arXiv:2601.12515v1 [stat.CO] 18 Jan 2026

具有完全分布依赖的部分观测麦基恩–弗拉索夫随机微分方程的贝叶斯推断

Ning Ning[a,*] 和 Amin Wu[b,*]

[a]统计系，德克萨斯农工大学，美国德克萨斯州大学城。[b]统计学项目，计算机、电气与数学科学及工程学部，阿卜杜拉国王科技大学，图瓦勒，23955-6900，沙特阿拉伯。电子邮箱：*patning@tamu.edu*，*amin.wu@kaust.edu.sa*

摘要

麦基恩–弗拉索夫随机微分方程（MVSDEs）描述的系统，其动力学既取决于个体状态，也取决于神经元群体分布，广泛见于神经科学、金融和流行病学领域。在许多应用中，系统只能被部分观测；当漂移和扩散系数均依赖于不断变化的经验分布时，推断便极具挑战性。本文为此类部分观测的麦基恩–弗拉索夫随机微分方程开发了一种贝叶斯框架，用于潜在状态推断和参数估计。我们将时间离散化与基于粒子的近似相结合，构造易于处理的似然估计量，并设计了两种粒子马尔可夫链蒙特卡罗（PMCMC）算法：单层 PMCMC 方法，以及一种在不同离散化层级间耦合粒子系统的多层 PMCMC（MLPMCMC）方法。多层构造产生相关的似然估计值，并以计算成本（$O(\varepsilon^{-6})$）达到均方误差（$O(\varepsilon^{2})$），优于单层方案的（$O(\varepsilon^{-7})$）复杂度。我们研究了完全依赖分布的扩散情形，这是 MVSDEs 最一般的表述，并在标准正则性假设下给出了理论保证。数值实验验证了所提方法论的效率和准确性。

关键词：贝叶斯估计、广义麦基恩–弗拉索夫随机微分方程、马尔可夫链蒙特卡洛、层级蒙特卡洛。

1 引言

我们首先在第1.1节介绍研究背景与动机，随后在第1.2节概述主要贡献，最后介绍论文的整体结构

*作者按字母顺序排列。

第 1.3节。

1.1 背景与动机

麦基恩–弗拉索夫随机微分方程（麦基恩–弗拉索夫随机微分方程）为描述相互作用的粒子或智能体系统提供了一个基本框架，其动力学既取决于各自的状态，也取决于整体神经元群体的概率分布。这类方程自然出现在多个领域，包括平均场博弈、统计物理学、神经科学和量化金融；在这些领域中，集体效应对个体行为起着至关重要的作用（Baldaron 等人 2012、Mogilner 与 Edelstein- Keshet 1999、Giesecke 等人 2020）。尽管表现力强，麦基恩–弗拉索夫随机微分方程的模拟仍面临重大挑战。对过程概率分布的依赖使所有粒子之间产生内在耦合，导致计算复杂度高，并在基于粒子的近似中引入混沌传播误差（Sznitman 1991、Ning 与 Wu 2026）。此外，对经验测度的非线性依赖往往使闭式解难以获得，并加剧数值不稳定性，尤其是在系数非全局利普希茨的情况下（Ning 等人 2024）。因此，如何准确且高效地模拟麦基恩–弗拉索夫随机微分方程，仍是一个活跃且技术要求很高的研究领域。

尽管在完全观测情形下，对麦基恩–弗拉索夫随机微分方程的分析与模拟已取得显著进展（Sharrock 等人 2023、Wen 等人 2016），但许多真实世界系统只能被部分观测。在这类情形中，隐状态随时间连续演化，但我们只能获得离散且含噪的测量结果，因而无法完整、直接地了解其底层动力学（Heng 等人 2024）。这种部分观测结构自然存在于神经科学、金融和流行病学等领域，因为系统的完整状态无法直接测量，或获取成本过高（Li 等人 2024、Gu 等人 2025）。在部分信息条件下对麦基恩–弗拉索夫随机微分方程进行建模会带来重大挑战：必须通过观测推断动力学对神经元群体分布的依赖关系，这使得参数估计成为概率测度空间上的非线性滤波问题。如果扩散项还依赖于过程的概率分布，这些困难便会进一步加剧，因为推断必须同时跟踪随机性和不断演变的分布相互作用，从而使理论分析和数值计算都更加复杂。

尽管存在这些复杂性，在扩散系数中纳入对概率分布的依赖，对于准确表征波动率本身具有平均场效应的系统至关重要。例如，在神经科学中，大型

神经元群体中的突触噪声强度取决于集体活动：总体放电模式会调节膜电导和背景波动（Dos Reis 等（2022））。因此，与概率分布无关的扩散模型会忽略噪声诱导同步和随机共振背后的群体层面波动现象。在金融领域，系统性风险建模中也会出现类似效应：机构的特质波动率取决于资本或风险敞口横截面分布的变化。例如，Giesecke 等人(2020）为漂移和扩散系数依赖于经验分布的跳扩散大型相互作用系统构建了渐近推断框架，从而刻画相关违约、银行间依赖以及其他分布驱动的波动机制。在这些应用中，依赖概率分布的扩散对于描述共同变动、波动率聚集和分布介导的相互作用不可或缺，而这些现象无法通过更简单的设定重现。

然而，纳入依赖分布的扩散后，统计和计算负担都会大幅加重。由于潜在轨迹、未知参数和不断演化的概率测度相互交织，后验分布变得极难处理。经典推断方法已不足以应对：无法获得闭式转移密度（Beskos 等（2006）），而数值方案必须处理概率测度空间上的高维积分。这些挑战引出了两个自然的问题：

- 能否为这类一般模型开发一种具有广泛适用性、由数据驱动的方法论?
- 一旦建立了这样的方法，能否通过降低估计变异性进一步提升其性能?

解决这些问题是本文的核心目标。

1.2 我们的贡献

本文针对具有依赖分布的漂移和扩散系数的部分观测均值场随机微分方程（PO-MVSDEs），构建了一套通用的贝叶斯方法，用于潜在状态推断和参数估计。我们首先将适当的时间离散化方案与基于粒子的近似相结合（Awadelkarim& Jasra 2024），从而为动力学的两个组成部分都依赖于不断演变的经验分布的模型，得到易于处理的似然估计量。在此基础上，我们设计了适用于该场景后验采样的马尔可夫链蒙特卡洛（MCMC）算法（Andrieu 等（2010））。由于精细的离散化层级会带来高昂的计算成本，我们进一步引入

多层蒙特卡洛（MLMC）技术（Giles 2008、2015、Jasra 等人 2018），以平衡不同分辨率下的方差和偏差。与经典的单层方案相比，多层构造显著提升了效率；相应的理论分析则为近似误差和整体复杂度提供了严格保证。

我们针对离散化的部分观测均值场随机微分方程，开发了两种用于贝叶斯推断的粒子马尔可夫链蒙特卡洛（PMCMC）方法。粒子滤波器（又称序贯蒙特卡洛方法）是非线性随机滤波中的重要工具（Doucet 等人 2001，Chopin 等人 2020）。首先，我们基于近似转移律构造单层 PMCMC 算法，该转移律由粒子滤波器构建 $\mu^{l}_{k\Delta_l,\ \theta}$。粒子滤波器提供无偏的边际似然估计值，并将其嵌入伪边际马尔可夫链蒙特卡洛方案中，从而对离散化模型实现精确推断。其次，为了提高效率，我们提出多层 PMCMC（MLPMCMC）算法，对两个相邻层级的近似概率律进行耦合 $(l-1,\ l)$。这种耦合使得能够通过耦合粒子滤波器构造相关的似然估计值，再通过加权差或加权平均将其汇总，形成一种随层级变化而方差递减的望远镜估计量。在整个层级体系中，粒子数N_l 和参数 θ 在每个层级内保持固定，而层间耦合则旨在最小化多层分解的方差。这些方法将Andrieu 等人 (2010)的基础性 PMCMC 结果推广到含有依赖分布噪声的均值场系统；我们还证明，即使只能获得有偏的概率律近似，精确性仍得以保留。

我们的方法论推广了Jasra 与 Wu (2025) 的框架。该框架研究的是具有依赖分布的漂移和仅依赖状态的扩散的部分观测 McKean–Vlasov 模型。在我们的框架中，他们的设定是以下特殊情形 $\zeta_{2,\ \theta} \equiv 0$。允许 $\zeta_{2,\ \theta} \neq 0$ 会带来诸多新的挑战：必须同时近似两个相互作用核$\zeta_{1,\ \theta}$（漂移）和 $\zeta_{2,\ \theta}$（扩散）；依赖分布的扩散会影响动力学的随机部分，使耦合过程的构造和方差分析都更加复杂；此外，正则性条件必须在整个测度空间 $\mathcal{P}(\mathbb{R}^d)$ 上一致成立，而不能仅针对状态变量。尽管面临这些挑战，我们证明，多层方差缩减仍然有效；对于均方误差 $O(\varepsilon^2)$，其计算成本为 $O(\varepsilon^{-6})$，这与 Jasra 与 Wu (2025) 中的最优复杂度相匹配，同时适用于严格更广泛的模型类别。

证明通过将均方误差分解为若干部分来展开，这些部分分别对应不同的近似误差来源。具体而言，我们区分基准层 MCMC 采样器产生的误差、额外误差

由双层 MCMC 更新引入的误差、基础离散化层级上的粒子近似误差、相邻层级间的耦合粒子近似误差，以及时间离散化导致的偏差。主要技术难点源于依赖分布的扩散项，它会影响与两个相互作用核相关的方差界，$\zeta_{1,\theta}$ 和$\zeta_{2,\theta}$。为解决这一问题，我们针对耦合系统推导了新的方差估计，并证明在假设 A1所规定的正则性条件下，这些误差项均可得到充分控制，从而确保整体算法达到最优的多层收敛速率。

1.3 论文结构

本文其余部分的安排如下。第2 节介绍整体模型表述，首先介绍 PO-MVSDE 框架，随后给出其时间离散化版本。第3 节针对离散化模型下的近似推断，构建计算方法论，并提出两种用于对 PO-MVSDE 进行后验采样的 PMCMC 方法。第4 节给出所提方法论的理论结果，包括必要的正则性假设和主要收敛结果。第5 节将所提方法应用于一个由具有非全局利普希茨结构的多维 MVSDE 所支配的神经元活动模型，并给出相应的数值结果。最后，第6 节总结主要发现，讨论更广泛的影响，并提出未来研究方向。所有理论结果的证明均见补充材料。

2 模型表述

本节介绍整体模型表述，首先在第2.1节介绍 PO-MVSDE 框架，随后在第2.2节介绍其时间离散化设置。

2.1 部分观测的 McKean–Vlasov随机微分方程

考虑一个完备的概率空间 $(\Omega\ \mathscr{F}, \mathbb{F}=\{\mathscr{F}_t\}^{t\in[0,T]}, P)$，该空间支持一个$\mathbb{F}$-适应的标准布朗运动 $W=\{W_t\}^{t\in[0,T]}$。我们考虑一个麦基恩 – 弗拉索夫型平均场随机微分方程。令 $X_0=x_0\in\mathbb{R}^d$ 为初始条件，$\theta\in\Theta\subseteq\mathbb{R}^{d_\theta}$

为一个固定参数。过程$\{X_t\}_{t\geq 0}$ 满足以下随机微分方程:

$$dX_t = a_\theta\left(X_t, \overline{\zeta}_{1,\theta}(X_t, \mu_{t,\theta})\right)dt + b_\theta\left(X_t, \overline{\zeta}_{2,\theta}(X_t, \mu_{t,\theta})\right)dW_t, \tag{2.1}$$

其中相互作用项定义为

$$\overline{\zeta}_{j,\theta}(X_t, \mu_{t,\theta}) = \int_{\mathbb{R}^d} \zeta_{j,\theta}(X_t, x)\mu_{t,\theta}(dx), \qquad j \in \{1, 2\}.$$

其中，$\mu_{t,\,\theta}$ 表示参数为 θ 且初始条件为 $\mu_{0,\,\theta}(dx) = \delta_{x_0}(dx)$时 X_t 的概率分布，其中 $\delta_{x_0}(dx)$是集中于固定起始点 x_0的狄拉克测度。对于每个 $\theta \in \Theta$，函数 $\zeta_{j,\,\theta}$: $\mathbb{R}^{2d} \to \mathbb{R}$ 表示相互作用核，a_θ: $\mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$ 是漂移系数函数，而 b_θ: $\mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^{d\times d}$ 是扩散系数函数。

我们考虑离散时间观测过程 Y_1，Y_2，……，Y_T，其取值于紧集 $\mathbf{Y} \subseteq \mathbb{R}^{d_Y}$，其中 $T \in \mathbb{N}$。为方便记号起见，假设观测以单位时间间隔收集。在给定时刻 $t \in \mathbb{N}$ 的位置 X_t 条件下，假设随机变量 Y_t 与所有其他随机变量相互独立，并具有有界且为正的概率密度函数 $g_\theta(x_t,\ y_t)$。令$P_{\mu_{t-1,\,\theta},\,t,\,\theta}(x_{t-1},\ dx_t)$表示给定 $\mathscr{F}_{t-1}$ 时 X_t 的条件分布，即单位时间内的转移核。令 ν 为参数空间 Θ 上关于勒贝格测度的概率密度。令 $d\theta$ 表示 $(\Theta,\ \mathscr{B}(\Theta))$ 上的勒贝格测度，其中 $\mathscr{B}(\Theta)$ 是 Θ 上的博雷尔 σ-域。我们的目标是从后验分布中进行采样

$$\pi\left(d(\theta, x_{1:T}) \mid y_{1:T}\right) = \frac{L(\theta, x_{1:T}; y_{1:T})\nu(\theta)d\theta}{\int_{\Theta\times\mathbb{R}^{Td}} L(\theta, x_{1:T}; y_{1:T})\nu(\theta)d\theta}, \tag{2.2}$$

似然函数为

$$L(\theta, x_{1:T}; y_{1:T}) = \prod_{k=1}^{T} g_\theta(x_k, y_k)P_{\mu_{k-1,\theta},k,\theta}(x_{k-1}, dx_k).$$

我们假设　在全文中，式（2.2）中的归一化常数（分母）是有限的。

2.2 时间离散化

由于（2.1）中的连续时间 formulation 不便于直接进行计算实现，我们采用标准的时间离散化方案。所得的离散时间近似还需要进一步的数值近似方法，我们将在第 3 节对此进行讨论。

给定 $l \in \mathbb{N}_0 = \mathbb{N}_+ \cup \{0\}$，定义时间步长为 $\Delta_l = 2^{-l}$。我们采用一阶 Euler-Maruyama格式对随机微分方程（2.1）进行离散化。对于$k \in \{0、1、\ldots\ldots、\Delta_l^{-1}T - 1\}$ 以及初始条件 $\widetilde{X}_0 = x_0$，离散过程满足：

$$\widetilde{X}_{(k+1)\Delta_l} = \widetilde{X}_{k\Delta_l} + a_\theta\left(\widetilde{X}_{k\Delta_l}, \overline{\zeta}_{1,\theta}(\widetilde{X}_{k\Delta_l}, \mu^l_{k\Delta_l,\theta})\right)\Delta_l + b_\theta\left(\widetilde{X}_{k\Delta_l}, \overline{\zeta}_{2,\theta}(\widetilde{X}_{k\Delta_l}, \mu^l_{k\Delta_l,\theta})\right)\Delta W_k, \tag{2.3}$$

其中，$\Delta W_k = W_{(k+1)\Delta_l} - W_{k\Delta_l}$ 表示时间区间 $[k\Delta_l,(k+1)\Delta_l]$上的布朗增量，而 $\mu^l_{k\Delta_l,\,\theta}$ 表示时间离散过程 $X_{k\Delta_l}$在时刻 $k\Delta_l$的概率分布。令$P^l_{\mu^l_{t-1,\,\theta},\,t,\,\theta}(x_{t-1},\ dx_t)$ 表示离散过程（2.3）中 X_t 的条件分布，它代表单位时间上的时间离散转移核。这对应于前文引入的连续时间转移核$P_{\mu_{t-1,\,\theta},\,t,\,\theta}(x_{t-1},\ dx_t)$的离散时间类比。

我们的计算目标转变为从离散化的后验分布中采样

$$\pi^l\left(d(\theta, x_{1:T}) \mid y_{1:T}\right) = \frac{L^l(\theta, x_{1:T}; y_{1:T})\nu(\theta)d\theta}{\int_{\Theta\times\mathbb{R}^{Td}} L^l(\theta, x_{1:T}; y_{1:T})\nu(\theta)d\theta}, \tag{2.4}$$

其中，离散似然函数定义为

$$L^l(\theta, x_{1:T}; y_{1:T}) = \prod_{k=1}^{T} g_\theta(x_k, y_k) P^l_{\mu^l_{k-1,\theta},k,\theta}(x_{k-1}, dx_k).$$

离散化公式（2.4）近似连续时间后验分布（2.2），且随着离散化参数 l 增大（等价地，随着 $\Delta_l \to 0$增大），近似误差会减小。然而，即使采用 Euler-Maruyama 时间离散化，由于均场依赖，计算转移核$P^l_{\mu^l_{k-1,\,\theta},\,k,\,\theta}(x_{k-1},\ dx_k)$仍然难以实现。因此，我们需要进一步采用基于模拟的近似方法，并将在下一节中介绍这些方法。

3 方法论

本节将针对前文介绍的离散化模型，建立近似推断所需的计算方法。我们提出两种用于后验采样的 PMCMC 方法：第3.1 节介绍单层 PMCMC 算法，第3.2节介绍其多层扩展。

3.1 单层 PMCMC 算法

我们使用$\mu^l_{t,\theta}$分别对每个 $t \in \{1, \ldots, T\}$ 采用Sznitman (1991)提出的基于粒子的方法进行近似，并固定参数 $\theta \in \Theta$。概率律近似过程对应于算法A.1的第 2(b) 步。对于每个时间区间 $[t-1, t]$，我们从前一时间步的经验分布中，或在 $t = 1$时从初始条件中，抽取 N 个粒子进行初始化$\{X^i_{t-1}\}_{i=1}^N$。随后，粒子在离散网格$\{(t-1) + j\Delta_l\}_{j=0}^{\Delta_l^{-1}}$上演化，并通过经验平均近似相互作用项。由此得到的经验测度$\{\mu^{l,N}_{t-1+j\Delta_l,\theta}\}_{j=1}^{\Delta_l^{-1}}$可近似真实的 McKean–Vlasov 定律，以供后续粒子滤波使用。

我们基于算法 A.1生成的基于粒子的概率律估计，构造（2.4）中的时间离散化后验分布。该后验分布作为马尔可夫链蒙特卡罗采样的目标分布。对于每个 $t \in \{1, \ldots, T\}$，单位时间间隔上的转移核为

$$P^l_{\mu^l_{t-1,\theta},t,\theta}(x_{t-1}, dx_t) = \int_{\mathbb{R}^{d(\Delta_l^{-1}-1)}} \prod_{k=1}^{\Delta_l^{-1}} Q^l_{\mu^l_{t-1+(k-1)\Delta_l,\theta},\theta}(x_{t-1+(k-1)\Delta_l}, dx_{t-1+k\Delta_l}),$$

其中，$Q^l_{\mu^l_{t-1+(k-1)\Delta_l,\theta},\theta}(x_{t-1+(k-1)\Delta_l}, dx_{t-1+k\Delta_l})$ 表示离散化随机微分方程（2.3）在 Δ_l 时间步上、定义于 $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ 上的高斯马尔可夫核，其概率分布为 $\mu^l_{t-1+(k-1)\Delta_l,\theta} \in \mathcal{P}(\mathbb{R}^d)$。

令 $u_t = (x_{t-1+\Delta_l}, \ldots, x_t)$ 表示时间区间$[t-1, t]$ 内、针对 $t \in \{1, \ldots, T\}$ 的中间状态。在离散化层级 l 下，我们引入以下记号：定义 $E_l := (\mathbb{R}^d)^{\Delta_l^{-1}}$ 为单位时间间隔上中间状态的离散化路径的状态空间，使得 $u_t \in E_l$；对于 N 个粒子，定义$E_l^N := (E_l)^N$ 为 N 粒子轨迹的状态空间；对于完整时间范围，定义 $E_l^T := (E_l)^T$。随后我们定义扩展转移核

$$\overline{P}^l_{\mu^l_{t-1,\theta},t,\theta}(x_{t-1}, du_t) := \prod_{k=1}^{\Delta_l^{-1}} Q^l_{\mu^l_{t-1+(k-1)\Delta_l,\theta},\theta}(x_{t-1+(k-1)\Delta_l}, dx_{t-1+k\Delta_l}).$$

$(\Theta \times \mathsf{E}_l^T, \mathscr{B}(\Theta \times \mathsf{E}_l^T))$ 上的目标后验分布为

$$\overline{\pi}^l\left(d(\theta, u_1, \ldots, u_T)\right) := \frac{\left\{\prod_{k=1}^T g_\theta(x_k, y_k) \overline{P}^l_{\mu^l_{k-1,\theta},k,\theta}(x_{k-1}, du_k)\right\} \nu(d\theta)}{\int_{\Theta \times \mathsf{E}_l^T} \left\{\prod_{k=1}^T g_\theta(x_k, y_k) \overline{P}^l_{\mu^l_{k-1,\theta},k,\theta}(x_{k-1}, du_k)\right\} \nu(d\theta)}. \tag{3.1}$$

