arXiv:2301.00900v1 [stat.ME] 2 Jan 2023

利用 PARIS 粒子Gibbs进行状态与参数学习

Gabriel Cardoso[†]、Yazid Janati El Idrissi[‡]、Sylvain Le Corff[⋆]、 ´Eric Moulines[†]和 Jimmy Olsson[⊤]

[†]CMAP， ´巴黎理工学院 École Polytechnique，巴黎理工学院，帕莱索。
[‡]Samovar，T´él´écom SudParis，CITI系，TIPIC，巴黎理工学院，帕莱索。[⋆]LPSM，索邦大学，CNRS 8001联合研究单位，居斯ieu广场4号，巴黎75005。[⊤]瑞典斯德哥尔摩，KTH皇家理工学院数学系。

摘要

非线性状态空间模型又称一般隐马尔可夫模型，在统计机器学习中无处不在，是序列数据及一般序列数据最经典的生成模型。基于粒子的快速增量平滑器（PARIS）是一种序贯蒙特卡洛（SMC）技术，可对这类模型中平滑分布下加性泛函的期望进行高效在线近似。这类期望自然出现在多种学习场景中，例如似然估计（最大似然估计）和马尔可夫得分爬升（MSC）。PARIS 具有线性计算复杂度、较低的内存需求，并具备非渐近界、收敛结果和稳定性保证。不过，由于采用自归一化重要性采样，PARIS估计量存在偏差。我们的首项贡献是设计一种新颖的加性平滑算法——Parisian 粒子 Gibbs（PPG）采样器。该采样器可视为由条件SMC移动驱动的 PARIS 算法，从而降低目标量估计值的偏差。我们通过理论结果为 PPG 算法提供支撑，包括新的偏差和方差界以及偏差不等式。我们的第二项贡献是在学习框架中应用 PPG，其中最大似然估计和 MSC 是其特例。在此背景下，我们在标准假设下建立了非渐近界，凸显了偏差降低和 PPG 隐式 Rao–Blackwell化的价值。这是该领域中首批此类非渐近结果。我们通过数值实验展示理论结果，并验证了相关论断。

1. 引言

序贯蒙特卡洛 (SMC)方法，或粒子滤波器，是基于模拟的方法，用于在状态空间模型的贝叶斯推断中在线近似后验分布。对于非线性隐马尔可夫模型 (HMM)，这些方法已成功用于在线近似给定观测值 $(Y_{t_1}, \ldots, Y_{t_2})$ 时，未观测状态序列 $(X_{s_1}, \ldots, X_{s_2})$ 的典型不可处理的后验分布，其中 $0 \leq s_1 \leq s_2$ 和 $0 \leq t_1 \leq t_2$。标准 SMC 方法使用通过序贯重要性采样和重采样步骤递归生成的蒙特卡洛样本。粒子滤波器通过一系列与蒙特卡洛样本序列 $\{\xi_t^i\}_{i=1}^N$, $t \in \mathbb{N}$ 相关联的占据测度，近似边缘后验分布的演化；其中每个粒子 ξ_t^i 都是隐过程状态空间中的一个随机抽样。粒子滤波器围绕两种操作展开：选择步骤，分别复制／丢弃重要性权重较大／较小的粒子；以及变异步骤，使所选粒子在状态空间中随机演化。交替并迭代地应用选择和变异，会产生在时间和空间上均相互依赖的粒子群。HMM 的联合状态后验分布也可以解释为与某类马尔可夫反向动力学相关联的概率律；这种解释很有用，

例如，在设计用于非线性平滑的基于后向采样的粒子算法时[Douc 等人，2011，Del Moral 等人， 2010]。

多年来，随着粒子数 N 趋于无穷，已有若干收敛结果；例如， [Del Moral，2004，Douc 和 Moulines，2008，Capp´e 等人，以及其中的参考文献。此外，也已有不少非渐近结果，包括关于 SMC L$_p$ 误差和偏差的时间一致界，以及描述粒子间混沌传播的界。基于后向采样的粒子算法的扩展也可见于例如 [Douc 等人，2011，Del Moral 等，2010，Dubarry 和 Le Corff， 2013]。

本文重点研究如何递归计算加性泛函 h_t 的平滑期望 $\eta_{0:t}h_t = \mathbb{E}[h_t(X_{0:t}) \mid Y_{0:t}]$ ，其形式为

$$h_t(x_{0:t}) \coloneqq \sum_{s=0}^{t-1} \tilde{h}_s(x_{s:s+1}), \tag{1.1}$$

其中， $X_{0:n}$ 和 $Y_{0:n}$ 分别表示状态向量和观测值向量（精确定义见下文）。此类期望在非线性 HMM 的最大似然参数估计中经常出现，例如在计算得分函数（对数似然函数的梯度）或期望最大化的中间量时；参见 [Capp´e，2001，Andrieu 和 Doucet，2003，Poyiadjis 等，2005，Capp´e，2011，Poyiadjis 等人， 2011]。Olsson 和 Westerborn， 2017] 提出的基于粒子的快速增量平滑器（PARIS） [专为在线求解这一加性平滑问题而设计。当潜在状态的转移密度有上下界时，可以证明，该算法的计算复杂度随粒子数 N 线性增长，且内存需求有限。PARIS 的一个有趣特点是，它会根据粒子滤波器诱导的后向动力学即时进行抽样。为应对采样轨迹的退化并确保长期数值稳定，每个粒子需要进行两次或更多次后向抽样；其渐近方差仅随时间线性增长。

本文介绍了一种降低 $\eta_{0:t}h_t$ 的 PARIS 估计量偏差的方法。其思路是通过引入条件 *PARIS* 算法，将 PARIS 算法与基于向后采样的粒子 *Gibbs* 采样器 [（Andrieu 等，2010b；Lindsten 等，2014a；Chopin 和 Singh，2015a；Del Moral 等人，2016；Del Moral 和 Jasra， 2018]）结合起来。由此得到一种批处理模式的 *PARIS* 粒子 *Gibbs*（PPG）采样器，并为其给出了偏差上界：该上界与粒子数 N 成反比，并随粒子 Gibbs 迭代索引呈指数级快速下降（假设粒子 Gibbs 采样器是一致遍历的）。

作为一项应用，我们研究使用随机梯度进行似然最大化的问题。在这一特定情境下，平滑估计量会被反复用于生成平均场估计，因此控制偏差至关重要。因此，在方差不会爆炸的前提下，自然的目标是在固定计算预算下尽可能减小偏差。正因如此，随机模拟中的偏差降低在过去几十年里一直受到广泛研究[（Jacob 等人，2020；Glynn 和 Rhee， 2014]）。本文也为这一研究方向作出了贡献。具体而言，我们证明，采用 PPG 的随机逼近（SA）达到 $\mathcal{O}(\log(n)/\sqrt{n})$ 收敛速率，其中n 为 SA 步数。这一结果优于 [**?**], 先前的研究；该研究证明了基于带祖先采样（PGAS）的粒子 Gibbs 的 SA期望最大化（EM）算法几乎必然收敛（至似然函数的一个驻点）。

本文结构如下。第 2 节回顾隐马尔可夫模型框架、粒子滤波器和 PARIS 算法。第 3 节介绍 PPG 算法，并给出本文的第一个核心结果：估计量偏差的上界，该上界是粒子数和吉布斯算法迭代索引的函数。此外，我们还给出均方误差（MSE）的上界。第 4 节研究学习问题，并给出本文的第二个结果：在随机索引 K 处计算的梯度范数平方期望的一个 $\mathcal{O}(\log(n)/\sqrt{n})$ 非渐近界。第 5.1 节通过数值实验展示我们的结果。所有证明均汇集于补充材料中。

记号。给定可测空间 $(\mathsf{X}, \mathcal{X})$，其中 $\mathcal{X}$ 是可数生成的 σ-代数，我们用 $\mathrm{F}(\mathcal{X})$ 表示 X 上有界的 $\mathcal{X}/\mathcal{B}(\mathbb{R})$-可测函数的集合。对于任意 $h \in \mathrm{F}(\mathcal{X})$，令$\|h\|_\infty := \sup_{x\in\mathsf{X}} |h(x)|$，并令 $\operatorname{osc}(h) := \sup_{(x,x')\in\mathsf{X}^2} |h(x) - h(x')|$ 分别表示 h 的上确界范数和振荡范数。令 $\mathrm{M}(\mathcal{X})$ 表示 $(\mathsf{X}, \mathcal{X})$ 上 σ-有限测度的集合，$\mathrm{M}_1(\mathcal{X}) \subset \mathrm{M}(\mathcal{X})$ 表示概率测度。对于任意 $h \in \mathrm{F}(\mathcal{X})$ 和 $\mu \in \mathrm{M}(\mathcal{X})$，记 $\mu(f) = \int h(x)\mu(\mathrm{d}x)$。对于从 $(\mathsf{X}, \mathcal{X})$ 到另一个可测空间 $(\mathsf{Y}, \mathcal{Y})$ 的马尔可夫核 K，定义可测函数 $Kh : \mathsf{X} \ni x \mapsto \int h(y)K(x, \mathrm{d}y)$。复合 μK 是 $(\mathsf{Y}, \mathcal{Y})$ 上的概率测度，满足 $\mu K : \mathcal{X} \ni A \mapsto \int \mu(\mathrm{d}x)K(x, \mathrm{d}y)\mathbb{1}_A(y)$。对于所有序列 $\{a_u\}_{u\in\mathbb{Z}}$ 和 $\{b^u\}_{u\in\mathbb{Z}}$，以及所有 $s \le t$，记作$a_{s:t} = \{a_s, \ldots, a_t\}$ 和 $b^{s:t} = \{b^s, \ldots, b^t\}$。

2. 背景

2.1 隐马尔可夫模型

隐马尔可夫模型 由不可观测的状态过程 $\{X_t\}_{t\in\mathbb{N}}$ 和观测值 $\{Y_t\}_{t\in\mathbb{N}}$组成，其中在每个时间 $t \in \mathbb{N}$，假设不可观测状态 X_t 和观测值 Y_t 分别取值于某些一般的可测空间 $(\mathsf{X}_t, \mathcal{X}_t)$ 和 $(\mathsf{Y}_t, \mathcal{Y}_t)$。假设 $\{X_t\}_{t\in\mathbb{N}}$是一个具有转移核 $\{M_{t+1}\}_{t\in\mathbb{N}}$ 和初始分布 η_0的马尔可夫链。给定状态 $\{X_t\}_{t\in\mathbb{N}}$后，假设观测值 $\{Y_t\}_{t\in N}$ 相互独立，并且对于所有 $t \in \mathbb{N}$，观测值 Y_t 的条件分布仅取决于当前状态 X_t。假设该分布相对于某个参考测度具有密度 $g_t(X_t, \cdot)$。下文假设给定了一个固定的观测值序列 $\{y_t\}_{t\in\mathbb{N}}$，并滥用记号，对每个 $t \in \mathbb{N}$定义 $g_t(\cdot) = g_t(\cdot, y_t)$。对于 $0 \le s \le t$，记 $\mathsf{X}_{s:t} := \prod_{u=s}^t \mathsf{X}_u$ 和 $\mathcal{X}_{s:t} := \bigotimes_{u=s}^t \mathcal{X}_u$。考虑非归一化转移核

$$Q_s : \mathsf{X}_s \times \mathcal{X}_{s+1} \ni (x, A) \mapsto g_s(x)M_s(x, A) \tag{2.2}$$

并令

$$\gamma_{0:t} : \mathcal{X}_{0:t} \ni A \mapsto \int \mathbb{1}_A(x_{0:t})\, \eta_0(\mathrm{d}x_0) \prod_{s=0}^{t-1} Q_s(x_s, \mathrm{d}x_{s+1}). \tag{2.3}$$

利用这些量，我们可以定义时间 $t \in \mathbb{N}$ 的联合平滑 和预测分布 如下:

$$\eta_{0:t} : \mathcal{X}_{0:t} \ni A \mapsto \frac{\gamma_{0:t}(A)}{\gamma_{0:t}(\mathsf{X}_{0:t})}, \tag{2.4}$$

$$\eta_t : \mathcal{X}_t \ni A \mapsto \eta_{0:t}(\mathsf{X}_{0:t-1} \times A), \tag{2.5}$$

分别。可以证明（参见 [Capp´e 等，2005 年，第 3] 节），$\eta_{0:t}$ 和 η_t 分别是给定 $Y_{0:t-1}$ 时 $X_{0:t}$ 和 X_t 在 $y_{0:t-1}$ 处的条件分布。遗憾的是，这些分布对于贝叶斯平滑和滤波至关重要，因为它们能够利用观测数据流估计隐状态，但只有在线性高斯模型或具有有限状态空间的模型中，才能以闭式形式得到；全面介绍请参见 [Capp´e 等，2009]。

2.2 粒子滤波器

对于实践中大多数值得关注的模型，联合平滑分布和预测分布都难以解析求解，与这些分布相关的任何期望也同样如此。不过，通常可以使用基于预测递推$\eta_{t+1} = \eta_t Q_t/\eta_t g_t$的粒子方法高效地估计这些期望。在时间 t，假设我们手头已有对 η_t的一致粒子近似，它由 N个随机抽取的$\{\xi_t^i\}_{i=1}^N$个所谓的粒子在 X_t中构成，记为$\eta_t^N = N^{-1}\sum_{i=1}^N \delta_{\xi_t^i}$。将 η_t^N代入连接 η_{t+1} 与 η_t 的递推式，可得到混合分布$\eta_t^N Q_t$，从中可以抽取 N个新粒子，以构造 η_{t+1}^N。为此，我们对所有 $1 \le i \le N$抽取祖先

索引 $\alpha_t^i \sim \mathrm{Categorical}(\{g_t(\xi_t^\ell)\}_{\ell=1}^N)$，然后传播 $\xi_{t+1}^i \sim M_t(\xi_t^{\alpha_t^i}, \cdot)$。这一过程通过从 η_0 中独立抽取初始粒子 $\{\xi_0^i\}_{i=1}^N$ 进行初始化，描述了采用多项式重采样的粒子滤波器，并产生一致的估计量：对于每个 $h \in \mathsf{F}(\mathsf{X}_t)$，当粒子数 N 趋于无穷大时，$\eta_t^N(h)$ 几乎必然收敛到 $\eta_t(h)$。

这一过程也可以扩展，用于生成联合平滑分布 $\{\eta_{0:t}\}_{t\in\mathbb{N}}$ 的粒子近似。请注意，前文所述的连续祖先选择步骤会为每个终端粒子生成一条祖先谱系 ξ_t^i，记作 $\xi_{0:t}^i$。可以很容易地证明，$\eta_{0:t}^N = N^{-1}\sum_{i=1}^N \delta_{\xi_{0:t}^i}$ 构成联合平滑分布 $\eta_{0:t}$ 的粒子近似。然而，众所周知，相同的选择操作也会使祖先谱系逐渐减少，因为在每一步中，两个不同粒子很可能源自上一代中的同一个父粒子。因此，最终所有粒子都会有很大一部分初始祖先相同。这意味着在实践中，这种朴素方法（我们称之为**简易平滑器**）在用于估计依赖于整个状态轨迹的目标泛函的联合平滑期望时，通常会产生较大的方差。

2.3 反向平滑与 PARIS 算法

现在，我们讨论如何通过所谓的**反向抽样**来避免平滑问题中的粒子退化。尽管这条研究路线的适用范围更广，但为简化起见，我们仅考虑形式如下的**加性状态泛函**

$$h_t(x_{0:t}) \coloneqq \sum_{s=0}^{t-1} \tilde{h}_s(x_{s:s+1}), \quad x_{0:t} \in \mathsf{X}_{0:t}. \tag{2.6}$$

值得一提的是，使用上一节所述的简易平滑器，可以在粒子滤波器运行的同时在线平滑加性泛函：对每个 s，令

$$\eta_{0:s}^N h_s \coloneqq N^{-1}\sum_{i=1}^N \beta_s^i, \tag{2.7}$$

其中，统计量 $\{\beta_s^i\}_{i=1}^N$ 满足以下递推

$$\beta_{s+1}^i = \beta_s^{\alpha_s^i} + \tilde{h}_s(\xi_s^{\alpha_s^i}, \xi_{s+1}^i), \tag{2.8}$$

其中，α_s^i 是如前所述的粒子 ξ_{s+1}^i 在时间 s 的祖先。

如上所述，当 s 相对于 N 较大时，前述估计量的方差较高。不过，现在假设模型在**完全支配**意义下成立，即每个状态过程核 M_s 相对于某个参考测度都具有转移密度 m_s；那么，有趣的是，不难看出，给定后代粒子 ξ_{s+1}^i 和祖先 $\{\xi_s^\ell\}_{\ell=1}^N$ 时，$\alpha_s^i = j$ 的条件概率为

$$\mathbf{\Lambda}_s(i,j) \coloneqq \frac{\omega_s^j m_s(\xi_s^j, \xi_{s+1}^i)}{\sum_{\ell=1}^N \omega_s^\ell m_s(\xi_s^\ell, \xi_{s+1}^i)}. \tag{2.9}$$

这里的 $\mathbf{\Lambda}_s$ 构成了 $[\![1,\ N]\!] \times [\![1,\ N]\!]$ 上的反向马尔可夫转移核。利用这一观察，我们只需将朴素更新 (2.8) 替换为 Rao–Blackwell化对应项，即可完全避免简易平滑器中的粒子路径退化

$$\beta_{s+1}^i = \sum_{j=1}^N \mathbf{\Lambda}_s(i,j)\{\beta_s^j + \tilde{h}_s(\xi_s^j, \xi_{s+1}^i)\}. \tag{2.10}$$

这一方法由 [Del Moral 等人提出，2010]，通过对粒子进行平均，消除粒子之间的祖先关联，从而巧妙地避免了路径退化。此外，它完全

可以在线运行，因为在步骤 s 中，只需使用粒子群 $\xi_s^{1:N}$和 $\xi_{s+1}^{1:N}$ 即可执行更新。不过，一个显著的缺点是计算 $\beta_t^{1:N}$的总体复杂度为 $\mathcal{O}(N^2)$，因为式（2.10）中每个 β_{s+1}^i 的计算都涉及 N^2 项的计算；当粒子数量 N 很大时，这可能会难以承受。因此，在 [Olsson 和 Westerborn 的研究中， 2017], 作者提出从分布 $\mathbf{\Lambda}_s(i, \cdot)$ 中抽取 $M \ll N$ 条件独立指标 $\{J_s^{i,j}\}_{j=1}^M$，并据此更新统计量

$$\beta_{s+1}^i = M^{-1} \sum_{j=1}^{M} \left(\beta_s^{J_s^{i,j}} + \tilde{h}_s(\xi_s^{J_s^{i,j}}, \xi_{s+1}^i) \right). \tag{2.11}$$

如果转移密度 m_s 有一致的上下界，那么使用接受-拒绝方法可以通过基于抽样的更新（2.11）实现 $i \in [\![1, N]\!]$，且总体复杂度为 $\mathcal{O}(N(M+1))$，前提是使用预先初始化的多项式采样器。这种方法的一个关键之处在于，每一步采样的索引数量 M 可以非常少；事实上，对于任意固定的 $M \geq 2$，该算法（称为 PARIS）可以证明具有随机稳定性，且方差为 $\mathcal{O}(t)$（详情见 [Olsson 和 Westerborn, 2017年，第 1] 节）；将 M 设为 2 或 3 通常就能获得完全令人满意的结果。

PARIS估计量可以看作 FFBSm 的一种替代方法，而非 FFBSi 的替代方法。尽管 PARIS 和 FFBSi 都是 FFBSm 估计量的随机化版本，但 PARIS 的本质与 FFBSi 截然不同。对于加性泛函，PARIS 通过额外的蒙特卡洛抽样来近似每个更新步骤，从而在线近似前向 FFBSm。样本量 M 是一个精度参数，它决定了这种近似的精度；随着 M 增大，PARIS 的统计性质会逐渐接近前向 FF BSm。另一方面，如 Douc 等人，2011年，推论 9], 所示，FFBSi 的渐近方差总是大于 FFBSm，其差值等于状态泛函在联合平滑分布下的方差。因此，我们预计，尤其是在信噪比较低的情况下，对于给定的计算预算，PARIS 会比 FFBSi 更准确。另一个选择关注 PARIS估计量而非 FFBSi 的重要原因，是前者具有出色的在线特性；这些特性与粒子马尔可夫链蒙特卡洛方法的相互作用及其相关性仍有待探究。我们的结果可以自然地推广到 FFBSi 和 PGAS，但由于 PARIS 在实践中更具优势，我们选择以它为核心展开工作，尽管本文的核心思想更为普遍。

3. PARIS粒子Gibbs

3.1 粒子Gibbs方法

Andrieu 等，2010a] 提出的条件粒子滤波器 (CPF) [构成了一种以联合平滑分布 $\eta_{0:t}$为目标的基于粒子的 MCMC 算法的基础。设 $\ell \in \mathbb{N}^*$ 为迭代索引， $\zeta_{0:t}[\ell]$ 为在 CPF 的第 ℓ 次迭代中使用的条件路径，用于构造 $\eta_{0:t}$ 的粒子近似，如下所示。在 CPF 的步骤 $s \in [\![1, t]\!]$ 以均匀概率$1/N$随机选择一个粒子，并将其设为 $\zeta_s[\ell]$，而其余 $N-1$ 个粒子均从混合分布 $\eta_{s-1}^N Q_{s-1}$中抽取。在最后一步，新的粒子路径 $\zeta_{0:t}[\ell+1]$ 通过以下任一种方式抽取：

- 再次以均匀概率 $1/N$从 CPF 生成的粒子祖先树 $\{\xi^{1:}{}_s{}^N\}_{s=0}^t$ 中随机选择一条谱系轨迹，如标准粒子吉布斯采样器中所做的那样；

- 或者通过反向抽样生成路径，即即，按照 $J_t \sim$分类分布$(\{1/N\}_{i=1}^N)$ 随时间倒序抽取索引 $J_{0:t}$，并在给定 J_{s+1} 的条件下，根据 $J_s \sim \mathbf{\Lambda}_s(J_{s+1}, \cdot)$、$s \in [\![0, t-1]\!]$ 抽取，然后令 $\zeta_{0:t}[\ell+1] = (\xi_0^{J_0}, \ldots \xi_t^{J_t})$（其中，转移核 $\{\mathbf{\Lambda}_s\}_{s=0}^t$ 由条件粒子滤波器生成的粒子诱导，定义见 (2.9)），如 [Whiteley, 2010] 所提出。

不同版本的粒子 Gibbs 采样器的理论性质已得到充分研究[Singh 等人，2017年，Chopin 和 Singh，2015b，Andrieu 等人， 2018]。简言之，生成的条件路径（$\zeta_{0:t}[\ell]$）$_{\ell\in\mathbb{N}}$ 构成一个马尔可夫链，其边缘分布以几何速度在全变差意义下收敛至目标分布 $\eta_{0:t}$。与平滑算法的情形一样，标准粒子吉布斯采样器会因粒子路径退化而混合效果不佳，而其反向抽样版本则表现更优，且随着 t 增大而更加明显 [Lee 等人， 2020]。

3.2 PPG算法

值得注意的是，为了让标准粒子吉布斯采样器输出一条条件路径，需要运行整个粒子滤波器，随后再将其丢弃，造成大量计算浪费。因此，我们现在介绍 PARIS 算法的一种变体，称为 PARIS粒子吉布斯（PPG）：该算法将带有反向采样的粒子 Gibbs 条件路径与中间粒子合并，从而减少计算浪费，并降低相对于标准 PARIS 算法的偏差。

以下设 $t \in \mathbb{N}$ 为固定的时间范围，并详细说明 PPG 如何通过迭代近似 $\eta_{0:t}h_t$，其中 h_t 是形式为（2.6）的加性泛函。以给定的条件路径 $\zeta_{0:t}[\ell-1]$ 为输入，PPG 的第 ℓ次迭代输出一个多体系统$\boldsymbol{\upsilon}_t[\ell] = ((\xi^1_{0:t},\ \beta^1_t)\text{.....},(\xi^N_{0:t},\ \beta^N_t))$，该系统包含 N条向后粒子路径 $\{\xi^i_{0:t}\}_{i=1}^N$及其对应的 PARIS统计量 $\{\beta^i_t\}_{i=1}^N$。这就是算法1中详述的所谓条件 *PARIS* 更新。之后，从 N 条粒子路径 $\{\xi^i_{0:t}\}_{i=1}^N$中以概率 $1/N$ 选取一条更新后的条件路径，并将其作为下一次条件 PARIS 操作的输入。在每次迭代中，所生成的统计量$\{\beta^i_t\}_{i=1}^N$ 根据（2.7）近似 $\eta_{0:t}h_t$ 。整体算法总结于算法 2。函数 CPF$_s$ 描述条件粒子滤波器的一步，详见补充材料。此外，PPG 算法定义了一个马尔可夫链，其马尔可夫转移核记为 $\mathbb{K}_t$，并在（A.41）中详述。

算法1 一次条件 PARIS 更新（CondPaRIS）

输入： $\{(\xi^i_{0:s},\beta^i_s)\}_{i=1}^N$, ζ_{s+1}, $\tilde{h}_{s-1}$
结果： $\{(\xi^i_{0:s+1},\beta^i_{s+1})\}_{i=1}^N$

抽取 $\xi^{1:N}_{s+1}\sim$ 条件粒子滤波器$_s(\zeta_{s+1},\xi^{1:N}_s)$
对 $i\leftarrow 1$ **到** N **执行**
  抽取 $\{J^{i,\ell}_s\}_{\ell=1}^M \sim \boldsymbol{\Lambda}(i,\ \cdot)^{\otimes M}$
  设置 $\beta^i_{s+1}\leftarrow M^{-1}\sum_{\ell=1}^M\left(\beta^{i,J^{i,\ell}_s}_s+\tilde{h}_s(\xi^{i}_{\cdot,\ J^{i\ \ell}_s\cdot},\ \xi^i_{s+1})\right)$
  设置 $\xi^i_{0:s+1}\leftarrow(\xi^{i}_{\cdot,\ 0:s}[J^{i\ 1}_s\cdot:],\xi^i_{s+1})$

算法 2 PPG 的一次迭代

输入： 初始路径 $\zeta_{0:t}$, $\{\tilde{h}_s\}_{s=0}^{t-1}$**结果：** $\{\beta^i_t\}_{i=1}^N$, $\zeta'_{0:t}$**5** 抽取 $\xi^{1:N}_0\sim$条件粒子滤波器$_0(\zeta_0)$**6** 设定 $\beta^i_0\leftarrow 0$对 $i\in[\![1,N]\!]$**7** **对** $s\leftarrow 0$ **至** $t-1$ **执行8** 设定 $\{(\xi^i_{0:s+1},\beta^i_{s+1})\}_{i=1}^N\leftarrow$CondPaRIS($\{(\xi^i_{0:s},\beta^i_s)\}_{i=1}^N,\zeta_{s+1},\tilde{h}_s$)**9** 抽取 $\zeta'_{0:t}\sim N^{-1}\sum_{i=1}^N\delta_{\xi^i_{0:t}}$

随着执行 k 步 PPG 会产生 k 个多体系统，因此自然会考虑

以下是结合了从步骤 $k_0 < k$ 到 k的反向统计量的滚动估计量:

$$\Pi_{(k_0,k),N}(h_t) = [N(k-k_0)]^{-1} \sum_{\ell=k_0+1}^{k} \sum_{j=1}^{N} \beta_t^j[\ell]. \tag{3.12}$$

该估计量在每个时间步使用的粒子总数为 $C = (N-1)k$ 。我们用$\upsilon = (k-k_0)/k$ 表示估计量中使用的粒子数与采样粒子总数之比。

现在，我们以理论界的形式给出本文的首要结果，这些理论界针对滚动估计量 (3.12) 的偏差和均方误差（MSE）。这些结果是在以下强混合 假设下得到的，这些假设现已成为文献中的标准（参见[Del Moral，2004，Douc 和 Moulines，2008，Del Moral，2013，Del Moral 等人， 2016]）。这些假设对于获得粒子平滑算法的定量界至关重要，参见 [Olsson 和 Westerborn， 2017] 或[Gloaguen 等人， 2022]；同时也适用于耦合条件后向采样粒子滤波器 [Lee 等人， 2020] .

