究采用更先进的度量（例如 [74]）来衡量不确定性。