该测度以（2.4）为边缘分布。然而，由于精确概率律 $\mu^l_{k,\theta}$ 未知且必须进行近似，直接采样难以实现。

使用步骤 2(b) 中的基于粒子的概率律近似过程，我们用基于 N 个粒子的经验近似 $\mu^{l,t,\theta}$ 替代真实概率律。令 $u_t = u_t^{1:N} \in \mathsf{E}_l^N$ 表示在时间 t 由概率律近似过程生成的所有模拟中间状态 $\{(X_{t-1+\Delta_l}^1, \ldots, X_{t-1+\Delta_l}^N) \ldots, (X_t^1, \ldots, X_t^N)\}$。完整轨迹序列 $(u_1, \ldots, u_T)$ 的联合分布记为 $\mathbb{P}_\theta^N$，相应的期望算子记为 $\mathbb{E}_\theta^N$。粒子近似转移核为

$$\overline{P}^l_{\mu^{l,N}_{t-1,\theta},t,\theta}(x_{t-1}, du_t) = \prod_{k=1}^{\Delta_l^{-1}} Q^l_{\mu^{l,N}_{t-1+(k-1)\Delta_l,\theta},\theta}(x_{t-1+(k-1)\Delta_l}, dx_{t-1+k\Delta_l}).$$

我们的马尔可夫链蒙特卡罗方法针对近似后验分布

$$\begin{aligned}&\overline{\pi}^{l,N}\left(d(\theta, u_1, \ldots, u_T)\right) \\ &= \frac{\left\{\prod_{k=1}^T g_\theta(x_k, y_k)\right\} \overline{\mathbb{E}}_\theta^N \left[\prod_{k=1}^T \overline{P}^l_{\mu^{l,N}_{k-1,\theta},k,\theta}(x_{k-1}, du_k)\right] \nu(\theta) d\theta}{\int_{\Theta \times \mathsf{E}_l^T} \left\{\prod_{k=1}^T g_\theta(x_k, y_k)\right\} \overline{\mathbb{E}}_\theta^N \left[\prod_{k=1}^T \overline{P}^l_{\mu^{l,N}_{k-1,\theta},k,\theta}(x_{k-1}, du_k)\right] \nu(\theta) d\theta}.\end{aligned} \tag{3.2}$$

3.1.1 单层粒子马尔可夫链蒙特卡罗采样算法

我们采用单层 PMCMC 方法从 π^l 中进行采样 N。算法 A.1 在每次马尔可夫链蒙特卡洛迭代中依次执行四个关键组件。完整的算法细节见补充材料的第 S.1 节。

1. 参数提议（步骤 2(a)）：从提议核 $r(\theta^{(k-1)}, \theta')$ 中生成候选参数 θ'。**2. 转移律近似**（步骤 2(b)）：对于提议参数 θ'，通过利用离散化动力学和经验相互作用项演化 N 个粒子，构造 McKean-Vlasov 概率分布的基于粒子的近似 $\{\mu^{l,N}_{t-1+j\Delta_l,\theta'}\}_{j=1}^{\Delta_l^{-1}}$。**3. 粒子滤波器似然估计**（步骤 2(c)）：在给定近似概率律的条件下，运行一个包含 M 个粒子的粒子滤波器，以估计边缘似然 $\hat{p}^{l,M,N}_{\theta'}(y_{1:T})$ 并生成轨迹样本。

4.Metropolis-Hastings 步骤（步骤 2(d)）：根据接受概率接受或拒绝所提议的参数

$$\alpha = \min\left\{1, \frac{\hat{p}^{l,M,N}_{\theta'}(y_{1:T})\nu(\theta')r(\theta', \theta^{(k-1)})}{\hat{p}^{l,M,N}_{\theta^{(k-1)}}(y_{1:T})\nu(\theta^{(k-1)})r(\theta^{(k-1)}, \theta')}\right\}. \tag{3.3}$$

该接受-拒绝过程确保马尔可夫链 $\{\theta^{(k)}, \xi_{1\,T:}^{(k)}\}_{k=0}^{K}$ 以 $_{\pi^{l,N}}$ （见（3.2））作为其平稳分布。该算法是对Andrieu 等人（2010）提出的原始 PMCMC 框架的改编。在该文确立的温和正则性条件下，所得到的马尔可夫链会生成来自近似后验分布 $\pi^{l,N}$ 的样本。

计算复杂度： 每次马尔可夫链蒙特卡洛迭代的计算复杂度为 $\mathcal{O}(\Delta_l^{-1}N^2T)$ $+\mathcal{O}(\Delta_l^{-1}NMT)$。第一项来自转移律近似，其中需要在每个离散化步长上计算经验平均时进行粒子相互作用。第二项对应于粒子滤波器操作，该操作涉及通过由 N 律近似粒子导出的转移核传播 M 滤波粒子。

实现细节： 初始化通过从先验分布$\nu(\theta)$中抽样 $\theta^{(0)}$ 并执行粒子滤波器以获得初始似然估计来完成。随后，对于 $k = 1, \ldots, K$，迭代应用 Metropolis-Hastings 步骤，以生成样本链 $\{\theta^{(k)}, \xi_{1\,T:}^{(k)}\}_{k=0}^{K}$。

对于任意可测函数 $\varphi:\ \Theta \times \mathbb{R}^{Td} \to \mathbb{R}$，我们将期望

$$\int_{\Theta \times \mathsf{E}_l^T} \varphi(\theta, x_{1:T}) \overline{\pi}^{l,N}\left(d(\theta, u_1, \ldots, u_T)\right)$$

用经验平均值进行近似

$$\overline{\pi}^{l,N,K}(\varphi) := \frac{1}{K+1}\sum_{k=0}^{K} \varphi(\theta^{(k)}, x_{1:T}^{(k)}). \tag{3.4}$$

该估计量为 $\pi^{l,N}$ 下的期望提供相合近似，其中$K \to \infty$，从而能够对 MVSDE 参数进行蒙特卡洛推断。

3.2 多层伪边际马尔可夫链蒙特卡罗算法

3.2.1 双层望远镜差

为了实现多层蒙特卡洛估计，我们需要近似双层望远镜差

$$\int_{\Theta \times \mathsf{E}_l^T} \varphi(\theta, x_{1:T}) \overline{\pi}^{l,N_l}\left(d(\theta, u_1, \ldots, u_T)\right) - \int_{\Theta \times \mathsf{E}_{l-1}^T} \varphi(\theta, x_{1:T}) \overline{\pi}^{l-1,N_l}\left(d(\theta, u_1, \ldots, u_T)\right), \tag{3.5}$$

其中 $l \in \mathbb{N}$ 为固定值，$\varphi:\ \Theta \times \mathbb{R}^{Td} \to \mathbb{R}$ 为测试函数，$N_l \in \mathbb{N}$ 表示用于近似概率分布的粒子数。有效的 MLMC 估计

需要从 π^{l,N_l} 与 π^{l-1,N_l}的耦合分布中采样。这里， π^{l,N_l} 的定义方式与式 (3.2) 中的单层后验分布类似，使用 N_l 个粒子来近似离散化层级 l上的概率分布。同样，π^{l-1,N_l} 对应于离散化层级 $l-1$上的后验分布，但关键在于，它使用相同的粒子数N_l 来近似概率分布。我们的双层马尔可夫链蒙特卡洛算法详见算法 A.2 的步骤 3，它在 $(\Theta \times \mathsf{E}_l^T \times \mathsf{E}_{l-1}^T, \mathscr{B}(\Theta \times \mathsf{E}_l^T \times \mathsf{E}_{l-1}^T))$ 上构造联合概率测度，在保持边缘分布正确的同时，引入不同离散化层级间的相关性。该构造利用测度变换技术，使得能够通过马尔可夫链蒙特卡洛近似式 (3.5)，并将多层蒙特卡罗框架扩展到具有依赖于概率律的扩散系数的 McKean–Vlasov 系统。

为实现双层马尔可夫链蒙特卡罗组件，我们同时近似测度序列 $\{\mu^l_{t-1+\Delta_l,\theta}, \ldots, \mu^l_{t,\theta}\}$ 和 $\{\mu^{l-1}_{t-1+\Delta_{l-1},\theta}, \ldots, \mu^{l-1}_{t\,\theta},\}$，对每个 $t \in \{1, \ldots, T\}$ 均如此。为避免符号混淆，我们将后者记为 $\{\widetilde{\mu}^{l-1}_{t-1+\Delta_{l-1},\theta}, \ldots, \widetilde{\mu}^{l-1}_{t\,\theta},\}$。我们通过算法 A.2 的步骤3(b)(ii)中详述的耦合概率律逼近过程实现这一点。该方案通过在细层级 (l) 和粗层级 ($l-1$) 离散化之间引入统计依赖，对算法 A.1 步骤 2(b) 中的标准概率律近似进行了推广。这种耦合设计受 MLMC 理论启发：在连续的离散化层级之间建立近似的相关性，可确保式 (3.5) 中望远镜差分估计量的方差随 l 增大而适当降低，从而保持有效多层方法所特有的计算效率。

3.2.2 耦合后验分布框架

我们提出双层马尔可夫链蒙特卡洛方法，以近似（3.5）。我们用 $\check{\mathbb{P}}^{N_l}_\theta$表示耦合轨迹序列的联合分布

$$(\overline{u}^l_1, \ldots, \overline{u}^l_T) \quad \text{and} \quad (\widetilde{\overline{u}}^{l-1}_1, \ldots, \widetilde{\overline{u}}^{l-1}_T)$$

在 $[0, T]$ 上，其对应的期望算子为 $\check{\mathbb{E}}^{N_l}_\theta$。此处， $u^l_t = u^{1:N_l}_t \in \mathsf{E}^{N_l}_l$表示细层轨迹集合

$$\overline{u}^l_t = \left\{(X^1_{t-1+\Delta_l}, \ldots, X^{N_l}_{t-1+\Delta_l}), \ldots, (X^1_t, \ldots, X^{N_l}_t)\right\},$$

而 $\widetilde{\overline{u}}^{l-1}_t = \widetilde{u}^{1:N_l}_t \in \mathsf{E}^{N_l}_{l-1}$ 表示粗层轨迹集合

$$\widetilde{\overline{u}}^{l-1}_t = \left\{(\widetilde{X}^1_{t-1+\Delta_{l-1}}, \ldots, \widetilde{X}^{N_l}_{t-1+\Delta_{l-1}}), \ldots, (\widetilde{X}^1_t, \ldots, \widetilde{X}^{N_l}_t)\right\},$$

由耦合概率律逼近过程（算法 2，步骤3(b)(ii)）生成。上标 N_l 明确表示该概率分布源自

使用 N_l 个粒子，在离散化层级 l 和 $l-1$上构造 McKean–Vlasov 定律的经验近似$\{\mu^{l,t}_{\theta}{}^{N_l},\}$ 和 $\{\tilde{\mu}^{l-1}_{\theta},_{t}{}^{N_l},\}$。

对于观测加权，我们定义耦合条件密度

$$H_{k,\theta}(x,x';y_k)=\tfrac{1}{2}\{g_\theta(x',y_k)+g_\theta(x,y_k)\},\tag{3.6}$$

对于 $(x,x',\theta)\in\mathbb{R}^{2d}\times\Theta$ 以及 $k\in\{1,\dots,T\}$。这种对称平均在不同离散化层级之间引入了相关性。算法 A.2 生成耦合核，用于同时模拟 $P^{l}_{\mu^{l},t-1}{}^{N_l},_{\theta},_{t},_{\theta}(x_{t-1},du_t)$ 和 $P^{l-1}_{\mu^{l-1},t-1}{}^{N_l}_{\theta},_{,t},_{\theta}(x_{t-1},du_t)$；其中，底层概率测度是通过算法 A.2 步骤3(b)(ii) 中预先计算的经验近似生成的。我们将生成时刻 t 的 $(u^l_t,\widetilde{u}^{l-1}_t)\in\mathsf{E}_l\times\mathsf{E}_{l-1}$、并以时刻 $t-1$ 的前一状态 $(u^l_{t-1},\widetilde{u}^{l-1}_{t-1})$ 为条件的耦合转移核记为 $\check{}\,P^l_\theta$，其初始条件为 $(u^l_0,\widetilde{u}^{l-1}_0)=(x_0,x_0)$。该核 $\check{}\,P^l_\theta$ 用于 Delta 粒子滤波器（算法 2，步骤 3(b)(iii)）。在此阶段，耦合经验测度 $\{\mu^{l},_{s}{}^{N_l}_{\theta},\}$ 和 $\{\tilde{\mu}^{l-1},_{s}{}^{N_l}_{\theta},\}$ 已固定，并作为确定性输入参数。

作为双层马尔可夫链蒙特卡罗采样器目标分布的联合概率测度定义在乘积空间 $\big(\Theta\times(\mathsf{E}_l\times\mathsf{E}_{l-1})^T,\ \mathscr{B}(\Theta\times(\mathsf{E}_l\times\mathsf{E}_{l-1})^T)\big)$ 上，其定义为

$$\begin{aligned}&\check{\pi}^{l,N_l}\big(d(\theta,u^l_{1:T},\widetilde{u}^{l-1}_{1:T})\mid y_{1:T}\big)\\&\quad=\frac{\{\prod_{k=1}^T H_{k,\theta}(x^l_k,\widetilde{x}^{l-1}_k;y_k)\}\,\check{\mathbb{E}}^{N_l}_\theta\Big[\prod_{k=1}^T\check{P}^{N_l}_\theta\big((x^l_{k-1},\widetilde{x}^{l-1}_{k-1}),d(u^l_k,\widetilde{u}^{l-1}_k)\big)\Big]\times\nu(\theta)d\theta}{\int_{\Theta\times(\mathsf{E}_l\times\mathsf{E}_{l-1})^T}\{\prod_{k=1}^T H_{k,\theta}(x^l_k,\widetilde{x}^{l-1}_k;y_k)\}\,\check{\mathbb{E}}^{N_l}_\theta\Big[\prod_{k=1}^T\check{P}^{N_l}_\theta\big((x^l_{k-1},\widetilde{x}^{l-1}_{k-1}),d(u^l_k,\widetilde{u}^{l-1}_k)\big)\Big]\times\nu(\theta)d\theta}.\end{aligned}\tag{3.7}$$

耦合测度 $\check{}\,\pi^{l,}{}^{N_l}$ 编码了离散化层级 l 和 $l-1$处轨迹之间的统计依赖，这对于多层方法中的有效方差缩减至关重要。为实现望远镜分解，我们定义辅助权重

$$\check{H}_{k,\theta}(x,x';y_k)=\frac{g_\theta(x,y_k)}{H_{k,\theta}(x,x';y_k)},\qquad(x,x',\theta)\in\mathbb{R}^{2d}\times\Theta,\ k\in\{1,\dots,T\}.$$

该比率将细层观测密度 $g_\theta(x,\ y_k)$ 与来自$H_{k,\ \theta}$ 式（3.6）的平均耦合密度联系起来。

确立这些定义后，式（3.5）中的双层望远镜差现在可以用关于耦合测度的期望来表示

$\check{\pi}^{l,\,N_l}$。关键的技术难点在于，$\check{\pi}^{l,\,N_l}$ 使用平均观测密度$H_{k,\,\theta}$ 在不同层级之间引入相关性，以实现方差缩减；而我们的目标量 $\bar{\pi}^{l,\,N_l}(\varphi)$和 $\bar{\pi}^{l-1,\,N_l}(\varphi)$涉及原始观测密度$\check{g}_\theta$。辅助权重$H_{k,\,\theta}$ 使这些分布之间进行必要的测度变换成为可能。具体而言，重要性权重 $\prod_{k=1}^{T}\check{H}_{k,\,\theta}(x_k^l,\;\widetilde{x}_k^{l-1};y_k)$具有一项关键性质：其分母$\prod_{k=1}^{T}H_{k,\,\theta}(x_k^l,\;\widetilde{x}_k^{l-1};y_k)$与耦合测度$\check{\pi}^{l,\,N_l}$中的耦合观测密度项完全一致，因此可以进行抵消：

$$\frac{\int_{\Theta\times(\mathsf{E}_l\times\mathsf{E}_{l-1})^T}\varphi(\theta,x_{1:T}^l)\left\{\frac{\prod_{k=1}^{T}g_\theta(x_k^l,y_k)}{\prod_{k=1}^{T}H_{k,\theta}(x_k^l,\widetilde{x}_k^{l-1};y_k)}\right\}\left\{\prod_{k=1}^{T}H_{k,\theta}(x_k^l,\widetilde{x}_k^{l-1};y_k)\right\}}{\times\,\check{\mathbb{E}}_\theta^{N_l}\left[\prod_{k=1}^{T}\check{P}_\theta^l\big((x_{k-1}^l,\widetilde{x}_{k-1}^{l-1}),d(u_k^l,\widetilde{u}_k^{l-1})\big)\right]\nu(\theta)d\theta}$$
$$\frac{}{\int_{\Theta\times(\mathsf{E}_l\times\mathsf{E}_{l-1})^T}\left\{\frac{\prod_{k=1}^{T}g_\theta(x_k^l,y_k)}{\prod_{k=1}^{T}H_{k,\theta}(x_k^l,\widetilde{x}_k^{l-1};y_k)}\right\}\left\{\prod_{k=1}^{T}H_{k,\theta}(x_k^l,\widetilde{x}_k^{l-1};y_k)\right\}}$$
$$\times\,\check{\mathbb{E}}_\theta^{N_l}\left[\prod_{k=1}^{T}\check{P}_\theta^l\big((x_{k-1}^l,\widetilde{x}_{k-1}^{l-1}),d(u_k^l,\widetilde{u}_k^{l-1})\big)\right]\nu(\theta)d\theta}.$$

抵消项恢复了细层后验分布 $\bar{\pi}^{l,\,N_l}$ 并使用原始观测密度 $\prod_{k=1}^{T}g_\theta(x_k^l,\;y_k)$。

因此，我们从耦合测度 $\check{\pi}^{l,\,N_l}$中抽取耦合轨迹 $((x_1^l,\;\ldots,\;x_T^l),\;(\widetilde{x}_1^{l-1},\;\ldots,\;\widetilde{x}_T^{l-1}))$。该测度采用平均观测密度 $H_{k,\,\theta}(x_k^l,\;\widetilde{x}_k^{l-1};y_k)$，以在不同离散化层级之间建立强相关性。随后，我们使用重要性权重 $\prod_{k=1}^{T}\check{H}_{k,\,\theta}(x_k^l,\;\widetilde{x}_k^{l-1};\;y_k)$ 和 $\prod_{k=1}^{T}\check{H}_{k,\,\theta}(\widetilde{x}_k^{l-1},\;x_k^l;y_k)$ 对这些轨迹重新加权，从而计算正确边缘后验分布 $\bar{\pi}^{l,\,N_l}$ 和 $\bar{\pi}^{l-1,\,N_l}$ 下的期望。这种测度变换技术是多层框架中实现方差缩减的关键：耦合抽样确保细层和粗层轨迹之间具有高度相关性，从而降低方差；而重要性权重则保证期望与目标分布一致。

望远镜差可改写为

$$\frac{\int_{\Theta\times(\mathsf{E}_l\times\mathsf{E}_{l-1})^T}\varphi(\theta,x_{1:T}^l)\left\{\prod_{k=1}^{T}\check{H}_{k,\theta}(x_k^l,\widetilde{x}_k^{l-1};y_k)\right\}\check{\pi}^{l,N_l}\left(d(\theta,u_{1:T}^l,\widetilde{u}_{1:T}^{l-1})\right)}{\int_{\Theta\times(\mathsf{E}_l\times\mathsf{E}_{l-1})^T}\left\{\prod_{k=1}^{T}\check{H}_{k,\theta}(x_k^l,\widetilde{x}_k^{l-1};y_k)\right\}\check{\pi}^{l,N_l}\left(d(\theta,u_{1:T}^l,\widetilde{u}_{1:T}^{l-1})\right)}$$
$$-\frac{\int_{\Theta\times(\mathsf{E}_l\times\mathsf{E}_{l-1})^T}\varphi(\theta,\widetilde{x}_{1:T}^{l-1})\left\{\prod_{k=1}^{T}\check{H}_{k,\theta}(\widetilde{x}_k^{l-1},x_k^l;y_k)\right\}\check{\pi}^{l,N_l}\left(d(\theta,u_{1:T}^l,\widetilde{u}_{1:T}^{l-1})\right)}{\int_{\Theta\times(\mathsf{E}_l\times\mathsf{E}_{l-1})^T}\left\{\prod_{k=1}^{T}\check{H}_{k,\theta}(\widetilde{x}_k^{l-1},x_k^l;y_k)\right\}\check{\pi}^{l,N_l}\left(d(\theta,u_{1:T}^l,\widetilde{u}_{1:T}^{l-1})\right)}. \tag{3.8}$$