假设 3.1 （强混合）。对于每个 $s \in \mathbb{N}$，存在 $\underline{\tau}_s$、 $\bar{\tau}_s$、 $\underline{\sigma}_s$ 和 $\bar{\sigma}_s$ 属于 $\mathbb{R}_+^*$，使得 (i) 对每个 $x_s \in \mathsf{X}_s$， $\underline{\tau}_s \le g_s(x_s) \le \bar{\tau}_s$； (ii) 对每个 $(x_s, x_{s+1}) \in \mathsf{X}_{s:s+1}$， $\underline{\sigma}_s \le m_s(x_s, x_{s+1}) \le \bar{\sigma}_s$。

在 **假设**3.1 下，对每个 $s \in \mathbb{N}$，定义

$$\rho_s \coloneqq \max_{m \in [\![0,s]\!]} \frac{\bar{\tau}_m \bar{\sigma}_m}{\underline{\tau}_m \underline{\sigma}_m} \tag{3.13}$$

并且，对于每个满足 $N > N_t$ 的 $N \in \mathbb{N}^*$ 和 $t \in \mathbb{N}$：$= (1 + 5\rho_t^2/2) \vee 2t(1+\rho_t^2)$,

$$\kappa_{N,t} \coloneqq 1 - \frac{1 - (1 + 5t\rho_t^2/2)/N}{1 + 4t(1+2\rho_t^2)/N}. \tag{3.14}$$

注意，对于上述所有 N 和 t，$\kappa_{N,t} \in (0,1)$。

定理 1。 假设 假设*3.1*。则对于每个 $t \in \mathbb{N}$、 $M \in \mathbb{N}^*$、 $\xi \in \mathsf{M}_1(\mathcal{X}_{0:t})$，以及满足 $N > N_t$ 的 $k_0 \in \mathbb{N}^*$、$k > k_0$ 和 $N \in \mathbb{N}^*$，

$$\begin{aligned} \left|\mathbb{E}_\xi[\Pi_{(k_0,k),N}(h_t)] - \eta_{0:t} h_t\right| &\le \sigma_{bias} \\ \mathbb{E}_\xi\left[\left(\Pi_{(k_0,k),N}(h_t) - \eta_{0:t} h_t\right)^2\right] &\le \sigma_{mse}^2, \end{aligned} \tag{3.15}$$

其中

$$\begin{aligned} \sigma_{bias} &\coloneqq \frac{\mathsf{c}_t^{bias} \kappa_{t,N}^{k_0} \sum_{m=0}^{t-1} \|\tilde{h}_m\|_\infty}{(k-k_0)(1-\kappa_{t,N})N}, \\ \sigma_{mse}^2 &\coloneqq \frac{(\sum_{m=0}^{t-1} \|\tilde{h}_m\|_\infty)^2}{N(k-k_0)} \left(\mathsf{c}_t^{mse} + \frac{2\mathsf{c}_t^{cov}}{N^{1/2}(1-\kappa_{t,N})}\right) \end{aligned}$$

且c_t^{bias}、c_t^{mse}和c_t^{cov}是与 N 无关的常数，而 $\mathbb{E}_\xi$ 表示当 *PPG* 按照 ξ初始化时，其所形成的马尔可夫链的概率分布下的期望。

证明见补充材料。值得注意的是，(3.15) 给出了滚动估计量偏差的一个界，该偏差会随着预热期 k_0 的延长而呈指数级下降，并与粒子数量 N 成反比。这意味着，通过更合理地分配计算资源，我们可以降低 PARIS 估计量的偏差。

4. 使用 PPG进行参数学习

现在我们转向使用 PPG 和基于梯度的方法进行参数学习。我们重点学习函数 $V(\theta)$ 的参数 θ；该函数的梯度是加性泛函 $s_{0:t,\ \theta}$ 的平滑期望，其形式见 (2.6)。算法 4 定义了一种随机逼近（SA）方案，其中噪声构成一个依赖于参数的马尔可夫链，其对应的不变测度为 π_θ。我们遵循 [Karimi 等人，2019] 的方法，为平均场$h(\theta)$ 建立一个非渐近界：$= \pi_\theta s_{0:t,\ \theta}$。例如，以下估计过程均属于这一设定。

(1) 得分上升。 对于完全支配的 HMM，我们通常希望优化给定观测值的对数似然，其形式为 $V(\theta) = \log\int \gamma_{0:t,\theta}(\mathrm{d}x_{0:t})$。应用费舍尔恒等式后，可将其梯度表示为加性泛函的平滑期望，如下所示:

$$\begin{aligned}\nabla_\theta V(\theta) &= \int \nabla_\theta \log \gamma_{0:t}(x_{0:t})\, \eta_{0:t,\theta}(\mathrm{d}x_{0:t}),\\ &= \int \sum_{\ell=0}^{t-1} s_{\ell,\theta}(x_\ell, x_{\ell+1})\, \eta_{0:t,\theta}(\mathrm{d}x_{0:t}),\end{aligned}$$

其中 $s_{\ell,\theta}\colon \mathrm{X}_{\ell:\ell+1} \ni (x, x') \mapsto \nabla_\theta \log\{g_{\ell,\theta}(x) m_{\ell,\theta}(x, x')\}$ 且 $s_{0:t,\theta} := \sum_{\ell=0}^{t-1} s_{\ell,\theta}$。

(2) 正向 *KL* 代理目标。 受 [Naesseth 等人 2020], 的启发，我们可以考虑学习 $\eta_{0:t,\theta}$ 的代理模型的问题，其形式为 $q_\phi(x_{0:t}) = q_\phi(x_0)\prod_{\ell=0}^{t-1} q_\phi(x_{\ell+1}, x_\ell)$，并通过最小化$V(\phi) = \mathrm{KL}(\eta_{0:t,\theta}, q_\phi)$ 来实现。

算法 3 使用滚动式 PPG 进行梯度估计（GdEst）

输入： θ, $\zeta_{0:t}[0]$, $\{s_{\ell,\theta}\}_{\ell=0}^{t-1}$，PPG 迭代次数为 k，预热期为 k_0。

结果： $\beta_t^{1:N}[k_0 : k]$, $\zeta_{0:t}[k]$**10 对于** $\ell \leftarrow 0$ **到** $k-1$ **执行**11 运行 ($\tilde{\beta}_t^{1:N}[\ell+1]$, $\zeta_{0:t}[\ell+1]$) $\leftarrow$ PPG(θ; $\zeta_{0:t}[\ell]$, $\{s_{\ell,\theta}\}_{\ell=0}^{t-1}$)**12 如果** $\ell \geq k_0 - 1$ **则**13 设置 $\beta_t^{1:N}[\ell+1] = \tilde{\beta}_t^{1:N}[\ell+1]$

算法 4 使用 PPG 的得分上升。

输入： θ_0, $\zeta_{0:t}[0]$, PPG 迭代次数 k，预热期 k_0，随机逼近迭代次数 n，学习率序列 $\{\gamma_\ell\}_{\ell\in\mathbb{N}}$。

Result: θ_n

14 **对于** $i \leftarrow 0$ **到** $n-1$，**执行**15 运行 ($\beta_t^{1:N}[k_0 : k], \zeta_{0:t}[i+1]$) $\leftarrow$ GdEst($\theta_i, \zeta_{0:t}[i], \{s_{\ell,\theta_i}\}_{\ell=0}^{t-1}, k, k_0$)16 设置 $\Pi_{(k_0,k)\ N}(s_{0:t,\theta_i}) = (N(k-k_0))^{-1}\sum_{\ell=k_0}^{k-1}\sum_{j=1}^{N}\beta_t^j[\ell]$17 设置 $\theta_{i+1} \leftarrow \theta_i + \gamma_{i+1}\Pi_{(k_0,k)\ N}(s_{0:t,\theta_i})$

注意，算法 3 定义了一个（折叠的）马尔可夫核 $\mathbb{P}_{\theta,t}$，它针对每条路径 $\zeta_{0:t}$ 定义一个测度$\mathbb{P}_{\theta,t}(\zeta_{0:t}, \mathrm{d}(\tilde{\zeta}_{0:t}, \tilde{\beta}_t^{1:N}[k_0 : k]))$，该测度定义在路径与充分统计量构成的扩展空间上。注意，在该核的一次实现上计算函数 $b_t^{1:N}[k_0 : k] \mapsto [N(k-k_0)]^{-1}\sum_{\ell=k_0+1}^{k}\sum_{j=1}^{N} b_t^j[\ell]$，即可得到滚动估计量，其性质在定理 1 中进行了分析。马尔可夫核 $\mathbb{P}_{\theta,t}$ 详见 (B.72)。

以下假设对于分析算法 4 的收敛性至关重要。

条件 A 4.1. (i) 函数 $\theta \mapsto V(\theta)$ 是 L^V-光滑的。

(ii) 函数 $\theta \mapsto \eta_{0:t,\ \theta}$ 关于全变差距离是 L^{η}-利普希茨的。(iii) 对于每条路径 $\zeta_{0:t} \in \mathsf{X}_{0:t}$，函数

$$\theta \mapsto K_{\theta,t}(\zeta_{0:t}, \mathrm{d}\tilde{\zeta}_{0:t}) \tag{4.16}$$

关于全变差距离是 L_1^P-利普希茨的，其中 $K_{\theta,\ t}$ 是模型由 θ 参数化时与 PPG 算法相关的路径边缘化马尔可夫转移核，见 (A.41)。(iv) 对于每条路径 $\zeta_{0:t} \in \mathsf{X}_{0:t}$，函数

$$\theta \mapsto \mathbb{P}_{\theta,t}\Pi_{k_0-1,k,N}(s_{0:t,\theta})(\zeta_{0:t}) \tag{4.17}$$

关于全变差距离是 L_2^P-利普希茨的。

对于得分上升的情况，我们在附录 B 中验证：若强混合假设 **A3.1** 关于 θ 一致成立，且模型满足额外假设，则这些假设成立。现在我们可以陈述算法 4 的均值场 $h(\theta)$ 的一个界。

定理 2。 假设 ***A3.1*** 关于 θ 一致成立，并且 ***A 4.1*** 成立；再假设步长 $\{\gamma_{\ell+1}\}_{\ell \in [\![0,\ n-1]\!]}$ 满足 $\gamma_{\ell+1} \leq \gamma_\ell$、$\gamma_\ell < a\gamma_{\ell+1}$、$\gamma_\ell - \gamma_{\ell+1} < a'\gamma_\ell^2$ 和 $\gamma_1 \leq 0.5(L^V + C_h)$，其中某些 $a > 0$、$a' > 0$ 以及所有$n \in \mathbb{N}$均满足条件。那么，

$$\mathbb{E}\left[\|h(\theta_\varpi)\|^2\right] \leq 2\frac{V_{0,n} + C_{0,n} + C_{0,\gamma}\sum_{k=0}^n \gamma_{k+1}^2}{\sum_{k=0}^n \gamma_{k+1}}\ , \tag{4.18}$$

其中 $V_{0,\ n} = \mathbb{E}\left[V(\theta)\ - V(\theta_n)\right]$ 以及

$$C_{0,n} \coloneqq \gamma_1 h(\theta_0)C_0 + \sigma_{bias}(\gamma_1 - \gamma_{n+1} + 1)\delta_{k,N,t}^{-1}\ , \tag{4.19}$$

$$C_{0,\gamma} \coloneqq \sigma_{mse}^2 L^V + \sigma_{mse}C_1 + \sigma_{mse}\sigma_{bias}\left(L^V + \frac{C_2}{1-\kappa_{N,t}}\right)\delta_{k,N,t}^{-1} + \sigma_{bias}L^V\delta_{k,N,t}^{-1}\ , \tag{4.20}$$

$$C_h \coloneqq \left(C_1 + \sigma_{bias}\frac{C_2}{(1-\kappa_{N,t})\delta_{k,N,t}}\right)[(a+1)/2 + a\sigma_{mse}] + (L^V + a' + 1)\sigma_{bias}\delta_{k,N,t}^{-1}\ , \tag{4.21}$$

$$C_1 = L_2^P\left[1 + \kappa_{N,t}^k\delta_{k,N,t}^{-1}\right] + L^V \tag{4.22}$$

$$C_2 = L_1^P\delta_{k,N,t}^{-1} + L^\eta\kappa_{N,t}^k\ . \tag{4.23}$$

其中 C_0 与 σ_{bias}、σ_{mse}、N 相互独立，并且 $\delta_{k,\ N,\ t} = 1 - \kappa_{N,\ t}^k$。

定理 2 不仅证明了算法 4 的收敛性，还展示了 PPG 的偏差和方差对收敛速率的影响。

注 1。 在模型的附加假设下（参见附录 B），如果我们考虑 $\gamma_1 \leq 0.5(L^V + C_h)$，$\gamma_\ell = \gamma_1\ell^{-1/2}$对所有 $\ell \in [\![1,\ n]\!]$均成立，那么 $\sum_{k=0}^n\gamma_{k+1}^2/\sum_{k=0}^n\gamma_{k+1} \sim \log n/\sqrt{n}$，这表明 $\mathbb{E}\left[\|h(\theta_\varpi)\|^2\right]$ 为 $\mathcal{O}(\log n/\sqrt{n})$，其中首项常数取决于 σ_{bias}和 σ_{mse}。

注 1 确立了算法 4 的收敛速度。原则上，我们可以尝试利用这些界来优化算法的参数 k，k_0 和 N，但这种方法面临的主要挑战之一是确定混合速率，而混合速率被 $\kappa_{N,\ t}$低估了。尽管如此，我们的界还是提供了关于偏差和均方误差所起作用的有趣信息。

5. 数值分析

本节重点对本文的两项主要结果进行数值分析，即定理 1 中推导的滚动估计量的偏差和均方误差界，以及在第 4 节提出的框架下使用 PPG 进行学习的效率。对于后者，我们将仅讨论以下情形

：通过得分上升进行参数学习。在这种情况下，与本文方法最为相近的竞争方法是按照算法 5 所述，使用标准粒子 Gibbs 采样器 Π_θ 代替 PPG。这类采样器中最常见的一种是 [Lindsten 等人，2014b] 提出的带祖先采样的粒子吉布斯（PGAS）。在 [Lindholm and Lindsten，2018], 中，PGAS 被用于通过期望最大化（EM）算法对隐马尔可夫模型进行参数学习。

算法 5 基于粒子Gibbs核的得分上升。

数据： $\zeta_{0:t}[0]$, θ_0，每条轨迹的路径数为 k，预热期为 k_0，随机逼近迭代次数为 n，学习率序列为 $\{\gamma_\ell\}_{\ell\in\mathbb{N}}$, Π_θ ($\zeta_{0:t}$，$\mathrm{d}\zeta_{0:t}$)，一个以 $\eta_{0:t}$ 为目标的马尔可夫核。

结果： θ_n **对于** $i \leftarrow 0$ **到** $n-1$ **执行** **对于** $j \leftarrow 0$ **到** $k-1$ **执行** $\zeta_{0:t}[j+1] \sim \Pi_\theta(\zeta_{0:t}[j], \cdot)$ 设置 $\theta_{i+1} \leftarrow \theta_i + \frac{\gamma_{i+1}}{k-k_0}\sum_{\ell=k_0+1}^{k} s_{0:t,\,\theta_i}(\tilde{\zeta}_{0:t}[\ell])$ 设置 $\zeta_{0:t}[i+1] = \zeta_{0:t}[k]$

5.1 PPG

线性高斯状态空间模型（LGSSM）。我们首先考虑一个线性高斯 HMM

$$X_{m+1} = AX_m + Q\epsilon_{m+1}, \quad Y_m = BX_m + R\zeta_m, \quad m \in \mathbb{N}, \tag{5.24}$$

其中，$\{\epsilon_m\}_{m\in\mathbb{N}^*}$ 和 $\{\zeta_m\}_{m\in\mathbb{N}}$ 是相互独立且服从标准正态分布的随机变量序列，并且与 X_0相互独立。假设系数 A、Q、B和 R 已知，且分别等于 0.97、0.60、0.54 和 0.33。使用这种参数化方式，我们通过模拟生成一条包含 $t = 999$ 个观测值的记录。

在此情形下，我们旨在计算状态单滞后协方差 $h_t(x_{0:t}) := \sum_{m=0}^{t-1} x_m x_{m+1}$的平滑期望。在线性高斯情形下，扰动平滑器 （参见 [Capp´e 等人，2005，算法 5.2.15]）可提供平滑充分统计量的精确值，这使我们能够研究给定计算预算 C下估计量的偏差。图 1 展示了三个不同总预算 C下，使用 PARIS 以及三种不同 PPG 配置（对应于 $k \in \{2, 4, 10\}$（和 $N = C/k$），其中 $k_0 = k/2$和 $k_0 = k/4$）得到的 $\eta_{0:n}h_n$ 估计值的分布。红色虚线表示参考值，每个分布的均值以黑色虚线表示。每个箱线图基于相应估计量的 1000 次独立重复实验。我们观察到，在此示例中，所有 PPG 配置的偏差都小于等效的 PARIS 估计量。关于定理 1 中界限的说明推迟至附录 D.1。

5.2 得分上升

线性高斯状态空间模型。我们考虑状态空间和观测空间分别为 $\mathbb{R}^5$ 的线性高斯状态空间模型。假设参数 R 和 Q 已知，并基于模拟得到的一系列 $n = 999$ 观测值，对 $\theta = (A, B)$ 进行推断。在这种情况下，借助扰动平滑器可以精确求解期望最大化算法的 M步 [Capp´e 等人，2005，第 11] 章。通过该过程得到的参数（记为 θ_{mle}）是任何似然最大化算法的参考值。表 1 显示了 θ_{mle} 的奇异值与算法 4 和算法 5 得到的参数的奇异值之间的 L_2 距离。中心极限定理置信区间基于 25 次重复实验得到。各配置均符合给定的粒子预算 $kN = C = 1024$。选择保留 $k_0 = k/2$ 是一条启发式规则，旨在实现良好的偏差-方差权衡；不过，对于不同问题，k_0 和 k 的其他组合可能会带来更好的表现。我们在附录D.2中针对线性高斯状态空间混合模型分析了这一点。两种算法的所有设置均相同，具体说明见附录D.2。PPG 达到

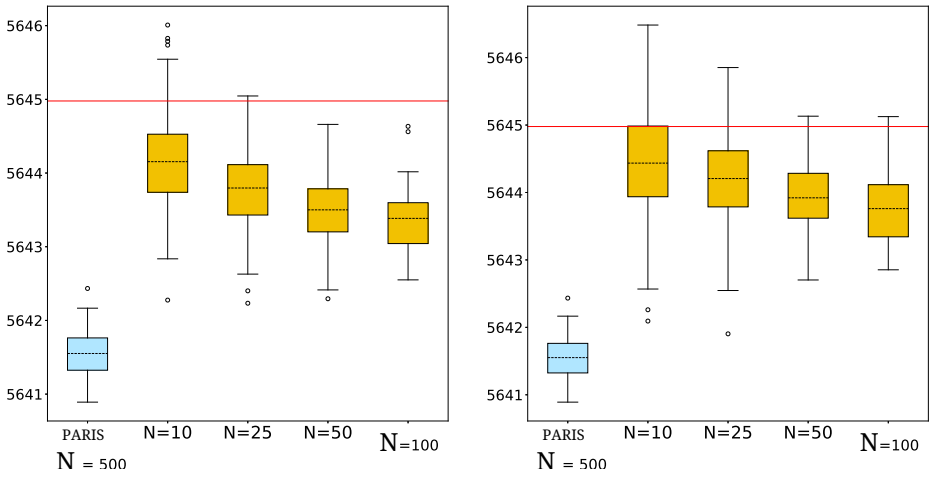

图 1： $C = 500$ 对应的线性高斯状态空间模型的 PARIS 和 PPG 输出，黄色框对应使用 $k \in \{50, 20, 10, 5\}$ 迭代和 $N \in \{C/50, C/20, C/10, C/5\}$ 粒子生成的 PPG 输出。左侧图像对应取 $k_0 = k/2$ ，右侧图像对应取 $k_0 = k/4$。

始终与 θ_{mle} 的距离更小。图 2 展示了每种估计量和配置下，距离随迭代索引变化的情况，其中距离是相对于最大似然估计量计算的。

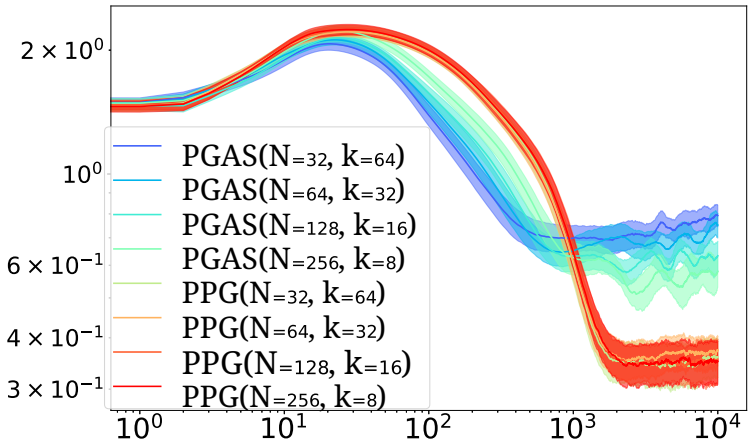

图 2：对于 25 个不同的随机种子，在 LGSSM 中固定粒子预算时，PGAS 和 PPG 在不同参数下的迭代步数与最大似然估计量之间的距离。

卷积循环神经网络。现在我们考虑非线性 HMM 中的推断问题，特别是由 [Zhao 等人 2021] 提出的混沌循环神经网络。我们采用与原论文相同的设置。状态方程和观测方程为

$$\begin{aligned} X_{m+1} &= X_m + \tau^{-1}\Delta\left(-X_m + \gamma W \tanh(X_m)\right) + \epsilon_{m+1}, \\ Y_m &= BX_m + \zeta_m, \quad m \in \mathbb{N}, \end{aligned}$$

其中 $\{\epsilon_m\}_{m\in\mathbb{N}^*}$ 是一列均值为零、协方差为 $0.01\mathbf{I}$ 的 20 维独立多元高斯随机变量，且 $\{\zeta_m\}_{m\in\mathbb{N}}$ 是一列相互独立的随机变量

表 1：线性高斯状态空间模型案例中各配置与 θ_{MLE} 的距离。

算法	N	k_0	k	D_{mle}
PGAS	32	32	64	0.793 ± 0.048
PGAS	64	16	32	0.751 ± 0.052
PGAS	128	8	16	0.633 ± 0.054
PGAS	256	4	8	0.580 ± 0.049
PPG	32	32	64	0.358 ± 0.038
PPG	64	16	32	0.373 ± 0.031
PPG	128	8	16	0.355 ± 0.043
PPG	256	4	8	0.351 ± 0.042

表2：卷积循环神经网络模型每种配置下的负对数似然。

算法	N	k_0	k	NLL
PGAS	32	16	32	31364.932 ± 173.708
PGAS	64	8	16	31083.408 ± 380.527
PGAS	128	4	8	30264.836 ± 265.880
PPG	32	16	32	22291.971 ± 47.683
PPG	64	8	16	22314.537 ± 25.028
PPG	128	4	8	22353.416 ± 39.443

其中每个分量均独立地服从尺度参数为 0.1、自由度为 2 的学生t分布。

在这种情况下，用于评估不同估计量的自然度量是负对数似然（NLL）。我们使用粒子滤波器产生的对数权重均值作为似然的无偏估计量 [Douc 等人，2014 年，第 12 节1]，使用 $N = 10^4$ 粒子。表2展示了在总粒子预算保持不变的情况下，对 PPG 和 PGAS 的多种不同配置分别进行 25 次重复实验所得的结果。更多数值细节见附录D.2。我们观察到，在所有配置下，PPG 的 NLL 都显著低于 PGAS。

6. 结论

我们提出了一种新算法，称为 PPG，并在定理 1 中给出了其偏差和均方误差的界。随后，我们提出了一种在学习框架中使用 PPG 的方法，并推导出使用 PPG 进行得分上升时更新梯度的非渐近界，其中明确体现了估计量的偏差和均方误差。我们通过数值模拟支持我们的论断，并展示了在所分析的两个不同示例中，我们的算法优于现有的竞争方法。

目录

1 引言 **1**

2 背景 **3**

2.1 隐马尔可夫模型.3 2.2 粒子滤波器.3 2.3 反向平滑与 PARIS 算法.4

3 PARIS 粒子Gibbs **5**

3.1 粒子吉布斯方法.5 3.2 PPG 算法.6

4 使用PPG 进行参数学习8

5 数值分析 **9**

5.1 PPG.10 5.2 得分上升.10

6 结论 12

APPG 14A.1 多体费曼–卡茨模型.14 A.2 Feynman–Kac 路径流的反向解释.15 A.3 条件对偶过程和粒子Gibbs.16 A.4 PARIS 算法.17 A.5 定理 1 的证明.21 A.6 中间结果的证明.24 A.6.1 命题 1 的证明.24 A.6.2 定理 3 的证明.24 A.6.3 定理 5 的证明.25 A.6.4 命题 2 的证明.29 A.6.5 定理 7 的证明.31 A.6.6 命题 4 的证明.34 A.6.7 命题 5 的证明.35

B 使用 PPG 进行学习 其中PPG **3 6**

B.1 非渐近界.36 B.2 定理 2 的应用.38 B.2.1 定理 8 的假设验证.38 B.2.2 定理 2 的证明.42 B.3 验证 A4.1 的模型条件.42

C 利普希茨性 质 4 **4**

C.1 $\mathbb{P}_{\theta,}$ 的 Lipschitz 连续性.44 C.1.1 $\theta \mapsto \mathbb{C}_{m,\theta}$ 是 Lipschitz 的.46 C.1.2 $\theta \mapsto \mathbb{B}_{t,\theta}(\boldsymbol{x}_{0:t,}\cdot)$ 是 Lipschitz 的.48 C.1.3 $\theta \mapsto \int \mathbb{S}_{t,\theta}(\boldsymbol{x}_{0:t,}\mathrm{d}\boldsymbol{b}_t)\mu(\boldsymbol{b}_t)(\mathrm{id})$ 是 Lipschitz 的.49 C.2 马尔可夫核的 Lipschitz 性质.51

D 补充数值结果 **53**

D.1 PPG.53 D.2 学习.53

A. PPG

在本节中，我们将构建确立定理 1 所需的理论框架。我们回顾*Feynman–Kac* 模型、多体 *Feynman–Kac* 模型、后向解释和条件对偶过程的概念。我们的阐述主要遵循 [Del Moral 等人， 2016] 但对多体扩展采用了不同且希望更清晰的定义。我们重述（见下文定理 3） [Del Moral 等人提出的一个对偶公式， 2016] 该公式将这些概念联系起来。此公式为算法 2 中所述的粒子 *Gibbs* 采样器 奠定了基础。

记号。 设 $(\mathsf{Z}\ \mathcal{Z})$ 是一个可测空间， L 是 $\mathsf{Y}\ \times \mathcal{Z}$ 上另一个可能未归一化的转移核。按照上文的定义，令 K，

$$KL : \mathsf{X} \times \mathcal{Z} \ni (x, A) \mapsto \int L(y, A)\, K(x, \mathrm{d}y)$$

and

$$K \otimes L : \mathsf{X} \times (\mathcal{Y} \otimes \mathcal{Z}) \ni (x, A) \mapsto \iint \mathbb{1}_A(y, z)\, K(x, \mathrm{d}y)\, L(y, \mathrm{d}z),$$

只要这些定义良好。这也定义了核 K （定义在 $\mathsf{X}\times\mathcal{Y}$ 上）与测度ν （定义在 $\mathcal{X}$ 上）的 $\otimes$ 乘积，以及核 L （定义在 $\mathsf{Y}\ \times\mathcal{X}$ 上）与测度 μ （定义在 $\mathcal{Y}$ 上）的乘积，它们分别是测度

$$\nu \otimes K : \mathcal{X} \otimes \mathcal{Y} \ni A \mapsto \iint \mathbb{1}_A(x, y)\, K(x, \mathrm{d}y)\, \nu(\mathrm{d}x),$$
$$L \otimes \mu : \mathcal{X} \otimes \mathcal{Y} \ni A \mapsto \iint \mathbb{1}_A(x, y)\, L(y, \mathrm{d}x)\, \mu(\mathrm{d}y).$$

A.1 多体 Feynman–Kac 模型

下文假设所有随机变量都定义在同一个概率空间 $(\Omega\ \mathcal{F},\ \mathbb{P})$ 上。式 (2.4) 中定义的分布流 $\{\eta_m\}_{m\in\mathbb{N}}$ 通常难以处理，但可以通过随机样本 $\boldsymbol{\xi}_m = \{\xi_m^i\}_{i=1}^N$, $m \in \mathbb{N}$ 来近似，这些样本称为粒子，其中 $N \in \mathbb{N}^*$ 是固定的蒙特卡洛样本量，且每个粒子 ξ_m^i 都是取值于 X_m 的随机变量。这种粒子近似基于递推 $\eta_{m+1} = \Phi_m(\eta_m)$, $m \in \mathbb{N}$，其中 Φ_m 表示映射

$$\Phi_m : \mathsf{M}_1(\mathcal{X}_m) \ni \eta \mapsto \frac{\eta Q_m}{\eta g_m} \tag{A.25}$$

取值于 $\mathsf{M}_1(\mathcal{X}_{m+1})$。为了递归描述粒子群的演化，令 $m \in \mathbb{N}$ ，并假设粒子 $\boldsymbol{\xi}_m$ 构成对 η_m 的一致近似，即 $\mu(\boldsymbol{\xi}_m)h$，其中 $\mu(\boldsymbol{\xi}_m) \coloneqq N^{-1}\sum_{i=1}^N \delta_{\xi_m^i}$， δ_x 表示位于x的狄拉克测度，是由 $\boldsymbol{\xi}_m$形成的占据测度，可作为 $\eta_m h$ 的代理，适用于所有 η_m-可积测试函数 h。在一般条件下， $\mu(\boldsymbol{\xi}_m)h$ 依概率收敛到 η_m ，其中 $N \to \infty$; 参见 [Del Moral，2004、Chopin 和 Papaspiliopoulos， 2020] 及其中的参考文献。然后，为了生成一个近似于 η_{m+1}的更新粒子样本，在给定 $\boldsymbol{\xi}_m$ 的条件下，按照以下方式独立抽取新粒子 $\boldsymbol{\xi}_{m+1} = \{\xi_{m+1}^i\}_{i=1}^N$

$$\xi_{m+1}^i \sim \Phi_m(\mu(\boldsymbol{\xi}_m)) = \sum_{\ell=1}^N \frac{g_m(\xi_m^\ell)}{\sum_{\ell'=1}^N g_m(\xi_m^{\ell'})} M_m(\xi_m^\ell, \cdot), \quad i \in [\![1, N]\!].$$

由于这一粒子更新过程涉及从混合分布 $\Phi_m(\mu(\boldsymbol{\xi}_m))$ 中采样，因此自然可以分解为两个子步骤：选择和变异。选择步骤包括以概率 $g_m(\xi_m^\ell)/\sum_{\ell'=1}^N g_m(\xi_m^{\ell'})$ 随机选择第 ℓ 个混合层，而变异步骤则包括从选定的混合层 $M_m(\xi_m^\ell,\ \cdot)$ 中抽取一个新粒子 ξ_{m+1}^i 。在

[Del Moral 等人， 2016], 术语多体 *Feynman–Kac* 模型与过程 $\{\boldsymbol{\xi}_m\}_{m\in\mathbb{N}}$ 的分布有关。对于所有 $m \in \mathbb{N}$, 令$\mathbf{X}_m \coloneqq \mathrm{X}_m^N$和$\boldsymbol{\mathcal{X}}_m \coloneqq \mathcal{X}_m^{\otimes N}$；则 $\{\boldsymbol{\xi}_m\}_{m\in\mathbb{N}}$ 是 $\{\mathbf{X}_m\}_{m\in\mathbb{N}}$ 上的非齐次马尔可夫链，其转移核为

$$\boldsymbol{M}_m : \mathbf{X}_m \times \boldsymbol{\mathcal{X}}_{m+1} \ni (\boldsymbol{x}_m, A) \mapsto \Phi_m(\mu(\boldsymbol{x}_m))^{\otimes N}(A)$$

以及初始分布 $\boldsymbol{\eta}_0 = \eta_0^{\otimes N}$。现在，记 $\mathbf{X}_{0:n} \coloneqq \prod_{m=0}^n \mathbf{X}_m$ 和 $\boldsymbol{\mathcal{X}}_{0:n} \coloneqq \bigotimes_{m=0}^n \boldsymbol{\mathcal{X}}_m$。下文中，我们用粗体符号强调某个量与多体过程相关。多体 *Feynman–Kac* 路径模型 指的是由马尔可夫核 $\{\boldsymbol{M}_m\}_{m\in\mathbb{N}}$、初始分布 $\boldsymbol{\eta}_0$、势函数以及相应的未归一化转移核，通过（2.4）和（2.3）在 $\{\boldsymbol{\mathcal{X}}_{0:m}\}_{m\in\mathbb{N}}$ 上生成的未归一化和归一化概率分布流 $\{\boldsymbol{\gamma}_m\}_{m\in\mathbb{N}}$ 和 $\{\boldsymbol{\eta}_m\}_{m\in\mathbb{N}}$。

$$\boldsymbol{g}_m : \mathbf{X}_m \ni \boldsymbol{x}_m \mapsto \mu(\boldsymbol{x}_m) g_m = \frac{1}{N}\sum_{i=1}^N g_m(x_m^i), \quad m \in \mathbb{N},$$

以及相应的未归一化转移核

$$\boldsymbol{Q}_m : \mathbf{X}_m \times \boldsymbol{\mathcal{X}}_{m+1} \ni (\boldsymbol{x}_m, A) \mapsto \boldsymbol{g}_m(\boldsymbol{x}_m)\boldsymbol{M}_m(\boldsymbol{x}_m, A), \quad m \in \mathbb{N}.$$

A.2 Feynman–Kac 路径流的反向解释

假设每个在（2.2）中定义的核 Q_n、 $n \in \mathbb{N}$ 都相对于某个支配测度 $\lambda_{n+1} \in \mathrm{M}(\mathcal{X}_{n+1})$ 具有转移密度 q_n。那么，对于 $n \in \mathbb{N}$ 和 $\eta \in \mathrm{M}_1(\mathcal{X}_n)$，我们可以定义反向核

$$\overleftarrow{Q}_{n,\eta} : \mathsf{X}_{n+1} \times \mathcal{X}_n \ni (x_{n+1}, A) \mapsto \frac{\int \mathbb{1}_A(x_n) q_n(x_n, x_{n+1})\, \eta(\mathrm{d}x_n)}{\int q_n(x_n', x_{n+1})\, \eta(\mathrm{d}x_n')}. \tag{A.26}$$

现在，对于 $n \in \mathbb{N}^*$，记作

$$B_n : \mathsf{X}_n \times \mathcal{X}_{0:n-1} \ni (x_n, A) \mapsto \int \cdots \int \mathbb{1}_A(x_{0:n-1}) \prod_{m=0}^{n-1} \overleftarrow{Q}_{m,\eta_m}(x_{m+1}, \mathrm{d}x_m), \tag{A.27}$$

我们可以陈述如下如今已成经典的 Feynman–Kac 路径测度反向分解，这一结果在本文中发挥着至关重要的作用。

命题 1. 对于每个 $n \in \mathbb{N}^*$，都有 $\gamma_{0:n} = \gamma_n \otimes B_n$ 且 $\eta_{0:n} = \eta_n \otimes B_n$.