借助（3.8），我们将开发一种高效的马尔可夫链蒙特卡罗算法，用于从耦合目标测度（3.7）中进行采样。关键挑战在于设计

离散化后验分布之间的有效耦合 π^{l,N_l} 和 π^{l-1,N_l}，并使用 MCMC 技术。除了简单的独立耦合外，目前不存在适用于此目的的标准马尔可夫链蒙特卡罗方法，而独立耦合无法提供必要的方差缩减。为解决这一问题，我们构造了联合概率测度（3.7），该测度使耦合轨迹 $(u^l_{1:T}, \widetilde{u}^{l-1}_{1:T})$之间产生统计依赖。

3.2.3 多层 PMCMC 采样算法

我们的多层 PMCMC 方法结合单层和双层组件，以估计 MLMC 近似。该框架将用于基准层计算的单层 PMCMC 与用于跨多个离散化层级进行望远镜差估计的双层马尔可夫链蒙特卡罗相结合。完整的算法细节见补充材料的第S.2节。下面我们概述算法A.2:

1. 基准层计算（步骤 2）：在基础离散化层级 $l_\star \in \mathbb{N}$ 运行算法 A.1，迭代 $K_{l_\star}$ 次，并使用 $N_{l_\star}$ 个粒子进行概率律近似，以获得基准层样本 $\{\theta^{(k)}_{l_\star}, x^{(k)}_{l_\star,1:T}\}_{k=0}^{K_{l_\star}}$及相应的似然估计。

2. 双层望远镜差分近似（步骤 3）：对于每个层级$l \in \{l_\star+1, \ldots, L\}$，我们独立运行双层马尔可夫链蒙特卡罗，以计算相应的望远镜差。

2.1. 参数提议（步骤 3(b)(i)）：对于每个层级 l，从提议核 $r(\theta^{(k-1)}, \theta')$中生成候选参数 θ'。

2.2. 耦合概率律近似（步骤3(b)(ii)）：对于提议参数θ'，通过使用相关布朗增量 W 以及 W、$\widetilde{}$ 演化 N_l 耦合粒子对，构造连续离散化层级上 McKean–Vlasov 律的同步粒子近似 $\{\mu^{l,N_l}_{t-1+j\Delta_l,\theta'}\}_{j=1}^{\Delta_l^{-1}}$ 和$\{\widetilde{\mu}^{l-1,N_l}_{t-1+j\Delta_{l-1},\theta'}\}_{j=1}^{\Delta_{l-1}^{-1}}$，其中

$$\Delta W^i_{t-1+j\Delta_{l-1}} = \Delta W^i_{t-1+(2j-1)\Delta_l} + \Delta W^i_{t-1+2j\Delta_l}。$$

2.3. Delta粒子滤波似然估计（步骤 3(b)(iii)）：对于给定参数 $\theta' \in \Theta$，在步骤 3(b)(ii) 中生成的耦合近似经验测度序列 $\{\mu^{l,N_l}_{t,\theta'}, \widetilde{\mu}^{l-1,N_l}_{t,\theta'}\}$ 条件下，我们使用包含 M 个耦合粒子对的 Delta 粒子滤波器来估计联合似然 $\hat{p}^{l,MN_l}_{\theta'}(y_{1:T})$，并生成耦合轨迹样本，以便随后在 Metropolis-Hastings 步骤中使用。

2.4. Metropolis-Hastings 步骤（步骤 3(b)(iv)）：接受或拒绝所提出的 d

参数，依据接受概率

$$\alpha = \min\left\{1,\ \frac{\hat{p}^{l,M,N_l}_{\theta'}(y_{1:T})\nu(\theta')r(\theta',\theta^{(k-1)})}{\hat{p}^{l,M,N_l}_{\theta^{(k-1)}}(y_{1:T})\nu(\theta^{(k-1)})r(\theta^{(k-1)},\theta')}\right\}. \tag{3.9}$$

2.5. 望远镜差的重要性权重校正：使用辅助权重函数 $\check{H}_{k,\theta}(x,x';y_k) = g_\theta(x,y_k)\,H_{k,\theta}(x,x';y_k)$ 进行重要性加权校正，以获得来自正确后验差 $\bar{\pi}^{l,N_l}(\varphi)-\bar{\pi}^{l-1,N_l}(\varphi)$ 的样本，具体论证见公式 (3.8)。

双层马尔可夫链蒙特卡罗计算复杂度：每次双层马尔可夫链蒙特卡罗迭代的计算复杂度为 $\mathcal{O}(\Delta_l^{-1}T\{MN_l+N_l^2\})$。其中，$\mathcal{O}(\Delta_l^{-1}N_l^2T)$ 项源于耦合概率律近似，而 $\mathcal{O}(\Delta_l^{-1}MN_lT)$ 对应于 Delta 粒子滤波器操作。

双层差分估计量：对于每个层级 $l\in\{l_\star+1,\dots,L\}$，使用由双层马尔可夫链蒙特卡罗生成的耦合样本 $\{\theta^l(k),\ u^l_{1:T}(k),\ \widetilde{u}^{l-1}_{1:T}(k)\}_{k=0}^{K_l}$，我们通过以下方式近似望远镜差：

$$\begin{aligned}
&\bar{\pi}^{l,N_l,K_l}(\varphi)-\bar{\pi}^{l-1,N_l,K_l}(\varphi) \\
&:=\frac{\frac{1}{K_l+1}\sum_{k=0}^{K_l}\varphi(\theta^l(k),x^l_{1:T}(k))\left\{\prod_{s=1}^{T}\check{H}_{s,\theta^l(k)}(x^l_s(k),\widetilde{x}^{l-1}_s(k);y_s)\right\}}{\frac{1}{K_l+1}\sum_{k=0}^{K_l}\left\{\prod_{s=1}^{T}\check{H}_{s,\theta^l(k)}(x^l_s(k),\widetilde{x}^{l-1}_s(k);y_s)\right\}} \\
&\quad-\frac{\frac{1}{K_l+1}\sum_{k=0}^{K_l}\varphi(\theta^l(k),\widetilde{x}^{l-1}_{1:T}(k))\left\{\prod_{s=1}^{T}\check{H}_{s,\theta^l(k)}(\widetilde{x}^{l-1}_s(k),x^l_s(k);y_s)\right\}}{\frac{1}{K_l+1}\sum_{k=0}^{K_l}\left\{\prod_{s=1}^{T}\check{H}_{s,\theta^l(k)}(\widetilde{x}^{l-1}_s(k),x^l_s(k);y_s)\right\}}.
\end{aligned} \tag{3.10}$$

完整多层框架：对于测试函数 $\varphi:\ \Theta\times\mathbb{R}^{Td}\to\mathbb{R}$，定义

$$\pi(\varphi) := \int_{\Theta\times\mathbb{R}^{Td}}\varphi(\theta,x_{1:T})\pi\left(d(\theta,x_{1:T})\,|\,y_{1:T}\right),$$

我们假设其有限。多层估计量为

$$\widehat{\pi}(\varphi) := \bar{\pi}^{l_\star,N_{l_\star},K_{l_\star}}(\varphi)+\sum_{l=l_\star+1}^{L}\left\{\bar{\pi}^{l,N_l,K_l}(\varphi)-\bar{\pi}^{l-1,N_l,K_l}(\varphi)\right\}, \tag{3.11}$$

其中，$\bar{\pi}^{l_\star,N_{l_\star},K_{l_\star}}(\varphi)$ 通过式（3.4）计算，而每个差值 $\pi^{l,N_l,K_l}(\varphi)-\pi^{l-1,N_l,K_l}(\varphi)$ 通过式（3.10）计算。该估计量受到六种误差来源的影响：基准层 MCMC 误差、多层 MCMC 误差、跨层交互误差、基准层粒子近似误差、多层粒子近似误差以及时间离散化偏差。通过适当选择参数 L、$\{K_{l_\star},\dots,K_L\}$ 和 $\{N_{l_\star},\dots,N_L\}$，可以系统地控制这些误差，下一节将对此进行分析。

4 理论结果

本节将阐述所提出方法论的理论结果。我们的分析依赖于三项正则性假设（A1）–（A3），详见补充材料第 S.6 节。证明基于附录 S.8中的技术，并将其扩展以适应依赖于概率律的扩散系数。在本节中， C 表示一个通用的有限常数，其取值可能随行而变化。

4.1 主要结果

基于上述正则性假设以及补充材料S.8 节中的辅助技术结果，我们给出本文的主要理论贡献。在整个分析过程中， $\mathbb{E}$表示关于（3.11）中定义的随机性的期望算子。为简化记号，我们在表达式中省略时间范围 T 。

定理 4.1 (多层 PMCMC 估计量的均方误差界)。 在假设 *(A1)–(A3)* 下，设 $\varphi \in C_b^2(\Theta \times \mathbb{R}^{dT}) \cap B_b(\Theta \times \mathbb{R}^{dT})$ 为测试函数，则存在常数 $C < \infty$ ，使得对于由整数 $l_\star < L$以及粒子数/迭代次数$(N_{l_\star}, \ldots, N_L, K_{l_\star}, \ldots, K_L) \in \mathbb{N}^{2(L-l_\star+1)}$指定的任意多层配置，均方误差满足

$$\mathbb{E}\left[(\widehat{\pi}(\varphi) - \pi(\varphi))^2\right] \leq C \left(\varepsilon^{(l_\star)}_{\text{MCMC}} + \varepsilon^{(\text{ML})}_{\text{MCMC}} + \varepsilon_{\text{cross}} + \varepsilon^{(l_\star)}_{\text{部分}} + \varepsilon^{(\text{ML})}_{\text{部分}} + \varepsilon_{\text{disc}}\right), \tag{4.1}$$

其中误差分量定义如下：

- 基准层 *MCMC* 误差： $\varepsilon_{(l_\star)\ \text{MCMC}}$ ： $= \dfrac{1}{K_{l_\star}+1}$，
- 多层 *MCMC* 误差： $\varepsilon_{(\text{ML})\ \text{MCMC}}$ ： $= \displaystyle\sum_{l=l_\star+1}^{L} \frac{\Delta_l}{K_l+1}$，
- 跨层交互误差： $\varepsilon_{\text{cross}}$ ： $= \displaystyle\sum_{l,q=l_\star+1}^{L} \frac{\sqrt{\Delta_l \Delta_q}}{K_l+1}\,\mathbb{1}_{(l \neq q)}(K_q+1)$，
- 基准层粒子近似误差： $\varepsilon_{(l_\star)\ \text{part}}$ ： $= \dfrac{1}{N_{l_\star}}\sqrt{\Delta_{l_\star}}$，
- 多层粒子近似误差： $\varepsilon_{(\text{ML})\ \text{部分}}$ ： $= \left(\displaystyle\sum_{l=l_\star+1}^{L} \sqrt{\frac{\Delta_l}{N_l}}\right)^2$，

- 时间离散化偏差：$\varepsilon_{\text{disc}}$ ： $= \Delta_L^2$。

定理 4.1 的证明见补充材料的 S.7 节。定理 4.1 的实际意义可量化如下。令 $\epsilon > 0$ 表示目标精度水平。在通常的假设下，即粒子滤波器组件（算法 A.1 的步骤 2(c) 和算法 A.2 的步骤 3(b)(iii)）中使用 $M = \mathrm{O}(T)$ 个粒子，这与实践中的典型做法一致（Andrieu 等（2010）），并忽略与观测时域 T相关的常数，我们的多层估计量（3.11）的总计算成本为 $\mathrm{O}\left(\sum_{l=l_\star}^{L} I_l \Delta_l^{-1} N_l^2\right)$。在多层蒙特卡罗框架中，当基准层 $l_\star$ 的计算负担可忽略不计时，最优参数选择可得 $I_l = \mathrm{O}(\epsilon^{-2}\Delta_l^{6/7})$、$N_l = \mathrm{O}(\epsilon^{-2}\Delta_l^{1/2})$ 和 $L = \mathrm{O}(|\log(\epsilon)|)$。此配置在计算成本为 $\mathrm{O}(\epsilon^{-6})$ 的同时，使均方误差 $\mathbb{E}\hat{\imath}(\hat{\pi}(\varphi) - \pi(\varphi))^2\acute{o}$ 的上界达到 $\mathrm{O}(\epsilon^2)$。相比之下，仅通过算法 A.1 使用层级 L 的传统单层方法，要达到相同的均方误差 $\mathrm{O}(\epsilon^2)$，需要计算成本 $\mathrm{O}(\epsilon^{-7})$；也就是说，我们将成本降低了一个数量级。因此，相较于标准单层 PMCMC 方法，我们的多层方法将计算复杂度降低了一个数量级。

5 数值模拟

本节介绍一种神经元活动模型，该模型采用具有非全局利普希茨结构的多维 MVSDE，以刻画复杂的神经动力学。模型的三维形式详见第 5.1 节。数值模拟结果见第 5.2 节。

5.1 三维神经元模型

本研究考察了 Dos Reis 等人 （2022）提出的神经元活动模型，该模型采用具有非全局利普希茨结构的 MVSDE。该模型包含三个维度，并采用专门设计的漂移项和扩散项来刻画神经元动力学。相应的 MVSDE 采用一般形式，

$$dX_t = a(t, x_t, \mu_t)dt + b(t, x_t, \mu_t)dW_t, \tag{5.1}$$

其中，漂移项和扩散系数表示如下：

$$a\left(t, x_t, \mu_t\right) := \begin{pmatrix} x_1 - (x_1)^3/3 - x_2 + I - \int_{\mathbb{R}^3} J\left(x_1 - V_{rev}\right) z_3 d\mu(z) \\ c\left(x_1 + a - bx_2\right) \\ a_r \frac{T_{max}(1-x_3)}{1+\exp(-\lambda(x_1 - V_T))} - a_d x_3 \end{pmatrix},$$

$$b\left(t, x_t, \mu_t\right) := \begin{pmatrix} b_{ext} & 0 & -\int_{\mathbb{R}^3} b_J\left(x_1 - V_{rev}\right) z_3 d\mu(z) \\ 0 & 0 & 0 \\ 0 & b_{32}(x) & 0 \end{pmatrix},$$

with

$$b_{32}(x) := \mathbb{1}_{\{x_3 \in (0,1)\}} \sqrt{a_r \frac{T_{max}(1-x_3)}{1+\exp(-\lambda(x_1 - V_T))} + a_d x_3}\ \Gamma \exp(-\Lambda/(1-(2x_3-1)^2)).$$

在漂移系数和扩散系数中出现的平均场相互作用项 $\int_{\mathbb{R}^3} J(x_1 - V_{\text{rev}}) z_3\, d\mu(z)$ 和 $\int_{\mathbb{R}^3} b_J(x_1 - V_{\text{rev}}) z_3\, d\mu(z)$ 中，变量 $z = (z_1,\ z_2,\ z_3)$ 表示时刻 t 神经元群体中任意一个神经元的状态，与焦点神经元状态 $x = (x_1,\ x_2,\ x_3)$ 不同。由于积分是关于 z 进行的，函数 $J(x_1 - V_{\text{rev}})$ 和 $b_J(x_1 - V_{\text{rev}})$ 可视为常数（仅取决于焦点神经元状态 x_1），而期望则是关于神经元群体分布的第三个分量 z_3 取的。这两个平均场相互作用项 $\int_{\mathbb{R}^3} J(x_1 - V_{rev}) z_3 d\mu(z)$ 和 $\int_{\mathbb{R}^3} b_J(x_1 - V_{rev}) z_3 d\mu(z)$ 表示对每个神经元的突触输入取群体平均，其中 z_3 表示网络中其他神经元的突触门控变量，从而形成 McKean–Vlasov 耦合结构。

有关状态变量和模型参数的物理解释的详细说明，请参见S.4 补充材料。

我们使用以单位时间间隔记录的离散观测$\{Y_k\}_{k=1}^T$，对上述模型进行了数值模拟。给定观测过程服从如下条件高斯分布：

$$Y_k | X_k = x_k \sim \mathcal{N}(x_k, \Sigma), \quad k \in \{1, 2, \ldots, T\}, \tag{5.2}$$

其中，$T = 50$ 选为终止时间，$\mathcal{N}(x_k, \Sigma)$表示均值向量为 $\mu = x_k$ 、协方差矩阵为对角协方差矩阵$\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \sigma_3^2)$的多元正态分布，其中 $\sigma_1^2, \sigma_2^2, \sigma_3^2$ 分别表示三个维度中观测噪声的方差。

初始条件规定如下：

$$X_0 \sim \mathcal{N}\left(\begin{pmatrix} V_0 \\ w_0 \\ y_0 \end{pmatrix}, \begin{pmatrix} \sigma_{V_0} & 0 & 0 \\ 0 & \sigma_{w_0} & 0 \\ 0 & 0 & \sigma_{y_0} \end{pmatrix}\right).$$

本模拟研究采用补充材料表 T.1 中列出的真实参数值生成合成数据，并将观测噪声水平

固定为 $(\sigma_1,\ \sigma_2,\ \sigma_3) = (0。2,\ 0。1,\ 0。02)$。这些不同的方差设置专门用于适应神经元模型中膜电位 V、恢复变量 w 和门控变量 y 各自固有的数值量级差异。

5.2 模拟结果

我们将参数 $\{I, J, c, \lambda, b_{ext}, \Gamma, \sigma_1, \sigma_2, \sigma_3\}$ 视为待推断的未知量，并根据参数间依赖关系较少这一特点选取这些参数。为参数 I 和 J 指定相互独立的高斯先验。由于 $\{c, \lambda, b_{ext}, \Gamma, \sigma_1, \sigma_2, \sigma_3\}$ 必须为正，我们对它们的对数变换量进行推断，即 $\log(c)$、$\log(\lambda)$、$\log(b_{ext})$、$\log(\Gamma)$、$\log(\sigma_1)$、$\log(\sigma_2)$ 和 $\log(\sigma_3)$，并分别为其指定相互独立的高斯先验分布。马尔可夫链蒙特卡洛采样采用具有对角协方差矩阵结构的高斯随机游走提议。我们对采样器进行调节，以优化混合特性并确保收敛诊断稳健。模拟参数遵循第 4 节中详述的设定。

为评估计算效率，我们针对三维神经元模型，对粒子马尔可夫链蒙特卡罗算法和 MLPMCMC 算法的计算成本与均方误差（MSE）之间的对数关系进行了比较分析。具体而言，我们通过考察 log(COST) 与 log(MSE) 之间的关系，分析均方误差如何随计算成本增加而降低。对于每种算法，我们通过数值计算拟合数据点（log(COST)，log(MSE)）所得线性回归直线的斜率，估计收敛速率；这些数据点来自在多个离散化层级范围内进行的多次模拟运行。该斜率量化了计算成本增加时均方误差降低的速率，可直接衡量算法效率。

补充材料表 T.2中给出的估计收敛速率表明，多层方法的收敛速率约为 -3，对应的计算复杂度为$\mathcal{O}(\epsilon^{-6})$；而单层方法的收敛速率接近 -3.5，对应的计算复杂度为 $\mathcal{O}(\epsilon^{-7})$。经验观察到的速率与理论分析吻合良好；由于估计底层分布本身存在不确定性，出现轻微差异是预期之中的。多层方法展现出更优的计算扩展性，证明了其相较于传统单层方法的效率优势。

所提出的方法应用于三维神经元模型时的收敛行为见补充材料：图F.1 展示了粒子马尔可夫链蒙特卡罗的轨迹图，图F.2 展示了 MLPMCMC 的轨迹图，图F.3

展示了 MLPMCMC 增量的运行均值。经过 10000 次迭代，粒子马尔可夫链蒙特卡罗和 MLPMCMC 在整个参数空间内均表现出令人满意的混合特性，这一点可从轨迹图中看出。此外，MLPMCMC 增量的运行均值逐渐收敛，证实了多层估计量的遍历性。

6 结论

本文开发了一种通用贝叶斯框架，用于估计具有完全依赖概率分布的漂移系数和扩散系数的部分观测均值场随机微分方程中的隐状态与参数。我们提出了首批能够处理这类一般均值场模型的粒子马尔可夫链蒙特卡罗和 MLPMCMC 算法。一项关键创新是多层耦合策略：该策略同步不同离散化层级间粒子系统概率律的演化，并通过耦合增量粒子滤波器构造相关的似然估计量。这一设计在适应依赖概率分布的随机性所引入的非线性的同时，保留了 MLMC 特有的方差缩减性质。

我们证明，所提出的 MLPMCMC 算法可达到 $O(\varepsilon^{2})$ 阶的最优均方误差，计算成本为 $O(\varepsilon^{-6})$，优于单层 PMCMC 的$O(\varepsilon^{-7})$复杂度。这些结果在依赖概率分布的扩散系数满足标准正则性和椭圆性假设时成立，表明即使存在分布依赖波动率，多层方法仍具有稳健性。针对带有群体层级噪声的三维神经元模型所进行的数值实验进一步验证了我们的理论发现，并凸显了本方法在模拟必须采用真实波动率结构的系统时的实用价值。

总而言之，这项工作显著拓展了多层粒子马尔可夫链蒙特卡洛方法在相互作用驱动的随机系统中的适用性，并为在漂移和扩散均具有均场依赖的模型中实现可扩展且准确的推断提供了有理有据的路径。未来的研究可将这些技术扩展到高维相互作用系统、跳跃扩散情形，或能够自动调整离散化和粒子分配的自适应多层方案（例如参见 Ning&Ionides (2023),Ionides 等人(2024)).

补充材料

补充材料详细介绍了单层和多层粒子马尔可夫链蒙特卡罗算法，提供了完整的伪代码、三维神经元模型的详细描述、补充表格和图示、正则性假设，以及理论结果的完整证明。

利益披露声明

作者声明，没有其他需要披露的竞争性利益。

数据可用性声明

用于复现数值结果的代码和数据已公开发布于https://github.com/aminwu-111/MV_随机微分方程_完全。

参考文献

Andrieu, C., Doucet, A. & Holenstein, R.（2010），《粒子马尔可夫链蒙特卡罗方法（附讨论）》，皇家统计学会会刊：*B*辑（统计方法学）**72**，269–342。

Awadelkarim, E. & Jasra, A.（2024），《用于部分观测 McKean–Vlasov 随机微分方程的多层粒子滤波器》，arXiv 预印本。

Baldaron, J., Fasoli, D., Faugeras, O. & Toubol, J.（2012），《Hodgkin-Huxely 神经元和 FitzHugh-Nagumo神经元网络中的均场描述与混沌传播》，数学神经科学杂志 **2**，10。

Beskos, A., Papaspiliopoulos, O., Roberts, G. O. & Fearnhead, P.（2006），《离散观测扩散过程的精确且计算高效的基于似然的估计（附讨论）》，皇家统计学会会刊：*B*辑（统计方法学）**68**，333–382。

Chopin, N., Papaspiliopoulos, O. 等人（2020），序贯蒙特卡罗简介，第4卷，Springer。

Dos Reis, G., Engelhardt, S. & Smith, G. (2022), ‘超线性增长下 McKean–Vlasov随机微分方程的模拟’, *IMA*数值分析杂志**42**(1), 874–922.

Doucet, A., De Freitas, N., Gordon, N. J. 等 (2001), 序贯蒙特卡罗方法实践, 第1卷, Springer. Giesecke, K., Schwenkler, G. & Sirignano, J. A. (2020), ‘大型金融系统的推断’,《数学金融》30, 3–46. Giles, M. B. (2008), ‘多层蒙特卡洛路径模拟’,《运筹学》56, 607–617. Giles, M. B. (2015), ‘多层蒙特卡洛方法’,《数值学报》24, 259–328. Gu, H., Li, J., Sun, W., Li, M., Leung, K., Wu, J. T., Yuan, H.-Y., Wang, M. H., Yang, B., McKay, M. R. 等 (2025), ‘优化全球基因组监测，以早期发现新兴 SARS-CoV-2变体’,*arXiv*预印本 *arXiv:2502.00934* . Heng, J., Houssineau, J. & Jasra, A. (2024), ‘部分观测扩散过程的无偏估计’,《机器学习研究杂志》25(66), 1–66. Ionides, E., Ning, N. & Wheeler, J. (2024), ‘具有共享参数和个体特定参数的耦合动态系统推断的迭代分块粒子滤波器’,《中华统计学志》34, 1–22. Jasra, A., Kamatani, K., Law, K. & Zhou, Y. (2018), ‘通过多层蒙特卡洛对部分观测扩散过程进行贝叶斯静态参数估计’,《*SIAM*科学计算杂志》40, A887–A902. Jasra, A. & Wu, A. (2025), ‘利用多层马尔可夫链蒙特卡罗对部分观测的 McKean–Vlasov扩散过程进行贝叶斯参数估计’, arXiv预印本. Li, J., Ionides, E. L., King, A. A., Pascual, M. & Ning, N. (2024), ‘耦合生物种群时空动态的推断’,《英国皇家学会界面》21(216), 20240217. Mogilner, A. & Edelstein-Keshet, L. (1999), ‘群集的非局部模型’,《数学生物学杂志》38, 534–570. Ning, N. & Ionides, E. (2023), ‘用于高维参数学习的迭代分块粒子滤波器：克服维数灾难’,《机器学习研究杂志》24(82), 1–76. Ning, N. & Wu, J. (2026), ‘McKean–Vlasov随机变分不等式的适定性与混沌传播’,《理论概率杂志》39(1), 5.

Ning, N.、Wu, J. 与 Zheng, J.（2024），《一维 McKean–Vlasov 随机变分不等式及带局部 Hölder 噪声系数的耦合 BSDE》，随机过程及其应用 **171**，104315。Roberts, G. 与 Rosenthal, J.（1997），《几何遍历性与混合马尔可夫链》。Sharrock, L.、Kantas, N.、Parpas, P. 与 Pavliotis, G.（2023），《McKean–Vlasov 随机微分方程的在线参数估计》，随机过程及其应用 **162**，481–546。Sznitman, A. S.（1991），混沌传播专题，载于《圣弗卢尔概率暑期学校第十九届》，第1464卷，第165–251页。Wen, J.、Wang, X.、Mao, S. 与 Xiao, X.（2016），《McKean–Vlasov 随机微分方程的最大似然估计及其应用》，应用数学与计算**274**，237–246。

S 补充材料

补充目录

S.1 单层 PMCMC 算法详情 .S1S.2 多层 PMCMC 算法详情 .S3S.3 完整算法伪代码 .S8S.4 三维神经元模型：详细描述 .S10S.5 补充表格和图 .S11S.6 正则性假设 .S14S.7 定理 4.1 的证明 .S15S.8 技术性证明.S18

S.1 单层 PMCMC 算法详情

本节给出了正文第 3.1.1 节所述单层 PMCMC 采样算法的完整算法细节。

1. **参数提议**（步骤 2(a)）：从提议核 $r(\theta^{(k-1)}, \theta')$ 中生成候选参数 θ'。

2. **转移律近似**（步骤 2(b)）：对于提议参数 θ'，构造粒子近似 $\{\mu^{l,N}_{t-1+j\Delta_l,\theta'}\}_{j=1}^{\Delta_l^{-1}}$，以近似每个时间步 $t \in \{1, \dots, T\}$处的 McKean–Vlasov 定律。将经验测度初始化为

$$\mu^{l,N}_{t-1,\theta'}(dx) = \begin{cases} \delta_{x_0}(dx), & \text{if } t = 1, \\ \frac{1}{N}\sum_{i=1}^{N} \delta_{X^i_{t-1}}(dx), & \text{if } t > 1. \end{cases} \tag{S.1}$$

然后通过离散化动力学，使 N粒子在 $[t-1, t]$ 的子区间内演化，$j = 1$、……、Δ_l^{-1}

$$\begin{aligned} X^i_{t-1+j\Delta_l} = X^i_{t-1+(j-1)\Delta_l} &+ a_{\theta'}\left(X^i_{t-1+(j-1)\Delta_l}, \overline{\zeta}_{1,\theta'}(X^i_{t-1+(j-1)\Delta_l}, \mu^{l,N}_{t-1+(j-1)\Delta_l,\theta'})\right)\Delta_l \\ &+ b_{\theta'}\left(X^i_{t-1+(j-1)\Delta_l}, \overline{\zeta}_{2,\theta'}(X^i_{t-1+(j-1)\Delta_l}, \mu^{l,N}_{t-1+(j-1)\Delta_l,\theta'})\right)\Delta W^i_{t-1+j\Delta_l}, \end{aligned} \tag{S.2}$$

其中 $\Delta W^i_{t-1+j\Delta_l} \overset{\text{独立同分布}}{\sim} \mathcal{N}_d(0, \Delta_l I_d)$ 是一个均值为 0、协方差为 $\Delta_l I_d$ 的 d 维高斯随机变量，其中 I_d 是 $d \times d$ 维单位矩阵。

每个子步的经验测度定义为

$$\mu^{l,N}_{t-1+(j-1)\Delta_l,\theta'} = \frac{1}{N}\sum_{n=1}^{N} \delta_{X^n_{t-1+(j-1)\Delta_l}}(dx), \tag{S.3}$$

且 $m \in \{1,2\}$ 的平均相互作用为

$$\overline{\zeta}_{m,\theta'}(X^i_{t-1+(j-1)\Delta_l}, \mu^{l,N}_{t-1+(j-1)\Delta_l,\theta'}) = \frac{1}{N}\sum_{n=1}^{N} \zeta_{m,\theta'}(X^i_{t-1+(j-1)\Delta_l}, X^n_{t-1+(j-1)\Delta_l}). \tag{S.4}$$

3. 粒子滤波器似然估计（步骤 2(c)）：给定参数 $\theta' \in \Theta$，并以步骤 2(b) 中生成的近似经验测度序列 $\{\mu^{l}_{,t\,\theta'}{}^{N}_{,}\}$为条件，我们采用一个包含 M 个粒子的粒子滤波器来估计边际似然。该过程针对 $s = 1, \ldots, T$递归执行：粒子从时间 $s-1$ 的重采样祖先出发，利用由 Δ_l^{-1}个 Euler-Maruyama 步构成的转移核进行传播，定义如下

$$\xi^m_s \sim \prod_{j=1}^{\Delta_l^{-1}} Q^l_{\mu^{l,N}_{s-1+(j-1)\Delta_l,\theta'},\theta'}(z_{j-1}, dz_j), \tag{S.5}$$

其中 $z_0 = \xi^{A^m_{s-1}}_{s-1}$，且动力学使用固定的经验分布。在每个时间步，我们计算归一化重要性权重

$$r^m_s = \frac{g_{\theta'}(\xi^m_s, y_s)}{\sum_{j=1}^{M} g_{\theta'}(\xi^m_j, y_s)}, \quad m = 1, \ldots, M, \tag{S.6}$$

随后利用这些权重对祖先索引$\{A^m_s\}_{m=1}^M$ 进行重采样，以用于下一次迭代。边际似然估计量通过以下方式递归更新

$$\hat{p}^{l,M,N}_{\theta'}(y_{1:s}) = \hat{p}^{l,M,N}_{\theta'}(y_{1:s-1}) \cdot \frac{1}{M}\sum_{m=1}^{M} g_{\theta'}(\xi^m_s, y_s), \tag{S.7}$$

最终估计量由所有时间步的未归一化权重平均值相乘得到:

$$\hat{p}^{l,M,N}_{\theta'}(y_{1:T}) := \prod_{s=1}^{T} \left\{ \frac{1}{M}\sum_{m=1}^{M} g_{\theta'}(\xi^m_s, y_s) \right\}. \tag{S.8}$$

最后，通过追溯从最终权重 $\{r^m_T\}$中选出的粒子的祖先，生成单条轨迹样本；该样本将作为 Metropolis-Hastings 更新所提议的潜在轨迹。

4. **Metropolis-Hastings 步骤** （步骤 2(d)）：此步骤采用标准的 Metropolis-Hastings 接受-拒绝机制，以决定接受提议的参数 θ'，还是保留当前参数 $\theta^{(k-1)}$。接受概率根据提议参数值与当前参数值处后验密度之比计算（忽略归一化常数）

$$\alpha = \min\left\{1,\ \frac{\hat{p}^{l,M,N}_{\theta'}(y_{1:T})\nu(\theta')r(\theta',\theta^{(k-1)})}{\hat{p}^{l,M,N}_{\theta^{(k-1)}}(y_{1:T})\nu(\theta^{(k-1)})r(\theta^{(k-1)},\theta')}\right\}. \tag{S.9}$$

接受-拒绝决策：生成一个均匀随机变量 $U \sim \mathcal{U}[0,1]$, 服从 $[0,1]$上的均匀分布。若 $U < \alpha$，则接受该提议：设置 $\theta^{(k)} = \theta'$，更新似然估计

$$\hat{p}^{l,M,N}_{\theta^{(k)}}(y_{1:T}) = \hat{p}^{l,M,N}_{\theta'}(y_{1:T}), \tag{S.10}$$

并存储相应的轨迹

$$(\xi_1^{(k)},\ldots,\xi_T^{(k)}) = (\xi_1^{m^*},\ldots,\xi_T^{m^*}), \tag{S.11}$$

其中，轨迹索引 m^* 根据归一化权重 $\{r_T^m\}_{m=1}^M$进行采样，定义如下：

$$r_T^m = \frac{g_{\theta'}(\xi_T^m, y_T)}{\sum_{j=1}^M g_{\theta'}(\xi_T^j, y_T)} \tag{S.12}$$

该值在步骤 2(c) 的最终时间步 T 计算得到。否则，拒绝该提议：保留$\theta^{(k)} = \theta^{(k-1)}$, 以及与其关联的似然估计和轨迹。

该接受-拒绝过程确保马尔可夫链 $\{\theta^{(k)}, \xi_{1\,T:}^{(k)}\}_{k=0}^K$具有$_{\pi^{l,N}}$ （见（3.2））作为其平稳分布，从而尽管使用了粒子滤波器产生的近似似然，仍能进行有效的后验推断。

该算法是对 Andrieu 等人 （2010）提出的原始粒子马尔可夫链蒙特卡罗框架的改进。在其中建立的温和正则性条件下，所得马尔可夫链从近似后验分布 $\pi^{l,N}$中生成样本。 —

S.2 多层伪边际马尔可夫链蒙特卡罗算法细节

本节给出正文第 3.2.3 节所述多层伪边际马尔可夫链蒙特卡罗采样算法的完整算法细节。

1. 基准层计算 （步骤 2）：执行算法A.1 在基础离散化层级 $l_\star \in \mathbb{N}$ 下进行 $K_{l_\star}$ 次迭代，使用 $N_{l_\star}$ 个粒子进行概率律近似

以获得基准层样本 $\{\theta_{l_\star}^{(k)},\ x_{l_\star 1,:T}^{(k)}\}_{k=0}^{K_{l_\star}}$ 以及对应的似然估计 $\{\hat{p}^{l_\star}_{\theta_{l_\star}^{(k)},M N_{l_\star}}(y_{1:T})\}_{k=1^\circ}^{K_{l_\star}}$

2. 双层望远镜差分近似（步骤 3）：对于每个层级 $l \in \{l_\star+1,\dots,L\}$，我们分别运行双层马尔可夫链蒙特卡罗，以计算相应的望远镜差。

2.1. 参数提议（步骤 3(b)(i)）：对于每个层级 l，根据提议核 $r(\theta^{(k-1)},\theta')$生成候选参数 θ'。

2.2. 耦合概率律近似（步骤3(b)(ii)）：对于提议参数θ'，在连续的离散化层级上构造同步粒子近似$\{\mu^{l,N_l}_{t-1+j\Delta_l,\theta'}\}_{j=1}^{\Delta_l^{-1}}$ 和$\{\widetilde{\mu}^{l-1,N_l}_{t-1+j\Delta_{l-1},\theta'}\}_{j=1}^{\Delta_{l-1}^{-1}}$，以近似 McKean–Vlasov 律。耦合经验测度初始化如下

$$\begin{aligned}
&\mu^{l,N_l}_{t-1,\theta'}(dx)=\widetilde{\mu}^{l-1,N_l}_{t-1,\theta'}(dx)=\delta_{x_0}(dx),\quad \text{if } t=1, \\
&\mu^{l,N_l}_{t-1,\theta'}(dx)=\frac{1}{N_l}\sum_{i=1}^{N_l}\delta_{X^{l,i}_{t-1}}(dx)\quad\text{and}\quad \widetilde{\mu}^{l-1,N_l}_{t-1,\theta'}(dx)=\frac{1}{N_l}\sum_{i=1}^{N_l}\delta_{\widetilde{X}^{l-1,i}_{t-1}}(dx),\quad \text{if } t>1.
\end{aligned} \tag{S.13}$$

随后，N_l 个耦合粒子对使用相关布朗增量在 $[t-1,t]$的子区间上演化。对于 $j=1,\dots,\Delta_l^{-1}$ 以及每个粒子 $i\in\{1,\dots,N_l\}$，细层粒子按以下方式演化

$$\begin{aligned}
X^{l,i}_{t-1+j\Delta_l} = X^{l,i}_{t-1+(j-1)\Delta_l} &+ a_{\theta'}\left(X^{l,i}_{t-1+(j-1)\Delta_l},\zeta_{1,\theta'}\left(X^{l,i}_{t-1+(j-1)\Delta_l},\mu^{l,N_l}_{t-1+(j-1)\Delta_l,\theta'}\right)\right)\Delta_l \\
&+ b_{\theta'}\left(X^{l,i}_{t-1+(j-1)\Delta_l},\zeta_{2,\theta'}\left(X^{l,i}_{t-1+(j-1)\Delta_l},\mu^{l,N_l}_{t-1+(j-1)\Delta_l,\theta'}\right)\right)\Delta W^i_{t-1+j\Delta_l}
\end{aligned} \tag{S.14}$$

其中 $\Delta W^i_{t-1+j\Delta_l}\overset{\text{独立同分布}}{\sim}\mathcal{N}_d(0,\ \Delta_l I_d)$。同时，对于 $j=1,\ \dots,\Delta_{l-1}^{-1}$，粗层粒子按如下方式演化

$$\begin{aligned}
&\widetilde{X}^{l-1,i}_{t-1+j\Delta_{l-1}} \\
&= \widetilde{X}^{l-1,i}_{t-1+(j-1)\Delta_{l-1}} \\
&\quad + a_{\theta'}\left(\widetilde{X}^{l-1,i}_{t-1+(j-1)\Delta_{l-1}},\zeta_{1,\theta'}\left(\widetilde{X}^{l-1,i}_{t-1+(j-1)\Delta_{l-1}},\widetilde{\mu}^{l-1,N_l}_{t-1+(j-1)\Delta_{l-1},\theta'}\right)\right)\Delta_{l-1} \\
&\quad + b_{\theta'}\left(\widetilde{X}^{l-1,i}_{t-1+(j-1)\Delta_{l-1}},\zeta_{2,\theta'}\left(\widetilde{X}^{l-1,i}_{t-1+(j-1)\Delta_{l-1}},\widetilde{\mu}^{l-1,N_l}_{t-1+(j-1)\Delta_{l-1},\theta'}\right)\right)\Delta\widetilde{W}^i_{t-1+j\Delta_{l-1}}
\end{aligned} \tag{S.15}$$

其中，$\widetilde{\Delta W^i_{t-1+j\Delta_{l-1}}}$ 由细层级的布朗增量构造而成，以确保通过聚合 $\Delta W^i_{t-1+j\Delta_{l-1}} = \Delta W^i_{t-1+(2j-1)\Delta_l}+\Delta W^i_{t-1+2j\Delta_l}$实现$\widetilde{\text{耦合}}$，

每个子步上的经验测度为

$$\mu^{l,N_l}_{t-1+(j-1)\Delta_l,\theta'} = \frac{1}{N_l}\sum_{n=1}^{N_l}\delta_{X^{l,n}_{t-1+(j-1)\Delta_l}}(dx),$$
$$\widetilde{\mu}^{l-1,N_l}_{t-1+(j-1)\Delta_{l-1},\theta'} = \frac{1}{N_l}\sum_{n=1}^{N_l}\delta_{\widetilde{X}^{l-1,n}_{t-1+(j-1)\Delta_{l-1}}}(dx), \tag{S.16}$$

并且，平均相互作用定义为，对于 $m \in \{1,2\}$，

$$\zeta_{m,\theta'}\left(X^{l,i}_{t-1+(j-1)\Delta_l}, \mu^{l,N_l}_{t-1+(j-1)\Delta_l,\theta'}\right) = \frac{1}{N_l}\sum_{n=1}^{N_l}\zeta_{m,\theta'}\left(X^{l,i}_{t-1+(j-1)\Delta_l}, X^{l,n}_{t-1+(j-1)\Delta_l}\right),$$
$$\zeta_{m,\theta'}\left(\widetilde{X}^{l,i}_{t-1+(j-1)\Delta_l}, \widetilde{\mu}^{l,N_l}_{t-1+(j-1)\Delta_l,\theta'}\right) = \frac{1}{N_l}\sum_{n=1}^{N_l}\zeta_{m,\theta'}\left(\widetilde{X}^{l,i}_{t-1+(j-1)\Delta_l}, \widetilde{X}^{l,n}_{t-1+(j-1)\Delta_l}\right). \tag{S.17}$$

2.3. Delta粒子滤波似然估计（步骤 3(b)(iii)）：对于给定参数 $\theta' \in \Theta$，在步骤 3(b)(ii)中生成的耦合近似经验测度序列 $\{\mu^{l}_{,t\,\theta'}{}^{N_l}_{,,}\widetilde{\mu}^{l-1}_{,t\,\theta'}{}^{N_l}_{,}\}$ 的条件下，我们采用包含 M 个耦合粒子对的 Delta 粒子滤波器来估计联合似然。该过程对于 $s = 1,\ldots,T$递归进行，其中耦合粒子对从时间 $s-1$ 的重采样祖先出发，使用同步耦合转移核进行传播，记为

$$(\xi^{m,l}_s, \widetilde{\xi}^{m,l-1}_s) \sim \check{P}^{l}_{\theta'}\left((\xi^{A^m_{s-1},l}_{s-1}, \widetilde{\xi}^{A^m_{s-1},l-1}_{s-1}), d(x,\tilde{x})\right), \tag{S.18}$$