尽管命题 1 中的分解广为人知（参见例如， [Del Moral 等，2010，Del Moral 等人， 2016]），为求完整起见，我们仍在附录 A.6.1 中给出证明。利用反向分解，可以通过以下方式获得给定 Feynman–Kac 路径测度 $\eta_{0:n}$ 的粒子近似：首先，在初始的前向遍历中，从 $\boldsymbol{\eta}_0 \otimes \boldsymbol{M}_0 \otimes \cdots \otimes \boldsymbol{M}_{n-1}$ 采样粒子云 $\{\boldsymbol{\xi}_m\}_{m=0}^n$；然后，在后续的反向遍历中，例如，从 $\mathbb{B}_n$有条件地独立采样路径$\{\tilde{\xi}_{0:n}^i\}_{i=1}^N$ N （$\boldsymbol{\xi}_0, \ldots, \boldsymbol{\xi}_n, \cdot$），其中

$$\mathbb{B}_n : \mathbf{X}_{0:n} \times \mathcal{X}_{0:n} \ni (\boldsymbol{x}_{0:n}, A) \mapsto \int \cdots \int \mathbb{1}_A(x_{0:n}) \left(\prod_{m=0}^{n-1} \overleftarrow{Q}_{m,\mu(\boldsymbol{x}_m)}(x_{m+1}, \mathrm{d}x_m) \right) \mu(\boldsymbol{x}_n)(\mathrm{d}x_n) \tag{A.28}$$

是一个描述由前向遍历生成的粒子近似所诱导的时间反向动力学的马尔可夫核。这里以及下文中，我们使用黑板粗体表示与多体路径空间相关的核。最后，对于任意 $\eta_{0:n}$可积测试函数 h，将 $\mu(\{\tilde{\xi}_{0:n}^i\}_{i=1}^N)h$ 作为 $\eta_{0:n}h$ 的估计量返回。文献中将该算法称为前向滤波–后向模拟（FFBSi）算法，由 [Godsill 等人于 2004]提出；另见[Capp´e 等人（2007 年）以及 Douc 等人 2011]。更准确地说，给定前向粒子 $\{\boldsymbol{\xi}_m\}_{m=0}^n$，每条路径

$\tilde{\xi}^i_{0:n}$ 的生成方式是：先从最后一代的粒子 $\boldsymbol{\xi}_n$ 中均匀抽取 $\tilde{\xi}^i_n$ ，然后递归地抽取，

$$\tilde{\xi}^i_m \sim \overleftarrow{Q}_{m,\mu(\boldsymbol{\xi}_m)}(\tilde{\xi}^i_{m+1}, \cdot) = \sum_{j=1}^{N} \frac{q_m(\xi^j_m, \tilde{\xi}^i_{m+1})}{\sum_{\ell=1}^{N} q_m(\xi^\ell_m, \tilde{\xi}^i_{m+1})} \delta_{\xi^j_m}(\cdot), \tag{A.29}$$

即，给定 $\tilde{\xi}^i_{m+1}$， $\tilde{\xi}^i_m$ 按照与 $\{q_m(\xi^j_m, \tilde{\xi}^i_{m+1})\}_{j=1}^N$成比例的权重，从 $\boldsymbol{\xi}_m$ 中随机选取。请注意，在 FFBSi 算法的这一基本表述中，每次反向采样操作 (A.29) 都需要计算归一化常数 $\sum_{\ell=1}^N q_m(\xi^\ell_m, \tilde{\xi}^i_{m+1})$，这意味着该算法的总体复杂度为二次复杂度。不过，可以通过附录A.4中讨论的有效接受–拒绝技术来减轻这一沉重的计算负担。

A.3 条件对偶过程与粒子 Gibbs

与给定的 Feynman–Kac 模型 (2.4–2.3) 以及给定轨迹 $\{z_n\}_{n\in\mathbb{N}}$（其中每个 $n \in \mathbb{N}$ 都满足 $z_n \in \mathsf{X}_n$ ）相关联的对偶过程 ，定义为具有以下核的典范马尔可夫链

$$\boldsymbol{M}_n\langle z_{n+1}\rangle : \mathbf{X}_n \times \boldsymbol{\mathcal{X}}_{n+1} \ni (\boldsymbol{x}_n, A) \mapsto \frac{1}{N}\sum_{i=0}^{N-1} \left(\Phi_n(\mu(\boldsymbol{x}_n))^{\otimes i} \otimes \delta_{z_{n+1}} \otimes \Phi_n(\mu(\boldsymbol{x}_n))^{\otimes(N-i-1)}\right)(A), \tag{A.30}$$

对于 $n \in \mathbb{N}$，以及初始分布

$$\boldsymbol{\eta}_0\langle z_0\rangle \coloneqq \frac{1}{N}\sum_{i=0}^{N-1} \left(\eta_0^{\otimes i} \otimes \delta_{z_0} \otimes \eta_0^{\otimes(N-i-1)}\right). \tag{A.31}$$

从（A.30–A.31）可以清楚地看出，给定 $\{z_n\}_{n\in\mathbb{N}}$，对偶过程的一个实现 $\{\boldsymbol{\xi}_n\}_{n\in\mathbb{N}}$ 按如下方式生成。在时间零点，通过将 z_0 插入向量 $\boldsymbol{\xi}_0$ 中随机选定的位置，并独立地从 η_0 抽取其余分量，从而初始化该过程。然后，给定步骤 n 时的 $\boldsymbol{\xi}_n$ ，将 z_{n+1} 插入 $\boldsymbol{\xi}_{n+1}$ 中随机选定的位置，并独立地从 $\Phi_n(\mu(\boldsymbol{\xi}_n))$ 抽取其余分量。

为了简洁地描述条件对偶过程的分布，我们定义马尔可夫核

$$\mathbb{C}_n : \mathsf{X}_{0:n} \times \boldsymbol{\mathcal{X}}_{0:n} \ni (z_{0:n}, A) \mapsto \boldsymbol{\eta}_0\langle z_0\rangle \otimes \boldsymbol{M}_0\langle z_1\rangle \otimes \cdots \otimes \boldsymbol{M}_{n-1}\langle z_n\rangle(A).$$

以下结果优雅地结合了底层模型（2.4–2.3）、多体费曼–卡茨模型、反向分解以及条件对偶过程。

定理 3 ([Del Moral 等人， 2016])。对于所有 $n \in \mathbb{N}$,

$$\mathbb{B}_n \otimes \boldsymbol{\gamma}_{0:n} = \gamma_{0:n} \otimes \mathbb{C}_n. \tag{A.32}$$

在 [Del Moral 等人， 2016], 多体过程的每个状态 $\boldsymbol{\xi}_n$ 都会将样本空间 Ω 中的一个结果 ω 映射为 X 中由 N 个元素组成的无序集合$_n$。不过，我们选择让每个$\boldsymbol{\xi}_n$ 取值于标准乘积空间X_n^N ，原因有二：首先，[Del Moral 等人， 2016] 的构造需要复杂的测度论论证，才能为这类无序集合赋予合适的 σ-域和适当的测度；其次，只要条件对偶过程的马尔可夫动力学（A.30–A.31）在粒子云上经过对称化，我们就没有必要忽略粒子的索引顺序。因此，在附录A.6.2中，为完整起见，我们给出了对偶性（A.32）的证明。请注意， $\mathcal{X}_{0:n} \otimes \boldsymbol{\mathcal{X}}_{0:n}$ 上的测度（A.32）是非归一化的，但由于核 $\mathbb{B}_n$ 和 $\mathbb{C}_n$ 都是马尔可夫的，用 $\gamma_{0:n}(\mathsf{X}_{0:n}) = \boldsymbol{\gamma}_{0:n}(\mathbf{X}_{0:n})$ 对该恒等式进行归一化后，便可立即得到

$$\mathbb{B}_n \otimes \boldsymbol{\eta}_{0:n} = \eta_{0:n} \otimes \mathbb{C}_n. \tag{A.33}$$

由于（A.33）两边给出了完整条件分布，因此自然的做法是采用数据增广方法，并使用两阶段确定性扫描吉布斯采样器 [Andrieu 等，2010b，Chopin 和 Singh，2015a]对目标分布（A.33）进行采样。具体而言，假设我们基于采样器的 $\ell \in \mathbb{N}$ 次迭代，已生成一个状态（$\boldsymbol{\xi}_{0:n}[\ell]$，$\zeta_{0:n}[\ell]$），其中包含一个对偶过程及其关联路径；那么，下一个状态（$\boldsymbol{\xi}_{0:n}[\ell+1]$，$\zeta_{0:n}[\ell+1]$）将以马尔可夫方式生成：首先对$\boldsymbol{\xi}_{0:n}[\ell+1] \sim \mathbb{C}_n(\zeta_{0:n}[\ell], \cdot)$进行采样，然后对 $\zeta_{0:n}[\ell+1] \sim \mathbb{B}_n(\boldsymbol{\xi}_{0:n}[\ell+1], \cdot)$进行采样。经过任意初始化（并舍弃可能的预热迭代）后，该过程会产生一条马尔可夫轨迹$\{(\boldsymbol{\xi}_{0:n}[\ell], \zeta_{0:n}[\ell])\}_{\ell\in\mathbb{N}}$；在一些较弱的附加技术条件下，这条马尔可夫链以（A.33）为其唯一不变分布。在这种情况下，该马尔可夫链是遍历的 [Douc 等人，2018，第 5], 章，且条件路径 $\zeta_{0:n}[\ell]$ 的边缘分布会收敛到目标分布 $\eta_{0:n}$。因此，对于每个 $h \in \mathrm{F}(\mathcal{X}_{0:n})$，

$$\lim_{L\to\infty} \frac{1}{L}\sum_{\ell=1}^{L} h(\zeta_{0:n}[\ell]) = \eta_{0:n}h, \quad \mathbb{P}\text{-a.s.}$$

A.4 PARIS PARIS算法

下文中，假设给定一列如式（2.6）所示的加性状态泛函 。在一般状态空间模型中，这个问题与基于最大似然的参数估计尤为相关，例如，通过 Fisher 恒等式计算得分函数（即对数似然函数的梯度），或计算期望最大化 （EM）算法的中间量。在后一种情况下， $\eta_{0:n}$ 和 h_n 分别对应联合状态后验分布和某个充分统计量的一个元素；参见 [Capp´e 和 Moulines，2005；Douc 等人，2011；Del Moral 等，2010；Poyiadjis 等，2011；Olsson 和 Westerborn， 2017] 以及其中的参考文献。有趣的是，正如 [Capp´e，2011，Del Moral 等人， 2010], 所指出的，将反向分解应用于加性状态泛函时，可以得到关于期望 $\{\eta_{0:n}h_n\}_{n\in\mathbb{N}}$ 的前向递推。更具体地，利用前向分解 $h_{n+1}(x_{0:n+1}) = h_n(x_{0:n}) + \tilde{h}_n(x_n, x_{n+1})$ 以及式（A.27）中定义的反向核 B_{n+1}，对于 $x_{n+1} \in \mathrm{X}_{n+1}$，可写为：

$$\begin{aligned} B_{n+1}h_{n+1}(x_{n+1}) &= \int \overleftarrow{Q}_{n,\eta_n}(x_{n+1}, \mathrm{d}x_n) \int \left(h_n(x_{0:n}) + \tilde{h}_n(x_n, x_{n+1})\right) B_n(x_n, \mathrm{d}x_{0:n-1}) \\ &= \overleftarrow{Q}_{n,\eta_n}(B_n h_n + \tilde{h}_n)(x_{n+1}), \end{aligned} \tag{A.34}$$

根据命题 1 可得

$$\eta_{0:n+1}h_{n+1} = \eta_{n+1}\overleftarrow{Q}_{n,\eta_n}(B_n h_n + \tilde{h}_n). \tag{A.35}$$

由于边缘流 $\{\eta_n\}_{n\in\mathbb{N}}$ 可通过映射 $\{\Phi_n\}_{n\in\mathbb{N}}$递归表示，因此（A.35）原则上为在线计算 $\{\eta_{0:n}h_n\}_{n\in\mathbb{N}}$ 提供了基础。为处理边缘分布通常难以计算这一事实，我们可以仿照 [Del Moral 等人 2010],的做法，将粒子近似 $\mu(\boldsymbol{\xi}_{n+1})$ 和 $\overleftarrow{Q}_{n,\,\mu(\boldsymbol{\xi}_n)}$（见（A.29））分别代入 η_{n+1}和 $\overleftarrow{Q}_{n,\,\mu(\eta_n)}$的递推式（A.35）。更准确地说，我们递归地进行计算，并假设在时间 n，我们手头有一个由粒子及其相应统计量构成的样本$\{(\xi_n^i, \beta_n^i)\}_{i=1}^N$，其中每个统计量 β_n^i 都作为 $B_n h_n(\xi_n^i)$的近似；随后按照 $\boldsymbol{\xi}_{n+1} \sim \boldsymbol{M}_n(\boldsymbol{\xi}_n, \cdot)$ 演化粒子云，并使用（A.34）更新统计量，同时将 $\overleftarrow{Q}_{n,\,\eta_n}$替换为 $\overleftarrow{Q}_{n,\,\mu(\boldsymbol{\xi}_n)}$，即可得到逐粒子的递推式

$$\beta_{n+1}^i = \sum_{\ell=1}^{N} \frac{q_n(\xi_n^\ell, \xi_{n+1}^i)}{\sum_{\ell'=1}^N q_n(\xi_n^{\ell'}, \xi_{n+1}^i)} \left(\beta_n^\ell + \tilde{h}_n(\xi_n^\ell, \xi_{n+1}^i)\right), \quad i \in [\![1, N]\!], \tag{A.36}$$

最后，估计量

$$\mu(\boldsymbol{\beta}_n)(\mathrm{id}) = \frac{1}{N}\sum_{i=1}^{N} \beta_n^i \tag{A.37}$$

关于 $\eta_{0:n}h_n$ 的估计量，其中 $\boldsymbol{\beta}_n \coloneqq (\beta_n^1, \ldots\ldots, \beta_n^N)$， $i \in [\![1, N]\!]$。该过程的初始化方法很简单：令所有 $i \in [\![1, N]\!]$ 的 $\beta_0^i = 0$。注意，(A.37) 为命题 1 中反向分解提供了粒子解释。该算法是前向滤波反向平滑 （FFBSm）算法 的一个特例（参见 [Andrieu 和 Doucet，2003；Godsill 等人，2004；Douc 等人，2011；S¨arkk¨a， 2013]），适用于满足 (2.6) 的加性泛函。该算法可对序列 $\{\eta_{0:n}h_n\}_{n\in\mathbb{N}}$ 进行在线处理，而且还有一个优点：只需存储当前粒子 $\boldsymbol{\xi}_n$ 和统计量 $\boldsymbol{\beta}_n$。然而，由于每次更新 (A.36) 都需要对 N 项求和，该方案的总体复杂度随粒子数呈二次 增长，这会成为复杂模型应用中的计算瓶颈，因为这类模型需要较大的粒子样本数 N。

为了避免采用这种 FFBSm 前向算法实现所带来的计算负担，PARIS 算法 [Olsson 和 Westerborn，2017] 通过用蒙特卡罗估计替换每个求和式 (A.36) 来更新统计量 $\boldsymbol{\beta}_n$

$$\beta_{n+1}^i = \frac{1}{M}\sum_{j=1}^{M}\left(\tilde{\beta}_n^{i,j} + \tilde{h}_n(\tilde{\xi}_n^{i,j}, \xi_{n+1}^i)\right), \quad i \in [\![1, N]\!], \tag{A.38}$$

其中， $\{(\tilde{\xi}^{i,}{}_n{}^j, \tilde{\beta}^{i,}{}_n{}^j)\}_{j=1}^M$ 从 $\{(\xi_n^i, \beta_n^i)\}_{i=1}^N$ 中有放回地随机抽取：以 $q_n(\xi_n^\ell, \xi_{n+1}^i)/\sum_{\ell'=1}^N q_n(\xi_n^{\ell'}, \xi_{n+1}^i)$ 的概率，将 $(\tilde{\xi}^{i,}{}_n{}^j, \tilde{\beta}^{i,}{}_n{}^j)$ 赋值为 $(\xi_n^\ell, \beta_n^\ell)$。蒙特卡洛样本量 $M \in \mathbb{N}^*$ 应远小于 N （例如，小于 5）。形式化地，

$$\{(\tilde{\xi}_n^{i,j}, \tilde{\beta}_n^{i,j})\}_{j=1}^M \sim \left(\sum_{\ell=1}^N \frac{q_n(\xi_n^\ell, \xi_{n+1}^i)}{\sum_{\ell'=1}^N q_n(\xi_n^{\ell'}, \xi_{n+1}^i)}\delta_{(\xi_n^\ell, \beta_n^\ell)}\right)^{\otimes M}, \quad i \in [\![1, N]\!].$$

算法1总结了由此得到的过程，该过程支持在线处理且内存需求恒定，因为每次迭代只需存储当前粒子云和估计得到的辅助统计量。此外，如果模型的马尔可夫转移密度具有一致上界，即对于每个 $n \in \mathbb{N}$，都存在一个上界 $\bar{\sigma}_n > 0$ ，使得对所有 $(x_n, x_{n+1}) \in \mathrm{X}_n \times \mathrm{X}_{n+1}$， $m_n(x_n, x_{n+1}) \leq \bar{\sigma}_n$ （这是一个大多数关注模型都满足的弱假设），则可以通过以下方式生成样本 $(\tilde{\xi}^{i,}{}_n{}^j, \beta^{i,}{}_n{}^j)$：根据归一化粒子权重 $\{g_n(\xi_n^\ell)/\sum_{\ell'} g_n(\xi_n^{\ell'})\}_{\ell=1}^N$，从 $\{(\xi_n^i, \beta_n^i)\}_{i=1}^N$ 中有放回地抽取候选项 $(\tilde{\xi}^{i,}{}_n{}^*, \tilde{\beta}^{i,}{}_n{}^*)$，直至接受为止；这些权重是在生成 $\boldsymbol{\xi}_{n+1}$ 时作为副产品得到的。候选项以概率 $m_n(\tilde{\xi}^{i,}{}_n{}^*, \xi_{n+1}^i)/\bar{\sigma}_n$ 被接受。由于这一抽样过程完全绕过了目标分类分布的归一化常数 $\sum_{\ell'=1}^N q_n(\xi_n^{\ell'}, \xi_{n+1}^i)$ 的计算，因此整个算法的复杂度为 $\mathcal{O}(MN)$；详情参见 [Douc 等人， 2011]。

增加 M 可提高算法的准确性，但代价是计算复杂度增加。如 [Olsson 和 Westerborn 2017], 所示，$M = 1$ 和 $M \geq 2$这两种情况存在定性差异，而事实证明，后一种情况是保持 PARIS 数值稳定所必需的。更确切地说，在后一种情况下，可以证明 PARIS估计量 $\mu(\boldsymbol{\beta}_n)$ 在 N趋于无穷、同时 M 保持固定时，满足收敛速率为$\sqrt{N}$ 的中心极限定理（CLT），其n归一化渐近方差的阶为 $\mathcal{O}(1 - 1/(M - 1))$. 从这个界可以清楚看出，取较大的$M$ 只会浪费计算工作，而在实践中，将 M 设为 2 或 3 通常效果良好。

现在我们介绍巴黎粒子Gibbs（PPG）算法。对于所有 $t \in \mathbb{N}^*$，令 $\mathrm{Y}_t \coloneqq \mathrm{X}_{0:t} \times \mathbb{R}$ 和 $\mathcal{Y}_t \coloneqq \mathcal{X}_{0:t} \otimes \mathcal{B}(\mathbb{R})$。此外，令 $\mathrm{Y}_0 \coloneqq \mathrm{X}_0 \times \{0\}$ 和 $\mathcal{Y}_0 \coloneqq \mathcal{X}_0 \otimes \{\{0\}, \emptyset\}$。$\mathrm{Y}_t$ 中的元素始终记作 $y_t = (x_{0:t|t}, b_t)$。巴黎粒子Gibbs采样器的一个关键组成部分是条件 *PARIS* 步骤，它递归地更新一组取值于 Y_t的随机变量 $\upsilon_t^i \coloneqq (\xi_{0:t|t}^i, \beta_t^i), i \in [\![1, N]\!]$。令 $(\boldsymbol{\upsilon}_t)_{t\in\mathbb{N}}$ 表示相应的多体过程，其中每个 $\boldsymbol{\upsilon}_t \coloneqq \{(\xi_{0:t|t}^i, \beta_t^i)\}_{i=1}^N$ 取值于空间 $\mathrm{Y}_t \coloneqq \mathrm{Y}_t^N$，并为其配备一个 σ-域 $\mathrm{Y}_t \coloneqq \mathcal{Y}_t^{\otimes N}$。空间 Y_0 及相应的 σ-域 Y_0 按此定义。对于每个 $t \in \mathbb{N}$，我们将 $\boldsymbol{\xi}_{0:t|t}$ 记作$\{\xi_{0:t|t}^i\}_{i=1}^N$ 中路径的集合，将 $\boldsymbol{\xi}_{t|t}$ 记作同一集合中终点的集合 $\{\xi_{t|t}^i\}_{i=1}^N$。

下文中，我们令 $t \in \mathbb{N}$ 为固定的时间范围，并详细说明 PPG 如何通过迭代近似 $\eta_{0:t}h_t$。简而言之，在每次迭代 ℓ 中，给定一个输入条件

路径 $\zeta_{0:t}[\ell]$，PPG 通过一系列条件 PARIS 操作生成多体系统 $\boldsymbol{\upsilon}_t[\ell+1]$；随后，从 $\boldsymbol{\upsilon}_t[\ell+1]$ 中的路径 $\boldsymbol{\xi}_{0:t|t}[\ell+1]$ 随机选取一条，生成更新后的路径 $\zeta_{0:t}[\ell+1]$，作为下一次迭代的输入。根据 (A.37)，每次迭代中在 $\boldsymbol{\upsilon}_t$ 得到的统计量 $\boldsymbol{\beta}_t$ 都近似于 $\eta_{0:t}h_t$。

更准确地说，给定路径 $\zeta_{0:t}[\ell]$，条件 PARIS 操作按如下方式执行。在初始步骤中，从 (A.31) 中定义的 $\boldsymbol{\eta}_0\langle\zeta_0[\ell]\rangle$ 抽取 $\boldsymbol{\xi}_{0|0}[\ell+1]$，并为所有 $i\in[\![1,N]\!]$ 抽取 $\upsilon_0^i[\ell+1]\leftarrow(\xi_{0|0}^i[\ell+1],0)$；随后，对于 $m\in[\![0,t]\!]$ 递归地进行，假设可以访问 $\boldsymbol{\upsilon}_m[\ell+1]$，

(1) 我们生成更新后的粒子云 $\boldsymbol{\xi}_{m+1}[\ell+1]\sim\boldsymbol{M}_m\langle\zeta_{m+1}[\ell]\rangle(\boldsymbol{\xi}_{m|m}[\ell+1],\cdot)$，(2) 对于每个 $i\in[\![1\ N]\!]$ $\xi_{0\ m}^{i\ 1}[\ell+,$，我们通过抽取，从以下对象中随机 $1]\ \tilde{\beta}_m^{i\ 1}[\ell+1]$ $\boldsymbol{\upsilon}_m[\ell+1]$ 选取一条带有关联统计量的祖先路径 $(.,.)$：

$$(\tilde{\xi}_{0:m}^{i,1}[\ell+1],\tilde{\beta}_m^{i,1}[\ell+1])\sim\sum_{s=1}^{N}\frac{q_m(\xi_{m|m}^s[\ell+1],\xi_{m+1}^i[\ell+1])}{\sum_{s'=1}^{N}q_m(\xi_{m|m}^{s'}[\ell+1],\xi_{m+1}^i[\ell+1])}\delta_{\upsilon_m^s[\ell+1]},\quad i\in[\![1,N]\!],$$

(3) 我们根据 $\{(\xi_{m|m}^s[\ell+1],\beta_m^s)[\ell+1]\}_{s=1}^N$，有放回地随机抽取 $M-1$ 个祖先粒子及其对应的统计量 $\{(\tilde{\xi}^{i}{}_{,m}^{\ j}[\ell+1],\tilde{\beta}^{i}{}_{,}^{\ j}{}_{m}[\ell+1])\}_{j=2}^M$

$$\{(\tilde{\xi}_m^{i,j}[\ell+1],\tilde{\beta}_m^{i,j}[\ell+1])\}_{j=2}^M\sim\left(\sum_{s=1}^{N}\frac{q_m(\xi_{m|m}^s[\ell+1],\xi_{m+1}^i[\ell+1])}{\sum_{s'=1}^{N}q_m(\xi_{m|m}^{s'}[\ell+1],\xi_{m+1}^i[\ell+1])}\delta_{(\xi_{m|m}^s[\ell+1],\beta_m^s[\ell+1])}\right)^{\otimes(M-1)},$$

(4) 对于所有 $i\in[\![1,N]\!]$，$\xi_{0:m+1|m+1}^i[\ell+1]\leftarrow(\tilde{\xi}^{i}{}_{,1\,0:m}[\ell+1],\xi_{m+1}^i[\ell+1])$ 和 $\upsilon_{m+1}^i[\ell+1]\leftarrow(\xi_{0:m+1|m+1}^i[\ell+1],\beta_{m+1}^i[\ell+1])$，我们设定，其中

$$\beta_{m+1}^i[\ell+1]\leftarrow M^{-1}\sum_{j=1}^{M}\left(\tilde{\beta}_m^{i,j}[\ell+1]+\tilde{h}_m(\tilde{\xi}_m^{i,j}[\ell+1],\xi_{m+1}^i[\ell+1])\right).$$

条件 PARIS 程序总结于算法1。

通过 n 次递归条件 PARIS 更新形成轨迹集合及其对应的统计量 $\boldsymbol{\upsilon}_t[\ell+1]$ 后，从 $\mu(\boldsymbol{\xi}_{0:t|t}[\ell+1])$ 中抽取一条更新后的路径 $\zeta_{0:t}[\ell+1]$。PPG 的一次完整扫描总结于算法 2。

以下马尔可夫核将在后续内容中发挥关键作用。对于给定路径$\{z_m\}_{m\in\mathbb{N}}$，算法1中的条件 PARIS 更新定义了空间 $\{(\mathbf{Y}_m,\ \boldsymbol{\mathcal{Y}}_m)\}_{m\in\mathbb{N}}$ 上的非齐次马尔可夫链，其核为

$$\mathbf{Y}_m\times\boldsymbol{\mathcal{Y}}_{m+1}\ni(\boldsymbol{y}_m,A)\mapsto\int\boldsymbol{M}_m\langle z_{m+1}\rangle(\boldsymbol{x}_{m|m},\mathrm{d}\boldsymbol{x}_{m+1})\,\boldsymbol{S}_m(\boldsymbol{y}_m,\boldsymbol{x}_{m+1},A),\quad m\in\mathbb{N},$$

其中

$$\begin{aligned}
&\boldsymbol{S}_m:\mathbf{Y}_m\times\mathbf{X}_{m+1}\times\boldsymbol{\mathcal{Y}}_{m+1}\ni(\boldsymbol{y}_m,\boldsymbol{x}_{m+1},A)\\
&\mapsto\int\cdots\int\mathbb{1}_A\left(\left\{\left((\tilde{x}_{0:m}^{i,1},x_{m+1}^i),\frac{1}{M}\sum_{j=1}^{M}\left(\tilde{b}_m^{i,j}+\tilde{h}_m(\tilde{x}_m^{i,j},x_{m+1}^i)\right)\right)\right\}_{i=1}^{N}\right)\\
&\times\prod_{i=1}^{N}\left(\sum_{\ell=1}^{N}\frac{q_m(x_{m|m}^{\ell},x_{m+1}^i)}{\sum_{\ell'=1}^{N}q_m(x_{m|m}^{\ell'},x_{m+1}^i)}\delta_{y_m^{\ell}}(\mathrm{d}(\tilde{x}_{0:m}^{i,1},\tilde{b}_m^{i,1}))\right.\\
&\left.\times\left(\sum_{\ell=1}^{N}\frac{q_m(x_{m|m}^{\ell},x_{m+1}^i)}{\sum_{\ell'=1}^{N}q_m(x_{m|m}^{\ell'},x_{m+1}^i)}\delta_{(x_{m|m}^{\ell},b_m^{\ell})}\right)^{\otimes(M-1)}(\mathrm{d}(\tilde{x}_m^{i,2},\tilde{b}_m^{i,2},\ldots,\tilde{x}_m^{i,M},\tilde{b}_m^{i,M}))\right).
\end{aligned}\tag{A.39}$$

此外，我们引入联合律

$$\mathbb{S}_t : \mathbf{X}_{0:t} \times \boldsymbol{\mathcal{Y}}_t \ni (\boldsymbol{x}_{0:t}, A) \mapsto \int \cdots \int \mathbb{1}_A(\boldsymbol{y}_t)\, \boldsymbol{S}_0(\mathbf{J}\boldsymbol{x}_0, \boldsymbol{x}_1, \mathrm{d}\boldsymbol{y}_1) \prod_{m=1}^{t-1} \boldsymbol{S}_m(\boldsymbol{y}_m, \boldsymbol{x}_{m+1}, \mathrm{d}\boldsymbol{y}_{m+1}), \quad \text{(A.40)}$$

其中我们定义了 $\mathbf{J} \coloneqq \mathrm{Id}_N \otimes (0,1)^\intercal$。

核 $\mathbb{S}_t$ 可视为一个超占优抽样核，它描述了在给定与底层序贯蒙特卡洛算法相关联的多体过程 $\{\boldsymbol{\xi}_m\}_{m=0}^t$时，由一系列 PARIS 迭代生成的输出 $\boldsymbol{\upsilon}_t$ 的分布。因此，我们也可以如下描述 PPG：给定 $\zeta_{0:t}[\ell]$，抽取 $\boldsymbol{\xi}_{0:t}[\ell+1] \sim \mathbb{C}_t(\zeta_{0:t}[\ell], \cdot)$；然后抽取 $\boldsymbol{\upsilon}_t[\ell+1] \sim \mathbb{S}_t(\boldsymbol{\xi}_{0:t}[\ell+1], \cdot)$，并从 $\boldsymbol{\xi}_{0:t|t}[\ell+1]$ 中随机选取一条轨迹 $\zeta_{0:t}[\ell+1]$。下面的命题将在后续推导中发挥重要作用，它证明了在给定 $\boldsymbol{\xi}_{0:t}[\ell+1]$ 的条件下，$\zeta_{0:t}[\ell+1]$的条件分布正如预期的那样，与粒子诱导的反向动力学 $\mathbb{B}_t$一致。

命题 2. 对于所有 $t \in \mathbb{N}^*$, $N \in \mathbb{N}^*$, $\boldsymbol{x}_{0:t} \in \mathbf{X}_{0:t}$，以及 $h \in \mathsf{F}(\mathcal{X}_{0:t})$，

$$\int \mathbb{S}_t(\boldsymbol{x}_{0:t}, \mathrm{d}\boldsymbol{y}_t)\, \mu(\boldsymbol{x}_{0:t|t}) h = \mathbb{B}_t h(\boldsymbol{x}_{0:t}).$$

最后，我们定义由 PPG 诱导的马尔可夫核，以及它所针对的扩展概率分布。为此，我们引入扩展可测空间 $(\mathbf{E}_t, \boldsymbol{\mathcal{E}}_t)$，其中

$$\mathbf{E}_t \coloneqq \mathbf{Y}_t \times \mathsf{X}_{0:t}, \quad \boldsymbol{\mathcal{E}}_t \coloneqq \boldsymbol{\mathcal{Y}}_t \otimes \mathcal{X}_{0:t}.$$

算法 2 中描述的 PPG 在 $(\mathbf{E}_t, \boldsymbol{\mathcal{E}}_t)$ 上定义了一个马尔可夫链，其马尔可夫转移核为

$$\mathbb{K}_t : \mathbf{E}_t \times \boldsymbol{\mathcal{E}}_t \ni (\boldsymbol{y}_t, z_{0:t}, A) \mapsto \iiint \mathbb{1}_A(\tilde{\boldsymbol{y}}_t, \tilde{z}_{0:t})\, \mathbb{C}_t(z_{0:t}, \mathrm{d}\tilde{\boldsymbol{x}}_{0:t})\, \mathbb{S}_t(\tilde{\boldsymbol{x}}_{0:t}, \mathrm{d}\tilde{\boldsymbol{y}}_t)\, \mu(\tilde{\boldsymbol{x}}_{0:t|t})(\mathrm{d}\tilde{z}_{0:t}). \quad \text{(A.41)}$$

请注意，上文定义的 $\mathbb{K}_t$ 不依赖于 $\boldsymbol{y}_t$，而仅依赖于 $(z_{0:t}, A)$。对于任意给定的初始分布 $\xi \in \mathsf{M}_1(\mathcal{X}_{0:t})$，令 $\mathbb{P}_\xi$ 表示由核 $\mathbb{K}_t$ 和初始分布 ξ 诱导的典范马尔可夫链的分布。对于特殊情形，即给定路径 $z_{0:t} \in \mathbf{X}_{0:t}$ 时 $\xi = \delta_{z_{0:t}}$，我们采用简写记号 $\mathbb{P}_{\delta_{z_{0:t}}} = \mathbb{P}_{z_{0:t}^\circ}$。此外，记

$$K_t : \mathsf{X}_{0:t} \times \mathcal{X}_{0:t} \ni (z_{0:t}, A) \mapsto \iiint \mathbb{1}_A(\tilde{z}_{0:t})\, \mathbb{C}_t(z_{0:t}, \mathrm{d}\tilde{\boldsymbol{x}}_{0:t})\, \mathbb{S}_t(\tilde{\boldsymbol{x}}_{0:t}, \mathrm{d}\tilde{\boldsymbol{y}}_t)\, \mu(\tilde{\boldsymbol{x}}_{0:t|t})(\mathrm{d}\tilde{z}_{0:t}) \quad \text{(A.42)}$$

$\mathbb{K}_t$的路径边缘化版本。根据命题 2，有 $K_t = \mathbb{C}_t \mathbb{B}_t$，这表明 K_t与附录 A.3 中讨论的基于反向抽样的粒子吉布斯采样器的马尔可夫转移核一致。也可以指定 $\mathbb{K}_t$的不变分布。

命题 3. 对于所有 $t \in \mathbb{N}^*$，有

$$\eta_{0:t} \mathbb{C}_t \mathbb{S}_t \mathbb{K}_t = \eta_{0:t} \mathbb{C}_t \mathbb{S}_t \,. \quad \text{(A.43)}$$

证明。设 $f \in \mathsf{M}(\mathbf{E}^{\otimes(k-k_0)_t})$。

$$\begin{aligned}
&\int f(\tilde{\boldsymbol{y}}_t, \tilde{z}_{0:t}) \eta_{0:t}(\mathrm{d}z_{0:t}) \mathbb{C}_t \mathbb{S}_t(z_{0:t}, \mathrm{d}(\boldsymbol{y}_t, z'_{0:t})) \mathbb{K}_t(z'_{0:t}, \boldsymbol{y}_t, \mathrm{d}(\tilde{\boldsymbol{y}}_t, \tilde{z}_{0:t})) \\
&= \int f(\tilde{\boldsymbol{y}}_t, \tilde{z}_{0:t}) \eta_{0:t}(\mathrm{d}z_{0:t}) \mathbb{C}_t \mathbb{S}_t(z_{0:t}, \mathrm{d}(\boldsymbol{y}_t, z'_{0:t})) \mathbb{C}_t \mathbb{S}_t(z'_{0:t}, \mathrm{d}(\tilde{\boldsymbol{y}}_t, \tilde{z}_{0:t})) \\
&= \int f(\tilde{\boldsymbol{y}}_t, \tilde{z}_{0:t}) \eta_{0:t}(\mathrm{d}z_{0:t}) K_t(z_{0:t}, \mathrm{d}z'_{0:t}) \mathbb{C}_t \mathbb{S}_t(z'_{0:t}, \mathrm{d}(\tilde{\boldsymbol{y}}_t, \tilde{z}_{0:t})) \\
&= \int f(\tilde{\boldsymbol{y}}_t, \tilde{z}_{0:t}) \eta_{0:t}(\mathrm{d}z'_{0:t}) \mathbb{C}_t \mathbb{S}_t(z'_{0:t}, \mathrm{d}(\tilde{\boldsymbol{y}}_t, \tilde{z}_{0:t})) \,.
\end{aligned}$$

□

最后，为陈述我们关于 PPG 的理论结果，我们需要引入以下带有冻结路径的 Feynman–Kac 路径模型。更确切地说，给定一条路径 $z_{0:t} \in \mathsf{X}_{0:t}$，对每个 $m \in [\![0, t-1]\!]$，定义未归一化核

$$Q_m\langle z_{m+1}\rangle : \mathsf{X}_m \times \mathcal{X}_{m+1} \ni (x_m, A) \mapsto \left(1 - \frac{1}{N}\right) Q_m(x_m, A) + \frac{1}{N} g_m(x_m)\, \delta_{z_{m+1}}(A)$$

以及初始分布 $\eta_0\langle z_0\rangle$：$\mathcal{X}_0 \ni A \mapsto (1-1/N)\eta_0(A) + \delta_{z_0}(A)/N$。给定这些量，对 $m \in [\![0,\ t]\!]$，$\gamma_m\langle z_{0:m}\rangle$：$= \eta_0\langle z_0\rangle Q_0\langle z_1\rangle \cdots Q_{m-1}\langle z_m\rangle$ 定义其归一化对应项$\eta_m\langle z_{0:m}\rangle \coloneqq \gamma_m\langle z_{0:m}\rangle / \gamma_m\langle z_{0:m}\rangle \mathbb{1}_{\mathsf{X}_{0:m}}$。最后，对于 $m \in [\![0,\ t]\!]$，我们引入核

$$B_m\langle z_{0:m-1}\rangle : \mathsf{X}_m \times \mathcal{X}_{0:m-1} \ni (x_m, A) \mapsto \int \cdots \int \mathbb{1}_A(x_{0:m-1}) \prod_{m=0}^{t-1} \overleftarrow{Q}_{m,\eta_m\langle z_{0:m}\rangle}(x_{m+1}, \mathrm{d}x_m),$$

以及路径模型 $\eta_{0:m}\langle z_{0:m}\rangle \coloneqq B_m\langle z_{0:m-1}\rangle \otimes \eta_m\langle z_{0:m}\rangle$。

A.5 定理 1 的证明

我们首先为 PPG 估计量在固定迭代下建立偏差、均方误差和协方差界。

定理 4. 假设 ***A3.1***。则对于每个 $t \in \mathbb{N}$，存在 c_t^{bias}、c_t^{mse} 和 c_t^{cov}，属于 $\mathbb{R}_+^*$，使得对于每个 $M \in \mathbb{N}^*$、$\xi \in \mathsf{M}_1(\mathcal{X}_{0:t})$、$\ell \in \mathbb{N}^*$、$s \in \mathbb{N}^*$ 和 $N \in \mathbb{N}^*$，满足 $N > N_t$，

$$|\mathbb{E}_\xi\left[\mu(\boldsymbol{\beta}_t[\ell])(\mathrm{id})\right] - \eta_{0:t}h_t| \leq \mathsf{c}_t^{bias} \left(\sum_{m=0}^{t-1} \|\tilde{h}_m\|_\infty\right) N^{-1} \kappa_{N,t}^{\ell}, \tag{A.44}$$

$$\mathbb{E}_\xi\left[\left(\mu(\boldsymbol{\beta}_t[\ell])(\mathrm{id}) - \eta_{0:t}h_t\right)^2\right] \leq \mathsf{c}_t^{mse} \left(\sum_{m=0}^{t-1} \|\tilde{h}_m\|_\infty\right)^2 N^{-1}, \tag{A.45}$$

$$|\mathbb{E}_\xi\left[\left(\mu(\boldsymbol{\beta}_t[\ell])(\mathrm{id}) - \eta_{0:t}h_t\right)\left(\mu(\boldsymbol{\beta}_t[\ell+s])(\mathrm{id}) - \eta_{0:t}h_t\right)\right]| \leq \mathsf{c}_t^{cov} \left(\sum_{m=0}^{t-1} \|\tilde{h}_m\|_\infty\right)^2 N^{-3/2} \kappa_{N,t}^{s}. \tag{A.46}$$

常数 c_t^{bias}、c_t^{mse} 和 c_t^{cov}在证明中有明确给出。由于本文关注的是对 N 及索引 ℓ的依赖，我们在证明中并未尝试优化这些常数对 t的依赖；不过，我们认为，在所述假设下，可以证明这种依赖是线性的。定理 4 中界的证明基于四个关键要素。首先是 PARIS 在多体 Feynman–Kac 路径模型下具有如下无偏性。

定理 5. 对于任意 $t \in \mathbb{N}$、$N \in \mathbb{N}^*$ 和 $\ell \in \mathbb{N}^*$，

$$\mathbb{E}_{\eta_{0:t}}\left[\mu(\boldsymbol{\beta}_t[\ell])(\mathrm{id})\right] = \int \eta_{0:t}\mathbb{C}_t\mathbb{S}_t(\mathrm{d}\boldsymbol{b}_t)\, \mu(\boldsymbol{b}_t)(\mathrm{id}) = \int \boldsymbol{\eta}_{0:t}\mathbb{S}_t(\mathrm{d}\boldsymbol{b}_t)\, \mu(\boldsymbol{b}_t)(\mathrm{id}) = \eta_{0:t}h_t.$$

定理 5 的证明推迟至附录 A.6.3。定理4证明的第二个要素是 [Del Moral 和 Jasra 在 2018] 中确立的带反向采样的粒子吉布斯的一致几何遍历性。

定理 6. 假设 ***A 3.1***。那么，对于任意 $t \in \mathbb{N}$，$(\mu, \nu) \in \mathsf{M}_1(\mathcal{X}_{0:t})^2$、$\ell \in \mathbb{N}^*$ 和 $N \in \mathbb{N}^*$，使得 $N > 1 + 5\rho_t^2 t/2$、$\|\mu K_t^\ell - \nu K_t^\ell\|_{\mathrm{TV}} \leq \kappa_{N,t}^\ell$，其中 $\kappa_{N,t}$ 定义于 (3.14)。

作为第三个要素，我们需要关于条件 PARIS 相对于上一节定义的冻结路径费曼–卡茨模型的一致指数集中不等式。

定理 7。 对于每个 $t \in \mathbb{N}$，存在 $\mathsf{c}_t > 0$ 和 $\mathsf{d}_t > 0$，使得对于每个 $M \in \mathbb{N}^*$、 $z_{0:t} \in \mathsf{X}_{0:t}$、$N \in \mathbb{N}^*$ 和 $\varepsilon > 0$,

$$\int \mathbb{C}_t \mathbb{S}_t(z_{0:t}, \mathrm{d}\boldsymbol{b}_t) \mathbb{1}\left\{|\mu(\boldsymbol{b}_t)(\mathrm{id}) - \eta_{0:t}\langle z_{0:t}\rangle h_t| \geq \varepsilon\right\} \leq \mathsf{c}_t \exp\left(-\frac{\mathsf{d}_t N \varepsilon^2}{2(\sum_{m=0}^{t-1} \|\tilde{h}_m\|_\infty)^2}\right).$$

定理 7 的证明推迟至附录 A.6.5；该定理进而给出以下条件方差界。

命题 4. 对于每个 $t \in \mathbb{N}$, $M \in \mathbb{N}^*$, $z_{0:t} \in \mathsf{X}_{0:t}$，以及 $N \in \mathbb{N}^*$,

$$\int \mathbb{C}_t \mathbb{S}_t(z_{0:t}, \mathrm{d}\boldsymbol{b}_t) \left|\mu(\boldsymbol{b}_t)(\mathrm{id}) - \eta_{0:t}\langle z_{0:t}\rangle h_t\right|^2 \leq \frac{\mathsf{c}_t}{\mathsf{d}_t} \left(\sum_{m=0}^{t-1} \|\tilde{h}_m\|_\infty\right)^2 N^{-1}.$$

利用命题 4，我们进而推得以下偏差界，其证明推迟至附录 A.6.7。

命题 5. 对于每个 $t \in \mathbb{N}$，存在 $\bar{\mathsf{c}}_t^{bias} > 0$，使得对于每个 $M \in \mathbb{N}^*$, $z_{0:t} \in \mathsf{X}_{0:t}$，以及$N \in \mathbb{N}^*$,

$$\left|\int \mathbb{C}_t \mathbb{S}_t(z_{0:t}, \mathrm{d}\boldsymbol{b}_t)\, \mu(\boldsymbol{b}_t)(\mathrm{id}) - \eta_{0:t}\langle z_{0:t}\rangle h_t\right| \leq \bar{\mathsf{c}}_t^{bias} N^{-1} \left(\sum_{m=0}^{t-1} \|\tilde{h}_m\|_\infty\right).$$

定理 4 证明中的第四个也是最后一个要素，是以下关于原始 Feynman–Kac 模型与冻结路径 Feynman–Kac 模型下加性期望之间差异的界。该界利用 [Gloaguen 等人 2022] 的新结果建立。更确切地说，由于对于每个 $m \in \mathbb{N}$，$(x,\ z) \in \mathsf{X}_m^2$, $N \in \mathbb{N}^*$，以及 $h \in \mathrm{F}(\mathcal{X}_{m+1})$，利用 **A3.1**,

$$|Q_m\langle z\rangle h(x) - Q_m h(x)| \leq \frac{1}{N}\|g_m\|_\infty \|h\|_\infty \leq \frac{1}{N}\bar{\tau}_m \|h\|_\infty,$$

应用 [Gloaguen 等，2022，定理 43] 可得如下结果。

命题 6。 假设***A3.1***。则存在$\mathsf{c} > 0$，使得对于每个 $t \in \mathbb{N}$、 $N \in \mathbb{N}$ 和$z_{0:t} \in \mathsf{X}_{0:t}$,

$$|\eta_{0:t}\langle z_{0:t}\rangle h_t - \eta_{0:t} h_t| \leq \mathsf{c} N^{-1} \sum_{m=0}^{t-1} \|\tilde{h}_m\|_\infty.$$

注意，若进一步假设 $\sup_{t\in\mathbb{N}} \|h_t\|_\infty < \infty$，则命题 6 中得到一个 $\mathcal{O}(n/N)$ 界。最后，结合这些结果，我们现在可以给出定理 4 的证明。

定理 *4* 的证明。 利用塔式性质，写作

$$\mathbb{E}_\xi\left[\mu(\boldsymbol{\beta}_t[\ell])(\mathrm{id})\right] = \mathbb{E}_\xi\left[\mathbb{E}_{\zeta_{0:t}[\ell]}\left[\mu(\boldsymbol{\beta}_t[0])(\mathrm{id})\right]\right] = \int \xi K_t^\ell \mathbb{C}_t \mathbb{S}_t(\mathrm{d}\boldsymbol{b}_t)\, \mu(\boldsymbol{b}_t)(\mathrm{id}).$$

因此，根据定理 5 中的无偏性,

$$\begin{aligned}\left|\mathbb{E}_\xi\left[\mu(\boldsymbol{\beta}_t[\ell])(\mathrm{id})\right] - \eta_{0:t} h_t\right| &= \left|\int \xi K_t^\ell \mathbb{C}_t \mathbb{S}_t(\mathrm{d}\boldsymbol{b}_t)\, \mu(\boldsymbol{b}_t)(\mathrm{id}) - \int \eta_{0:t} \mathbb{C}_t \mathbb{S}_t(\mathrm{d}\boldsymbol{b}_t)\, \mu(\boldsymbol{b}_t)(\mathrm{id})\right| \\ &\leq \left\|\xi K_t^\ell - \eta_{0:t}\right\|_{\mathrm{TV}} \operatorname{osc}\left(\int \mathbb{C}_t \mathbb{S}_t(\cdot, \mathrm{d}\boldsymbol{b}_t)\, \mu(\boldsymbol{b}_t)(\mathrm{id})\right),\end{aligned}$$

其中，根据定理 6，$\|\xi K_t^\ell - \eta_{0:t}\|_{\mathrm{TV}} \leq \kappa_{N,\ t}^\ell$。此外，为了推导振荡的上界，我们考虑分解

$$\operatorname{osc}\left(\int \mathbb{C}_t\mathbb{S}_t(\cdot, \mathrm{d}\boldsymbol{b}_t)\,\mu(\boldsymbol{b}_t)(\mathrm{id})\right) \leq 2\left(\left\|\int \mathbb{C}_t\mathbb{S}_t(\cdot, \mathrm{d}\boldsymbol{b}_t)\,\mu(\boldsymbol{b}_t)(\mathrm{id}) - \eta_{0:t}\langle\cdot\rangle h_t\right\|_\infty + \|\eta_{0:t}\langle\cdot\rangle h_t - \eta_{0:t}h_t\|_\infty\right),$$

右端的两项分别可利用命题 6 和命题 5 进行界定。这就完成了 (A.44) 的证明。现在我们考虑 (A.45) 的证明。写作

$$\mathbb{E}_\xi\left[\left(\mu(\boldsymbol{\beta}_t[\ell])(\mathrm{id}) \ - \eta_{0:t}h_t\right)^2\right] = \int \xi K_t^\ell(\mathrm{d}z_{0:t})\mathbb{C}_t\mathbb{S}_t(z_{0:t}, \mathrm{d}\boldsymbol{b}_t)\left(\mu(\boldsymbol{b}_t)(\mathrm{id}) - \eta_{0:t}h_t\right)^2,$$

我们可以利用命题 4 和命题 6 得到 (A.45)。最后考虑 (A.46)。根据马尔可夫性质，我们得到

$$\begin{aligned}\mathbb{E}_\xi\left[\left(\mu(\boldsymbol{\beta}_t[\ell])(\mathrm{id}) \ - \eta_{0:t}h_t\right)\left(\mu(\boldsymbol{\beta}_t[\ell+s])(\mathrm{id}) \ - \eta_{0:t}h_t\right)\right]& \\ = \mathbb{E}_\xi\left[\left(\mu(\boldsymbol{\beta}_t[\ell])(\mathrm{id}) \ - \eta_{0:t}h_t\right)\left(\mathbb{E}_{\zeta_{0:t}[\ell]}[\mu(\boldsymbol{\beta}_t[s])(\mathrm{id})] \ - \eta_{0:t}h_t\right)\right],&\end{aligned}$$

由此，根据 (A.44) 和 (A.45) 即可推出 (A.46)。 □

现在我们终于可以证明定理 1 了。

定理 *1* 的证明。 我们首先考虑偏差，它可以根据以下不等式进行界定:

$$\begin{aligned}\left|\mathbb{E}_\xi[\Pi_{(k_0,k),N}(f)] - \eta_{0:t}h_t\right| &\leq (k-k_0)^{-1}\sum_{\ell=k_0+1}^{k}\left|\mathbb{E}_\xi\mu(\boldsymbol{\beta}_t[\ell])(\mathrm{id}) - \eta_{0:t}h_t\right| \\ &\leq (k-k_0)^{-1}N^{-1}\mathsf{c}_t^{bias}\left(\sum_{m=0}^{t-1}\|\tilde{h}_m\|_\infty\right)\sum_{\ell=k_0+1}^{k}\kappa_{N,t}^\ell,\end{aligned}$$

由此立即可得界 (3.15)。

接下来讨论均方误差。利用分解

$$\begin{aligned}\mathbb{E}_\xi[(\Pi_{(k_0,k),N}(f) - \eta_{0:t}h_t)^2] \leq (k-k_0)^{-2}\Bigg\{&\sum_{\ell=k_0+1}^{k}\mathbb{E}_\xi[(\mu(\boldsymbol{\beta}_t[\ell])(\mathrm{id}) - \eta_{0:t}h_t)^2] \\ &+ 2\sum_{\ell=k_0+1}^{k}\sum_{j=\ell+1}^{k}\mathbb{E}_\xi[(\mu(\boldsymbol{\beta}_t[\ell])(\mathrm{id}) - \eta_{0:t}h_t)(\mu(\boldsymbol{\beta}_t[j])(\mathrm{id}) - \eta_{0:t}h_t)]\Bigg\},\end{aligned}$$

定理 4 中的均方误差界表明

$$\sum_{\ell=k_0+1}^{k}\mathbb{E}_\xi[(\mu(\boldsymbol{\beta}_t[\ell])(\mathrm{id}) - \eta_{0:t}h_t)^2] \leq \mathsf{c}_t^{mse}\left(\sum_{m=0}^{t-1}\|\tilde{h}_m\|_\infty\right)^2 N^{-1}(k-k_0).$$

此外，利用定理 4 中的协方差界，我们推得

$$\sum_{\ell=k_0+1}^{k}\sum_{j=\ell+1}^{k}\mathbb{E}_\xi[(\mu(\boldsymbol{\beta}_t[\ell])(\mathrm{id}) - \eta_{0:t}h_t)(\mu(\boldsymbol{\beta}_t[j])(\mathrm{id}) - \eta_{0:t}h_t)] \leq \mathsf{c}_t^{cov}\left(\sum_{m=0}^{t-1}\|\tilde{h}_m\|_\infty\right)^2 N^{-3/2}\left(\sum_{\ell=k_0+1}^{k}\sum_{j=\ell+1}^{k}\kappa_{N,t}^{(j-\ell)}\right).$$

Thus，注意到 $\sum_{\ell=k_0+1}^{k}\sum_{j=\ell+1}^{k}\kappa_{N}^{(j-\ell)},_t \leq (k-k_0)/(1-\kappa_{N,t}$，证明即告完成). □

A.6 中间结果的证明

A.6.1 命题 1 的证明

利用恒等式

$$\eta_0 Q_0 \cdots Q_{t-1} \mathbb{1}_{\mathsf{X}_t} = \prod_{m=0}^{t-1} \eta_m Q_m \mathbb{1}_{\mathsf{X}_{m+1}}$$

并且由于每个核 Q_m 都有转移密度，对于 $h \in \mathsf{F}(\mathcal{X}_{0:t})$，可写为：

$$\begin{aligned}
\eta_{0:t} h &= \int \cdots \int h(x_{0:t})\, \eta_0(\mathrm{d}x_0) \prod_{m=0}^{t-1} \left(\frac{\eta_m[q_m(\cdot, x_{m+1})]\, \lambda_{m+1}(\mathrm{d}x_{m+1})}{\eta_m Q_m \mathbb{1}_{\mathsf{X}_{m+1}}} \right) \left(\frac{q_m(x_m, x_{m+1})}{\eta_m[q_m(\cdot, x_{m+1})]} \right) \\
&= \int \cdots \int h(x_{0:t})\, \eta_t(\mathrm{d}x_t) \prod_{m=0}^{t-1} \frac{\eta_m(\mathrm{d}x_m)\, q_m(x_m, x_{m+1})}{\eta_m[q_m(\cdot, x_{m+1})]} \qquad \text{(A.47)} \\
&= \left(\overleftarrow{Q}_{0,\eta_0} \otimes \cdots \otimes \overleftarrow{Q}_{n-1,\eta_{t-1}} \otimes \eta_t \right) h,
\end{aligned}$$

这正是需要证明的结论。

A.6.2 定理3的证明

引理 1. 对于所有 $t \in \mathbb{N}$, $\boldsymbol{x}_t \in \mathbf{X}_t$，以及 $h \in \mathsf{F}(\boldsymbol{\mathcal{X}}_{t+1} \otimes \mathcal{X}_{t+1})$,

$$\iint h(\boldsymbol{x}_{t+1}, z_{t+1})\, \boldsymbol{Q}_t(\boldsymbol{x}_t, \mathrm{d}\boldsymbol{x}_{t+1})\, \mu(\boldsymbol{x}_{t+1})(\mathrm{d}z_{t+1}) = \iint h(\boldsymbol{x}_{t+1}, z_{t+1})\, \mu(\boldsymbol{x}_t) Q_t(\mathrm{d}z_{t+1})\, \boldsymbol{M}_t\langle z_{t+1}\rangle(\boldsymbol{x}_t, \mathrm{d}\boldsymbol{x}_{t+1}). \tag{A.48}$$

此外，对于所有 $h \in \mathsf{F}(\boldsymbol{\mathcal{X}}_0 \otimes \mathcal{X}_0)$,

$$\iint h(\boldsymbol{x}_0, z_0)\, \boldsymbol{\eta}_0(\mathrm{d}\boldsymbol{x}_0)\, \mu(\boldsymbol{x}_0)(\mathrm{d}z_0) = \iint h(\boldsymbol{x}_0, z_0)\, \boldsymbol{\eta}_0\langle z_0\rangle(\mathrm{d}\boldsymbol{x}_0)\, \eta_0(\mathrm{d}z_0). \tag{A.49}$$

证明。由于 $\mu(\boldsymbol{x}_t)\, Q_t(\mathrm{d}z_{t+1}) = \boldsymbol{g}_t(\boldsymbol{x}_t)\ \Phi_t(\mu(\boldsymbol{x}_t))(\mathrm{d}z_{t+1})$，我们可以根据下式重写 (A.48) 的右端

$$\begin{aligned}
&\iint h(\boldsymbol{x}_{t+1}, z_{t+1})\, \mu(\boldsymbol{x}_t) Q_t(\mathrm{d}z_{t+1})\, \boldsymbol{M}_t\langle z_{t+1}\rangle(\boldsymbol{x}_t, \mathrm{d}\boldsymbol{x}_{t+1}) \\
&\qquad = \boldsymbol{g}_t(\boldsymbol{x}_t) \frac{1}{N} \sum_{i=0}^{N-1} \iint h(\boldsymbol{x}_{t+1}, z_{t+1})\, \Phi_t(\mu(\boldsymbol{x}_t))(\mathrm{d}z_{t+1}) \\
&\qquad\qquad \times \left(\Phi_t(\mu(\boldsymbol{x}_t))^{\otimes i} \otimes \delta_{z_{t+1}} \otimes \Phi_t(\mu(\boldsymbol{x}_t))^{\otimes(N-i-1)} \right) (\mathrm{d}\boldsymbol{x}_{t+1}) \\
&\qquad = \boldsymbol{g}_t(\boldsymbol{x}_t) \frac{1}{N} \sum_{i=1}^{N} \int \cdots \int h((x_{t+1}^1, \ldots, x_{t+1}^{i-1}, z_{t+1}, x_{t+1}^{i+1}, \ldots, x_{t+1}^N), z_{t+1}) \\
&\qquad\qquad \times \Phi_t(\mu(\boldsymbol{x}_t))(\mathrm{d}z_{t+1}) \prod_{\ell \neq i} \Phi_t(\mu(\boldsymbol{x}_t))(\mathrm{d}x_{t+1}^\ell) \\
&\qquad = \boldsymbol{g}_t(\boldsymbol{x}_t) \frac{1}{N} \sum_{i=1}^{N} \int h(\boldsymbol{x}_{t+1}, x_{t+1}^i)\, \boldsymbol{M}_t(\boldsymbol{x}_t, \mathrm{d}\boldsymbol{x}_{t+1}).
\end{aligned}$$

另一方面，注意到 (A.48) 的左侧可以表示为

$$\iint h(\boldsymbol{x}_{t+1}, z_{t+1})\, \boldsymbol{Q}_t(\boldsymbol{x}_t, \mathrm{d}\boldsymbol{x}_{t+1})\, \mu(\boldsymbol{x}_{t+1})(\mathrm{d}z_{t+1}) = \boldsymbol{g}_t(\boldsymbol{x}_t) \frac{1}{N} \sum_{i=1}^{N} \int h(\boldsymbol{x}_{t+1}, x_{t+1}^i)\, \boldsymbol{M}_t(\boldsymbol{x}_t, \mathrm{d}\boldsymbol{x}_{t+1}),$$

由此得出该恒等式。恒等式 (A.49) 的证明思路类似 . □

我们通过归纳法证明定理 3；因此，假设该命题对 n 成立，并证明对于所有 $h \in \mathsf{F}(\boldsymbol{\mathcal{X}}_{0:t+1} \otimes \mathcal{X}_{0:t+1})$,

$$\iint h(\boldsymbol{x}_{0:t+1}, z_{0:t+1})\, \boldsymbol{\gamma}_{0:t+1}(\mathrm{d}\boldsymbol{x}_{0:t+1})\, \mathbb{B}_{t+1}(\boldsymbol{x}_{0:t+1}, \mathrm{d}z_{0:t+1}) = \iint h(\boldsymbol{x}_{0:t+1}, z_{0:t+1})\, \gamma_{0:t+1}(\mathrm{d}z_{0:t+1})\, \mathbb{C}_{t+1}(z_{0:t+1}, \mathrm{d}\boldsymbol{x}_{0:t+1}). \quad \text{(A.50)}$$

为证明这一点，根据定义（C.85），我们对（A.50）的左侧进行如下处理:

$$\begin{aligned} &\iint h(\boldsymbol{x}_{0:t+1}, z_{0:t+1})\, \boldsymbol{\gamma}_{0:t+1}(\mathrm{d}\boldsymbol{x}_{0:t+1})\, \mathbb{B}_{t+1}(\boldsymbol{x}_{0:t+1}, \mathrm{d}z_{0:t+1}) \\ &= \iint \boldsymbol{\gamma}_{0:t}(\mathrm{d}\boldsymbol{x}_{0:t})\, \mathbb{B}_t(\boldsymbol{x}_{0:t}, \mathrm{d}z_{0:t}) \\ &\qquad \times \iint \bar{h}(\boldsymbol{x}_{0:t+1}, z_{0:t+1})\, \boldsymbol{Q}_t(\boldsymbol{x}_t, \mathrm{d}\boldsymbol{x}_{t+1})\, \mu(\boldsymbol{x}_{t+1})(\mathrm{d}z_{t+1}), \end{aligned} \quad \text{(A.51)}$$