其中，动力学利用固定耦合经验分布。在每个时间步，我们计算归一化耦合重要性权重

$$r^m_s = \frac{H_{s,\theta'}(\xi^{m,l}_s, \widetilde{\xi}^{m,l-1}_s; y_s)}{\sum_{j=1}^{M} H_{s,\theta'}(\xi^{m,l}_j, \widetilde{\xi}^{m,l-1}_j; y_s)}, \quad m = 1,\ldots,M, \tag{S.19}$$

其中$H_{s,\theta'}$ 是式（3.6）中定义的耦合观测密度。随后使用这些权重对祖先索引 $\{A^m_s\}_{m=1}^{M}$ 进行重采样，以进行下一次迭代。耦合似然估计量通过以下方式递归更新：

$$\hat{p}^{l,M,N_l}_{\theta'}(y_{1:s}) = \hat{p}^{l,M,N_l}_{\theta'}(y_{1:s-1}) \cdot \frac{1}{M}\sum_{m=1}^{M} H_{s,\theta'}(\xi^{m,l}_s, \widetilde{\xi}^{m,l-1}_s; y_s), \tag{S.20}$$

从而得到最终估计量，该估计量构造为所有时间步的未归一化权重的平均值之乘积：

$$\hat{p}^{l,M,N_l}_{\theta'}(y_{1:T}) := \prod_{s=1}^{T}\left\{\frac{1}{M}\sum_{m=1}^{M} H_{s,\theta'}(\xi^{m,l}_s, \widetilde{\xi}^{m,l-1}_s; y_s)\right\}. \tag{S.21}$$

最后，通过追溯从最终权重 $\{r_T^m\}$中选出的粒子对的祖先关系，生成一个耦合轨迹样本，作为 Metropolis-Hastings 更新所提议的耦合潜在轨迹。

2.4. Metropolis-Hastings 步骤（步骤 3(b)(iv)）：此步骤采用标准的 Metropolis-Hastings 接受-拒绝机制，以决定接受提议的参数 θ'，还是保留当前参数 $\theta^{(k-1)}$。根据耦合似然估计，使用提议参数值和当前参数值处的后验密度（至归一化常数）之比计算接受概率:

$$\alpha = \min\left\{1,\ \frac{\hat{p}_{\theta'}^{l,M,N_l}(y_{1:T})\nu(\theta')r(\theta',\theta^{(k-1)})}{\hat{p}_{\theta^{(k-1)}}^{l,M,N_l}(y_{1:T})\nu(\theta^{(k-1)})r(\theta^{(k-1)},\theta')}\right\}. \tag{S.22}$$

接受-拒绝决策：生成一个均匀随机变量 $U \sim \mathcal{U}[0,1]$。若 $U < \alpha$，则接受提议：令 $\theta^{(k)} = \theta'$，更新耦合似然估计

$$\hat{p}_{\theta^{(k)}}^{l,M,N_l}(y_{1:T}) = \hat{p}_{\theta'}^{l,M,N_l}(y_{1:T}), \tag{S.23}$$

并存储对应的耦合轨迹

$$\left((\xi_1^{(k)},\ldots,\xi_T^{(k)}),(\widetilde{\xi}_1^{(k)},\ldots,\widetilde{\xi}_T^{(k)})\right) = \left((\xi_1^{m^*,l},\ldots,\xi_T^{m^*,l}),(\widetilde{\xi}_1^{m^*,l-1},\ldots,\widetilde{\xi}_T^{m^*,l-1})\right), \tag{S.24}$$

其中，轨迹索引 m^* 根据终止时刻的归一化权重$\{r_T^m\}_{m=1}^M$进行重采样,

$$r_T^m = \frac{H_{T,\theta'}(\xi_T^{m,l},\widetilde{\xi}_T^{m,l-1};y_T)}{\sum_{j=1}^M H_{T,\theta'}(\xi_T^{j,l},\widetilde{\xi}_T^{j,l-1};y_T)}$$

这些权重是在步骤 3(b)(iii) 的最终时间步 T 计算得到的。否则，拒绝该提议：保留 $\theta^{(k)} = \theta^{(k-1)}$ 及其对应的耦合似然估计和轨迹对。

2.5. 望远镜差分的重要性权重校正：第 3 步(b)(iii)中引入的耦合观测密度 $H_{k,\theta}(x,x';y_k)$ 会诱导从方程 (3.7) 中定义的耦合测度 $\check{\pi}^{l,N_l}$ 进行采样，但这并不直接对应于各个后验分布 $\bar{\pi}^{l,N_l}$ 和 $\bar{\pi}^{l-1,N_l}$ 之差。为了从正确的后验差中获取样本，我们使用辅助权重函数 $\check{H}_{k,\theta}(x,x';\ y_k) = {}^{g_\theta(x,y_k)}{}_{H_{k,\theta}(x,x';y_k)}$ 进行重要性加权校正，具体定义见第 3.2.2 节。对于由双层马尔可夫链蒙特卡罗生成的耦合轨迹样本 $\{(\theta^l(k)\ x_{1:T}^l(k)\ \widetilde{x}_{1:T}^{l-1}(k))\}_{k=0}^{K_l}$，~~望远镜~~

差分估计量见方程 (3.10)，它使用这些校正权重来恰当处理分布不匹配问题。具体而言，计算 $\varphi(\theta_l^{(k)}, x_{1:T}^l(k))$ 时，细层期望由 $\prod_{s=1}^T \check{H}_{s,\theta_l^{(k)}}(x_s^l(k)\ \widetilde{x}_s^{l-1}(k);\ y_s)$ 加权；计算 $\varphi(\theta_l^{(k)}, \widetilde{x}_{1:T}^{l-1}(k))$ 时，粗层期望则由 $\prod_{s=1}^T \check{H}_{s,\theta_l^{(k)}}(\widetilde{x}_s^{l-1}(k)\ x_s^l(k); y_s)$ 加权。通过方程 (3.8) 得到严格论证的这种测度变换技术，确保归一化的加权经验平均收敛到真实的后验差 $\bar{\pi}^{l,N_l}(\varphi) - \bar{\pi}^{l-1,N_l}(\varphi)$，从而在通过耦合采样机制实现有效方差缩减的同时，保持 MLMC 框架的有效性。

S.3 完整算法伪代码

A算法 A.1 用于 MVSDE 参数估计的粒子马尔可夫链蒙特卡罗算法 1. 初始化：输入离散化层级 $l \in \mathbb{N}_0$（设置为 $\Delta_l = 2^{-l}$）、粒子数 $N \in \mathbb{N}$（用于转移律近似）和 $M \in \mathbb{N}$（用于滤波）、初始参数 $\theta^{(0)} \in \Theta$、马尔可夫链蒙特卡罗迭代次数 $K \in \mathbb{N}$，以及数据 $y_{1:T}$。 2. 马尔可夫链蒙特卡罗迭代循环：对于 $k = 1, \ldots, K$: (a) 参数提议：从提议核 $r(\theta^{(k-1)}, \theta')d\theta'$ 生成 θ'。 (b) 概率律近似：对于每个时间步 $t \in \{1, \ldots, T\}$：i. 根据式 (S.1) 初始化经验测度。ii. 对于 $j = 1, \ldots, \Delta_l^{-1}$以及每个粒子 $i \in \{1, \ldots, N\}$，根据式 (S.2) 演化粒子 $X^i_{t-1+j\Delta_l}$，并使用式 (S.4) 和 (S.3) 中定义的相互作用项和经验测度。增量 $\Delta W^i_{t-1+j\Delta_l}$ 独立同分布于 $\sim \mathrm{N}_d(0, \Delta_l I_d)$。iii. 存储近似概率律 $\{\mu^{l,N}_{t-1+j\Delta_l,\theta'}\}_{j=1}^{\Delta_l^{-1}}$。(c) 粒子滤波器似然估计：i. 根据式 (S.5)，使用 $s = 1$ 从转移核初始化 M 个粒子 $\{\xi_1^m\}_{m=1}^M$。设置祖先索引 $A_0^m = m$ 和边际似然 $\hat{p}^{l,MN}_{\theta'}(y_0) = 1$。ii. 对于 $s = 1, \ldots, T$:

- **重采样：** 通过公式(S.6)计算归一化权重，并采样祖先索引 A_s^m依据 $\{r_s^m\}_{m=1}^M$。
- **似然更新：** 通过公式(S.7)更新似然。

• 传播：若 $s < T$，对于 $m \in \{1, \cdots M\}$，使用转移核并根据公式 (S.5) 采样 $\xi^m_{s+1} | \xi^{A_s^m}_s$，其中 $s = s + 1$，且 $x_s^m = \xi_s^{A_s^m}$和 $(\xi_1^m, \cdots \xi_{s+1}^m) = (\xi_1^{A_s^m}, \cdots \xi_s^{A_s^m}, \xi_{s+1}^m)$。iii. 从最终权重 $\{r_T^m\}_{m=1}^M$ 中采样最终轨迹索引 m^*。(d) Metropolis-Hastings 步骤：i. 根据公式 (S.9) 计算接受概率。ii. 生成 $U \sim \mathrm{U}[0, 1]$。若 $U < \alpha$，则接受：$\theta(k) = \theta'$，并根据公式 (S.10) 和 (S.11) 更新似然和轨迹。否则，拒绝：$\theta(k) = \theta(k-1)$，似然和轨迹保持不变。3. 输出：返回参数样本链 $\{\theta(k)\}_{k=1}^K$、对应的似然估计值 $\{\hat{p}^{l,MN}_{\theta^{(k)}}(y_{1:T})\}_{k=1}^K$，以及轨迹样本 $\{(\xi(k)_1, \ldots, \xi(k)_T)\}_{k=1}^K$。

算法 A.2 McKean–Vlasov随机微分方程的多层伪边际马尔可夫链蒙特卡罗 - 第一部分

1. **初始化：** 给定基准层 $l^* \in \mathbb{N}_0$ 和最细层 $L \in \mathbb{N}$ ($L > l^*$)，其中 $\Delta_l = 2^{-l}$。设置基础层概率律近似的粒子数 $N_{l^*} \in \mathbb{N}$，$\{N_{l^*+1}, \ldots, N_L\} \subset \mathbb{N}$ 用于耦合概率律近似，$M \in \mathbb{N}$ 用于滤波。设置基础层马尔可夫链蒙特卡洛的迭代次数 $K_{l^*} \in \mathbb{N}$，以及双层马尔可夫链蒙特卡洛的迭代次数 $\{K_{l^*+1}, \ldots, K_L\} \subset \mathbb{N}$。输入初始参数 $\theta^{(0)} \in \Theta$ 和数据 $y_{1:T}$。

2. **基准层计算：** 在层级 l^* 执行算法A.1，进行 K_{l^*} 次迭代，并在步骤 2(b) 的概率律近似中使用 N_{l^*} 个粒子。获得基准层级样本$\{\theta_{l^*}^{(k)}, x_{l^*}^{(k)}{}_{1:T}\}_{k=0}^{K_{l^*}}$ 以及似然估计值 $\{\hat{p}^{l^*}_{\theta^{(k)}, M, N_{l^*}}(y_{1:T})\}_{k=0}^{K_{l^*}}$。

3. **望远镜差计算：** 对于每个离散化层级 $l \in \{l^*+1, \ldots, L\}$：

   (a) **双层马尔可夫链蒙特卡罗设置：** 设置 $\Delta_l = 2^{-l}$, $\Delta_{l-1} = 2^{-(l-1)}$。初始化马尔可夫链蒙特卡罗计数器 $k = 0$。

   (b) **马尔可夫链蒙特卡罗迭代循环：** 对于 $k = 1, \ldots, K_l$：

      i. **参数提议：** 从核 $r(\theta^{(k-1)}, \theta')d\theta'$生成 θ'。

      ii. **耦合概率律近似：** 对于每个时间步 $t \in \{1, \ldots, T\}$：

         1. 根据式（S.13）初始化耦合经验测度。

         2. 对于 $j = 1, \ldots, \Delta_l^{-1}$以及粒子 $i \in \{1, \ldots, N_l\}$，使用式（S.14）演化细层粒子，并使用式（S.17）和（S.16）中定义的相互作用项和经验测度。增量$\Delta W_{t-1+j\Delta_l}^i$是服从 $\mathcal{N}_d(0, \Delta_l I_d)$分布的独立同分布随机向量。

         3. 类似地，对于 $j = 1, \ldots, \Delta_{l-1}^{-1}$，根据式（S.15）演化粗层粒子，并分别通过式（S.17）和（S.16）采用相互作用项和经验测度。粗层增量$\sim \Delta W_{t-1+j\Delta_{l-1}}^i$ 由细层增量构造。

         4. 存储近似的耦合定律

$$\left\{\left(\mu_{t-1+\Delta_l,\theta'}^{l,N_l}, \widetilde{\mu}_{t-1+\Delta_{l-1},\theta'}^{l-1,N_l}\right), \ldots, \left(\mu_{t,\theta'}^{l,N_l}, \widetilde{\mu}_{t,\theta'}^{l-1,N_l}\right)\right\}.$$

3. **望远镜差分计算（续）：** 对于每个离散化层级 $l \in \{l^*+1, \ldots, L\}$:

(b) 马尔可夫链蒙特卡罗迭代循环（续）： 对于 $k = 1, \ldots, K_l$: iii. **Delta 粒子滤波器似然估计：** i. 使用转移核 $\check{P}^l_{\theta'}\left((x_0, \widetilde{x}_0),\ d\left(u^l_1, \widetilde{u}^{l-1}_1\right)\right)$初始化 M 耦合粒子对 $\left\{\left(\xi^m_1, \widetilde{\xi}^m_1\right)\right\}_{m=1}^M$。设置$A^m_0 = m$ 和 $p^l_{\theta', M, N_l}(y_0) = 1$。 ii. 对于 $s = 1, \ldots, T$:

- 根据式(S.19)计算耦合归一化权重，并从 A^m_s中$\{r^m_s\}_{m=1}^M$采样。
- 根据式(S.20)更新似然。

- 若 $s < T$，根据式 (S.18)，使用 $s = s+1$ 从耦合转移核中抽取 $\left(\xi^m_{s+1}, \widetilde{\xi}^m_{s+1}\right) \mid \left(\xi^{A^m_s}_s, \widetilde{\xi}^{A^m_s}_s\right)$，其中输入测度已在步骤 (b)(ii) 中生成。对于$m \in \{1, \ldots, M\}$，$\left(\xi^m_1, \ldots, \xi^m_s, \widetilde{\xi}^m_1, \ldots, \widetilde{\xi}^m_{s-1}\right) = \left(\xi^{A^m_s}_1, \ldots, \xi^{A^m_s}_s, \widetilde{\xi}^{A^m_s}_1, \ldots, \widetilde{\xi}^{A^m_s}_{s-1}\right)$。 iii. 从最终权重 $\{r^m_T\}_{m=1}^M$中抽取最终耦合轨迹索引 m^*。 iv. Metropolis-Hastings 步骤：i. 根据式 (S.22) 计算接受概率 α： ii. 生成 $U \sim \mathrm{U}[0,1]$。若 $U < \alpha$，则接受：$\theta(k) = \theta'$，并根据式 (S.23) 和式 (S.24) 更新似然和轨迹。否则，拒绝：$\theta(k) = \theta(k-1)$，似然和轨迹保持不变。

4. **输出：**
   - 基准层样本：$\{\theta^{(k)}_{l^*}\}_{k=0}^{K_{l^*}}$, $\{\hat{p}^{l^*}_{\theta^{(k)}, M, N_{l^*}, 0}(y_{1:T})\}_{k=0}^{K_{l^*}}$, $\{(\xi^{(k)}_1, \ldots, \xi^{(k)}_T)\}_{k=0}^{K_{l^*}}$。
   - 双层耦合样本：对于每个层级 $l \in \{l^*+1, \ldots, L\}$，返回 $\{\theta^{(k)}_l\}_{k=0}^{K_l}$, $\{\hat{p}^l_{\theta^{(k)}, M, N_l}(y_{1:T})\}_{k=0}^{K_l}$, $\left\{\left(\xi^{(k)}_1, \ldots, \xi^{(k)}_T, \widetilde{\xi}^{(k)}_1, \ldots, \widetilde{\xi}^{(k)}_T\right)\right\}_{k=0}^{K_l}$.

S.4 三维神经元模型：详细描述

三维状态向量 $x = (x_1, x_2, x_3)$ 表示不同的神经元变量：x_1 对应膜电位 V，x_2 表示恢复变量 w，x_3 刻画突触门控变量 y。漂移项$a(t, x, \mu)$ 通过三个相互耦合的分量控制确定性动力学。

第一个分量描述带有外部输入I 和平均场突触耦合的 FitzHugh-Nagumo 神经元动力学，该耦合受反转电位 V_{rev} 和耦合强度 J调制。第二个分量刻画恢复动力学，其中包含时间尺度参数 c 以及耦合系数 a 和 b。第三个分量描述突触门控动力学，其中包含上升速率和衰减速率 a_r 和 a_d、最大传递 T_{max}，以及由陡峭度参数 λ 和阈值电位 V_T控制的 S形激活函数。扩散矩阵 $b(t,\ x,\ \mu)$ 将随机性引入系统。强度为 b_{ext} 的外部噪声直接影响膜电位，而强度为 b_J 的突触噪声则通过平均场项产生作用。恢复变量不受直接噪声影响。突触门控变量通过函数 $b_{32}(x)$ 引入状态依赖噪声；该函数将突触动力学与由参数 Γ 和 Λ控制的指数阻尼因子相结合，并包含一个示性函数，以确保 x_3 保持在具有物理意义的区间 $(0,\ 1)$ 内。

S.5 补充表格和图

$V_0 = 0$	$\sigma_{V_0} = 0.4$	$a = 0.7$	$b = 0.8$	$c = 0.08$	$I = 0.5$	$b_{ext} = 0.5$
$w_0 = 0.5$	$\sigma_{w_0} = 0.4$	$V_{rev} = 1$	$a_r = 1$	$a_d = 1$	$T_{max} = 1$	$\lambda = 0.2$
$y_0 = 0.3$	$\sigma_{y_0} = 0.05$	$J = 1$	$b_J = 0.2$	$V_T = 2$	$\Gamma = 0.1$	$\Lambda = 0.5$

表 T.1：模型参数及其取值

参数	粒子马尔可夫链蒙特卡罗	MLPMCMC
I	-3.32	-3.10
J	-3.39	-2.99
log(c)	-3.92	-2.18
log(λ)	-4.50	-2.76
log(b_{ext})	-3.48	-2.83
log(Γ)	-3.69	-2.47
log(σ_1)	-3.74	-2.38
log(σ_2)	-3.64	-3.26
log(σ_3)	-3.24	-3.17