其中，我们定义了函数

$$\bar{h}(\boldsymbol{x}_{0:t+1}, z_{0:t+1}) \coloneqq \frac{q_t(z_t, z_{t+1}) h(\boldsymbol{x}_{0:t+1}, z_{0:t+1})}{\mu(\boldsymbol{x}_t)[q_t(\cdot, z_{t+1})]}.$$

现在，将引理 1 应用于内积分，并利用

$$\mu(\boldsymbol{x}_t) Q_t(\mathrm{d}z_{t+1}) = \mu(\boldsymbol{x}_t)[q_t(\cdot, z_{t+1})]\, \lambda_{t+1}(\mathrm{d}z_{t+1})$$

可得，对于每个 $\boldsymbol{x}_{0:t}$ 和 $z_{0:t}$,

$$\begin{aligned} &\iint \bar{h}(\boldsymbol{x}_{0:t+1}, z_{0:t+1})\, \boldsymbol{Q}_t(\boldsymbol{x}_t, \mathrm{d}\boldsymbol{x}_{t+1})\, \mu(\boldsymbol{x}_{t+1})(\mathrm{d}z_{t+1}) \\ &= \iint \bar{h}(\boldsymbol{x}_{0:t+1}, z_{0:t+1})\, \mu(\boldsymbol{x}_t) Q_t(\mathrm{d}z_{t+1})\, \boldsymbol{M}_t\langle z_{t+1}\rangle(\boldsymbol{x}_t, \mathrm{d}\boldsymbol{x}_{t+1}) \\ &= \iint h(\boldsymbol{x}_{0:t+1}, z_{0:t+1})\, Q_t(z_t, \mathrm{d}z_{t+1})\, \boldsymbol{M}_t\langle z_{t+1}\rangle(\boldsymbol{x}_t, \mathrm{d}\boldsymbol{x}_{t+1}). \end{aligned}$$

将前述恒等式代入（A.51），并利用归纳假设，可得

$$\begin{aligned} &\iint h(\boldsymbol{x}_{0:t+1}, z_{0:t+1})\, \boldsymbol{\gamma}_{0:t+1}(\mathrm{d}\boldsymbol{x}_{0:t+1})\, \mathbb{B}_{t+1}(\boldsymbol{x}_{0:t+1}, \mathrm{d}z_{0:t+1}) \\ &= \iint \gamma_{0:t}(\mathrm{d}z_{0:t})\, \mathbb{C}_t(z_{0:t}, \mathrm{d}\boldsymbol{x}_{0:t}) \\ &\qquad \times \iint h(\boldsymbol{x}_{0:t+1}, z_{0:t+1})\, Q_t(z_t, \mathrm{d}z_{t+1})\, \boldsymbol{M}_t\langle z_{t+1}\rangle(\boldsymbol{x}_t, \mathrm{d}\boldsymbol{x}_{t+1}) \\ &= \iint h(\boldsymbol{x}_{0:t+1}, z_{0:t+1})\, \gamma_{0:t+1}(\mathrm{d}z_{0:t+1})\, \mathbb{C}_{t+1}(z_{0:t+1}, \mathrm{d}\boldsymbol{x}_{0:t+1}), \end{aligned}$$

这就证明了（A.50）。

A.6.3 定理5的证明

首先，对 $m \in \mathbb{N}$ 定义:

$$\boldsymbol{P}_m : \mathbf{Y}_m \times \boldsymbol{\mathcal{Y}}_{m+1} \ni (\boldsymbol{y}_m, A) \mapsto \int \boldsymbol{M}_m(\boldsymbol{x}_{m|m}, \mathrm{d}\boldsymbol{x}_{m+1})\, \boldsymbol{S}_m(\boldsymbol{y}_m, \boldsymbol{x}_{m+1}, A). \quad \text{(A.52)}$$

对于任意给定的初始分布 $\psi_0 \in \mathsf{M}_1(\mathcal{Y}_0)$，令 $\mathbb{P}^{\boldsymbol{P}}_{\psi_0}$ 表示由马尔可夫核 $\{\boldsymbol{P}_m\}_{m\in\mathbb{N}}$ 和初始分布 ψ_0 所诱导的典范马尔可夫链的分布。为简化记法，对于 $\boldsymbol{\eta}_0 \in \mathsf{M}_1(\boldsymbol{\mathcal{X}}_0)$，我们将 $\mathbb{P}^{\boldsymbol{P}}_{\psi_0[\boldsymbol{\eta}_0]}$ 记作 $\mathbb{P}^{\boldsymbol{P}}_{\boldsymbol{\eta}_0}$，其中扩展 $\psi_0[\boldsymbol{\eta}_0](A) = \int \mathbb{1}_A(\mathbf{J}\boldsymbol{x}_0)\,\boldsymbol{\eta}_0(\mathrm{d}\boldsymbol{x}_0)$，$A \in \boldsymbol{\mathcal{Y}}_0$ 定义如下。我们先介绍若干技术引理和一个命题，再证明定理 5。

引理 2。对于所有 $t \in \mathbb{N}$ 和 $(f_{t+1}, \tilde{f}_{t+1}) \in \mathsf{F}(\mathcal{X}_{t+1})^2$，

$$\gamma_{t+1}(f_{t+1}B_{t+1}h_{t+1} + \tilde{f}_{t+1}) = \gamma_t\{Q_t f_{t+1}B_t h_t + Q_t(\tilde{h}_t f_{t+1} + \tilde{f}_{t+1})\}.$$

证明。任取 $\varphi \in \mathsf{F}(\mathcal{X}_{t:t+1})$，并利用定义 (A.27) 以及 Q_t 具有转移密度这一事实，写出：

$$\begin{aligned}
&\iint \varphi(x_{t:t+1})\,\gamma_t(\mathrm{d}x_t)\,Q_t(x_t, \mathrm{d}x_{t+1}) \\
&\quad= \iint \varphi(x_{t:t+1})\gamma_t[q_t(\cdot, x_{t+1})]\,\lambda_{t+1}(\mathrm{d}x_{t+1})\,\frac{\gamma_t(\mathrm{d}x_t)q_t(x_t, x_{t+1})}{\gamma_t[q_t(\cdot, x_{t+1})]} \\
&\quad= \iint \varphi(x_{t:t+1})\,\gamma_{t+1}(\mathrm{d}x_{t+1})\,\overleftarrow{Q}_{n,\eta_t}(x_{t+1}, \mathrm{d}x_t). \qquad \text{(A.53)}
\end{aligned}$$

现在，由 (A.34) 可得

$$B_{t+1}h_{t+1}(x_{t+1}) = \int \overleftarrow{Q}_{n,\eta_t}(x_{t+1}, \mathrm{d}x_t)\left(\tilde{h}_t(x_{t:t+1}) + \int h_t(x_{0:t})\,B_t(x_t, \mathrm{d}x_{0:t-1})\right);$$

因此，应用 (A.53)，并取

$$\varphi(x_{t:t+1}) \coloneqq f_{t+1}(x_{t+1})\left(\tilde{h}_t(x_{t:t+1}) + \int h_t(x_{0:t})\,B_t(x_t, \mathrm{d}x_{0:t-1})\right)$$

我们得到

$$\begin{aligned}
\gamma_{t+1}(f_{t+1}B_{t+1}h_{t+1}) &= \iint \varphi(x_{t:t+1})\,\gamma_{t+1}(\mathrm{d}x_{t+1})\,\overleftarrow{Q}_{n,\eta_t}(x_{t+1}, \mathrm{d}x_t) \\
&= \iint \varphi(x_{t:t+1})\,\gamma_t(\mathrm{d}x_t)\,Q_t(x_t, \mathrm{d}x_{t+1}) \\
&= \gamma_t(Q_t f_{t+1}B_t h_t + Q_t\tilde{h}_t f_{t+1}).
\end{aligned}$$

现在注意到，由于 $\gamma_{t+1} = \gamma_t Q_t$，$\gamma_{t+1}f_{t+1} = \gamma_t Q_t f_{t+1}$，证明得证。 □

引理 3. 对于每个 $t \in \mathbb{N}^*$, $h_t \in \mathsf{F}(\mathcal{Y}_t)$，以及 $\boldsymbol{\eta}_0 \in \mathsf{M}_1(\boldsymbol{\mathcal{X}}_0)$，都有

$$\mathbb{E}^{\boldsymbol{P}}_{\boldsymbol{\eta}_0}[h_t(\boldsymbol{v}_t) \mid \boldsymbol{\xi}_{0|0}, \ldots, \boldsymbol{\xi}_{t|t}] = \mathbb{S}_t h_t(\boldsymbol{\xi}_{0|0}, \ldots, \boldsymbol{\xi}_{t|t}), \quad \mathbb{P}^{\boldsymbol{P}}_{\boldsymbol{\eta}_0}\text{-}a.s.$$

证明。任取 $v_t \in \mathsf{F}(\mathcal{X}_{0:t})$。我们证明

$$\mathbb{E}^{\boldsymbol{P}}_{\boldsymbol{\eta}_0}[v_t(\boldsymbol{\xi}_{0|0}, \ldots, \boldsymbol{\xi}_{t|t})h_t(\boldsymbol{v}_t)] = \mathbb{E}^{\boldsymbol{P}}_{\boldsymbol{\eta}_0}[v_t(\boldsymbol{\xi}_{0|0}, \ldots, \boldsymbol{\xi}_{t|t})\mathbb{S}_t h_t(\boldsymbol{\xi}_{0|0}, \ldots, \boldsymbol{\xi}_{t|t})], \qquad \text{(A.54)}$$

由此得 该结论随之得证。根据定义 (A.52)，前一个恒等式的左侧

可改写为

$$\begin{aligned}
&\int\cdots\int \psi_0[\boldsymbol{\eta}_0](\mathrm{d}\boldsymbol{y}_0)\prod_{m=0}^{t-1}\boldsymbol{P}_m(\boldsymbol{y}_m,\mathrm{d}\boldsymbol{y}_{m+1})\,h_t(\boldsymbol{y}_t)v_t(\boldsymbol{x}_{0|0},\ldots,\boldsymbol{x}_{t|t})\\
&\quad=\int\cdots\int \boldsymbol{\eta}_0(\mathrm{d}\boldsymbol{x}_{0|0})\prod_{m=0}^{t-1}\boldsymbol{M}_m(\boldsymbol{x}_{m|m},\mathrm{d}\boldsymbol{x}_{m+1})\,\boldsymbol{S}_0(\mathbf{J}\boldsymbol{x}_{0|0},\boldsymbol{x}_1,\mathrm{d}\boldsymbol{y}_1)\\
&\qquad\times\prod_{m=0}^{t-1}\boldsymbol{S}_m(\boldsymbol{y}_m,\boldsymbol{x}_{m+1},\mathrm{d}\boldsymbol{y}_{m+1})\,h_t(\boldsymbol{y}_t)v_t(\boldsymbol{x}_{0|0},\ldots,\boldsymbol{x}_{t|t})\\
&\quad=\int\cdots\int \boldsymbol{\eta}_0(\mathrm{d}\boldsymbol{x}_0)\prod_{m=0}^{t-1}\boldsymbol{M}_m(\boldsymbol{x}_m,\mathrm{d}\boldsymbol{x}_{m+1})\,\boldsymbol{S}_0(\mathbf{J}\boldsymbol{x}_0,\boldsymbol{x}_1,\mathrm{d}\boldsymbol{y}_1)\\
&\qquad\times\prod_{m=0}^{t-1}\boldsymbol{S}_m(\boldsymbol{y}_m,\boldsymbol{x}_{m+1},\mathrm{d}\boldsymbol{y}_{m+1})\,h_t(\boldsymbol{y}_t)v_t(\boldsymbol{x}_0,\ldots,\boldsymbol{x}_t).
\end{aligned}$$

因此，我们可以 结合 $\mathbb{S}_t$ 的定义 (A.40) 和富比尼定理，得出证明结论 .

□

引理4。 对于每个 $t\in\mathbb{N}^*$ 和 $h_t\in\mathsf{F}(\mathcal{Y}_t)$,

$$\mathbb{E}_{\boldsymbol{\eta}_0}\left[\left(\prod_{m=0}^{t-1}\boldsymbol{g}_m(\boldsymbol{\xi}_{m|m})\right)h_t(\boldsymbol{v}_t)\right]=\int\boldsymbol{\gamma}_{0:t}\mathbb{S}_t(\mathrm{d}\boldsymbol{y}_t)\,h_t(\boldsymbol{y}_t).$$

证明。 该引理的结论可由引理 3 直接推出；事实上，应用塔式性质和后者可得

$$\begin{aligned}
&\mathbb{E}^{\boldsymbol{P}}_{\boldsymbol{\eta}_0}\left[\left(\prod_{m=0}^{t-1}\boldsymbol{g}_m(\boldsymbol{\xi}_{m|m})\right)h_t(\boldsymbol{v}_t)\right]\\
&\qquad=\mathbb{E}^{\boldsymbol{P}}_{\boldsymbol{\eta}_0}\left[\left(\prod_{m=0}^{t-1}\boldsymbol{g}_m(\boldsymbol{\xi}_{m|m})\right)\mathbb{S}_t h_t(\boldsymbol{\xi}_{0|0},\ldots,\boldsymbol{\xi}_{t|t})\right]\\
&\qquad=\int\cdots\int\boldsymbol{\eta}_0(\mathrm{d}\boldsymbol{x}_0)\prod_{m=0}^{t-1}\boldsymbol{g}_m(\boldsymbol{x}_m)\,\boldsymbol{M}_m(\boldsymbol{x}_m,\mathrm{d}\boldsymbol{x}_{m+1})\,\mathbb{S}_t h_t(\boldsymbol{x}_{0:t})\\
&\qquad=\int\boldsymbol{\gamma}_{0:t}\mathbb{S}_t(\mathrm{d}\boldsymbol{y}_t)\,h_t(\boldsymbol{y}_t).
\end{aligned}$$

□

命题 7。 对于所有 $t\in\mathbb{N}^*,(N,M)\in(\mathbb{N}^*)^2$, 以及 $(f_t,\tilde{f}_t)\in\mathsf{F}(\mathcal{X}_t)^2$,

$$\int\boldsymbol{\gamma}_{0:t}\mathbb{S}_t(\mathrm{d}\boldsymbol{y}_t)\left(\frac{1}{N}\sum_{i=1}^{N}\{b_t^i f_t(x_{t|t}^i)+\tilde{f}_t(x_{t|t}^i)\}\right)=\gamma_t(f_t B_t h_t+\tilde{f}_t).$$

证明。 应用引理4可得

$$\int\boldsymbol{\gamma}_{0:t}\mathbb{S}_t(\mathrm{d}\boldsymbol{y}_t)\left(\frac{1}{N}\sum_{i=1}^{N}\{b_t^i f_t(x_{t|t}^i)+\tilde{f}_t(x_{t|t}^i)\}\right)=\mathbb{E}^{\boldsymbol{P}}_{\boldsymbol{\eta}_0}\left[\left(\prod_{m=0}^{t-1}\boldsymbol{g}_m(\boldsymbol{\xi}_{m|m})\right)\frac{1}{N}\sum_{i=1}^{N}\{\beta_t^i f_t(\xi_{t|t}^i)+\tilde{f}_t(\xi_{t|t}^i)\}\right]. \tag{A.55}$$

下文将反复使用以下滤过。令 $\tilde{\mathcal{F}}_t \coloneqq \sigma(\{\boldsymbol{\upsilon}_m\}_{m=0}^t)$ 为由 PARIS（算法1）在前 t 次迭代中的输出生成的 σ-域。此外，令 $\tilde{\mathcal{F}}_t \coloneqq \mathcal{F}_{t-1} \vee \sigma(\boldsymbol{\xi}_{t|t})$。

我们通过归纳法进行证明。因此，假设该命题对于给定的 $t \in \mathbb{N}^*$ 成立，并考虑任意选取的 $(f_{t+1},\ \tilde{f}_{t+1}) \in \mathsf{F}(\mathcal{X}_{t+1})^2$，

$$\mathbb{E}_{\boldsymbol{\eta}_0}^{\boldsymbol{P}}\left[\left(\prod_{m=0}^{t} \boldsymbol{g}_m(\boldsymbol{\xi}_{m|m})\right) \frac{1}{N}\sum_{i=1}^{N}\{\beta_{t+1}^i f_{t+1}(\xi_{t+1|t+1}^i) + \tilde{f}_{t+1}(\xi_{t+1|t+1}^i)\} \mid \tilde{\mathcal{F}}_t\right] = \left(\prod_{m=0}^{t} \boldsymbol{g}_m(\boldsymbol{\xi}_{m|m})\right) \mathbb{E}_{\boldsymbol{\eta}_0}^{\boldsymbol{P}}[\beta_{t+1}^1 f_{t+1}(\xi_{t+1|t+1}^1) + \tilde{f}_{t+1}(\xi_{t+1|t+1}^1) \mid \tilde{\mathcal{F}}_t],$$

这里用到了变量 $\{\beta_{t+1}^i f_{t+1}(\xi_{t+1|t+1}^i) + \tilde{f}_{t+1}(\xi_{t+1|t+1}^i)\}_{i=1}^N$ 在给定$\tilde{\mathcal{F}}_t$的条件下条件独立同分布这一点。注意，由对称性，

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\eta}_0}^{\boldsymbol{P}}\left[\beta_{t+1}^1 \mid \mathcal{F}_{t+1}\right] &= \int \boldsymbol{S}_t(\boldsymbol{\upsilon}_t, \boldsymbol{\xi}_{t+1|t+1}, \mathrm{d}\boldsymbol{y}_{t+1})\, b_{t+1}^1 \\
&= \int \cdots \int \left(\prod_{j=1}^{M}\sum_{\ell=1}^{N} \frac{q_t(\xi_{t|t}^\ell, \xi_{t+1|t+1}^1)}{\sum_{\ell'=1}^N q_t(\xi_{t|t}^{\ell'}, \xi_{t+1|t+1}^1)} \delta_{(\xi_{t|t}^\ell, \beta_t^\ell)}(\mathrm{d}\tilde{x}_t^{1,j}, \mathrm{d}\tilde{b}_t^{1,j})\right) \\
&\qquad \times \frac{1}{M}\sum_{j=1}^{M}\left(\tilde{b}_t^{1,j} + \tilde{h}_t(\tilde{x}_t^{1,j}, \xi_{t+1|t+1}^1)\right) \\
&= \sum_{\ell=1}^{N} \frac{q_t(\xi_{t|t}^\ell, \xi_{t+1|t+1}^1)}{\sum_{\ell'=1}^N q_t(\xi_{t|t}^{\ell'}, \xi_{t+1|t+1}^1)} \left(\beta_t^\ell + \tilde{h}_t(\xi_{t|t}^\ell, \xi_{t+1|t+1}^1)\right). \qquad \text{(A.56)}
\end{aligned}$$

因此，利用塔式性质，

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\eta}_0}^{\boldsymbol{P}}\left[\beta_{t+1}^1 f_{t+1}(\xi_{t+1|t+1}^1) \mid \tilde{\mathcal{F}}_t\right] \\
&\qquad = \int \Phi_t(\mu(\boldsymbol{\xi}_{t|t}))(\mathrm{d}x_{t+1})\, f_{t+1}(x_{t+1}) \sum_{\ell=1}^{N} \frac{q_t(\xi_{t|t}^\ell, x_{t+1})}{\sum_{\ell'=1}^N q_t(\xi_{t|t}^{\ell'}, x_{t+1})} \left(\beta_t^\ell + \tilde{h}_t(\xi_{t|t}^\ell, x_{t+1})\right),
\end{aligned}$$

因此，根据定义（A.25），

$$\begin{aligned}
&\left(\prod_{m=0}^{t} \boldsymbol{g}_m(\boldsymbol{\xi}_{m|m})\right) \mathbb{E}_{\boldsymbol{\eta}_0}^{\boldsymbol{P}}\left[\beta_{t+1}^1 f_{t+1}(\xi_{t+1|t+1}^1) \mid \tilde{\mathcal{F}}_t\right] \\
&= \left(\prod_{m=0}^{t-1} \boldsymbol{g}_m(\boldsymbol{\xi}_{m|m})\right) \int \frac{1}{N}\sum_{i=1}^{N} q_t(\xi_{t|t}^i, x_{t+1}) \\
&\qquad \times f_{t+1}(x_{t+1}) \sum_{\ell=1}^{N} \frac{q_t(\xi_{t|t}^\ell, x_{t+1})}{\sum_{\ell'=1}^N q_t(\xi_{t|t}^{\ell'}, x_{t+1})} \left(\beta_t^\ell + \tilde{h}_t(\xi_{t|t}^\ell, x_{t+1})\right) \lambda_{t+1}(\mathrm{d}x_{t+1}) \\
&= \left(\prod_{m=0}^{t-1} \boldsymbol{g}_m(\boldsymbol{\xi}_{m|m})\right) \frac{1}{N}\sum_{\ell=1}^{N}\left(\beta_t^\ell Q_t f_{t+1}(\xi_{t|t}^\ell) + Q_t(\tilde{h}_t f_{t+1})(\xi_{t|t}^\ell)\right).
\end{aligned}$$

因此，应用归纳假设，

$$\begin{aligned}
&\mathbb{E}^{\boldsymbol{P}}_{\boldsymbol{\eta}_0}\left[\left(\prod_{m=0}^{t}\boldsymbol{g}_m(\boldsymbol{\xi}_{m|m})\right)\frac{1}{N}\sum_{i=1}^{N}\beta^i_{t+1}f_{t+1}(\xi^i_{t+1|t+1})\right]\\
&=\mathbb{E}^{\boldsymbol{P}}_{\boldsymbol{\eta}_0}\left[\left(\prod_{m=0}^{t-1}\boldsymbol{g}_m(\boldsymbol{\xi}_{m|m})\right)\frac{1}{N}\sum_{\ell=1}^{N}\left(\beta^\ell_t Q_t f_{t+1}(\xi^\ell_{t|t})+Q_t(\tilde{h}_t f_{t+1})(\xi^\ell_{t|t})\right)\right]\\
&=\gamma_t\left(Q_t f_{t+1}B_t h_t+Q_t(\tilde{h}_t f_{t+1})\right).
\end{aligned}\tag{A.57}$$

同样可以证明

$$\mathbb{E}^{\boldsymbol{P}}_{\boldsymbol{\eta}_0}\left[\left(\prod_{m=0}^{t}\boldsymbol{g}_m(\boldsymbol{\xi}_{m|m})\right)\frac{1}{N}\sum_{i=1}^{N}\tilde{f}_{t+1}(\xi^i_{t+1|t+1})\right]=\gamma_t Q_t\tilde{f}_{t+1}.\tag{A.58}$$

现在，由 (A.57–A.58) 和引理 2,

$$\begin{aligned}
&\mathbb{E}^{\boldsymbol{P}}_{\boldsymbol{\eta}_0}\left[\left(\prod_{m=0}^{t}\boldsymbol{g}_m(\boldsymbol{\xi}_{m|m})\right)\frac{1}{N}\sum_{i=1}^{N}\{\beta^i_{t+1}f_{t+1}(\xi^i_{t+1|t+1})+\tilde{f}_{t+1}(\xi^i_{t+1|t+1})\}\right]\\
&=\gamma_t\left(Q_t f_{t+1}B_t h_t+Q_t(\tilde{h}_t f_{t+1}+Q_t\tilde{f}_{t+1})\right)\\
&=\gamma_{t+1}(f_{t+1}B_{t+1}h_{t+1}+\tilde{f}_{t+1}),
\end{aligned}$$

这表明该命题的结论在时间 $n+1$ 成立。

剩下只需检验基例 $n=0$。由于按照约定 $\boldsymbol{\beta}_0=\mathbf{0}$, $B_0h_0=0$，且初始粒子 $\boldsymbol{\xi}_{0|0}$ 取自 $\boldsymbol{\eta}_0$，因此基例显然成立。证明至此完成。 □

定理 5 的证明。恒等式 $\int\boldsymbol{\eta}_{0:t}(\mathrm{d}\boldsymbol{x}_{0:t})\,\mathbb{S}_t(\boldsymbol{x}_{0:t},\mathrm{d}\boldsymbol{b}_t)\,\mu(\boldsymbol{b}_t)(\mathrm{id})=\eta_{0:t}h_t$ 可通过在命题 7 中令 $f_t\equiv1$ 和 $f_t\equiv0$，并利用 $\boldsymbol{\gamma}_{0:t}(\mathbf{X}_{0:t})=\gamma_{0:t}(\mathrm{X}_{0:t})$，立即得到。此外，应用定理 3 可得

$$\begin{aligned}
\int\eta_{0:t}\mathbb{C}_t\mathbb{S}_t(\mathrm{d}\boldsymbol{b}_t)\,\mu(\boldsymbol{b}_t)(\mathrm{id})&=\iint\eta_{0:t}(\mathrm{d}z_{0:t})\,\mathbb{C}_t(z_{0:t},\mathrm{d}\boldsymbol{x}_{0:t})\int\mathbb{S}_t(\boldsymbol{x}_{0:t},\mathrm{d}\boldsymbol{b}_t)\,\mu(\boldsymbol{b}_t)(\mathrm{id})\\
&=\iint\boldsymbol{\eta}_{0:t}(\mathrm{d}\boldsymbol{x}_{0:t})\,\mathbb{B}_t(\boldsymbol{x}_{0:t},\mathrm{d}z_{0:t})\int\mathbb{S}_t(\boldsymbol{x}_{0:t},\mathrm{d}\boldsymbol{b}_t)\,\mu(\boldsymbol{b}_t)(\mathrm{id})\\
&=\int\boldsymbol{\eta}_{0:t}\mathbb{S}_t(\mathrm{d}\boldsymbol{b}_t)\,\mu(\boldsymbol{b}_t)(\mathrm{id}).
\end{aligned}$$

最后，第一个恒等式成立，因为 K_t 使 $\eta_{0:t}$ 保持不变。 □

A.6.4 命题 2 的证明

首先注意，根据定义 (A.39) 和 (A.40),

$$\begin{aligned}
H_t(\boldsymbol{x}_{0:t})&\coloneqq\int\mathbb{S}_t(\boldsymbol{x}_{0:t},\mathrm{d}\boldsymbol{y}_t)\,\mu(\boldsymbol{x}_{\lceil0:n|n\rceil})h\\
&=\int\cdots\int\left(\frac{1}{N}\sum_{j_t=1}^{N}h(x^{j_t}_{0:t-1|t},x^{j_t}_t)\right)\\
&\quad\times\prod_{m=0}^{t-1}\prod_{i_{m+1}=1}^{N}\int\sum_{j_m=1}^{N}\frac{q_m(x^{j_m}_m,x^{i_{m+1}}_{m+1})}{\sum_{j'_m=1}^{N}q_m(x^{j'_m}_m,x^{i_{m+1}}_{m+1})}\delta_{x^{j_m}_{0:m|m}}(\mathrm{d}x^{i_{m+1}}_{0:m|m+1}),
\end{aligned}$$

其中，对于所有 $i \in [\![1, N]\!]$，按约定有 $x_{0:-1|0}^{i} = \emptyset$。我们将证明，对于每个 $k \in [\![0, t]\!]$，都有 $H_{k,t} \equiv H_t$，其中

$$H_{k,n}(\boldsymbol{x}_{0:t}) \coloneqq \frac{1}{N} \sum_{j_t=1}^{N} \cdots \sum_{j_k=1}^{N} \prod_{\ell=k}^{t-1} \frac{q_\ell(x_\ell^{j_\ell}, x_{\ell+1}^{j_{\ell+1}})}{\sum_{j'_\ell=1}^{N} q_\ell(x_\ell^{j'_\ell}, x_{\ell+1}^{j_{\ell+1}})} a_{k,n}(\boldsymbol{x}_0, \dots, \boldsymbol{x}_{k-1}, x_k^{j_k}, \dots, x_t^{j_t})$$

with

$$\begin{aligned} &a_{k,n}(\boldsymbol{x}_0, \dots, \boldsymbol{x}_{k-1}, x_k^{j_k}, \dots, x_t^{j_t}) \\ &\quad = \int \prod_{m=0}^{k-1} \prod_{i_{m+1}=1}^{N} \sum_{j_m=1}^{N} \frac{q_m(x_m^{j_m}, x_{m+1}^{i_{m+1}})}{\sum_{j'_m=1}^{N} q_m(x_m^{j'_m}, x_{m+1}^{i_{m+1}})} \delta_{x_{0:m|m}^{j_m}}(\mathrm{d}x_{0:m|m+1}^{i_{m+1}}) h(x_{0:k-1|k}^{j_k}, x_k^{j_k}, \dots, x_t^{j_t}). \end{aligned}$$

由于按照约定，$\prod_{\ell=n}^{t-1} \dots = 1$, $H_{n,n}(\boldsymbol{x}_{0:t}) = N^{-1} \sum_{j_t=1}^{N} a_{n,n}(\boldsymbol{x}_0, \dots, \boldsymbol{x}_{[}n-1], x_t^{j_t})$，并且注意到 $H_t \equiv H_{n,n_\circ}$ 现在我们证明，对每个 $k \in [\![1, t]\!]$，都有 $H_{k,n} \equiv H_{k-1,n}$；为此，注意到

$$\begin{aligned} &a_{k,n}(\boldsymbol{x}_0, \dots, \boldsymbol{x}_{k-1}, x_k^{j_k}, \dots, x_t^{j_t}) \\ &\qquad = \int \prod_{m=0}^{k-2} \prod_{i_{m+1}=1}^{N} \sum_{j_m=1}^{N} \frac{q_m(x_m^{j_m}, x_{m+1}^{i_{m+1}})}{\sum_{j'_m=1}^{N} q_m(x_m^{j'_m}, x_{m+1}^{i_{m+1}})} \delta_{x_{0:m|m}^{j_m}}(\mathrm{d}x_{0:m|m+1}^{i_{m+1}}) \\ &\quad \times \int \prod_{i_k=1}^{N} \sum_{j_{k-1}=1}^{N} \frac{q_{k-1}(x_{k-1}^{j_{k-1}}, x_k^{i_k})}{\sum_{j'_{k-1}=1}^{N} q_{k-1}(x_{k-1}^{j'_{k-1}}, x_k^{i_k})} \delta_{x_{0:k-1|k-1}^{j_{k-1}}}(\mathrm{d}x_{0:k-1|k}^{i_k})\, h(x_{0:k-1|k}^{j_k}, x_k^{j_k}, \dots, x_t^{j_t}), \end{aligned}$$

并且由于 $x_0^{j_{k-1}}{}_{:k-1|k-1} = (x_0^{j_{k-1}}{}_{:k-2|k-1}, x_{k-1}^{j_{k-1}})$，因此成立：

$$\begin{aligned} &\int \prod_{i_k=1}^{N} \sum_{j_{k-1}=1}^{N} \frac{q_{k-1}(x_{k-1}^{j_{k-1}}, x_k^{i_k})}{\sum_{j'_{k-1}=1}^{N} q_{k-1}(x_{k-1}^{j'_{k-1}}, x_k^{i_k})} \delta_{x_{0:k-1|k-1}^{j_{k-1}}}(\mathrm{d}x_{0:k-1|k}^{i_k})\, h(x_{0:k-1|k}^{j_k}, x_k^{j_k}, \dots, x_t^{j_t}) \\ &\qquad = \sum_{j_{k-1}=1}^{N} \frac{q_{k-1}(x_{k-1}^{j_{k-1}}, x_k^{j_k})}{\sum_{j'_{k-1}=1}^{N} q_{k-1}(x_{k-1}^{j'_{k-1}}, x_k^{j_k})} h(x_{0:k-2|k-1}^{j_{k-1}}, x_{k-1}^{j_{k-1}}, x_k^{j_k}, \dots, x_t^{j_t}). \end{aligned}$$

因此，我们得到

$$\begin{aligned} &a_{k,n}(\boldsymbol{x}_0, \dots, \boldsymbol{x}_{k-1}, x_k^{j_k}, \dots, x_t^{j_t}) \\ &\qquad = \int \prod_{m=0}^{k-2} \prod_{i_{m+1}=1}^{N} \sum_{j_m=1}^{N} \frac{q_m(x_m^{j_m}, x_{m+1}^{i_{m+1}})}{\sum_{j'_m=1}^{N} q_m(x_m^{j'_m}, x_{m+1}^{i_{m+1}})} \delta_{x_{0:m|m}^{j_m}}(\mathrm{d}x_{0:m|m+1}^{i_{m+1}}) \\ &\qquad\quad \times \sum_{j_{k-1}=1}^{N} \frac{q_{k-1}(x_{k-1}^{j_{k-1}}, x_k^{j_k})}{\sum_{j'_{k-1}=1}^{N} q_{k-1}(x_{k-1}^{j'_{k-1}}, x_k^{j_k})} h(x_{0:k-2|k-1}^{j_{k-1}}, x_{k-1}^{j_{k-1}}, x_k^{j_k}, \dots, x_t^{j_t}). \end{aligned}$$

现在，交换前一个公式右侧关于 j_{k-1} 的求和与积分的次序，可得

$$\begin{aligned} &a_{k,n}(\boldsymbol{x}_0, \dots, \boldsymbol{x}_{k-1}, x_k^{j_k}, \dots, x_t^{j_t}) \\ &\qquad = \sum_{j_{k-1}=1}^{N} \frac{q_{k-1}(x_{k-1}^{j_{k-1}}, x_k^{j_k})}{\sum_{j'_{k-1}=1}^{N} q_{k-1}(x_{k-1}^{j'_{k-1}}, x_k^{j_k})} a_{k-1,n}(\boldsymbol{x}_0, \dots, \boldsymbol{x}_{k-2}, x_{k-1}^{j_{k-1}}, \dots, x_t^{j_t}). \end{aligned}$$

因此，

$$
\begin{aligned}
&H_{k,n}(\boldsymbol{x}_{0:t}) \\
&= \frac{1}{N}\sum_{j_t=1}^{N}\cdots\sum_{j_k=1}^{N}\prod_{\ell=k}^{t-1}\frac{q_\ell(x_\ell^{j_\ell}, x_{\ell+1}^{j_{\ell+1}})}{\sum_{j'_\ell=1}^{N} q_\ell(x_\ell^{j'_\ell}, x_{\ell+1}^{j_{\ell+1}})} \\
&\quad\times\sum_{j_{k-1}=1}^{N}\frac{q_{k-1}(x_{k-1}^{j_{k-1}}, x_k^{j_k})}{\sum_{j'_{k-1}=1}^{N} q_{k-1}(x_{k-1}^{j'_{k-1}}, x_k^{j_k})}a_{k-1,n}(\boldsymbol{x}_0,\ldots,\boldsymbol{x}_{k-2}, x_{k-1}^{j_{k-1}},\ldots,x_t^{j_t}) \\
&= \frac{1}{N}\sum_{j_t=1}^{N}\cdots\sum_{j_{k-1}=1}^{N}\prod_{\ell=k-1}^{t-1}\frac{q_\ell(x_\ell^{j_\ell}, x_{\ell+1}^{j_{\ell+1}})}{\sum_{j'_\ell=1}^{N} q_\ell(x_\ell^{j'_\ell}, x_{\ell+1}^{j_{\ell+1}})}a_{k-1,n}(\boldsymbol{x}_0,\ldots,\boldsymbol{x}_{k-2}, x_{k-1}^{j_{k-1}},\ldots,x_t^{j_t}) \\
&= H_{k-1,n}(\boldsymbol{x}_{0:t}),
\end{aligned}
$$

这就得到了递推。因此，$H_t \equiv H_{0,\ n}$ 现在我们只需注意到 $\mathbb{B}_t h \equiv H_{0,\ n}$，即可完成证明。

A.6.5 定理 7 的证明

为了证明定理 7，我们将证明以下更一般的结果，而定理 7 可由此直接推出。

命题 8。对于每个 $t \in \mathbb{N}$ 和 $M \in \mathbb{N}^*$，都存在 $\mathsf{c}_t > 0$ 和 $\mathsf{d}_t > 0$，使得对于每个 $N \in \mathbb{N}^*$，$z_{0:t} \in \mathsf{X}_{0:t}$，$(f_t,\ \tilde{f}_t) \in \mathsf{F}(\mathcal{X}_t)^2$，和 $\varepsilon > 0$，

$$
\begin{aligned}
\int \mathbb{C}_t\mathbb{S}_t(z_{0:t}, \mathrm{d}\boldsymbol{b}_t)\mathbb{1}\left\{\left|\frac{1}{N}\sum_{i=1}^{N}\{b_t^i f_t(x_{t|t}^i) + \tilde{f}_t(x_{t|t}^i)\} - \eta_t\langle z_{0:t}\rangle(f_t B_t\langle z_{0:t-1}\rangle h_t + \tilde{f}_t)\right| \geq \varepsilon\right\} \\
\leq \mathsf{c}_t \exp\left(-\frac{\mathsf{d}_t N\varepsilon^2}{2\kappa_t^2}\right),
\end{aligned}
$$

其中

$$
\kappa_t \coloneqq \|f_t\|_\infty \sum_{m=0}^{t-1}\|\tilde{h}_m\|_\infty + \|\tilde{f}_t\|_\infty. \tag{A.59}
$$

为了证明命题 8，我们需要以下技术引理。

引理 5。对于每个 $t \in \mathbb{N}$，$(f_{t+1}, \tilde{f}_{t+1}) \in \mathsf{F}(\mathcal{X}_{t+1})^2$，$z_{0:t+1} \in \mathsf{X}_{0:t+1}$，以及 $N \in \mathbb{N}^*$，

$$
\begin{aligned}
&\gamma_{t+1}\langle z_{0:t+1}\rangle(f_{t+1}B_{t+1}\langle z_{0:t}\rangle h_{t+1} + \tilde{f}_{t+1}) \\
&\quad= \left(1 - \frac{1}{N}\right)\gamma_t\langle z_{0:t}\rangle\{Q_t f_{t+1}B_t\langle z_{0:t-1}\rangle h_t + Q_t(\tilde{h}_t f_{t+1} + \tilde{f}_{t+1})\} \\
&\qquad\quad+ \frac{1}{N}\gamma_t\langle z_{0:t}\rangle g_t\left(f_{t+1}(z_{t+1})B_{t+1}\langle z_{0:t}\rangle h_{t+1}(z_{t+1}) + \tilde{f}_{t+1}(z_{t+1})\right).
\end{aligned}
$$

证明。由于引理 2 对具有冻结路径的 Feynman–Kac 模型同样成立，我们得到

$$
\gamma_{t+1}\langle z_{0:t+1}\rangle(f_{t+1}B_{t+1}\langle z_{0:t}\rangle h_{t+1} + \tilde{f}_{t+1}) = \gamma_t\langle z_{0:t}\rangle\{Q_t\langle z_{t+1}\rangle f_{t+1}B_t\langle z_{0:t}\rangle h_t + Q_t\langle z_{t+1}\rangle(\tilde{h}_t f_{t+1} + \tilde{f}_{t+1})\}.
$$

因此，只需注意对于每个 $x_t \in \mathsf{X}_t$ 和 $h \in \mathsf{F}(\mathcal{X}_{t:t+1})$，证明便告完成。

$$
Q_t\langle z_{t+1}\rangle h(x_t) = \left(1 - \frac{1}{N}\right)Q_t h(x_t) + \frac{1}{N}g(x_t)h(x_t, z_{t+1}).
$$

□

最后，在开始证明命题 8 之前，我们先介绍 PARIS 在给定冻结路径 $z=\{z_m\}_{m\in\mathbb{N}}$ 条件下演化的规律。定义，对于 $m\in\mathbb{N}$ 和 $z_{m+1}\in\mathsf{X}^{m+1}$,

$$\boldsymbol{P}_m\langle z_{m+1}\rangle : \mathbf{Y}_m\times\boldsymbol{\mathcal{Y}}_{m+1}\ni(\boldsymbol{y}_m,A)\mapsto\int \boldsymbol{M}_m\langle z_{m+1}\rangle(\boldsymbol{x}_{m|m},\mathrm{d}\boldsymbol{x}_{m+1})\,\boldsymbol{S}_m(\boldsymbol{y}_m,\boldsymbol{x}_{m+1},A).$$

对于任意给定的初始分布 $\boldsymbol{\psi}_0\in\mathsf{M}_1(\boldsymbol{\mathcal{Y}}_0)$，令 $\mathbb{P}^{\boldsymbol{P},z}_{\boldsymbol{\psi}_0}$ 表示由马尔可夫核 $\{\boldsymbol{P}_m\langle z_{m+1}\rangle\}_{m\in\mathbb{N}}$ 和初始分布 $\boldsymbol{\psi}_0$所诱导的典范马尔可夫链的分布。为简化记法，我们用 $\mathbb{P}^{\boldsymbol{P},z}_{\boldsymbol{\eta}_0}$代替 $\mathbb{P}^{\boldsymbol{P},z}_{\boldsymbol{\psi}_0[\boldsymbol{\eta}_0\langle z_0\rangle]}$，其中扩展 $\boldsymbol{\psi}_0[\boldsymbol{\eta}_0]$定义见附录 A.6.3。

˜命题 *8* 的证明。 我们对 t进行正向归纳。令 σ-域 $\mathcal{F}_t$ 和 $\tilde{\mathcal{F}}_t$ 按定理 5 的证明中的方式定义，但针对条件 PARIS 对偶过程。于是，在分布 $\mathbb{P}^{\boldsymbol{P},z}_{\boldsymbol{\eta}_0}$下，沿用 (A.56),

$$\begin{aligned}
&\mathbb{E}^{\boldsymbol{P},z}_{\boldsymbol{\eta}_0}\Big[\ \beta_t^1 f_t(\xi_t^1)+\tilde{f}_t(\xi_t^1)\mid\tilde{\mathcal{F}}_{t-1}\Big]\\
&=\mathbb{E}^{\boldsymbol{P},z}_{\boldsymbol{\eta}_0}\Big[\mathbb{E}^{\boldsymbol{P},z}_{\boldsymbol{\eta}_0}\big[\ \beta_t^1\mid\mathcal{F}_t\big]\ \ f_t(\xi_t^1)+\tilde{f}_t(\xi_t^1)\mid\tilde{\mathcal{F}}_{t-1}\Big]\\
&=\mathbb{E}^{\boldsymbol{P},z}_{\boldsymbol{\eta}_0}\left[\ f_t(\xi_t^1)\sum_{\ell=1}^N\frac{q_{t-1}(\xi_{t-1}^\ell,\xi_t^1)}{\sum_{\ell'=1}^N q_{t-1}(\xi_{t-1}^{\ell'},\xi_t^1)}\Big(\beta_{t-1}^\ell+\tilde{h}_{t-1}(\xi_{t-1}^\ell,\xi_t^1)\Big)\ +\tilde{f}_t(\xi_t^1)\mid\tilde{\mathcal{F}}_{t-1}\right].
\end{aligned}$$

利用 (A.30)，可得

$$\begin{aligned}
&\mathbb{E}^{\boldsymbol{P},z}_{\boldsymbol{\eta}_0}\Big[\beta_t^1 f_t(\xi_t^1)+\tilde{f}_t(\xi_t^1)\mid\tilde{\mathcal{F}}_{t-1}\Big]\\
&\qquad=\left(1-\frac{1}{N}\right)\frac{\sum_{\ell=1}^N\{\beta_{t-1}^\ell Q_{t-1}f_t(\xi_{t-1}^\ell)+Q_{t-1}(\tilde{h}_{t-1}f_t+\tilde{f}_t)(\xi_{t-1}^\ell)\}}{\sum_{\ell'=1}^N g_{t-1}(\xi_{t-1}^{\ell'})}\\
&\qquad\quad+\frac{1}{N}\left(f_t(z_t)\sum_{\ell=1}^N\frac{q_{t-1}(\xi_{t-1}^\ell,z_t)}{\sum_{\ell'=1}^N q_{t-1}(\xi_{t-1}^{\ell'},z_t)}\Big(\beta_{t-1}^\ell+\tilde{h}_t(\xi_{t-1}^\ell,z_t)\Big)+\tilde{f}_t(z_t)\right).
\end{aligned}\tag{A.60}$$

为将归纳假设应用于前一个恒等式右端的每一项，注意到

$$B_t\langle z_{0:t-1}\rangle h_t(z_t)=\frac{\eta_{t-1}\langle z_{0:t-1}\rangle[q_{t-1}(\cdot,z_t)\{B_{t-1}\langle z_{0:t-2}\rangle h_{t-1}(\cdot)+\tilde{h}_{t-1}(\cdot,z_t)\}]}{\eta_{t-1}\langle z_{0:t-1}\rangle[q_{t-1}(\cdot,z_t)]}.$$

因此，利用引理 5，并注意到 $\gamma_t\langle z_{0:t}\rangle\mathbb{1}_{\mathsf{X}_t}/\gamma_{t-1}\langle z_{0:t}\rangle\mathbb{1}_{\mathsf{X}_{t-1}}=\eta_{t-1}\langle z_{0:t-1}\rangle g_{t-1}$ 可得

$$\begin{aligned}
\eta_t\langle z_{0:t}\rangle(f_tB_t\langle z_{0:t-1}\rangle h_t+\tilde{f}_t)&=\frac{1}{N}\Big(f_t(z_t)B_t\langle z_{0:t-1}\rangle h_t(z_t)+\tilde{f}_t(z_t)\Big)\\
&\quad+\left(1-\frac{1}{N}\right)\frac{\eta_{t-1}\langle z_{0:t-1}\rangle\{Q_{t-1}f_tB_{t-1}\langle z_{0:t-2}\rangle h_t+Q_{t-1}(\tilde{h}_{t-1}f_t+\tilde{f}_t)\}}{\eta_{t-1}\langle z_{0:t-1}\rangle g_{t-1}}.
\end{aligned}\tag{A.61}$$

结合 (A.60) 和 (A.61)，我们将误差分解为

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^N\{\beta_t^i f_t(\xi_{t|t}^i)+\tilde{f}_t(\xi_{t|t}^i)\}-\eta_t\langle z_{0:t}\rangle(f_tB_t\langle z_{0:t-1}\rangle h_t+\tilde{f}_t)\\
&\quad=\frac{1}{N}\sum_{i=1}^N\{\beta_t^i f_t(\xi_{t|t}^i)+\tilde{f}_t(\xi_{t|t}^i)\}-\mathbb{E}^{\boldsymbol{P},z}_{\boldsymbol{\eta}_0}\Big[\ \beta_t^1 f_t(\xi_t^1)+\tilde{f}_t(\xi_t^1)\mid\tilde{\mathcal{F}}_{t-1}\Big]\\
&\qquad+\mathbb{E}^{\boldsymbol{P},z}_{\boldsymbol{\eta}_0}\Big[\beta_t^1 f_t(\xi_t^1)+\tilde{f}_t(\xi_t^1)\mid\tilde{\mathcal{F}}_{t-1}\Big]-\eta_t\langle z_{0:t}\rangle(f_tB_t\langle z_{0:t-1}\rangle h_t+\tilde{f}_t)\\
&\quad=\mathrm{I}_N^{(1)}+\left(1-\frac{1}{N}\right)\mathrm{I}_N^{(2)}+\frac{1}{N}\mathrm{I}_N^{(3)},
\end{aligned}\tag{A.62}$$

其中

$$\mathrm{I}_N^{(1)} \coloneqq \frac{1}{N}\sum_{i=1}^{N}\{\beta_t^i f_t(\xi_t^i) + \tilde{f}_t(\xi_t^i)\} - \mathbb{E}_{\boldsymbol{\eta}_0}^{\boldsymbol{P},z}\left[\beta_t^1 f_t(\xi_t^1) + \tilde{f}_t(\xi_t^1) \mid \tilde{\mathcal{F}}_{t-1}\right],$$

$$\mathrm{I}_N^{(2)} \coloneqq \frac{\sum_{\ell=1}^{N}\{\beta_{t-1}^{\ell} Q_{t-1} f_t(\xi_{t-1}^{\ell}) + Q_{t-1}(\tilde{h}_{t-1} f_t + \tilde{f}_t)(\xi_{t-1}^{\ell})\}}{\sum_{\ell'=1}^{N} g_{t-1}(\xi_{t-1}^{\ell'})}$$

$$- \frac{\eta_{t-1}\langle z_{0:t-1}\rangle\{Q_{t-1} f_t B_t\langle z_{0:t-1}\rangle h_t + Q_{t-1}(\tilde{h}_{t-1} f_t + \tilde{f}_t)\}}{\eta_{t-1}\langle z_{0:t-1}\rangle g_{t-1}}, \tag{A.63}$$

and

$$\mathrm{I}_N^{(3)} \coloneqq f_t(z_t)\sum_{\ell=1}^{N}\frac{q_{t-1}(\xi_{t-1}^{\ell}, z_t)}{\sum_{\ell'=1}^{N} q_{t-1}(\xi_{t-1}^{\ell'}, z_t)}\left(\beta_{t-1}^{\ell} + \tilde{h}_{t-1}(\xi_{t-1}^{\ell}, z_t)\right)$$

$$- f_t(z_t)\frac{\eta_{t-1}\langle z_{0:t-1}\rangle[q_{t-1}(\cdot, z_t)\{B_{t-1}\langle z_{0:t-2}\rangle h_{t-1}(\cdot) + \tilde{h}_{t-1}(\cdot, z_t)\}]}{\eta_{t-1}\langle z_{0:t-1}\rangle[q_{t-1}(\cdot, z_t)]}. \tag{A.64}$$

现在分别处理项 $\mathrm{I}_N^{(1)}$ 、$\mathrm{I}_N^{(2)}$ 和 $\mathrm{I}_N^{(3)}$ ，并使用 Hoeffding不等式及其在 [Douc 等人，2011，引理 4] 中的推广，即可完成证明。取 $\varepsilon > 0$；则根据 Hoeffding不等式，

$$\mathbb{P}_{\boldsymbol{\eta}_0}^{\boldsymbol{P},z}\left(|\,\mathrm{I}_N^{(1)}\,| \geq \varepsilon\right) \leq 2\exp\left(-\frac{1}{2}\frac{\varepsilon^2}{\kappa_t^2}N\right). \tag{A.65}$$

对于 $\mathrm{I}_N^{(2)}$ ，我们对分子和分母分别应用归纳假设，并将两者都除以 $1/N$ 进行归一化；由于对所有 $h \in \mathsf{F}(\mathcal{X}_{t-1} \otimes \mathcal{X}_t)$ 都有 $\|Q_{t-1}h\|_\infty \leq \bar{\tau}_{t-1}\|h\|_\infty$ ，因此得到

$$\mathbb{P}_{\boldsymbol{\eta}_0}^{\boldsymbol{P},z}\left(\left|\frac{1}{N}\sum_{\ell=1}^{N}\{\beta_{t-1}^{\ell} Q_{t-1} f_t(\xi_{t-1}^{\ell}) + Q_{t-1}(\tilde{h}_{t-1} f_t + \tilde{f}_t)(\xi_{t-1}^{\ell})\}\right.\right.$$

$$\left.\left. -\eta_{t-1}\langle z_{0:t-1}\rangle\{Q_{t-1} f_t B_t\langle z_{0:t-1}\rangle h_t + Q_{t-1}(\tilde{h}_{t-1} f_t + \tilde{f}_t)\}\right| \geq \varepsilon\right)$$

$$\leq \mathsf{c}_{t-1}\exp\left(-\mathsf{d}_{t-1}\frac{\varepsilon^2}{\bar{\tau}_{t-1}^2\kappa_t^2}N\right)$$

and

$$\mathbb{P}_{\boldsymbol{\eta}_0}^{\boldsymbol{P},z}\left(\left|\frac{1}{N}\sum_{\ell=1}^{N} g_{t-1}(\xi_{t-1}^{\ell}) - \eta_{t-1}\langle z_{0:t-1}\rangle g_{t-1}\right| \geq \varepsilon\right) \leq \mathsf{c}_{t-1}\exp\left(-\mathsf{d}_{t-1}\frac{\varepsilon^2}{\bar{\tau}_{t-1}^2}N\right).$$

将前两个界与 [Douc 等人，2011，引理 4] 中的广义Hoeffding不等式结合，并同时使用以下界，可得

$$\frac{\sum_{\ell=1}^{N}\{\beta_{t-1}^{\ell} Q_{t-1} f_t(\xi_{t-1}^{\ell}) + Q_{t-1}(\tilde{h}_{t-1} f_t + \tilde{f}_t)(\xi_{t-1}^{\ell})\}}{\sum_{\ell'=1}^{N} g_{t-1}(\xi_{t-1}^{\ell'})} \leq \kappa_t$$

以及 $\eta_{t-1}\langle z_{0:t-1}\rangle g_{t-1} \geq \tau_{t-1}$，该不等式

$$\mathbb{P}_{\boldsymbol{\eta}_0}^{\boldsymbol{P},z}\left(|\,\mathrm{I}_N^{(2)}\,| \geq \varepsilon\right) \leq \mathsf{c}_{t-1}\exp\left(-\mathsf{d}_{t-1}\frac{\tau_{t-1}^2\varepsilon^2}{\bar{\tau}_{t-1}^2\kappa_t^2}N\right). \tag{A.66}$$

最后一项 $\mathrm{I}_N^{(3)}$ 的处理方式类似；事实上，根据归纳假设，由于 $\|q_{t-1}\|_\infty \leq$ $\tau_{t-1}\sigma_{t-1}$,

$$\mathbb{P}_{\boldsymbol{\eta}_0}^{\boldsymbol{P},z}\left(\left|\frac{1}{N}\sum_{\ell=1}^{N} q_{t-1}(\xi_{t-1}^{\ell}, z_t)\left(\beta_{t-1}^{\ell} + \tilde{h}_{t-1}(\xi_{t-1}^{\ell}, z_t)\right)\right.\right.$$
$$\left.\left. - \eta_{t-1}\langle z_{0:t-1}\rangle[q_{t-1}(\cdot, z_t)\{B_{t-1}\langle z_{0:t-1}\rangle h_{t-1}(\cdot) + \tilde{h}_{t-1}(\cdot, z_t)\}]\right| \geq \varepsilon\right)$$
$$\leq \mathsf{c}_{t-1}\exp\left(-\mathsf{d}_{t-1}\left(\frac{\varepsilon}{\bar{\tau}_{t-1}\bar{\sigma}_{t-1}\sum_{m=0}^{t-1}\|\tilde{h}_m\|_\infty}\right)^2 N\right)$$

and

$$\mathbb{P}_{\boldsymbol{\eta}_0}^{\boldsymbol{P},z}\left(\left|\frac{1}{N}\sum_{\ell=1}^{N} q_{t-1}(\xi_{t-1}^{\ell}, z_t) - \eta_{t-1}\langle z_{0:t-1}\rangle[q_{t-1}(\cdot, z_t)]\right| \geq \varepsilon\right) \leq \mathsf{c}_{t-1}\exp\left(-\mathsf{d}_{t-1}\left(\frac{\varepsilon}{\bar{\tau}_{t-1}\bar{\sigma}_{t-1}}\right)^2 N\right).$$

因此，由于

$$\sum_{\ell=1}^{N}\frac{q_{t-1}(\xi_{t-1}^{\ell}, z_t)}{\sum_{\ell'=1}^{N} q_{t-1}(\xi_{t-1}^{\ell'}, z_t)}\left(\beta_{t-1}^{\ell} + \tilde{h}_{t-1}(\xi_{t-1}^{\ell}, z_t)\right) \leq \sum_{m=0}^{t-1}\|\tilde{h}_m\|_\infty$$

且 $\eta_{t-1}\langle z_{0:t-1}\rangle[q_{t-1}(\cdot, z_t)] \geq \tau_{t-1}$，广义Hoeffding不等式给出

$$\mathbb{P}_{\boldsymbol{\eta}_0}^{\boldsymbol{P},z}\left(|\,\mathrm{I}_N^{(3)}\,| \geq \varepsilon\right) \leq \mathsf{c}_{t-1}\exp\left(-\mathsf{d}_{t-1}\left(\frac{\underline{\tau}_{t-1}\varepsilon}{2\bar{\tau}_{t-1}\bar{\sigma}_{t-1}\|f_t\|_\infty\sum_{m=0}^{t-1}\|\tilde{h}_m\|_\infty}\right)^2 N\right). \tag{A.67}$$

最后，结合界 (A.65–A.67) 即完成证明。 □

A.6.6 命题 4 的证明

命题 4 的结论可由以下更一般的结果推出，我们将在下文证明该结果。

命题 9. 对于每个 $t \in \mathbb{N}$, $M \in \mathbb{N}^*$, $N \in \mathbb{N}^*$, $z_{0:t} \in \mathsf{X}_{0:t}$, $(f_t, \tilde{f}_t) \in \mathsf{F}(\mathcal{X}_t)^2$, 和 $p \geq 2$，有

$$\int \mathbb{C}_t\mathbb{S}_t(z_{0:t}, \mathrm{d}\boldsymbol{b}_t)\left|\frac{1}{N}\sum_{i=1}^{N}\{b_t^i f_t(x_{t|t}^i) + \tilde{f}_t(x_{t|t}^i)\} - \eta_t\langle z_{0:t}\rangle(f_t B_t\langle z_{0:t-1}\rangle h_t + \tilde{f}_t)\right|^p \leq \mathsf{c}_t(p/\mathsf{d}_t)^{p/2}N^{-p/2}\kappa_t^p,$$

其中，$\mathsf{c}_t > 0$、$\mathsf{d}_t > 0$ 和 κ_t 分别在命题 *8* 和 (A.59) 中定义。

在证明命题 9 之前，我们先建立以下结果。

引理 6. 设 X 是一个取值于 $\mathbb{R}^d$ 的随机变量，定义在某个概率空间 $(\Omega\ \mathcal{F},\ \mathbb{P})$ 上，且满足 $\mathbb{P}(|X| \geq t) \leq c\exp(-t^2/(2\sigma^2))$，对每个 $t \geq 0$ 以及某个 $c > 0$ 和 $\sigma > 0$ 均成立。则对每个 $p \geq 2$，都有 $\mathbb{E}[|X|^p] \leq cp^{p/2}\sigma^p$。

证明。 利用富比尼定理和变量替换公式,

$$\mathbb{E}\left[|X|^p\right] = \int_0^\infty pt^{p-1}\mathbb{P}(|X| \geq t)\,\mathrm{d}t = cp2^{p/2-1}\sigma^p\Gamma(p/2),$$

其中 Γ 是伽马函数。接下来应用界 $\Gamma(p/2) \leq (p/2)^{p/2-1}$ (参见 [Anderson 和 Qiu, 1997])，该界由 [2, 定理 1.5]对 $p \geq 2$ 成立。 □

命题 *9* 的证明。 结合命题 8 和引理 6，我们得到

$$N\int \mathbb{C}_t\mathbb{S}_t(z_{0:t},\mathrm{d}\boldsymbol{b}_t)\left|\frac{1}{N}\sum_{i=1}^{N}\{b_t^i f_t(x_{t|t}^i)+\tilde{f}_t(x_{t|t}^i)\}-\eta_t\langle z_{0:t}\rangle(f_tB_t\langle z_{0:t-1}\rangle h_t+\tilde{f}_t)\right|^2$$
$$\leq \mathsf{c}_t(p/\mathsf{d}_t)^{p/2}N^{-p/2}\left(\|f_t\|_\infty\sum_{m=0}^{t-1}\|\tilde{h}_m\|_\infty+\|\tilde{f}_t\|_\infty\right)^p,$$

这正是需要证明的。 □

A.6.7 命题 5 的证明

与之前一样，我们通过一个更一般的结果来证明命题 5，具体如下。

命题 10。 对于每个 $t\in\mathbb{N}$，存在 $\bar{\mathsf{c}}_t^{bias}<\infty$，使得对于每个 $M\in\mathbb{N}^*$、$N\in\mathbb{N}^*$、$z_{0:t}\in$ $_{0:t}$ f_t $\tilde{f}_t$ $\in$ $\mathcal{X}_t$ ^{2}, X，以及(、) F()

$$\left|\int \mathbb{C}_t\mathbb{S}_t(z_{0:t},\mathrm{d}\boldsymbol{b}_t)\frac{1}{N}\sum_{i=1}^{N}\{b_t^i f_t(x_{t|t}^i)+\tilde{f}_t(x_{t|t}^i)\}-\eta_t\langle z_{0:t}\rangle(f_tB_t\langle z_{0:t-1}\rangle h_t+\tilde{f}_t)\right|\leq \bar{\mathsf{c}}_t^{bias}\kappa_t N^{-1},$$

其中， κ_t 定义于(A.59)。

在证明命题 10 之前，我们先给出一个技术性引理，用于界定随机变量之比的偏差。

引理 7。 设 α 和 β 是定义在某个概率空间$(\Omega\ \mathcal{F},\mathbb{P})$ 上的（可能相依的）随机变量，并且满足 $\mathbb{E}[\alpha^2]<\infty$ 和 $\mathbb{E}[\beta^2]<\infty$。此外，假设存在 $c>0$ 和 $d>0$，使得 $|\alpha/\beta|\leq c$, $\mathbb{P}$-*a.s.*, $|a/b|\leq c$, $\mathbb{E}[(\alpha-a)^2]\leq c^2d^2$，且 $\mathbb{E}[(\beta-b)^2]\leq d^2$。那么

$$|\mathbb{E}[\alpha/\beta]-a/b|\leq 2c(d/b)^2+c|\mathbb{E}[\beta-b]|/|b|+|\mathbb{E}[\alpha-a]|/|b|. \tag{A.68}$$

证明。 利用恒等式

$$\mathbb{E}[\alpha/\beta]-a/b=\mathbb{E}[(\alpha/\beta)(b-\beta)^2]/b^2+\mathbb{E}[(\alpha-a)(b-\beta)]/b^2+a\mathbb{E}[b-\beta]/b^2+\mathbb{E}[\alpha-a]/b,$$

根据引理的假设应用柯西–施瓦茨不等式，可得所述结论：

$$\begin{aligned}&|\mathbb{E}[\alpha/\beta]-a/b|\\&\quad\leq c\mathbb{E}[(\beta-b)^2]/b^2+\{\mathbb{E}[(\alpha-a)^2]\mathbb{E}[(\beta-b)^2]\}^{1/2}/b^2+|a||\mathbb{E}[\beta-b]|/b^2+|\mathbb{E}[\alpha-a]|/b^2\\&\quad\leq 2c(d/b)^2+c|\mathbb{E}[\beta-b]|/|b|+|\mathbb{E}[\alpha-a]|/|b|.\end{aligned}$$