表T.2：估计收敛速率比较

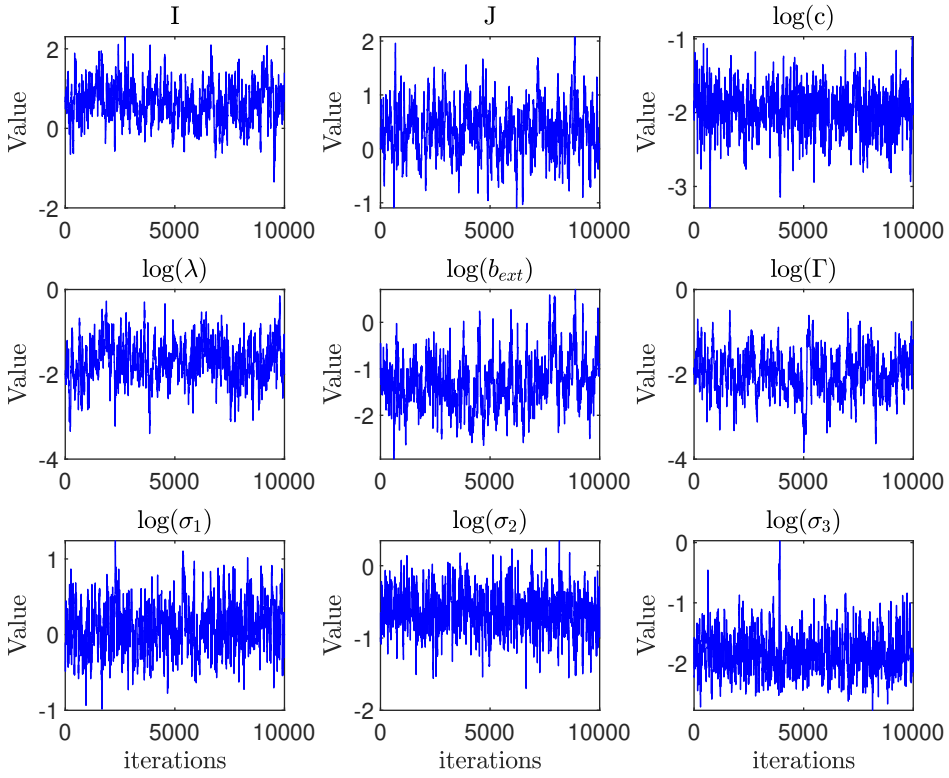
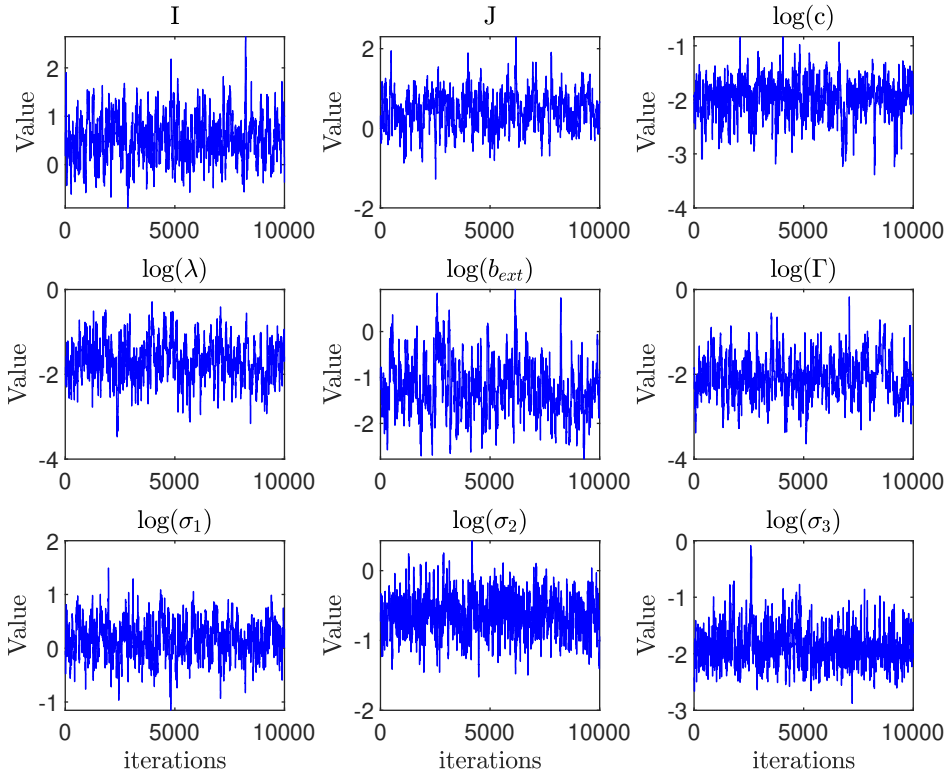


图 F.1：使用粒子马尔可夫链蒙特卡罗估计三维神经元模型参数的轨迹图。



图F.2：使用 MLPMCMC 对三维神经元模型估计参数的轨迹图。

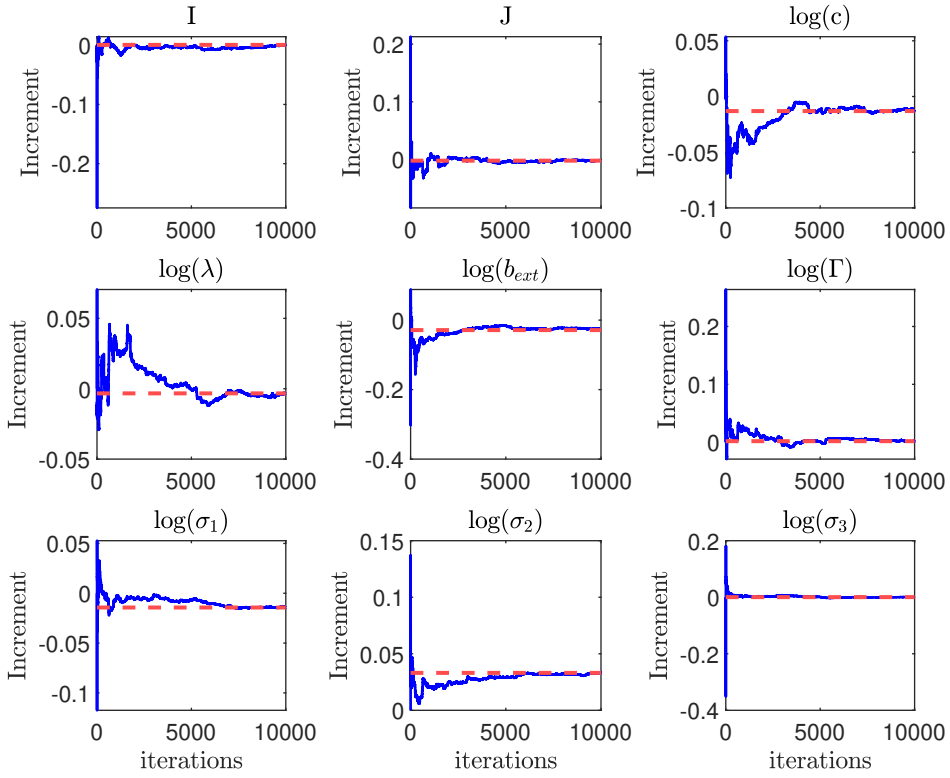


图 F.3：使用 MLPMCMC 估计三维神经元模型参数时，增量的运行均值。

S.6 正则性假设

令$\mathcal{P}(\mathbb{R}^d)$ 表示可测空间 $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ 上概率测度的集合，其中$\mathcal{B}(\mathbb{R}^d)$ 表示 $\mathbb{R}^d$上的 Borel σ-域。我们用$\mathcal{C}_b^k(\mathbb{R}^{d_1}, \mathbb{R}^{d_2})$ 表示函数空间，其中函数 $f\colon \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$ 具有最高至 k 阶的有界导数；用$\mathcal{B}_b(\mathbb{R}^{d_1}, \mathbb{R}^{d_2})$ 表示从 $\mathbb{R}^{d_1}$ 到 $\mathbb{R}^{d_2}$的有界可测函数全体。我们假设所有马尔可夫链均以各自的平稳分布为初始分布，并在理论分析中采用以下条件。

(A1) 对每个 $(\theta, \mu) \in \Theta \times \mathrm{P}(\mathbb{R}^d)$ 及固定的 $x \in \mathbb{R}$，有 $(a_\theta(\cdot, \zeta_{1,\theta}(\cdot\ \mu))\, \zeta_{1,\theta}(\cdot\ x)) \in \mathrm{C}_b^2(\mathbb{R}^{d+1}, \mathbb{R}^d) \cap \mathrm{B}_b(\mathbb{R}^{d+1}, \mathbb{R}^d) \times \mathrm{C}_b^2(\mathbb{R}^{2d}, \mathbb{R}) \cap \mathrm{B}_b(\mathbb{R}^{2d}, \mathbb{R})$，且 $(b_\theta(\cdot, \zeta_{2,\theta}(\cdot\ \mu))\, \zeta_{2,\theta}(\cdot\ x)) \in \mathrm{C}_b^2(\mathbb{R}^{d+1}, \mathbb{R}^{d\times d}) \cap \overline{\mathrm{B}}_b(\mathbb{R}^{d+1}, \mathbb{R}^{d\times d}) \times \mathrm{C}_b^2(\mathbb{R}^{2d}, \mathbb{R}) \cap \mathrm{B}_b(\mathbb{R}^{2d}, \mathbb{R})$。这些泛函的所有上界均关于 $\theta \in \Theta$ 一致成立。此外，存在 $k > 0$

使得一致椭圆性条件成立：

$$\inf_{(x,y)\in\mathbb{R}^d\times\mathbb{R}}\inf_{v\in\mathbb{R}^d\setminus\{0\}}\frac{v^\top b_\theta(x,y)^\top b_\theta(x,y)v}{\|v\|^2}\geq k,\qquad\forall\theta\in\Theta.$$

(A2) 对每个 $y\in\mathsf{Y}$，函数 $(\theta,x)\mapsto g_\theta(x,y)$ 属于 $\mathcal{C}_b^2(\Theta\times\mathbb{R}^d,\mathbb{R})\cap\mathcal{B}_b(\Theta\times\mathbb{R}^d,\mathbb{R})$。此外，$\inf_{(\theta,x,y)\in\Theta\times\mathbb{R}^d\times\mathsf{Y}}g_\theta(x,y)>0$。
(A3) 对每个 $l\in\mathbb{N}$，马尔可夫核 R_l (由算法 A.1) 产生）和 $\check{}R_l$ (由算法 A.2) 产生）均存在相对于其可逆的不变测度。此外，它们均一致遍历，且满足关于 l 的一致的一步小化条件。

S.7 定理 4.1 的证明

证明。 证明分为以下 4 个步骤。

步骤 1：误差分解。 对于 $\pi^{l,N_l,K_l}(\varphi)-\pi^{l-1,N_l,K_l}(\varphi)$ 在 (3.10) 中定义的差，以及 π^l 在 (3.1) 中定义的分布，记

$$\Delta_l^{N_l,K_l}[\varphi]:=\overline{\pi}^{l,N_l,K_l}[\varphi]-\overline{\pi}^{l-1,N_l,K_l}[\varphi]\quad\text{and}\quad\Delta_l[\varphi]:=\overline{\pi}^l[\varphi]-\overline{\pi}^{l-1}[\varphi].$$

对于 $\widehat{\pi}$（定义于（3.11）中），从层级 $l_\star$ 到 L 逐项相消并应用柯西–施瓦茨不等式，可得

$$\begin{aligned}\widehat{\pi}(\varphi)-\pi(\varphi)&=\left[\overline{\pi}^{l_\star,N_{l_\star},K_{l_\star}}[\varphi]-\overline{\pi}^{l_\star}[\varphi]\right]+\sum_{l=l_\star+1}^{L}\left[\Delta_l^{N_l,K_l}[\varphi]-\Delta_l[\varphi]\right]+\left[\overline{\pi}^L[\varphi]-\pi[\varphi]\right]\\&=:\mathcal{T}_1+\mathcal{T}_2+\mathcal{T}_3.\end{aligned}$$

将上述等式两边平方后取期望，利用柯西–施瓦茨不等式 $\mathbb{E}[(\sum_{i=1}^n X_i)^2]\leq n\sum_{i=1}^n\mathbb{E}[X_i^2]$，其中 $n=3$，可得

$$\mathbb{E}\left[\left(\widehat{\pi}(\varphi)-\pi(\varphi)\right)^2\right]\leq 3\left\{\mathbb{E}[\mathcal{T}_1^2]+\mathbb{E}[\mathcal{T}_2^2]+\mathcal{T}_3^2\right\}.\tag{S.25}$$

步骤 2：界定 $\mathcal{T}_1$。 我们将项 $\mathcal{T}_1$ 分解为两个不同的部分：一是由马尔可夫链蒙特卡罗采样产生的随机误差，二是由粒子滤波近似引入的系统性偏差。通过引入中间量 $\pi^{l_\star,N_{l_\star}}[\varphi]$（它表示相对于粒子近似后验分布的精确期望，即 $K_{l_*}\to\infty$ 时的极限），我们可以写出

$$\mathcal{T}_1=\overline{\pi}^{l_\star,N_{l_\star},K_{l_\star}}[\varphi]-\overline{\pi}^{l_\star}[\varphi]$$

$$= \underbrace{\left[\overline{\pi}^{l_\star,N_{l_\star},K_{l_\star}}[\varphi] - \overline{\pi}^{l_\star,N_{l_\star}}[\varphi]\right]}_{\text{MCMC Error}} + \underbrace{\left[\overline{\pi}^{l_\star,N_{l_\star}}[\varphi] - \overline{\pi}^{l_\star}[\varphi]\right]}_{\text{Particle Approximation Error}}. \tag{S.26}$$

两边平方并取期望后，我们应用不等式 $(a+b)^2 \leq 2a^2 + 2b^2$ 来分离各项的贡献：

$$\mathbb{E}[\mathcal{T}_1^2] \leq 2\mathbb{E}\left[\left(\overline{\pi}^{l_\star,N_{l_\star},K_{l_\star}}[\varphi] - \overline{\pi}^{l_\star,N_{l_\star}}[\varphi]\right)^2\right] + 2\left(\overline{\pi}^{l_\star,N_{l_\star}}[\varphi] - \overline{\pi}^{l_\star}[\varphi]\right)^2. \tag{S.27}$$

我们分别对这两项进行界定。首先，应用一致遍历条件（假设A3）以及Roberts & Rosenthal(1997中的推论 2.1，MCMC 估计量的方差随迭代次数线性衰减 $K_{l_\star}$。因此，存在一个常数 C 使得：

$$\mathbb{E}\left[\left(\overline{\pi}^{l_\star,N_{l_\star},K_{l_\star}}[\varphi] - \overline{\pi}^{l_\star,N_{l_\star}}[\varphi]\right)^2\right] \leq \frac{C}{K_{l_\star}+1}. \tag{S.28}$$

其次，根据引理 S.2，由粒子近似引起的偏差满足 $N_{l_\star}$ 个粒子对应的 $|\pi^{l_\star,N_{l_\star}}[\varphi] - \pi^{l_\star}[\varphi]| \leq CN_{l_\star}^{-1/2}$。将其平方可得

$$\left(\overline{\pi}^{l_\star,N_{l_\star}}[\varphi] - \overline{\pi}^{l_\star}[\varphi]\right)^2 \leq \frac{C}{N_{l_\star}}. \tag{S.29}$$

结合这些估计，可得基准层误差的最终上界：

$$\mathbb{E}[\mathcal{T}_1^2] \leq C\left(\frac{1}{K_{l_\star}+1} + \frac{1}{N_{l_\star}}\right) = C\left(\varepsilon_{\text{MCMC}}^{(l_\star)} + \varepsilon_{\text{part}}^{(l_\star)}\right). \tag{S.30}$$

步骤 3：界定 $\mathcal{T}_2$。 我们首先定义中间量

$$\Delta_l^{N_l}[\varphi] := \overline{\pi}^{l,N_l}[\varphi] - \overline{\pi}^{l-1,N_l}[\varphi],$$

该中间量表示层级 l 和 $l-1$下粒子近似后验分布所对应的精确期望之差，其中 π^{l,N_l} 和 π^{l-1,N_l} 是我们在式（3.5）中提到的分布。接着，我们将总和分解为随机 MCMC 误差项和确定性粒子偏差项：

$$\mathcal{T}_2 = \sum_{l=l_\star+1}^{L}\left(\Delta_l^{N_l,K_l}[\varphi] - \Delta_l^{N_l}[\varphi]\right) + \sum_{l=l_\star+1}^{L}\left(\Delta_l^{N_l}[\varphi] - \Delta_l[\varphi]\right) := \mathcal{T}_{2,\text{MCMC}} + \mathcal{T}_{2,\text{part}}.$$

利用不等式 $(a+b)^2 \leq 2a^2 + 2b^2$，我们有

$$\mathbb{E}[\mathcal{T}_2^2] \leq 2\mathbb{E}[\mathcal{T}_{2,\text{MCMC}}^2] + 2\mathcal{T}_{2,\text{part}}^2.$$

项 $\mathcal{T}_{2,\mathrm{part}}$ 表示各层累计的粒子近似偏差。根据引理 S.1,

$$|\mathcal{T}_{2,\mathrm{part}}| = \left| \sum_{l=l_\star+1}^{L} \left(\Delta_l^{N_l}[\varphi] - \Delta_l[\varphi] \right) \right| \leq C \sum_{l=l_\star+1}^{L} \frac{\sqrt{\Delta_l}}{\sqrt{N_l}}.$$

将该结果平方即可得到多层粒子误差界

$$\mathcal{T}_{2,\mathrm{part}}^2 \leq C \left(\sum_{l=l_\star+1}^{L} \frac{\sqrt{\Delta_l}}{\sqrt{N_l}} \right)^2 = C\varepsilon_{\mathrm{part}}^{(\mathrm{ML})}.$$

接下来，我们将 MCMC 误差和的平方 $\mathcal{T}_{2,\mathrm{MCMC}}$ 展开为对角（方差）项和非对角（协方差）项:

$$\begin{aligned}
\mathbb{E}[\mathcal{T}_{2,\mathrm{MCMC}}^2] = & \sum_{l=l_\star+1}^{L} \mathbb{E}\left[\left(\Delta_l^{N_l,K_l}[\varphi] - \Delta_l^{N_l}[\varphi] \right)^2 \right] \\
& + \sum_{\substack{l,q=l_\star+1 \\ l\neq q}}^{L} \mathbb{E}\left[\left(\Delta_l^{N_l,K_l}[\varphi] - \Delta_l^{N_l}[\varphi] \right) \left(\Delta_q^{N_q,K_q}[\varphi] - \Delta_q^{N_q}[\varphi] \right) \right] \\
=: & \ \mathcal{T}_{2,\mathrm{MCMC}}^{(1)} + \mathcal{T}_{2,\mathrm{MCMC}}^{(2)}.
\end{aligned}$$

为界定 $\mathcal{T}_2^{(1)}$,马尔可夫链蒙特卡洛，我们应用引理 S.4。在平稳性（假设 3）条件下，MCMC 误差的均值为零，而引理 S.4 将耦合估计量的方差界定为 $O(\Delta_l(K_l+1)^{-1})$。对各层级求和:

$$\sum_{l=l_\star+1}^{L} \mathbb{E}\left[\left(\Delta_l^{N_l,K_l}[\varphi] - \Delta_l^{N_l}[\varphi] \right)^2 \right] \leq C \sum_{l=l_\star+1}^{L} \frac{\Delta_l}{K_l+1} = C\varepsilon_{\mathrm{MCMC}}^{(\mathrm{ML})}.$$

为界定 $\mathcal{T}_2^{(2)}$,马尔可夫链蒙特卡洛，引理 S.5 给出

$$\begin{aligned}
\sum_{\substack{l,q=l_\star+1 \\ l\neq q}}^{L} \mathbb{E}\left[\left(\Delta_l^{N_l,K_l}[\varphi] - \Delta_l^{N_l}[\varphi] \right) \left(\Delta_q^{N_q,K_q}[\varphi] - \Delta_q^{N_q}[\varphi] \right) \right] &\leq C \sum_{\substack{l,q=l_\star+1 \\ l\neq q}}^{L} \frac{\sqrt{\Delta_l}}{K_l+1} \frac{\sqrt{\Delta_q}}{K_q+1} \\
&= C\varepsilon_{\mathrm{cross}}.
\end{aligned}$$

通过结合上述两个不等式，可得 $\mathbb{E}[\mathcal{T}_{22,\mathrm{MCMC}}]$ 的上界，即 $C(\varepsilon^{(\mathrm{ML})}$ 马尔可夫链蒙特卡洛 $+ \varepsilon_{\mathrm{cross}})$。

步骤 4：界定 $\mathcal{T}_3$。 对于 $\mathcal{T}_3$，，根据假设 A1下的标准弱收敛分析，并结合 Awadelkarim &Jasra(2024) 的引理 A.12，可得

$$\mathcal{T}_3^2 = (\overline{\pi}^L[\varphi] - \pi[\varphi])^2 \leq C(\Delta_L^2) = C(\varepsilon_{\text{disc}}). \tag{S.31}$$

代入上一步得到的 $\mathbb{E}[\mathcal{T}_{22\text{ 的上界，MCMC}}]$ 以及（S.30）和（S.31），代入（S.25）即可完成证明。 □

S.8 技术性证明

引理 S.1。 在假设（*A1*）–（*A2*）下，对于任意测试函数 $\varphi \in \mathcal{C}_b^2(\Theta \times \mathbb{R}^{dT}) \cap \mathcal{B}_b(\Theta \times \mathbb{R}^{dT})$，存在一个有限常数 C，使得对于任意配置（$l_\star$、 L、 $N_{l_\star}$、……、 N_L）$\in \mathbb{N}^{L-l_\star+3}$ 满足 $l_\star < L$，

$$\left| \sum_{l=l_\star+1}^{L} \left[\left(\overline{\pi}^{l,N_l}[\varphi] - \overline{\pi}^{l-1,N_l}[\varphi]\right) - \left(\overline{\pi}^{l}[\varphi] - \overline{\pi}^{l-1}[\varphi]\right) \right] \right| \leq C \sum_{l=l_\star+1}^{L} \frac{\sqrt{\Delta_l}}{\sqrt{N_l}}. \tag{S.32}$$

证明。 证明分为以下两个步骤进行。

步骤 1：后验表示。 对于每个层级 l，回顾一下，$P^l_{\mu^l_{k-1,\,\theta},\,k,\,\theta}(x_{k-1},\ du_k)$ 表示离散化层级 l 上时间区间 $[k-1,k]$ 的扩展转移核：

$$\overline{P}^l_{\mu^l_{k-1,\theta},k,\theta}(x_{k-1}, du_k) = \prod_{j=1}^{\Delta_l^{-1}} Q^l_{\mu^l_{k-1+(j-1)\Delta_l,\theta},\theta}(x_{k-1+(j-1)\Delta_l}, dx_{k-1+j\Delta_l}),$$

其中，$Q^l_{\mu^l_{t-1+(k-1)\Delta_l,\theta},\theta}(x_{t-1+(k-1)\Delta_l}, dx_{t-1+k\Delta_l})$ 表示离散化随机微分方程（2.3）在概率分布为 $\mu^l_{t-1+(k-1)\Delta_l,\theta} \in \mathcal{P}(\mathbb{R}^d)$ 时的第 Δ_l 个时间步上、定义于 $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ 的高斯马尔可夫核，而 $u_k = (x_{k-1+\Delta_l}, \ldots, x_k) \in E_l = (\mathbb{R}^d)^{\Delta_l^{-1}}$ 表示 $[k-1,k]$ 内的中间状态。

对于任意有界可测函数 $\psi\colon \Theta \times \mathbb{R}^{dT} \to \mathbb{R}$，定义路径泛函

$$\mathcal{P}^l_{T,\theta}[\psi] := \int_{\Theta \times E_l^T} \psi(\theta, x_{1:T}) \prod_{k=1}^{T} \overline{P}^l_{\mu^l_{k-1,\theta},k,\theta}(x_{k-1}, du_k),$$

其中，$E_l^T = (E_l)^T$ 是离散化层级 l 下、完整时间范围 $[0,T]$ 上的路径空间。同样，粒子近似路径泛函定义为

$$\mathcal{P}^{l,N_l}_{T,\theta}[\psi] := \int_{\Theta \times E_l^T} \psi(\theta, x_{1:T}) \prod_{k=1}^{T} \overline{P}^l_{\mu^{l,N_l}_{k-1,\theta},k,\theta}(x_{k-1}, du_k),$$

其中 $\mu^{l,N_l}_{k,\theta}$ 表示由 N_l 个粒子构造的经验测度 .