□

命题 *5* 的证明。 我们采用归纳法，并假设对于 $n-1$，该结论成立。重用误差分解 (A.62)，只需分别界定 (A.63) 和 (A.64) 中给出的 $\mathrm{I}_N^{(2)}$ 和 $\mathrm{I}_N^{(3)}$ 的期望（因为 $\mathbb{E}^{\boldsymbol{P}_{\cdot}}{}_{,\boldsymbol{\eta}_0}{}^z[\mathrm{I}_N^{(1)}]=0$）。我们将利用归纳假设、引理 7 和命题 9 来完成这一步。更具体地说，为了界定 $\mathrm{I}_N^{(2)}$ 的期望，我们在 $\alpha\leftarrow\alpha_t$、$\beta\leftarrow\beta_t$、$a\leftarrow a_t$ 和 $b\leftarrow b_t$ 的条件下使用引理 7，其中

$$\alpha_t\coloneqq\frac{1}{N}\sum_{\ell=1}^{N}\{\beta_{t-1}^\ell Q_{t-1}f_t(\xi_{t-1}^\ell)+Q_{t-1}(\tilde{h}_{t-1}f_t+\tilde{f}_t)(\xi_{t-1}^\ell)\},\quad \beta_t\coloneqq\frac{1}{N}\sum_{\ell=1}^{N}g_{t-1}(\xi_{t-1}^\ell),$$
$$a_t\coloneqq\eta_{t-1}\langle z_{0:t-1}\rangle\{Q_{t-1}f_tB_t\langle z_{0:t-1}\rangle h_t+Q_{t-1}(\tilde{h}_{t-1}f_t+\tilde{f}_t)\},\quad b_t\coloneqq\eta_{t-1}\langle z_{0:t-1}\rangle g_{t-1}.$$

为此，注意 $|\alpha_t/\beta_t| \le \kappa_t$ 和 $|a_t/b_t| \le \kappa_t$，其中 κ_t 在 (A.59) 中定义。另一方面，使用命题 9（应用于 $p=2$），我们得到

$$\mathbb{E}^{\boldsymbol{P},z}_{\boldsymbol{\eta}_0}[(\alpha_t - a_t)^2] \le d_t^2 \kappa_t^2 \quad \text{and} \quad \mathbb{E}^{\boldsymbol{P},z}_{\boldsymbol{\eta}_0}[(\beta_t - b_t)^2] \le d_t^2,$$

其中 $d_t^2 := \mathsf{c}_t \bar{\tau}_{t-1}^2/(\mathsf{d}_t N)$。根据归纳假设，我们得到

$$|\mathbb{E}^{\boldsymbol{P},z}_{\boldsymbol{\eta}_0}[\alpha_t] - a_t| \le \bar{\mathsf{c}}_{t-1}^{bias} N^{-1} \bar{\tau}_{t-1} \kappa_t \quad \text{and} \quad |\mathbb{E}^{\boldsymbol{P},z}_{\boldsymbol{\eta}_0}[\beta_t] - b_t| \le \bar{\mathsf{c}}_{t-1}^{bias} N^{-1} \bar{\tau}_{t-1}.$$

因此，引理 7 的条件得到满足，我们可推出

$$|\mathbb{E}^{\boldsymbol{P},z}_{\boldsymbol{\eta}_0}[\mathrm{I}_N^{(2)}]| = |\mathbb{E}^{\boldsymbol{P},z}_{\boldsymbol{\eta}_0}[\alpha_t/\beta_t] - a_t/b_t| \le 2\kappa_t \frac{\mathsf{c}_t}{\mathsf{d}_t N} \frac{\bar{\tau}_{t-1}^2}{\underline{\tau}_{t-1}^2} + 2\bar{\mathsf{c}}_{t-1}^{bias} \kappa_t \frac{\bar{\tau}_{t-1}}{\underline{\tau}_{t-1} N}.$$

$|\mathbb{E}^{\boldsymbol{P},z}_{\boldsymbol{\eta}_0}[\mathrm{I}_N^{(2)}]|$ 上的界也可用类似方法得到。 □

B. 学习 PPG

本节分为三个小节。附录 B.1 紧随 Karimi 等人 [的工作，在一般假设下为随机逼近方案建立了一个非渐近界。附录 B.2 说明假设 **A4.1** 和 **A 3.1** 如何蕴含附录 B.1 中给出的假设，从而能够证明定理 2。最后，附录 B.3 给出了确保模型满足 **A4.1** 的充分假设。

B.1 非渐近界

我们紧密遵循 [Karimi 等人， 2019]。考虑如下递推

$$\theta_{n+1} = \theta_n - \gamma_{n+1} H_{\theta_n}(X_{n+1}), \quad n \in \mathbb{N},$$

其中，对于某个 $\theta_n \in \Theta \subset \mathbb{R}^d$ $d \in \mathbb{N}^*$，且 $\{X_n\}_{n\in\mathbb{N}}$ 是定义在某个可测空间 $(\mathsf{X}, \mathcal{X})$ 上的依赖状态的马尔可夫链，即 $X_{n+1} \sim \mathbb{P}_{\theta_n}(X_n, \cdot)$，其中 $\mathbb{P}_\theta$ 是定义在 $(\mathsf{X}, \mathcal{X})$ 上的某个马尔可夫核。令 $h(\theta) = \int H_\theta(x)\, \pi_\theta(\mathrm{d}x)$，其中 π_θ 是 $\mathbb{P}_\theta$ 的不变测度，且 $e_{n+1} := H_{\theta_n}(X_{n+1}) - h(\theta_n)$。由于有限维向量空间中的所有范数都等价，我们用 $\|\cdot\|$ 表示一个泛指的范数。我们用 $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ 表示马尔可夫链 $\{X_n\}_{n\in\mathbb{N}}$ 的自然滤过。

A B.1. 存在一个博雷尔可测函数 $V\colon \Theta \to \mathbb{R}$，使得对每个 $\theta \in \Theta$，$\nabla V(\theta) = h(\theta)$。

A B.2. 存在 $L^V \in \mathbb{R}_{\ge 0}$，使得对于每个 $(\theta, \theta') \in \Theta^2$，

$$\|\nabla V(\theta) - \nabla V(\theta')\| \le L^V \|\theta - \theta'\|.$$

A B.3. 存在一个博雷尔可测函数 $\widehat{H}\colon \Theta \times \mathsf{X} \to \Theta$，使得对于每个 $\theta \in \Theta$ 和 $x \in \mathsf{X}$，

$$\widehat{H}_\theta(x) - \mathbb{P}_\theta \widehat{H}_\theta(x) = H_\theta(x) - h(\theta)\,.$$

A B.4. 存在 $L^{\mathbb{P}H} \in \mathbb{R}_{\ge 0}$，使得对于每个 $(\theta_0, \theta_1) \in \Theta^2$，

$$\sup_{x\in\mathsf{X}} \|\mathbb{P}_{\theta_0} \widehat{H}_{\theta_0}(x) - \mathbb{P}_{\theta_0} \widehat{H}_{\theta_1}(x)\| \le L^{\mathbb{P}\widehat{H}} \|\theta_0 - \theta_1\|\,.$$

假设 B.5. 存在 $L_0^{\mathbb{P}H} \in \mathbb{R}_{\ge 0}$，使得

$$\sup_{\theta\in\Theta} \|\mathbb{P}_\theta \widehat{H}_\theta\| \le L_0^{\mathbb{P}\widehat{H}}\,.$$

假设 B.6. 存在 $\sigma_{mse} \in \mathbb{R}_{\geq 0}$，使得对于每个 $x \in$X 和 $\theta \in \Theta$,

$$\int \|H_\theta(x') - h(\theta)\|^2 \, \mathbb{P}_\theta(x, \mathrm{d}x') \leq \sigma_{mse}^2 \, .$$

假设 B.7. 存在 $L^H \in \mathbb{R}_{\geq 0}$，使得对于每个 $x \in$X,

$$\sup_{\theta \in \Theta} \int \|\widehat{H}_\theta\| \, \mathbb{P}_\theta(x, \mathrm{d}x') \leq L^{\widehat{H}} \, .$$

定理 8. 假设 假设*B.1*–假设*B.7* 均成立。此外，假设存在 $a > 0$ 和 $a' > 0$，使得对于所有 $n \in \mathbb{N}$,

$$\gamma_{n+1} \leq \gamma_n \leq a\gamma_{n+1} \, , \quad \gamma_n - \gamma_{n+1} \leq a'\gamma_n^2 \, , \quad \gamma_1 \leq (L^V + C_h)^{-1}/2 \, .$$

此外，对于任意 $n \in \mathbb{N}^*$，令 ϖ 为一个 $[\![0, n]\!]$-值随机变量，它与 $\{\mathcal{F}_\ell\}_{\ell \geq 0}$ 独立，且满足 $\mathbb{P}(\varpi = k) = \gamma_{k+1} / \sum_{\ell=0}^n \gamma_{\ell+1}$，其中 $k \in [\![0, n]\!]$。于是，

$$\mathbb{E}\left[\|h(\theta_\varpi)\|^2\right] \leq 2 \frac{V_{0,n} + C_{0,n} + (\sigma_{mse}^2 L^V + C_\gamma) \sum_{k=0}^n \gamma_{k+1}^2}{\sum_{k=0}^n \gamma_{k+1}} \, ,$$

其中 $V_{0,\ n} \coloneqq \mathbb{E}\left[V(\theta) - V(\theta_n)\right]$ 且

$$C_{0,n} \coloneqq \gamma_1 h(\theta_0) L^{\widehat{H}} + L_0^{\mathbb{P}\widehat{H}} (\gamma_1 - \gamma_{n+1} + 1) \, , \tag{B.69}$$

$$C_\gamma \coloneqq \sigma_{mse} L^{\mathbb{P}\widehat{H}} + (1 + \sigma_{mse}) L^V L_0^{\mathbb{P}\widehat{H}} \, , \tag{B.70}$$

$$C_h \coloneqq L^{\mathbb{P}\widehat{H}} \left((a+1)/2 + a\sigma_{mse}\right) + (L^V + a' + 1) L_0^{\mathbb{P}\widehat{H}} \, . \tag{B.71}$$

证明. 我们沿用 Karimi 等人，2019，定理 2] 的证明，并将其调整到我们的设定中。首先注意，根据 **假设**B.1，Karimi 等人，2019，定理 2] 中的假设 **A1** 和 **A2** 在$c_0 = d_0 = 0$ 和 $c_1 = d_1 = 1$下成立。此外，Karimi 等人，2019，引理 1] 中的结论也成立，因为根据 **假设**B.2，**A3**成立。此外，在 **假设** B.6下也可以建立 Karimi 等人，2019，公式 17]，因为我们可以将其改写为

$$\sum_{\ell=0}^n \gamma_{\ell+1}^2 \mathbb{E}\left[\|e_{\ell+1}\|^2\right] = \sum_{\ell=0}^n \gamma_{\ell+1}^2 \mathbb{E}\left[\mathbb{E}\left[\|e_{\ell+1}\|^2 \mid \mathcal{F}_\ell\right]\right] \leq \sigma_{mse}^2 \sum_{\ell=0}^n \gamma_{\ell+1}^2 \, .$$

按照 [Karimi 等人，2019 年的证明， 2], 考虑分解

$$\mathbb{E}\left[-\sum_{\ell=0}^n \gamma_{\ell+1} \left\langle \nabla V(\theta_\ell), e_{\ell+1} \right\rangle\right] = \mathbb{E}\left[A_1 + A_2 + A_3 + A_4 + A_5\right] ,$$

其中

$$A_1 \coloneqq -\sum_{\ell=1}^n \gamma_{\ell+1} \left\langle \nabla V(\theta_\ell), \widehat{H}_{\theta_\ell}(X_{\ell+1}) - \mathbb{P}_{\theta_\ell} \widehat{H}_{\theta_\ell}(X_\ell) \right\rangle ,$$

$$A_2 \coloneqq -\sum_{\ell=1}^n \gamma_{\ell+1} \left\langle \nabla V(\theta_\ell), \mathbb{P}_{\theta_\ell} \widehat{H}_{\theta_\ell}(X_\ell) - \mathbb{P}_{\theta_{\ell-1}} \widehat{H}_{\theta_{\ell-1}}(X_\ell) \right\rangle ,$$

$$A_3 \coloneqq -\sum_{\ell=1}^n \gamma_{\ell+1} \left\langle \nabla V(\theta_\ell) - \nabla V(\theta_{\ell-1}), \mathbb{P}_{\theta_{\ell-1}} \widehat{H}_{\theta_{\ell-1}}(X_\ell) \right\rangle ,$$

$$A_4 \coloneqq -\sum_{\ell=1}^n (\gamma_{\ell+1} - \gamma_\ell) \left\langle \nabla V(\theta_{\ell-1}), \mathbb{P}_{\theta_{\ell-1}} \widehat{H}_{\theta_{\ell-1}}(X_\ell) \right\rangle ,$$

$$A_5 \coloneqq -\gamma_1 \left\langle \nabla V(\theta_0), \widehat{H}_{\theta_0}(X_1) \right\rangle + \gamma_{n+1} \left\langle \nabla V(\theta_n), \mathbb{P}_{\theta_n} \widehat{H}_{\theta_n}(X_{n+1}) \right\rangle .$$

$\widehat{\ }$ $\widehat{\ }$由于 $H_{\theta_\ell}(X_{\ell+1}) - \mathbb{P}_{\theta_\ell} H_{\theta_\ell}(X_\ell)$ 是鞅差，因此有 $\mathbb{E}[A_1] = 0$。 A_2、 A_3 和 A_4 的期望上界可按照 [Karimi 等人， 2019] 中的方法类似地得到。利用 **AB.4**,

$$A_2 \leq L^{\mathbb{P}\widehat{H}} \left(\sigma_{mse} \sum_{k=1}^{n} \gamma_k^2 + \frac{1}{2}(1 + 2a\sigma_{mse} + a) \sum_{k=0}^{n} \gamma_{k+1}^2 \|h(\theta_k)\|^2 \right) .$$

根据**A** B.2 和 B.5,

$$A_3 \leq L^V L_0^{\mathbb{P}\widehat{H}} \left((1 + \sigma_{mse}) \sum_{k=1}^{n} \gamma_k^2 + \sum_{k=1}^{n} \gamma_k^2 \|h(\theta_k)\|^2) \right) .$$

另一方面,

$$A_4 \leq L_0^{\mathbb{P}\widehat{H}} \left(\gamma_1 - \gamma_{n+1} + a' \sum_{k=1}^{n} \gamma_k^2 \|h(\theta_{k-1})\|^2 \right) .$$

现在我们聚焦于 A_5。如同 [Karimi 等人，2019，引理 2], 的证明一样，第一、项的期望可以使用柯西–施瓦茨不等式和**AB.7**，通过 $\gamma_1 \|h(\theta_0)\| L^H$ 直接进行界定。利用**A** B.5 和 $\gamma_{n+1}\|h(\theta_n)\| \leq 1 + \gamma_{n+1}^2 \|h(\theta_n)\|^2$，第二项可以用同样的方式界定为

$$\begin{aligned} \gamma_{n+1} \left\langle \nabla V(\theta_n), \mathbb{P}_{\theta_n} \widehat{H}_{\theta_n}(X_{n+1}) \right\rangle \leq L_0^{\mathbb{P}\widehat{H}} \gamma_{n+1} \|h(\theta_n)\| \leq L_0^{\mathbb{P}\widehat{H}} \left(1 + \gamma_{n+1}^2 \|h(\theta_n)\|^2\right) \\ \leq L_0^{\mathbb{P}\widehat{H}} \left(1 + \sum_{\ell=0}^{n} \gamma_{\ell+1}^2 \|h(\theta_\ell)\|^2 \right) . \end{aligned}$$

证明的其余部分遵循 [Karimi 等人，2019，定理 2]的证明。 □

B.2 定理 2 的应用

本节旨在证明，定理 2 的假设确保附录 B.1 中的所有假设均得到满足，从而可以应用定理 8。首先，我们明确地根据附录 A 中给出的核来定义核 $\mathbb{P}_\theta$ and 函数 h。我们写作 $\mathbb{P}_{\theta,t}$，而不是 $\mathbb{P}_\theta$，以明确核对固定 观测值数量 t 的依赖关系。

B.2.1 定理 8 的假设验证

对于满足 $k_0 < k$ 的 $(k_0, k) \in (\mathbb{N}^*)^2$，定义

$$\mathbb{P}_{\theta,t} : \mathsf{E}_t^{k-k_0} \times \boldsymbol{\mathcal{E}}_t^{\otimes(k-k_0)} \ni (\boldsymbol{y}_t[k_0 : k], z_{0:t}[k_0 : k], A) \mapsto \mathbb{K}_{\theta,t}^{k_0} \otimes \mathbb{K}_{\theta,t}^{\otimes(k-k_0)}(z_{0:t}[k], A), \tag{B.72}$$

其中，$\mathbb{K}_{\theta,t}$ 是 (A.41) 中定义的 PPG 核。注意，$\mathbb{P}_{\theta,t}$ 仅取决于最后一条冻结路径，即 $z_{0:t}[k]$。还要注意，由于 $\mathbb{K}_{\theta,t}$ 仅取决于路径，因此 $\boldsymbol{y}_{t,\ell}[k_0 : k]$ 与 $\boldsymbol{y}_{t,\ell+1}[k_0 : k]$ 之间不存在依赖关系。得分上升算法（算法 4）可表述如下。

1. Sample $(z_{0:t,\ell}[k_0 : k], \boldsymbol{y}_{t,\ell}[k_0 : k]) \sim \mathbb{P}_{\theta_\ell,t}\big((z_{0:t,\ell-1}[k_0 : k], \boldsymbol{y}_{t,\ell-1}[k_0 : k]), \cdot\big)$.
2. 根据 $\eta_{\ell+1} = \eta_\ell + \gamma_{\ell+1} H(z_{0:t,\ell}[k_0 : k], \boldsymbol{y}_{t,\ell}[k_0 : k])$ 更新参数，其中

$$H(z_{0:t,\ell}[k_0 : k], \boldsymbol{y}_{t,\ell}[k_0 : k]) = \frac{1}{k - k_0 + 1} \sum_{i=k_0}^{k} \mu(\boldsymbol{\beta}_{t,\ell}[i])(\mathrm{id}) = \Pi_{(k_0-1,k),N}(h_t),$$

其中，$\Pi_{(k_0-1,k)\ N}(h_t)$ 在 (3.12) 中定义。我们用 $\pi_{\theta,t}$ 表示 $\mathbb{P}_{\theta,t}$ 的不变分布；根据命题 3，该分布为 $\pi_{\theta,t} = (\eta_{0:t} \otimes \mathbb{C}_t \mathbb{S}_t)^{\otimes(k-k_0)}$。

我们还要求强混合假设关于 θ一致成立。

条件 B.8 （关于 θ 一致的强混合）。 对于每个 $s \in \mathbb{N}$，都存在 $\underline{\tau}_s$、 $\bar{\tau}_s$、 $\underline{\sigma}_s$和 $\bar{\sigma}_s$ 属于 $\mathbb{R}_+^*$，使得对所有 $\theta \in \Theta$,

(i) $\underline{\tau}_s \leq g_{s,\theta}(x_s) \leq \bar{\tau}_s$ 对每个 $x_s \in \mathsf{X}_s$ 均成立，(ii) $\underline{\sigma}_s \leq m_{s,\theta}(x_s, x_{s+1}) \leq \bar{\sigma}_s$ 对每个 $(x_s, x_{s+1}) \in \mathsf{X}_{s:s+1}$ 均成立。

注意，上述假设意味着 $\kappa_{N,t}$ 关于 θ 也是一致的。

A B.1 成立的证明。

命题 11. 对于所有 $\theta \in \Theta$, $h(\theta) = \nabla V(\theta)$，其中 $V(\theta) =$ 对数$\gamma_{0:t,\theta}(\mathsf{X}_{0:t})$ 是对数似然函数。

证明。 由定理 5,

$$\begin{aligned} h(\theta) &= \int H(\tilde{\boldsymbol{y}}_t[k_0:k], \tilde{x}_{0:t}[k_0:k])\, \pi_{\theta,t}(\mathrm{d}(\tilde{\boldsymbol{y}}_t[k_0:k], \tilde{x}_{0:t}[k_0:k])) \\ &= \frac{1}{k-k_0+1} \sum_{i=k_0}^{k} \int [\eta_{0:t,\theta} \otimes \mathbb{C}_{t,\theta} \mathbb{S}_{t,\theta}]\, (\mathrm{d}(\tilde{\boldsymbol{y}}_t[i], \tilde{x}_{0:t}[i])) \mu(\tilde{\boldsymbol{\beta}}_{t,\ell}[i])(\mathrm{id}) \\ &= \eta_{0:t,\theta}\left(s_{0:t,\theta}\right) = \nabla V(\theta). \end{aligned}$$

□

A B.2 成立的证明。AB.2 可由 **A 4.1(i)** 直接推出。

证明 A B.3 和 B.5 成立。令 H_θ 定义为

$$\widehat{H}_\theta : \mathsf{E}_t^{k-k_0} \ni (\boldsymbol{y}_t[k_0:k], z_{0:t}[k_0:k]) \mapsto \sum_{r=0}^{\infty} \{\mathbb{P}_{\theta,t}^r H(\boldsymbol{y}_t[k_0:k], z_{0:t}[k_0:k]) - h(\theta)\}. \tag{B.73}$$

则以下结论成立。

引理 8。 假设 ***A***满足条件 *B.8*。则对于所有 $\theta \in \Theta$ 和 $t \in \mathbb{N}^*$,

$$\|\mathbb{P}_{\theta,t} \widehat{H}_\theta\|_\infty \leq \sigma_{bias} (1 - \kappa_{N,t}^k)^{-1} .$$

证明。 根据定理 1，对于任意 $r > 0$，我们有

$$\left|\mathbb{P}_{\theta,t}^r H(\boldsymbol{y}_t[k_0:k], z_{0:t}[k_0:k]) - h(\theta)\right| \leq \sigma_{bias} \kappa_{N,t}^{(r-1)k}$$

从而

$$\|\mathbb{P}_{\theta,t} \widehat{H}_\theta\|_\infty \leq \sum_{r=1}^{\infty} \left\|\mathbb{P}_{\theta,t}^r H - h(\theta)\right\|_\infty \leq \sigma_{bias} \sum_{r=0}^{\infty} \kappa_{N,t}^{rk} \leq \sigma_{bias} (1 - \kappa_{N,t}^k)^{-1} ,$$

其中 $\kappa_{N,t} \in (0, 1)$。 □

引理 8 证明了 **A** B.3 和 B.5，其中 $L_0^{\mathbb{P}H} := \sigma_{bias}(1 - \kappa_{N,t}^k)^{-1}$。

A B.4 成立的证明。

定理 9. 假设 ***A*** 条件 ***B.8*** 和 ***A4.1***。那么对于每个满足 $N > 1 + 5\rho_t^2 t/2$ 的 $t \in \mathbb{N}$、$\theta \in \Theta$ 和 $N \in \mathbb{N}^*$，

$$\left\| \mathbb{P}_{\theta_1,t} \widehat{H}_{\theta_1} - \mathbb{P}_{\theta_2,t} \widehat{H}_{\theta_2} \right\|_\infty \leq L^{\mathbb{P}\widehat{H}} \|\theta_1 - \theta_2\| \,,$$

其中

$$L^{\mathbb{P}\widehat{H}} \coloneqq \|L_2^P\|_\infty \left[1 + \kappa_{N,t}^k (1 - \kappa_{N,t}^k)\right] + L^V + \sigma_{bias}(1 - \kappa_{N,t})^{-1}(1 - \kappa_{N,t}^k)^{-1} \left[\|L_1^P\|_\infty (1 - \kappa_{N,t}^k)^{-1} + L^\eta \kappa_{N,t}^k\right] \,. \quad \text{(B.74)}$$

证明。我们通过改写 [Karimi 等人，2019，引理 7] 的证明来确立这一结论。首先，回顾一下，在 (A.42) 中定义的核 $K_{\theta,\,t}$ 是 (A.41) 中给出的 $\mathbb{K}_{\theta,\,t}$ 的路径边缘化版本。注意，对每个 $x \in \mathsf{E}_t^{k-k_0}$，

$$\mathbb{P}_{\theta_1,t} \widehat{H}_{\theta_1}(x) = \sum_{n=0}^{\infty} \delta_x \mathbb{P}_{\theta_1,t} \left\{ \mathbb{P}_{\theta_1,t}^n H - h(\theta_1) \right\} = \sum_{n=0}^{\infty} \delta_x K_{\theta_1,t}^{kn} \left\{ \mathbb{P}_{\theta_1,t} H - \eta_{0:t,\theta_1} \mathbb{P}_{\theta_1,t} H \right\} \,,$$

其中我们用到了 (i) 由 $\mathbb{P}_{\theta,\,t}$ 输出的反向统计量与输入反向统计量相互独立这一事实，以及 (ii) 上文计算 $h(\theta)$ 时倒数第二行的结果。我们沿用 [Fort 等，2011，引理4.2] 的证明，并考虑如下分解：对于 $n \in \mathbb{N}^*$，

$$\begin{aligned}
&\delta_x K_{\theta_1,t}^{kn} \left(\mathbb{P}_{\theta_1,t} H - \eta_{0:t,\theta_1} \mathbb{P}_{\theta_1,t} H\right) - \delta_x K_{\theta_2,t}^{kn} \left(\mathbb{P}_{\theta_2,t} H - \eta_{0:t,\theta_2} \mathbb{P}_{\theta_2,t} H\right) \qquad \text{(B.75)} \\
&\quad = \sum_{j=0}^{n-1} \left(\delta_x K_{\theta_1,t}^{kj} - \eta_{0:t,\theta_1}\right) \left(K_{\theta_1,t}^{kj} - K_{\theta_2,t}^{kj}\right) \left(K_{\theta_2,t}^{k(n-j-1)} \mathbb{P}_{\theta_1,t} H - \eta_{0:t,\theta_2} \mathbb{P}_{\theta_1,t} H\right) \\
&\quad\quad - \left(\delta_x K_{\theta_2,t}^{kn} \mathbb{P}_{\theta_2,t} H - \eta_{0:t,\theta_2} \mathbb{P}_{\theta_2,t} H\right) + \left(\delta_x K_{\theta_2,t}^{kn} \mathbb{P}_{\theta_1,t} H - \eta_{0:t,\theta_2} \mathbb{P}_{\theta_1,t} H\right) \\
&\quad\quad - \eta_{0:t,\theta_1} \left(K_{\theta_2,t}^{kn} \mathbb{P}_{\theta_1,t} H - \eta_{0:t,\theta_2} \mathbb{P}_{\theta_1,t} H\right) \,.
\end{aligned}$$

将定理 6 应用于 $\mu = \delta_x$ 和 $\nu = \eta_{0:t,\,\theta}$，并利用对所有 $\ell \in \mathbb{N}$ 均有 $\eta_{0:t,\,\theta} K_{\theta,\,t}^\ell = \eta_{0:t,\,\theta}$ 这一事实，我们得到，对所有 $\ell \in \mathbb{N}$ 和所有 $\theta \in \Theta$，均有 $\left\| \delta_x K_{\theta,\,t}^\ell - \eta_{0:t,\,\theta} \right\|_{\mathrm{TV}} \leq \kappa_{N,\,t}^\ell$。注意，根据 **A4.1(iii)**，$K_{\theta,\,t}$ 是利普希茨连续的；因此，对所有 $r \in \mathbb{N}^*$，由引理18可知，$K_{\theta,\,t}^r$ 是利普希茨连续的，其常数为 $\|L_1^P\|_\infty (1 - \kappa_{N,\,t})^{-1}$。综上，我们得到

$$\begin{aligned}
&\left| \left(\delta_x K_{\theta_1,t}^{kj} - \eta_{0:t,\theta_1}\right) \left(K_{\theta_1,t}^{kj} - K_{\theta_2,t}^{kj}\right) \left(K_{\theta_2,t}^{k(n-j-1)} \mathbb{P}_{\theta_1,t} H - \eta_{0:t,\theta_2} \mathbb{P}_{\theta_1,t} H\right) \right| \\
&\quad = \left| \left(\delta_x K_{\theta_1,t}^{kj} - \eta_{0:t,\theta_1}\right) \left(K_{\theta_1,t}^{kj} - K_{\theta_2,t}^{kj}\right) \left\{ K_{\theta_2,t}^{k(n-j-1)} \left[\mathbb{P}_{\theta_1,t} H - h(\theta_1)\right] - \eta_{0:t,\theta_2} \left[\mathbb{P}_{\theta_1,t} H - h(\theta_1)\right] \right\} \right| \\
&\quad \leq \|L_1^P\|_\infty (1 - \kappa_{N,t})^{-1} \kappa_{N,t}^{kj} \kappa_{N,t}^{k(n-j-1)} \|\mathbb{P}_{\theta_1,t} H - h(\theta_1)\|_\infty \|\theta_1 - \theta_2\| \\
&\quad \leq \sigma_{bias} \|L_1^P\|_\infty (1 - \kappa_{N,t})^{-1} \kappa_{N,t}^{k(n-1)} \|\theta_1 - \theta_2\| \,,
\end{aligned}$$

其中，最后一个不等式由定理 1 得出。因此，(B.75) 右侧的第一项的上界为 $\sigma_{bias} \|L_1^P\|_\infty (1 - \kappa_{N,t})^{-1} n \kappa_{N,\,t}^{k(n-1)} \|\theta_1 - \theta_2\|$。(B.75) 的第二项可写为

$$\begin{aligned}
&- \left(\delta_x K_{\theta_2,t}^{kn} \mathbb{P}_{\theta_2,t} H - \eta_{0:t,\theta_2} \mathbb{P}_{\theta_2,t} H\right) + \left(\delta_x K_{\theta_2,t}^{kn} \mathbb{P}_{\theta_1,t} H - \eta_{0:t,\theta_2} \mathbb{P}_{\theta_1,t} H\right) \\
&\qquad = \left(\delta_x K_{\theta_2,t}^{kn} - \eta_{0:t,\theta_2}\right) \left(\mathbb{P}_{\theta_1,t} H - \mathbb{P}_{\theta_2,t} H\right) \,,
\end{aligned}$$

再次利用 $K_{\theta,t}$ 的遍历性，以及根据 **A4.1(iv)**，$\theta \mapsto \mathbb{P}_{\theta,t} H$ 是一致 Lipschitz 的这一事实，可得其上界为 $\|L_2^P\|_\infty \kappa_{N,\,t}^{kn} \|\theta_1 - \theta_2\|$。最后，对于最后一项，利用

根据 $K^k_{\theta,t}$ 是 $\eta_{0:t,\theta}$-不变且几何遍历的这一事实，以及根据 **A4.1(iv)**，$\theta \mapsto \eta_{0:t,\theta}$ 是 Lipschitz 的这一事实，可得

$$\begin{aligned}
&\left|\eta_{0:t,\theta_1}\left(K^{kn}_{\theta_2,t}\mathbb{P}_{\theta_1,t}H - \eta_{0:t,\theta_2}\mathbb{P}_{\theta_1,t}H\right)\right| \\
&\quad = \left|(\eta_{0:t,\theta_1} - \eta_{0:t,\theta_2})\left\{K^{kn}_{\theta_2,t}\left[\mathbb{P}_{\theta_1,t}H - h(\theta_1)\right] - \eta_{0:t,\theta_2}\left[\mathbb{P}_{\theta_1,t}H - h(\theta_1)\right]\right\}\right| \\
&\quad \leq L^{\eta}\kappa^{kn}_{N,t}\|\mathbb{P}_{\theta_1,t}H - h(\theta_1)\|_{\infty}\|\theta_1 - \theta_2\| \\
&\quad \leq L^{\eta}\sigma_{bias}(1-\kappa_{N,t})^{-1}\kappa^{kn}_{N,t}\|\theta_1 - \theta_2\| \,.
\end{aligned}$$