令$\mathcal{G}_\theta := \prod_{k=1}^{T} g_\theta(x_k, y_k)$表示观测似然泛函。于是，后验表示形式如下：

$$\overline{\pi}^{l}[\varphi] = \frac{\int_\Theta \mathcal{P}^{l}_{T,\theta}[\varphi \cdot \mathcal{G}_\theta]\,\nu(d\theta)}{\int_\Theta \mathcal{P}^{l}_{T,\theta}[\mathcal{G}_\theta]\,\nu(d\theta)} \quad \text{and} \quad \overline{\pi}^{l,N_l}[\varphi] = \frac{\int_\Theta \mathbb{E}^{N_l}_{\theta}[\mathcal{P}^{l,N_l}_{T,\theta}[\varphi \cdot \mathcal{G}_\theta]]\,\nu(d\theta)}{\int_\Theta \mathbb{E}^{N_l}_{\theta}[\mathcal{P}^{l,N_l}_{T,\theta}[\mathcal{G}_\theta]]\,\nu(d\theta)},$$

其中， $\mathbb{E}^{N_l}_{\theta}$表示关于 McKean–Vlasov 定律的 N_l粒子近似的期望。

步骤 2：比值展开与耦合构造。 根据三角不等式，我们分别对每个层级的差异进行界定。对于固定的 $l \in \{l_\star + 1, \dots, L\}$，可利用标准代数恒等式展开比值差。对于正实数 A, B, C, D 以及任意实数 a, b, c, d，有如下分解：

$$\frac{a}{A} - \frac{b}{B} - \frac{c}{C} + \frac{d}{D} = \frac{1}{A}[(a-b)-(c-d)] - \frac{b}{AB}(A - B - C + D) + \mathcal{R}_{\text{rem}}, \quad \text{(S.33)}$$

其中

$$\begin{aligned}\mathcal{R}_{\text{rem}} = &-\frac{1}{AC}(A-C)(c-d) - \frac{1}{AB}(C-D)(b-d) \\ &+ \frac{d}{CBD}(B-D)(C-D) + \frac{d}{ACB}(A-C)(C-D). \end{aligned} \quad \text{(S.34)}$$

分解式（S.33）中的前两项是主要贡献项，且具有相同的收敛阶。我们将详细分析第一项。第二项的分析方法类似，注意到测试函数 φ 有界（$\|\varphi\|_\infty < \infty$），且该项涉及归一化常数的耦合差异。余项 $\mathcal{R}_{\text{rem}}$ 表示高阶误差相互作用。我们将分析$\mathcal{R}_{\text{rem}}$中的第一项，因为其余各项也可以类似地控制。

我们重点关注第一项 $\frac{1}{A}[(a-b) - (c-d)]$, ，它是主要贡献项，对应于

$$\begin{aligned}&\left(\int_\Theta \mathbb{E}^{N_l}_{\theta}[\mathcal{P}^{l,N_l}_{T,\theta}[\mathcal{G}_\theta]]\,\nu(d\theta)\right)^{-1} \\ &\qquad \times \left\{ \int_\Theta \mathbb{E}^{N_l}_{\theta}[\mathcal{P}^{l,N_l}_{T,\theta}[\varphi \cdot \mathcal{G}_\theta]]\,\nu(d\theta) - \int_\Theta \mathbb{E}^{N_l}_{\theta}[\mathcal{P}^{l-1,N_l}_{T,\theta}[\varphi \cdot \mathcal{G}_\theta]]\,\nu(d\theta) \right.\end{aligned}$$

$$- \left(\int_\Theta \mathcal{P}^l_{T,\theta}[\varphi \cdot \mathcal{G}_\theta]\, \nu(d\theta) - \int_\Theta \mathcal{P}^{l-1}_{T,\theta}[\varphi \cdot \mathcal{G}_\theta]\, \nu(d\theta) \right) \Bigg\}. \quad \text{(S.35)}$$

为了控制（S.35），构造一个概率空间，支持由共同布朗增量驱动的四个同步 Euler-Maruyama 过程：

- $X^{l,\ 1\ T}_{:}{}^{N_l}$：采用 N_l 粒子律近似的细层路径，
- $X^{l-1,\ 1\ T}_{:}{}^{N_l}$：采用 N_l 粒子律近似的粗层路径，
- $X^l_{1:\ T}$：采用精确 McKean–Vlasov 定律的细层路径，
- $X^{l-1}_{1\ T:}$：采用精确 McKean–Vlasov 定律的粗层路径。

记 $\widetilde{\mathbb{E}}^{N_l}_\theta$为该联合耦合下的期望。(S.35) 中的分子可改写为

$$\int_\Theta \widetilde{\mathbb{E}}^{N_l}_\theta \Big[\varphi(\theta, X^{l,N_l}_{1:T}) \mathcal{G}_\theta(X^{l,N_l}_{1:T}) - \varphi(\theta, X^{l-1,N_l}_{1:T}) \mathcal{G}_\theta(X^{l-1,N_l}_{1:T}) \\ - \{ \varphi(\theta, X^l_{1:T}) \mathcal{G}_\theta(X^l_{1:T}) - \varphi(\theta, X^{l-1}_{1:T}) \mathcal{G}_\theta(X^{l-1}_{1:T}) \} \Big] \nu(d\theta). \quad \text{(S.36)}$$

由于 $\varphi \cdot \mathcal{G}_\theta \in \mathcal{C}^2_b \cap \mathcal{B}_b$，泰勒展开（参见 Awadelkarim & Jasra(2024) 中的引理 A.5）给出 (S.36) 中该量的上界如下

$$C \int_\Theta \widetilde{\mathbb{E}}^{N_l}_\theta \Big[\| (X^{l,N_l}_{1:T} - X^{l-1,N_l}_{1:T}) - (X^l_{1:T} - X^{l-1}_{1:T}) \| \\ + \| X^l_{1:T} - X^{l-1}_{1:T} \| \cdot \left(\| X^{l,N_l}_{1:T} - X^l_{1:T} \| + \| X^{l-1,N_l}_{1:T} - X^{l-1}_{1:T} \| \right) \Big] \nu(d\theta).$$

在假设 (A2) 下，结合柯西–施瓦茨不等式和Awadelkarim & Jasra (2024) 中的引理 A.7，(S.35) 中的量可由以下式子界定： $O\left(\frac{\sqrt{\Delta_l}}{\sqrt{N_l}}\right)$。

接下来，我们对$\mathcal{R}_{\text{rem}}$, $-\frac{1}{AC}(A-C)(c-d)$中的第一项进行界定，该项对应于

$$- \left(\int_\Theta \mathbb{E}^{N_l}_\theta [\mathcal{P}^{l,N_l}_{T,\theta}[\mathcal{G}_\theta]]\, \nu(d\theta) \right)^{-1} \left(\int_\Theta \mathcal{P}^l_{T,\theta}[\mathcal{G}_\theta]\, \nu(d\theta) \right)^{-1} \quad \text{(S.37)}$$

$$\times \left| \int_\Theta \left(\mathbb{E}^{N_l}_\theta [\mathcal{P}^{l,N_l}_{T,\theta}[\mathcal{G}_\theta]] - \mathcal{P}^l_{T,\theta}[\mathcal{G}_\theta] \right) \nu(d\theta) \right|$$

$$\times \left| \int_\Theta \left(\mathcal{P}^l_{T,\theta}[\varphi \cdot \mathcal{G}_\theta] - \mathcal{P}^{l-1}_{T,\theta}[\varphi \cdot \mathcal{G}_\theta] \right) \nu(d\theta) \right|.$$

在假设 (A2) 下，$\int_\Theta \mathbb{E}_\theta^{N_l}[\mathrm{P}_T^{l,N_l},_\theta[\mathrm{G}_\theta]]\,\nu(d\theta)$ 和 $\int_\Theta \mathrm{P}_{T,\theta}^{l}[\mathrm{G}_\theta]\,\nu(d\theta)$ 均远离零。这里，$\int_\Theta \ddot{\mathrm{A}}\mathbb{E}_\theta^{N_l}[\mathrm{P}_T^{l,N_l},_\theta[\mathrm{G}_\theta]] - \mathrm{P}_{T,\theta}^{l}[\mathrm{G}_\theta]\ddot{\mathrm{a}}\nu(d\theta)$ 表示粒子近似误差，根据 Awadelkarim & Jasra (2024) 的引理 A.7，可将其界定为 $O(N_l^{-1/2})$；$\int_\Theta \left(\mathrm{P}_{T,\theta}^{l}[\varphi\cdot\mathrm{G}_\theta] - \mathrm{P}_T^{l-1},_\theta[\varphi\cdot\mathrm{G}_\theta]\right)\nu(d\theta)$ 表示时间离散弱误差，在精确概率律下可将其界定为 $O(\Delta_l)$。因此，(S.37) 中的量为 $O(1)\cdot O(N_l^{-1/2})\cdot O(\Delta_l) = O\ddot{\mathrm{A}}\frac{\Delta_l}{\sqrt{N_l}}\ddot{\mathrm{a}}$ 阶。

$\mathcal{R}_{\mathrm{rem}}$ 中的另外 3 项可以用类似的方法处理，每一项的贡献为$O(\Delta_l/\sqrt{N_l})$。当$\Delta_l \to 0$时，它们的阶 $O(\sqrt{\Delta_l}/\sqrt{N_l})$ 均由主项的阶所支配。对 $l \in \{l_\star+1,\dots,L\}$ 求和即完成证明。□

引理 S.2。在假设 *(A1)–(A2)* 下，对于任意测试函数 $\varphi \in \mathcal{C}_b^2(\Theta\times\mathbb{R}^{dT}) \cap \mathcal{B}_b(\Theta\times\mathbb{R}^{dT})$，存在一个有限常数 C 使得对于任意 $(l,N_l)\in\mathbb{N}^2$，

$$\left|\overline{\pi}^{l,N_l}[\varphi] - \overline{\pi}^{l}[\varphi]\right| \le \frac{C}{\sqrt{N_l}}. \tag{S.38}$$

证明。回顾引理 S.1 的证明中的后验表示：

$$\overline{\pi}^{l}[\varphi] = \frac{\int_\Theta \mathcal{P}_{T,\theta}^{l}[\varphi\cdot\mathcal{G}_\theta]\,\nu(d\theta)}{\int_\Theta \mathcal{P}_{T,\theta}^{l}[\mathcal{G}_\theta]\,\nu(d\theta)} \quad\text{and}\quad \overline{\pi}^{l,N_l}[\varphi] = \frac{\int_\Theta \mathbb{E}_\theta^{N_l}[\mathcal{P}_{T,\theta}^{l,N_l}[\varphi\cdot\mathcal{G}_\theta]]\,\nu(d\theta)}{\int_\Theta \mathbb{E}_\theta^{N_l}[\mathcal{P}_{T,\theta}^{l,N_l}[\mathcal{G}_\theta]]\,\nu(d\theta)},$$

其中，$\mathcal{G}_\theta = \prod_{k=1}^T g_\theta(x_k,y_k)$ 和 $\mathcal{P}_{T,\theta}^{l}[\cdot], \mathcal{P}_T^{l,N_l},_{,\theta}[\cdot]$ 是其中定义的路径泛函。证明分为以下三个步骤进行。

步骤 1：比率分解。为分析差值 $\pi^{l,\ N_l}[\varphi] - \pi^{l}[\varphi]$，我们引入简化记号。设

$$\begin{aligned} a &:= \int_\Theta \mathbb{E}_\theta^{N_l}[\mathcal{P}_{T,\theta}^{l,N_l}[\varphi\cdot\mathcal{G}_\theta]]\,\nu(d\theta), \quad A := \int_\Theta \mathbb{E}_\theta^{N_l}[\mathcal{P}_{T,\theta}^{l,N_l}[\mathcal{G}_\theta]]\,\nu(d\theta),\\ c &:= \int_\Theta \mathcal{P}_{T,\theta}^{l}[\varphi\cdot\mathcal{G}_\theta]\,\nu(d\theta), \quad \text{and} \quad C := \int_\Theta \mathcal{P}_{T,\theta}^{l}[\mathcal{G}_\theta]\,\nu(d\theta). \end{aligned}$$

于是 $\pi^{l,\ N_l}[\varphi] = a/A$ 以及 $\pi^{l}[\varphi] = c/C$。利用代数恒等式

$$\frac{a}{A} - \frac{c}{C} = \frac{a-c}{C} + \frac{a(C-A)}{AC},$$

我们将差值分解为两项：

$$\pi^{l,N_l}[\varphi] - \pi^{l}[\varphi] = \underbrace{\frac{1}{C}(a-c)}_{\text{Term 1}} + \underbrace{\frac{a}{AC}(C-A)}_{\text{Term 2}}. \tag{S.39}$$

步骤 2：界定策略。 (S.39) 中的两项都由同一个量控制。根据假设 (A2)，分母 C 和 AC 均远离零。由于 φ 有界（$\|\varphi\|_\infty < \infty$），因此对于某个常数 $|a| \le C'$，有 C'。因此，两项都可以用归一化常数的粒子近似误差来界定：

$$|C - A| = \left| \int_\Theta \mathcal{P}^l_{T,\theta}[\mathcal{G}_\theta]\, \nu(d\theta) - \int_\Theta \mathbb{E}^{N_l}_\theta[\mathcal{P}^{l,N_l}_{T,\theta}[\mathcal{G}_\theta]]\, \nu(d\theta) \right|. \tag{S.40}$$

步骤 3：通过耦合得到的粒子近似误差。为了界定 (S.40)，我们构造一个概率空间，使其上存在两个耦合过程：

- $X^{l,\,1:N_l}_{1:\,T}$：层级 l 上含有 N_l 个粒子的粒子近似路径，
- $X^l_{1:\,T}$：层级 l 上的精确分布路径，

由共同的布朗增量驱动。将联合期望记为 $\widetilde{\mathbb{E}}^{N_l}_\theta$。那么

$$|C - A| \le C \left| \int_\Theta \widetilde{\mathbb{E}}^{N_l}_\theta \left[\prod_{k=1}^T g_\theta(X^{l,N_l}_k, y_k) - \prod_{k=1}^T g_\theta(X^l_k, y_k) \right] \nu(d\theta) \right| = O\left(\frac{1}{\sqrt{N_l}} \right),$$

由于 $\mathcal{G}_\theta \in \mathcal{C}^2_b \cap \mathcal{B}_b$，根据Awadelkarim & Jasra(2024)中的引理 A.7，即得所需界。

将项 1 和项 2 的结果相加，得到 $O\ddot{A}^{1}_{\sqrt{N_l}}\ddot{a} + \underline{O\ddot{A}}^{1}_{\sqrt{N_l}}\ddot{a} = \underline{O\ddot{A}}^{1}_{\sqrt{N_l}}\ddot{a}$，从而完成证明。

□

引理 S.3. 在假设 *(A1)–(A2)* 下，对于任意测试函数 $\varphi \in \mathcal{C}^2_b(\Theta \times \mathbb{R}^{dT}) \cap \mathcal{B}_b(\Theta \times \mathbb{R}^{dT})$，存在一个有限常数 C，使得对于任意 $(l, N_l) \in \mathbb{N}^2$，

$$\int_{\Theta \times (E_l \times E_{l-1})^T} \left(\Psi(\theta, x^l_{1:T}, \tilde{x}^{l-1}_{1:T}) \right)^2 \check{\pi}^{l,N_l}(d(\theta, u^l_{1:T}, \tilde{u}^{l-1}_{1:T})) \le C\Delta_l, \tag{S.41}$$

其中 Ψ 是定义如下的加权差分泛函

$$\Psi(\theta, x^l_{1:T}, \tilde{x}^{l-1}_{1:T}) := \varphi(\theta, x^l_{1:T}) \prod_{k=1}^T \check{H}_{k,\theta}(x^l_k, \tilde{x}^{l-1}_k) - \varphi(\theta, \tilde{x}^{l-1}_{1:T}) \prod_{k=1}^T \check{H}_{k,\theta}(\tilde{x}^{l-1}_k, x^l_k), \tag{S.42}$$

其中，$\check{H}_{k,\theta}(x, x') = \underline{g_\theta(x, y_k)}\, H_{k,\theta}(x, x')$以及$H_{k,\theta}(x, x') = \frac{1}{2}(g_\theta(x, y_k) + g_\theta(x', y_k))$在正文中定义。

证明。 耦合后验 $\check{\pi}^{l,N_l}$ 在（3.7）中定义，由层级 l 和 $l-1$ 上由共同布朗增量驱动的同步欧拉–丸山离散化构造而成。这种耦合机制确保轨迹 $(x^l_{1:T}, \tilde{x}^{l-1}_{1:T})$ 呈现强相关性，这对于多层框架中的方差缩减至关重要。证明分为以下两个步骤。

步骤 1. 由于 $\varphi \in \mathcal{C}^2_b$ 和 $\check{H}_{k,\theta}$ 在假设 (A2) 下有界，复合函数 Ψ 存在一阶泰勒展开。对于耦合轨迹 $(x^l_{1:T}, \tilde{x}^{l-1}_{1:T})$，标准的中值估计给出

$$|\Psi(\theta, x^l_{1:T}, \tilde{x}^{l-1}_{1:T})| \le C \sum_{k=1}^{T} \|x^l_k - \tilde{x}^{l-1}_k\|,$$

其中常数取决于 $\|\varphi\|_{\mathcal{C}^1_b}$ 和 $\sup_{k,\theta} \|\check{H}_{k,\theta}\|_{\mathcal{C}^1_b}$。

将两边平方，并对 $\check{\pi}^{l,N_l}$ 取期望：

$$\mathbb{E}_{\check{\pi}^{l,N_l}}[\Psi^2] \le C \mathbb{E}_{\check{\pi}^{l,N_l}} \left[\left(\sum_{k=1}^{T} \|x^l_k - \tilde{x}^{l-1}_k\| \right)^2 \right].$$

根据柯西–施瓦茨不等式，

$$\begin{aligned}
\mathbb{E}_{\check{\pi}^{l,N_l}} \left[\left(\sum_{k=1}^{T} \|x^l_k - \tilde{x}^{l-1}_k\| \right)^2 \right] &\le T \sum_{k=1}^{T} \mathbb{E}_{\check{\pi}^{l,N_l}}[\|x^l_k - \tilde{x}^{l-1}_k\|^2] \\
&= T \sum_{k=1}^{T} \frac{\mathbb{E}_{l,N_l}[\|x^l_k - \tilde{x}^{l-1}_k\|^2 \prod_{k=1}^{T} H_{k,\theta}(\tilde{x}^{l-1}_k, x^l_k)]}{\mathbb{E}_{l,N_l}[\prod_{k=1}^{T} H_{k,\theta}(\tilde{x}^{l-1}_k, x^l_k)]} \\
&\le CT \sum_{k=1}^{T} \mathbb{E}_{l,N_l}[\|x^l_k - \tilde{x}^{l-1}_k\|^2].
\end{aligned}$$

其中常数 C 由 $\prod_{k=1}^{T} H_{k,\theta}(\tilde{x}^{l-1}_k, x^l_k)$ 在假设 (A2) 下满足的界得出，期望是关于耦合轨迹序列 $\prod_{k=1}^{T} \check{P}^{N_l}_{\theta}\big((x^l_{k-1}, \tilde{x}^{l-1}_{k-1}),\ d(u^l_k, \tilde{u}^{l-1}_k)\big)$的联合分布取的。

步骤 2. 对于相邻层级 l 和 $l-1$ 上共享相同布朗增量的同步 Euler–Maruyama 离散化，以及相邻层级 l 和 $l-1$ 上共享相同布朗增量的同步 Euler–Maruyama 离散化，我们有

$$\mathbb{E}_{l,N_l}[\|x^l_t - \tilde{x}^{l-1}_t\|^2] \le C\Delta_l, \qquad t \in \{1, \dots, T\},$$

这可通过采用建立Awadelkarim & Jasra(2024）引理 A.7 时所用的证明策略得出。该结果考虑了：

- 层级 l 和 $l-1$上的耦合 Euler-Maruyama 离散化，其中两个过程由同一个布朗运动驱动；
- 采用N_l 个粒子对 McKean–Vlasov 定律 $\mu^{l,N_l}_{t,\theta}$和 $\mu^{l-1,N_l}_{\theta,t}$进行粒子近似，其中相互作用核 $\zeta_{1,\theta}$ （漂移项中）和 $\zeta_{2,\theta}$ （扩散项中）均采用同步近似；

- 耦合离散化的标准强收敛速率 $O(\Delta_l)$ ，在假设（1）和耦合结构成立的条件下适用。

结合上述两个步骤可得

$$\mathbb{E}_{\bar{\pi}^{l,N_l}}[\Psi^2] \le C \cdot T \sum_{t=1}^{T} \Delta_l = CT^2\Delta_l = O(\Delta_l),$$

由此完成证明。 □

引理 S.4.在假设（*A1*）–（*A3*）下，对于任意测试函数 $\varphi \in C_b^2(\Theta \times \mathbb{R}^{dT}) \cap B_b(\Theta \times \mathbb{R}^{dT})$，存在有限常数 C，使得对于任意$(l_\star, L, N_{l_\star}, K_{l_\star}, \dots, N_L, K_L) \in \mathbb{N}^{2(L-l_\star)+4}$且满足 $l_\star < L$，

$$\sum_{l=l_\star+1}^{L} \mathbb{E}\left[\left(\overline{\pi}^{l,N_l,K_l}[\varphi] - \overline{\pi}^{l-1,N_l,K_l}[\varphi] - \{\overline{\pi}^{l,N_l}[\varphi] - \overline{\pi}^{l-1,N_l}[\varphi]\}\right)^2\right] \le C \sum_{l=l_\star+1}^{L} \frac{\Delta_l}{K_l+1}. \tag{S.43}$$

证明。 固定层级 $l \in \{l_\star+1, \dots, L\}$。证明分以下四步进行。

步骤 1：比率恒等式与分解。双层 MCMC 估计量$\pi^{l,\ N_l,\ K_l}[\varphi] - \pi^{l-1,\ N_l,\ K_l}[\varphi]$，由方程（3.10）可知，该估计量包含四个比率项。应用引理S.1中的代数恒等式（方程S.33），我们将差值分解为六项（两个主项加四个余项）

$$\frac{a}{A} - \frac{b}{B} - \frac{c}{C} + \frac{d}{D},$$

其中

$$\frac{a}{A} = \overline{\pi}^{l,N_l,K_l}[\varphi], \quad \frac{b}{B} = \overline{\pi}^{l-1,N_l,K_l}[\varphi], \quad \frac{c}{C} = \overline{\pi}^{l,N_l}[\varphi], \quad \text{and} \quad \frac{d}{D} = \overline{\pi}^{l-1,N_l}[\varphi].$$

对于**主项**，我们考虑

$$\frac{1}{A}[(a-b)-(c-d)],$$

它表示耦合后验下的差分之差 $\check{\pi}^{l,\,N_l}$ 定义见方程（3.7）。

步骤 2：通过遍历理论控制主项。回顾式（S.42）中定义的加权差分泛函：

$$\Psi(\theta, x^l_{1:T}, \tilde{x}^{l-1}_{1:T}) = \varphi(\theta, x^l_{1:T}) \prod_{k=1}^{T} \check{H}_{k,\theta}(x^l_k, \tilde{x}^{l-1}_k) - \varphi(\theta, \tilde{x}^{l-1}_{1:T}) \prod_{k=1}^{T} \check{H}_{k,\theta}(\tilde{x}^{l-1}_k, x^l_k).$$

主项对应于使用双层马尔可夫链蒙特卡罗链 $\{V_i\}_{i=1}^{K_l}$ 估计 $\mathbb{E}_{\check{\pi}^{l,\,N_l}}[\Psi]$ 时的马尔可夫链蒙特卡罗采样误差。该链的状态空间为 $\Theta \times (E_l \times E_{l-1})^T$，其中每个 $V_i = (\theta_i,\ x^l_{i,\,1:T},\ \tilde{x}^{l-1}_{i,\,1:T})$ 表示第 i 次迭代时的一个耦合样本。

根据假设（A3），双层马尔可夫链蒙特卡罗核 $\check{R}_l$ 是一致遍历的，其不变测度为 $\check{\pi}^{l,\,N_l}$。应用一致遍历马尔可夫链的标准结果（Jasra 等人 (2018)，命题A.1），经过 K_l 次迭代可得

$$\mathbb{E}\left[\left(\frac{1}{K_l}\sum_{i=1}^{K_l}\Psi(V_i) - \mathbb{E}_{\check{\pi}^{l,N_l}}[\Psi]\right)^2\right] \le C\frac{\mathrm{Var}_{\check{\pi}^{l,N_l}}(\Psi)}{K_l+1},$$

其中 $\mathrm{Var}_{\check{\pi}^{l,\,N_l}}(\Psi) = \mathbb{E}_{\check{\pi}^{l,\,N_l}}[\Psi^2] - (\mathbb{E}_{\check{\pi}^{l,\,N_l}}[\Psi])^2$ 表示在 $\check{\pi}^{l,\,N_l}$ 下的方差。根据引理S.3，有 $\mathbb{E}_{\check{\pi}^{l,\,N_l}}[\Psi^2] \le C\Delta_l$，因此主项的二阶矩可由 $C\frac{\Delta_l}{K_l+1}$. 控制

步骤 3：余项控制。考虑方程（S.33）中的一个代表性余项，

$$\mathcal{R}_{\text{example}} = -\frac{1}{AC}(A-C)(c-d),$$

其对应的显式形式为

$$\begin{aligned}\mathcal{R}_{\text{example}} = &\left(\frac{1}{K_l+1}\sum_{k=0}^{K_l}\mathcal{P}^{l,N_l}_{T,\theta^{(k)}}[\mathcal{G}_{\theta^{(k)}}]\right)^{-1}\left(\int_\Theta \mathbb{E}^{N_l}_\theta[\mathcal{P}^{l,N_l}_{T,\theta}[\mathcal{G}_\theta]]\,\nu(d\theta)\right)^{-1}\\ &\times\left(\frac{1}{K_l+1}\sum_{k=0}^{K_l}\mathcal{P}^{l,N_l}_{T,\theta^{(k)}}[\mathcal{G}_{\theta^{(k)}}] - \int_\Theta \mathbb{E}^{N_l}_\theta[\mathcal{P}^{l,N_l}_{T,\theta}[\mathcal{G}_\theta]]\,\nu(d\theta)\right)\\ &\times\left(\int_\Theta\left(\mathbb{E}^{N_l}_\theta[\mathcal{P}^{l,N_l}_{T,\theta}[\varphi\cdot\mathcal{G}_\theta]] - \mathbb{E}^{N_l}_\theta[\mathcal{P}^{l-1,N_l}_{T,\theta}[\varphi\cdot\mathcal{G}_\theta]]\right)\nu(d\theta)\right).\end{aligned}$$

该项涉及三个因素：

- **分母**：根据假设 (A2)，$(AC)^{-1}$ 的阶为 $O(1)$。
- **MCMC误差 ($A-C$)**：MCMC经验平均值（经过 K_l 次迭代）与基于 N_l 粒子近似概率律的精确期望之间的归一化常数差。根据一致遍历性假设，该项的MCMC抽样方差为 $O((K_l+1)^{-1})$。
- **耦合粒子近似弱误差($c-d$)**：两个离散化层级 l 和 $l-1$上的期望之差，两者均使用N_l 个粒子进行近似。这表示离散化的弱误差；在 N_l粒子近似概率律下，该误差有界于 $O(\Delta_l)$。

应用柯西–施瓦茨不等式可得：

$$\begin{aligned}\mathbb{E}[|\mathcal{R}_{\text{example}}|^2] &= \cdot O(1)\mathbb{E}[(A-C)^2]\cdot|c-d|^2 \\ &= O(1)\cdot O\left(\frac{1}{K_l+1}\right)\cdot O(\Delta_l^2) \\ &= O\left(\frac{\Delta_l^2}{K_l+1}\right).\end{aligned}$$

由于在多层框架中，$\Delta_l \to 0$ 随 l 增大，因此相较于主项的界 $O!\,(\Delta_l(K_l+1)^{-1})$，该项在 Δ_l 中的阶更高，故可由后者控制。

通过类似分析，所有其他余项也可得到类似的界，每项至多为 $O(\Delta_l^2(K_l+1)^{-1})$，与主项相比均可忽略不计。

步骤4：对层级求和。

结合步骤2和步骤3，并应用六次柯西–施瓦茨不等式（主项应用一次，五个余项各应用一次），我们得到对每个层级 $l \in l_\star+1, \ldots\ldots, L$：

$$\mathbb{E}\left[\left(\overline{\pi}^{l,N_l,K_l}[\varphi]-\overline{\pi}^{l-1,N_l,K_l}[\varphi]-\{\overline{\pi}^{l,N_l}[\varphi]-\overline{\pi}^{l-1,N_l}[\varphi]\}\right)^2\right] \le C\frac{\Delta_l}{K_l+1}.$$

对 $l=l_\star+1, \ldots\ldots, L$ 求和，即完成证明。 □

引理 S.5. 在假设 *(A1)–(A3)*下，对于任意测试函数 $\varphi \in C_b^2(\Theta\times\mathbb{R}^{dT})\cap B_b(\Theta\times\mathbb{R}^{dT})$，存在一个有限常数 $C<+\infty$，使得对于任意$(l_\star, L, N_{l_\star}, K_{l_\star}, \ldots, N_L, K_L)\in\mathbb{N}^{2(L-l_\star)+4}$ 满足 $l_\star<L$，

$$\left|\sum_{(l,q)\in\mathcal{A}_{l_\star,L}}\mathbb{E}\left[\left\{\overline{\pi}^{l,N_l,K_l}[\varphi]-\overline{\pi}^{l-1,N_l,K_l}[\varphi]-\left(\overline{\pi}^{l,N_l}[\varphi]-\overline{\pi}^{l-1,N_l}[\varphi]\right)\right\}\right.\right.$$

$$\times \left\{ \overline{\pi}^{q,N_q,K_q}[\varphi] - \overline{\pi}^{q-1,N_q,K_q}[\varphi] - \left(\overline{\pi}^{q,N_q}[\varphi] - \overline{\pi}^{q-1,N_q}[\varphi] \right) \right\} \Bigg] \Bigg|$$

$$\leq C \sum_{(l,q)\in\mathcal{A}_{l_\star,L}} \frac{\sqrt{\Delta_l}}{K_l+1} \cdot \frac{\sqrt{\Delta_q}}{K_q+1},$$

其中 $\mathcal{A}_{l_\star,L} = \{(l,q) \in \{l_\star+1,\ldots,L\}^2 : l \neq q\}$。

证明。 我们控制多层估计量中产生的跨层协方差项。对于不同的离散化层级，相应双层马尔可夫链蒙特卡洛链的独立性使这些项可以因式分解，从而便于直接进行界定。证明分以下四步进行。

步骤 1：问题设置与独立性结构。 对于任意 $(l,q) \in \mathcal{A}_{l_\star,L}$，满足 $l \neq q$，考虑以下跨层期望：

$$\mathbb{E}\Bigg[\left\{ \overline{\pi}^{l,N_l,K_l}[\varphi] - \overline{\pi}^{l-1,N_l,K_l}[\varphi] - \left(\overline{\pi}^{l,N_l}[\varphi] - \overline{\pi}^{l-1,N_l}[\varphi] \right) \right\}$$
$$\times \left\{ \overline{\pi}^{q,N_q,K_q}[\varphi] - \overline{\pi}^{q-1,N_q,K_q}[\varphi] - \left(\overline{\pi}^{q,N_q}[\varphi] - \overline{\pi}^{q-1,N_q}[\varphi] \right) \right\} \Bigg].$$

根据算法 A.2，层级 l 和 q 上的双层马尔可夫链蒙特卡洛链独立运行，并使用相互独立的随机性（布朗运动和重采样索引）。由独立性可知，期望可因式分解为：

$$\mathbb{E}\left[\text{Cross-term}_{l,q}\right] = \mathbb{E}\left[\overline{\pi}^{l,N_l,K_l}[\varphi] - \overline{\pi}^{l-1,N_l,K_l}[\varphi] - \left(\overline{\pi}^{l,N_l}[\varphi] - \overline{\pi}^{l-1,N_l}[\varphi] \right) \right] \quad \text{(S.44)}$$
$$\times \mathbb{E}\left[\overline{\pi}^{q,N_q,K_q}[\varphi] - \overline{\pi}^{q-1,N_q,K_q}[\varphi] - \left(\overline{\pi}^{q,N_q}[\varphi] - \overline{\pi}^{q-1,N_q}[\varphi] \right) \right].$$

S骤 2：界定各个偏差项。定义层级 l a上的偏差 s:

$$B_l := \mathbb{E}\left[\overline{\pi}^{l,N_l,K_l}[\varphi] - \overline{\pi}^{l-1,N_l,K_l}[\varphi] \right] - \left(\overline{\pi}^{l,N_l}[\varphi] - \overline{\pi}^{l-1,N_l}[\varphi] \right).$$

按照引理 S.1证明中的比率分解方法，我们采用与公式（S.33）一致的记号。对于层级 l，定义

- $A :=$ $\frac{1}{K_l}\sum_{i=1}^{K_l} \prod_{k=1}^{T} \check{}Hk, \theta_{(i)}(x^{l,i}k, \tilde{x}_{,l-1,}{}^{i}k)$，细层马尔可夫链蒙特卡洛归一化常数；
- $a :=$ $\frac{1}{K_l}\sum i = 1^{K_l}\varphi(\theta_{(i)}\ {}_{x}{}^{l,i}1 : T)\ \prod k = 1^{T}\check{}H_{k,\theta_{(i)}}(x^{l}{}_{,k}{}^{i}, \tilde{x}_{,l-1\,i}{}_{k},)$, 细层马尔可夫链蒙特卡洛加权期望；
- B， b：粗层级（$l-1$）马尔可夫链蒙特卡洛对应的量；

- C， c：$\check{}\pi^{l,\ N_l}$下对应的细层精确期望；
- D， d：相应的粗层精确期望，在$\check{}\pi^{l,\ N_l}$下 .

双层差值可表示为：

$$\overline{\pi}^{l,N_l,K_l}[\varphi] - \overline{\pi}^{l-1,N_l,K_l}[\varphi] = \frac{a}{A} - \frac{b}{B}.$$

应用引理 S.1中的比值恒等式，结合等式（S.33），并注意到$\frac{c}{C} = \pi^{l,\ N_l}[\varphi]$ 以及$\frac{-d}{D} = \overline{\pi}^{l-1,\ N_l}[\varphi]$，可将差值分解为

$$\frac{a}{A} - \frac{b}{B} - \frac{c}{C} + \frac{d}{D}$$

六项——两个主项和四个余项。

P主项分析： 以等式（S.33）为例，我们取主项

$$T_1 = \frac{1}{A}[(a-b) - (c-d)]$$

为例。我们通过减去再加上 $\frac{1}{C}[(a-b) - (c-d)]$ 来重写该项：

$$T_1 = \left(\frac{1}{A} - \frac{1}{C}\right)[(a-b) - (c-d)] + \frac{1}{C}[(a-b) - (c-d)].$$

取期望，并注意到在平稳性下 $\mathbb{E}[(a-b) - (c-d)] = 0$，第二项为零：

$$\mathbb{E}[T_1] = \mathbb{E}\left[\left(\frac{1}{A} - \frac{1}{C}\right)[(a-b) - (c-d)]\right].$$

应用柯西–施瓦茨不等式可得

$$|\mathbb{E}[T_1]| \le \sqrt{\mathbb{E}\left[\left(\frac{1}{A} - \frac{1}{C}\right)^2\right]} \cdot \sqrt{\mathbb{E}[((a-b) - (c-d))^2]}.$$

第一个因子：项 $\left(\frac{1}{A} - \frac{1}{C}\right)$ 表示对归一化常数进行估计时产生的 MCMC误差。根据一致遍历马尔可夫链的标准结果（假设 A3）以及归一化常数的统一界 C' （假设 A2），

$$\sqrt{\mathbb{E}\left[\left(\frac{1}{A} - \frac{1}{C}\right)^2\right]} \le \frac{C'}{\sqrt{K_l + 1}}.$$

第二个因子：该项表示通过马尔可夫链蒙特卡罗采样得到的耦合双层差的方差。结合引理 S.4 （耦合轨迹的方差界）与假设 A3中的 MCMC混合性质，可得

$$\mathbb{E}[((a-b)-(c-d))^2] \leq C\frac{\Delta_l}{K_l+1}.$$

因此，

$$\sqrt{\mathbb{E}[((a-b)-(c-d))^2]} \leq C\frac{\sqrt{\Delta_l}}{\sqrt{K_l+1}}.$$

结合这两个因素可得

$$|\mathbb{E}[T_1]| \leq C' \cdot \frac{1}{\sqrt{K_l+1}} \cdot \frac{\sqrt{\Delta_l}}{\sqrt{K_l+1}} = C'\frac{\sqrt{\Delta_l}}{K_l+1}.$$

余项分析： 我们考虑余项

$$R_3 = -\frac{1}{AC}(A-C)(c-d).$$

我们通过减去并加上 $\frac{1}{C^2}(A-C)(c-d)$ 来改写：

$$R_3 = \left(\frac{1}{AC} - \frac{1}{C^2}\right)(A-C)(c-d) + \frac{1}{C^2}(A-C)(c-d).$$

取期望，并注意到 c， d，且 C 相对于 MCMC随机性是确定的，而 $\mathbb{E}[A-C]=0$ 在平稳性下成立，第二项为零：

$$\mathbb{E}\left[\frac{1}{C^2}(A-C)(c-d)\right] = \frac{c-d}{C^2}\mathbb{E}[A-C] = 0.$$

因此，期望由第一项决定。利用恒等式 $\frac{1}{AC} - \frac{1}{C^2} = \frac{C-A}{AC^2}$：

$$\mathbb{E}[R_3] = \mathbb{E}\left[-\frac{1}{AC^2}(A-C)^2(c-d)\right].$$

取绝对值，并注意到归一化常数与零保持一定距离（假设 A2）：

$$|\mathbb{E}[R_3]| \leq C|c-d|\mathbb{E}[(A-C)^2].$$

弱误差因子： 项 $|c-d|$ 表示时间离散化的弱误差。根据 Euler–Maruyama格式的标准误差分析，

$$|c-d| \le C\Delta_l.$$

方差因子：项 $\mathbb{E}[(A-C)^2]$ 表示马尔可夫链蒙特卡洛估计量的方差。根据一致遍历性（假设 A3）以及 Roberts & Rosenthal(1997）中的推论 2.1，

$$\mathbb{E}[(A-C)^2] \le \frac{C'}{K_l+1}.$$

结合这些界：

$$|\mathbb{E}[R_3]| \le C \cdot \Delta_l \cdot \frac{1}{K_l+1}.$$

将此与主项界 $O(\frac{\sqrt{\Delta_l}}{K_l+1})$ 比较，并注意到当 $\Delta_l \to 0$ 时 $\Delta_l \ll \sqrt{\Delta_l}$，余项严格占优。$\mathcal{R}_{\text{rem}}$ 中的其余项也可以用类似方法分析。

*Co*合并界：将主项与所有余项相加，得到

$$|B_l| \le O\left(\frac{\sqrt{\Delta_l}}{K_l+1}\right).$$

类似地，对于层级 q，

$$|B_q| \le C\frac{\sqrt{\Delta_q}}{K_q+1}.$$

步骤 3：合并界。

$$|\mathbb{E}\left[\text{Cross-term}_{l,q}\right]| = |B_l| \cdot |B_q| \le C\frac{\sqrt{\Delta_l}}{K_l+1} \cdot C\frac{\sqrt{\Delta_q}}{K_q+1} = C\frac{\sqrt{\Delta_l\Delta_q}}{(K_l+1)(K_q+1)}.$$

对所有不同的对 $(l,q) \in \mathcal{A}_{l_\star,L}$ 求和，

$$\sum_{(l,q)\in\mathcal{A}_{l_\star,L}} |\mathbb{E}\left[\text{Cross-term}_{l,q}\right]| \le C \sum_{(l,q)\in\mathcal{A}_{l_\star,L}} \frac{\sqrt{\Delta_l}}{K_l+1} \cdot \frac{\sqrt{\Delta_q}}{K_q+1},$$

由此得到所述界。 □