因此，我们有

$$\begin{aligned}
&\delta_x K^{kn}_{\theta_1,t}\left(\mathbb{P}_{\theta_1,t}H - \eta_{0:t,\theta_1}\mathbb{P}_{\theta_1,t}H\right) - \delta_x K^{kn}_{\theta_2,t}\left(\mathbb{P}_{\theta_2,t}H - \eta_{0:t,\theta_2}\mathbb{P}_{\theta_2,t}H\right) \\
&\qquad \leq \left\{\sigma_{bias}\|L^P_1\|_{\infty}(1-\kappa_{N,t})^{-1}n\kappa^{k(n-1)}_{N,t} + \left[\|L^P_2\|_{\infty} + L^{\eta}\sigma_{bias}(1-\kappa_{N,t})^{-1}\right]\kappa^{kn}_{N,t}\right\}\|\theta_1 - \theta_2\| \,.
\end{aligned}$$

因此，我们得到

$$\begin{aligned}
&\left|\mathbb{P}_{\theta_1,t}\widehat{H}_{\theta_1}(x) - \mathbb{P}_{\theta_2,t}\widehat{H}_{\theta_2}(x)\right| \\
&\leq \left|\delta_x\mathbb{P}_{\theta_1,t}H - \delta_x\mathbb{P}_{\theta_2,t}H\right| + \left|\eta_{0:t,\theta_1}\mathbb{P}_{\theta_1,t}H - \eta_{0:t,\theta_2}\mathbb{P}_{\theta_2,t}H\right| \\
&\qquad + \left|\sum_{n=1}^{\infty}\delta_x K^{kn}_{\theta_1,t}\left(\mathbb{P}_{\theta_1,t}H - \eta_{0:t,\theta_1}\mathbb{P}_{\theta_1,t}H\right) - \delta_x K^{kn}_{\theta_2,t}\left(\mathbb{P}_{\theta_2,t}H - \eta_{0:t,\theta_2}\mathbb{P}_{\theta_2,t}H\right)\right| \\
&\leq \left|\delta_x\mathbb{P}_{\theta_1,t}H - \delta_x\mathbb{P}_{\theta_2,t}H\right| + \left|\eta_{0:t,\theta_1}\mathbb{P}_{\theta_1,t}H - \eta_{0:t,\theta_2}\mathbb{P}_{\theta_2,t}H\right| \\
&+ \left\{\sigma_{bias}\|L^P_1\|_{\infty}(1-\kappa_{N,t})^{-1}(1-\kappa^k_{N,t})^{-2}\right. \\
&\left. + \left[\|L^P_2\|_{\infty} + L^{\eta}\sigma_{bias}(1-\kappa_{N,t})^{-1}\right]\kappa^k_{N,t}(1-\kappa^k_{N,t})^{-1}\right\}\|\theta_1 - \theta_2\| \,.
\end{aligned}$$

最后，请注意，根据 **A 4.1(iv)**，$\|\delta_x\mathbb{P}_{\theta_1,t}H - \delta_x\mathbb{P}_{\theta_2,t}H\| \leq \|L^P_2\|_{\infty}\|\theta_1 - \theta_2\|$。此外，请注意，根据定理 5，对于所有 $\theta \in \Theta$, $\eta_{0:t,\theta}\mathbb{P}_{\theta,t}H = \eta_{0:t,\theta}s_{0:t,\theta} = \nabla V(\theta)$，我们得到。因此，根据**A4.1(i)**，我们得到 $\|\eta_{0:t,\theta_1}\mathbb{P}_{\theta_1,t}H - \eta_{0:t,\theta_2}\mathbb{P}_{\theta_2,t}H\| \leq L^V\|\theta_1 - \theta_2\|$，证明得证。 □

假设 B.6 成立的证明。假设 B.6只是定理 1 给出的滚动式 PPG 估计量均方误差的一个界。

假设 B.7 成立的证明。

命题 12。 对所有 $\theta \in \Theta$ 以及所有 $\ell \in [\![1,\ t-1]\!]$

$$\mathbb{E}\left[\|\widehat{H}_{\theta}\| \mid \mathcal{F}_{\ell}\right] \leq 2\|s_{0:t,\theta}\|_{\infty} + \sigma_{bias}(1-\kappa^k_{N,t})^{-1} \,.$$

证明。 注意，对所有 $x \in \mathsf{E}^{k-k_0}_t$以及所有 $\theta \in \Theta$,

$$\widehat{H}_{\theta}(x) = H(x) - h(\theta) + \mathbb{P}_{\theta,t}\widehat{H}_{\theta}(x) \,. \tag{B.76}$$

Le引理 8 表明 $\|\mathbb{P}_{\theta,\ t}\widehat{H}_{\theta}\|_{\infty} \leq \sigma_{bias}(1-\kappa^k_{N,\ t})^{-1}$。注意，$h(\theta) \leq \|s_{0:t,\ \theta}\|_{\infty}$ 我们记 e

$$\mathbb{E}\left[\|H\| \mid \mathcal{F}_{\ell}\right] \leq \frac{1}{(k-k_0+1)N}\sum_{i=k_0}^{k}\sum_{j=1}^{N}\mathbb{E}\left[\|\beta^j_{t,\ell}[i]\| \mid \mathcal{F}_{\ell}\right] \,.$$

根据命题 14，$\mathbb{E}\left[\|\beta^j_{t,\ell}[i]\| \mid \mathcal{F}_{\ell}\right] \leq \|s_{0:t,\ \theta}\|_{\infty}$，证明得证。

□

A条件 **B.7** 可直接由命题 12 得出，并通过考虑 $\sup_{\theta\in\Theta}\|s_{0:t,\ \theta}\|_{\infty}$ 得到。

B.2.2 定理 2 的证明

我们已在附录 B.2.1 中证明，在**A** 4.1 和条件 B.8 下，可以应用定理 8。要完成定理 2 的证明，我们只需重新整理常数。我们首先改写定理 9 中的常数

$$L^{\mathbb{P}\widehat{H}} = C_1 + \sigma_{bias}(1-\kappa_{N,t})^{-1}(1-\kappa_{N,t}^k)^{-1}C_2,$$

with

$$\begin{aligned} C_1 &= \left\|L_2^P\right\|_\infty \left[1+\kappa_{N,t}^k(1-\kappa_{N,t}^k)^{-1}\right] + L^V \\ C_2 &= \left\|L_1^P\right\|_\infty (1-\kappa_{N,t}^k)^{-1} + L^\eta \kappa_{N,t}^k \ . \end{aligned}$$

由 (B.70) 和引理 8,

$$\begin{aligned} C_\gamma &= \sigma_{mse} L^{\mathbb{P}\widehat{H}} + (1+\sigma_{mse}) L^V L_0^{\mathbb{P}\widehat{H}} \\ &= \sigma_{mse}\left[C_1 + \sigma_{bias}(1-\kappa_{N,t})^{-1}(1-\kappa_{N,t}^k)^{-1}C_2\right] + (1+\sigma_{mse})L^V\sigma_{bias}(1-\kappa_{N,t}^k)^{-1} \\ &= \sigma_{mse}C_1 + \sigma_{mse}\sigma_{bias}(1-\kappa_{N,t}^k)^{-1}\left[L^V + (1-\kappa_{N,t})^{-1}C_2\right] + \sigma_{bias}L^V(1-\kappa_{N,t}^k)^{-1} \ . \end{aligned}$$

因此,

$$\begin{aligned} C_{0,\gamma} &\coloneqq \sigma_{mse}^2 L^V + C_\gamma \\ &= \sigma_{mse}^2 L^V + \sigma_{mse}C_1 + \sigma_{mse}\sigma_{bias}(1-\kappa_{N,t}^k)^{-1}\left[L^V + (1-\kappa_{N,t})^{-1}C_2\right] + \sigma_{bias}L^V(1-\kappa_{N,t}^k)^{-1} \ . \end{aligned}$$

同理，我们可以将 (B.71) 改写为

$$\begin{aligned} C_h &= L^{\mathbb{P}\widehat{H}}\left[(a+1)/2 + a\sigma_{mse}\right] + (L^V + a' + 1)L_0^{\mathbb{P}\widehat{H}} \\ &= \left[C_1 + \sigma_{bias}(1-\kappa_{N,t})^{-1}(1-\kappa_{N,t}^k)^{-1}C_2\right]\left[(a+1)/2 + a\sigma_{mse}\right] + (L^V + a' + 1)\sigma_{bias}(1-\kappa_{N,t}^k)^{-1} \ . \end{aligned}$$

~定理 2 中的常数 C_0 为 $L^H = 2\sup_{\theta\in\Theta}\|s_{0:t,\theta}\|_\infty + \sigma_{bias}(1-\kappa_{N,t}^k)^{-1}$ 由此完成证明。

B.3 验证 A4.1 的模型条件

在我们将得分上升应用于具体问题时，采用以下假设。

假设 B.9（利普希茨）。(i) 对所有 $t\in\mathbb{N}$，存在 $L_t^s \in \mathrm{M}(\mathrm{X}^{t:t+1})$，使得对于所有 $(x_t, x_{t+1}) \in \mathrm{X}^{t:t+1}$，函数 $\theta \mapsto s_{t,\theta}(x_t, x_{t+1})$ 是 $L_t^s(x_t, x_{t+1})$-利普希茨函数，且 $\mathrm{X}_{t:t+1} \ni (x_t, x_{t+1}) \mapsto s_{t,\theta}(x_t, x_{t+1})$ 被 $\|s_t(\theta)\|_\infty$ 所界，对所有 $\theta\in\Theta$ 均成立。此外，$\|L_k^s\|_\infty < \infty$。(ii) 对所有 $t\in\mathbb{N}$，存在 $L_t^q \in \mathrm{X}^{t:t+1}$，使得 $\|L_t^q\|_\infty < \infty$，并且对于所有 $(x_t, x_{t+1}) \in \mathrm{X}^{t:t+1}$，$\theta \mapsto q_{t,\theta}(x_t, x_{t+1})$ 是 $L_t^q(x_t, x_{t+1})$-利普希茨函数。

引理 9 (**A**条件 B.2(i) 成立)。 假设***A*** 条件 ***B.8*** 和***A 4.1*** 成立。存在常数 L^V，使李雅普诺夫函数 V 对所有 $(\theta_1, \theta_2) \in \Theta^2$ 均满足,

$$\|\nabla V(\theta_1) - \nabla V(\theta_2)\| \le L^V\|\theta_1 - \theta_2\|.$$

证明。 对所有 θ_1, θ_2,

$$\begin{aligned} \|\nabla V(\theta_1) - \nabla V(\theta_2)\| &= \|\eta_{0:t,\theta_1}(s_{0:t,\theta_1}) - \eta_{0:t,\theta_2}(s_{0:t,\theta_2})\| \\ &\le \|\eta_{0:t,\theta_1}(s_{0:t,\theta_1}) - \eta_{0:t,\theta_1}(s_{0:t,\theta_2})\| + \|\eta_{0:t,\theta_1}(s_{0:t,\theta_2}) - \eta_{0:t,\theta_2}(s_{0:t,\theta_2})\| \ . \end{aligned}$$

根据 (3.1) 以及 [Gloaguen 等，2022，定理 4.10] ，存在一个常数 c ，使得

$$\|\eta_{0:t,\theta_1}(s_{0:t,\theta_2}) - \eta_{0:t,\theta_2}(s_{0:t,\theta_2})\| \le ct\|\theta_1 - \theta_2\| \sup_\theta \sup_k \|s_k(\theta)\|_\infty \ ,$$

利用 **A** 3.1 和 **A**4.1[i],，我们可以写成:

$$\begin{aligned}\|\eta_{0:t,\theta_1}(s_{0:t,\theta_1}) - \eta_{0:t,\theta_1}(s_{0:t,\theta_2})\| &\leq \sum_{u=0}^{t-1} \eta_{0:t,\theta_1}\left[\|s_{u,\theta_1}(x_{u:u+1}) - s_{u,\theta_2}(x_{u:u+1})\|\right], \\ &\leq \sum_{u=0}^{t-1} \eta_{0:t,\theta_1}\left[L_u^s(x_{u:u+1})\right] \|\theta_1 - \theta_2\|, \\ &\leq \frac{\sigma_+}{\sigma_-} \sup_{u\in[\![0,t-1]\!]}\left[L_u^s\right] \|\theta_1 - \theta_2\| t.\end{aligned}$$

□

定理 10 （带反向采样的粒子Gibbs 的 Lipschitz 连续性）。 假设 ***AB.9***。对于每个 $t \in \mathbb{N}$， $\theta \in \Theta$ 和 $N \in \mathbb{N}^*$

$$\sup_{x_{0:t}\in \mathsf{X}_{0:t}} \|K_{\theta_1,t}(x_{0:t},.) - K_{\theta_2,t}(x_{0:t},.)\|_{\mathrm{TV}} \leq L_{t,N}^K \|\theta_1 - \theta_2\|,$$

其中

$$L_{t,N}^K \coloneqq \sum_{\ell=0}^{t-1} \bar{\tau}_\ell^{-1}\left[\bar{\sigma}_\ell^{-1} + (N-1)\right] \|L_\ell^q\|_\infty . \tag{B.77}$$

证明。 我们知道 $K_{\theta,\ t} = \mathbb{C}_{m,\ \theta}\mathbb{B}_{t,\ \theta}$。因此，根据引理14、16和19，我们有 $K_{\theta,\ t}$ 是 Lipschitz 的，其常数等于 $L_t^{\mathbb{C}} + \sup_\theta \mathbb{C}_{t,\ \theta} L_t^{\mathbb{B}}$。 □

推论1 （**A**4.1(iii)成立。）。 假设 ***A*** *B.9*。对于每个满足 $N > 1 + 5\rho_t^2 t/2$ 的 $t \in \mathbb{N}$、 $\theta \in \Theta$、 $r \in \mathbb{N}^*$ 和 $N \in \mathbb{N}^*$

$$\sup_{x_{0:t}\in \mathsf{X}_{0:t}} \left\|K_{\theta_1,t}^r(x_{0:t},.) - K_{\theta_2,t}^r(x_{0:t},.)\right\|_{\mathrm{TV}} \leq L_{t,N}^P \|\theta_1 - \theta_2\|$$

其中

$$L_{t,N}^P \coloneqq (1-\kappa_{t,N})^{-1} \|L_{t,N}^K\|_\infty \tag{B.78}$$

其中 $L_{t,\ N}^K$ 定义见（B.77）。

证明。 在条件 B.8 下，带反向抽样的 ParticleGibbs 是几何遍历的，其收缩率为 $\kappa_{t,\ N}$，因此 $L_{t,\ N}^K$ 有界，结论由引理18得出 □

推论2 (**A**4.1(i))。 假设 ***A*** 条件 *B.8* 和 ***A*** 条件 *B.9*。对于所有 $t \in \mathbb{N}^*$, $(\theta_0, \theta_1) \in \Theta^2$,

$$\|\eta_{0:t,\theta_0} - \eta_{0:t,\theta_1}\|_{\mathrm{TV}} \leq L^\eta \|\theta_0 - \theta_1\|,$$

其中

$$L^\eta \coloneqq L_{t,N^*}^P , \tag{B.79}$$

且 $L_{t,\ N}^P$在 (B.78) 中定义，且 $N^* = \lceil 1 + 5\rho_t^2/2 \rceil$.

证明。 考虑 以下分解对所有 $k \in \mathbb{N}^*$ 和 $N \geq 1+5\rho_t^2/2$, 以及所有 $x_{0:t} \in \mathsf{X}_{0:t}$均成立 ,

$$\begin{aligned}\|\eta_{0:t,\theta_1} - \eta_{0:t,\theta_2}\|_{\mathrm{TV}} &\leq \left\|\eta_{0:t,\theta_1} - K_{\theta_1,t}^k(x_{0:t},\cdot)\right\|_{\mathrm{TV}} + \left\|\eta_{0:t,\theta_2} - K_{\theta_2,t}^k(x_{0:t},\cdot)\right\|_{\mathrm{TV}} + \left\|K_{\theta_1,t}^k(x_{0:t},\cdot) - K_{\theta_2,t}^k(x_{0:t},\cdot)\right\|_{\mathrm{TV}} \\ &\leq \left\|\eta_{0:t,\theta_1} - K_{\theta_1,t}^k(x_{0:t},\cdot)\right\|_{\mathrm{TV}} + \left\|\eta_{0:t,\theta_2} - K_{\theta_2,t}^k(x_{0:t},\cdot)\right\|_{\mathrm{TV}} + L_{t,N}^P \|\theta_1 - \theta_2\| ,\end{aligned}$$

这里应用了推论1。由于 $K_{\theta,\ t}$ 的 Lipschitz 常数与 k无关，且对所有 θ, $K_{\theta,\ t}$ 都是几何遍历的，因此在固定 N 的情况下，令 k 趋于无穷，可得,

$$\|\eta_{0:t,\theta_1} - \eta_{0:t,\theta_2}\|_{\mathrm{TV}} \leq \frac{\|L_{t,N}^K\|_\infty}{1-\kappa_{t,N}} \|\theta_1 - \theta_2\| ,$$

对于所有 $N \geq 1 + 5\rho_t^2/2$, 其中对 N 的依赖隐藏在 $L_{t,\ N}^P$ 中。选择 $N = \lceil 1 + 5\rho_t^2/2 \rceil$ 即可得出该结果。 □

注 2. 正如 [Lindholm and Lindsten 2018]，所指出的，推论 1 中出现的 Lipschitz 常数对 $N-1$ 存在出人意料的依赖。人们原本会认为这并不成立，因为我们知道 $\mathbb{K}_{\theta,\,t}$ 以几何速度且一致地快速收敛到 $\eta_{0:t}$，并且 N 越大，收敛速度越快。因此，对于较大的 N，Lipschitz 常数预计会收敛到 $\eta_{0:t}$ 的 Lipschitz 常数，而后者与 N 无关。

命题 13 （$\theta \mapsto \mathbb{K}_{\theta,t}\mu(\boldsymbol{\beta}_t)(\mathrm{id})$ 的利普希茨连续性）**.** 假设 ***AB.9***。对于每个 $t \in \mathbb{N}$, $\theta \in \Theta$和 $N \in \mathbb{N}^*$,

$$\left\|\mathbb{K}_{\theta_1,t}\mu(\boldsymbol{\beta}_t)(\mathrm{id}) - \mathbb{K}_{\theta_2,t}\mu(\boldsymbol{\beta}_t)(\mathrm{id})\right\|_\infty \leq L_t^{\mathbb{K}}\|\theta_1 - \theta_2\| \,,$$

其中

$$L_t^{\mathbb{K}} \coloneqq (N-1)\sum_{\ell=0}^{t-1}\bar{\tau}_\ell\|L_\ell^q\|_\infty + \sum_{j=1}^{m}\|L_j^{\overleftarrow{Q}}\|_\infty\left[\sum_{\ell=0}^{m-1}s_\ell^\infty\right] + \sum_{j=1}^{m}\|L_j^s\|_\infty \,. \tag{B.80}$$

证明。考虑 $e = (x_{0:t}, \boldsymbol{y}_{0:t}) \in \mathsf{E}_t$ 和 $f_\theta(e) \coloneqq \int \mathbb{S}_{m,\theta}(x_{0:t}, \mathrm{d}\tilde{\boldsymbol{y}}_t)\mu(\boldsymbol{b}_t)(id)$。那么 $\mathbb{K}_{\theta,t}\mu(\boldsymbol{b}_t)(id) = \mathbb{C}_{m,\theta}f_\theta(x_{0:t})$ 是马尔可夫核与利普希茨函数的复合，因此是利普希茨的。 □

推论 3 （**A4.1(iv)** 成立。**.** 假设 ***A B.9***。对于每个 $t \in \mathbb{N}$, $\theta \in \Theta$ 和 $N \in \mathbb{N}^*$

$$\sup_{x_{0:t}\in\mathsf{X}_{0:t}}\left\|\mathbb{P}_{\theta_1,t}H - \mathbb{P}_{\theta_2,t}H\right\| \leq L_2^P\|\theta_1 - \theta_2\| \,,$$

其中

$$L_2^P = L_{t,N}^P + L_t^{\mathbb{K}} \,, \tag{B.81}$$

其中 L^P 和 $L_t^{\mathbb{K}}$ 定义于 （B.80）和 （B.78）中。

证明。令 $\tilde{f}\colon \mathsf{E}^{k-k_0} \ni (x_{0:t}[k_0\colon k],\ \boldsymbol{x}_{0:t|t}[k_0\colon k],\ \boldsymbol{b}_t[k_0\colon k]) \mapsto (k-k_0)^{-1}\sum_{\ell=k_0+1}^{k}\mu(\boldsymbol{b}_t[\ell])(\mathrm{id})$。由于 $\mathbb{K}_{\theta,\,t}$仅依赖于路径，为了稍微简化记号，我们可以定义 $f_\theta(x_{0:t}) \coloneqq \mathbb{K}_{\theta\,t}^{\otimes k-k_0}(\tilde{f})(x_{0:t})$。根据命题 13，$f_\theta$ 是利普希茨连续的，常数为 $L^f = L_t^{\mathbb{K}}$。注意，$\mathbb{P}_{\theta,\,t}H(x_{0:t},\ \boldsymbol{y}_t) = K_{\theta,\,t}^{k_0}f_\theta(x_{0:t})$，因此，根据引理 19，其为利普希茨连续的，常数为 $L^P + L_t^{\mathbb{K}}$。 □

C. Lipschitz 性质

C.1 $\mathbb{P}_\theta$的利普希茨连续性

本节将证明以下结论：

- $\mathbb{C}_{m,\,\theta}(z_{0:\,m},\ \cdot)$ 是利普希茨连续的，见附录 C.1.1
- $\mathbb{B}_{m,\,\theta}(\boldsymbol{x}_{0:\,m},\ \cdot)$ 是利普希茨连续的，见附录 C.1.2
- $\int \mathbb{S}_{m,\,\theta}(\boldsymbol{x}_{0:\,m},\ \mathrm{d}\boldsymbol{b}_m)\mu(\boldsymbol{b}_m)(\mathrm{Id})$ 是利普希茨连续的，见附录 C.1.3

以下技术引理将会用到。

引理 10。设 $\alpha \in]0,1]$, $x \in \mathbb{R}_{\geq 0}$ 和 $\ell \in \mathbb{N}$。则对所有满足$\alpha \geq \prod_{i=0}^{\ell}(1-\lambda_i x)$ 的 $\lambda_i \in \mathbb{R}_{\geq 0}$, $i \in [\![0,\ell]\!]$, 均有 $\alpha \geq 1 - x\sum_{i=0}^{\ell}\lambda_i$。

证明。首先考虑对所有 $i \in [\![0,\ell]\!]$, $x\lambda_i \leq 1$ 的情形。我们通过归纳法证明该结果。$\ell = 0$ 的情形显而易见。现在假设该结果对某个 $r \in [\![0,\ell-1]\!]$ 成立。那么，

$$\begin{aligned}\prod_{i=0}^{r+1}(1-\lambda_i x) = (1-\lambda_{r+1}x)\prod_{i=0}^{r}(1-\lambda_i x) &\geq (1-\lambda_{r+1}x)(1-x\sum_{i=0}^{r}\lambda_i) \\ &= 1 - x\sum_{i=0}^{r+1}\lambda_i + x^2\sum_{i=0}^{r}\lambda_i\lambda_{r+1} \geq 1 - x\sum_{i=0}^{r+1}\lambda_i \,.\end{aligned}$$

现在考虑存在一个指标 $j \in [\![0, \ell]\!]$ 使得 $x\lambda_j \geq 1$ 的情形。那么 $\alpha \geq 0 \geq 1 - (\sum_{i=0}^{\ell} \lambda_i)\ x$。

□

我们先介绍一些重要定义。设 P 和 Q 是某个共同可测空间 $(\mathbb{X}, \mathcal{X})$ 上的概率分布，并假设这些分布相对于某个共同参考测度 λ存在密度 p 和 q。令 $\mathbb{M}[P, Q]$ 表示 P与 Q之间的最大耦合。按照 [Lindholm and Lindsten, 2018, 定理 2], ，可以显式构造这样一个最大耦合:

$$\mathbb{M}[P,Q](\mathrm{d}(x,y)) \coloneqq \min\{p(x), g(x)\}\lambda(\mathrm{d}x)\delta_x(\mathrm{d}y) + \frac{\big[P(\mathrm{d}x) - \min\{p(x), g(x)\}\lambda(\mathrm{d}x)\big]\big[Q(\mathrm{d}y) - \min\{p(y), g(y)\}\lambda(\mathrm{d}y)\big]}{1 - \lambda\big(\min\{p, q\}\big)}. \quad \text{(C.82)}$$

由此定义可知，对于连续型和离散型支配测度 λ ，

$$\int \mathbb{1}_{\{x=y\}} \mathbb{M}[P,Q]\,\mathrm{d}(x,y) = \int \min\{p(x), g(x)\}\lambda(\mathrm{d}x)\,.$$

此外，对于两个马尔可夫转移核 K_1 和 K_2 ，它们定义在 $(\mathbb{X}, \mathcal{X})$ 上，并假设相对于某个共同支配测度存在转移密度；对于 $(x_1, x_2) \in \mathbb{X}^2$,$\mathbb{M}[K_1, K_2]((x_1, x_2), \cdot)$，我们将其定义为测度 $K_1(x_1, \cdot)$ 与 $K_2(x_2, \cdot)$ 之间的最大耦合。如此定义的 $\mathbb{M}[K_1, K_2]$ 在乘积空间 $(\mathbb{X}^2, \mathcal{X}^{\otimes 2})$ 上定义了一个马尔可夫转移核。

以下引理将在后续内容中发挥关键作用。

引理 11. *(i)* 设(μ_1, μ_2) 为两个概率测度，它们相对于同一个支配测度具有密度；设(K_1, K_2) 为两个马尔可夫转移核，它们也相对于某个支配测度具有转移密度。那么，该概率测度

$$\mathbb{M}[\mu_1,\mu_2]\,\mathbb{M}[K_1,K_2]\,(\mathrm{d}(x_1,x_2)) = \int \mathbb{M}[\mu_1,\mu_2]\,(\mathrm{d}(z_1,z_2))\mathbb{M}[K_1,K_2]\,((z_1,z_2),\mathrm{d}(x_1,x_2)),$$

是$(\mu_1 K_1, \mu_2 K_2)$的一个耦合，并且有

$$\int \mathbb{1}_{x_1=x_2}\mathbb{M}[\mu_1K_1,\mu_2K_2]\,(\mathrm{d}(x_1,x_2)) \geq \int\int \mathbb{1}_{z_1=z_2}\mathbb{1}_{x_1=x_2}\mathbb{M}[\mu_1,\mu_2]\,(\mathrm{d}(z_1,z_2))\mathbb{M}[K_1,K_2]\,((z_1,z_2),\mathrm{d}(x_1,x_2)).$$

(ii) 设$(\mu_1, \cdots, \mu_n)$ 和$(\nu_1, \cdots, \nu_n)$ 为概率测度，且对所有 $i \in [\![1、 n]\!]、$ μ_i 和 ν_i，它们都相对于同一个支配测度具有密度。那么 $\bigotimes_{i=1}^n \mathbb{M}[\mu_i, \nu_i]$ 是 $\bigotimes_{i=1}^n \mu_i$ 和 $\bigotimes_{i=1}^n \nu_i$ 的一个耦合，因此

$$\int \prod_{i=1}^n \mathbb{1}_{x_i=y_i}\mathbb{M}\left[\bigotimes_{i=1}^n \mu_i, \bigotimes_{i=1}^n \nu_i\right](\mathrm{d}(x_1,\ldots,x_n,y_1,\ldots,y_n)) \geq \int \prod_{i=1}^n \mathbb{1}_{x_i=y_i} \bigotimes_{i=1}^n \mathbb{M}[\mu_i,\nu_i]\,(\mathrm{d}(x_1,\ldots,x_n,y_1,\ldots,y_n)).$$

证明。 只需证明$\mathbb{M}[\mu_1, \mu_2]\,\mathbb{M}[K_1, K_2]$ 以 $\mu_1 K_1$和 $\mu_2 K_2$ 为边缘分布。这直接来自以下事实：$\mathbb{M}[\mu_1, \mu_1]$和 $\mathbb{M}[K_1, K_2]$ 具有正确的边缘分布

分布；确实，

$$\begin{aligned}
&\mathbb{M}\left[\mu_1,\mu_2\right]\mathbb{M}\left[K_1,K_2\right](\mathsf{X}\times A)\\
&\qquad=\int\mathbb{M}\left[\mu_1,\mu_2\right](\mathrm{d}z_1,\mathrm{d}_2)\,\mathbb{M}\left[K_1,K_2\right](z_1,z_2,\mathrm{d}(x_1,x_2))\,\mathbb{1}_{\mathsf{X}\times A}(x_1,x_2)\mathbb{1}_{\mathsf{X}^2}(z_1,z_2)\\
&\qquad=\int\mathbb{M}\left[\mu_1,\mu_2\right](\mathrm{d}z_1,\mathrm{d}_2)K_2(z_2,A)\\
&\qquad=\int\mu_2(\mathrm{d}z_2)K_2(z_2,A)\\
&\qquad=\mu_2K_2(A).
\end{aligned}$$

第一个边缘分布的推导类似。对于第二点，由于 $\mathbb{M}\left[\mu_1,\mu_2\right]\mathbb{M}\left[K_1,K_2\right]$是 (μ_1K_1,μ_2K_2) 的一个耦合，而 $\mathbb{M}\left[\mu_1K_1,\mu_2K_2\right]$ 是最大耦合，因此我们有

$$\begin{aligned}
&\int\mathbb{1}_{x_1=x_2}\mathbb{M}\left[\mu_1K_1,\mu_2K_2\right](\mathrm{d}(x_1,x_2))\\
&\qquad\geq\iint\mathbb{1}_{x_1=x_2}\mathbb{M}\left[\mu_1,\mu_2\right](\mathrm{d}(z_1,z_2))\,\mathbb{M}\left[K_1,K_2\right](z_1,z_2;\mathrm{d}(x_1,x_2))\\
&\qquad\geq\iint\mathbb{1}_{x_1=x_2}\mathbb{1}_{z_1=z_2}\mathbb{M}\left[\mu_1,\mu_2\right](\mathrm{d}(z_1,z_2))\,\mathbb{M}\left[K_1,K_2\right](z_1,z_2;\mathrm{d}(x_1,x_2)).
\end{aligned}$$

第二项的证明类似。 □

C.1.1 $\theta\mapsto\mathbb{C}_{m,\theta}$ **是利普希茨连续的。**

我们采用一种受 [Lindholm and Lindsten, 2018, Theorem 2] 启发的耦合方法。我们考虑的耦合方式是：粒子滤波器的选择 步骤和变异 步骤分别采用最大耦合。

算法 6 耦合$\mathbb{C}_{m,\theta}$

数据： θ_1, θ_2, $\zeta_{0:m}$结果： $\boldsymbol{x}_{0:m}$,1,$\boldsymbol{x}_{0:m}$,1 23 抽取 $\boldsymbol{x}_0$,1,$\boldsymbol{x}_{0,2}\sim\mathrm{M}$ $[\boldsymbol{\eta}_0\langle\zeta_0\rangle,\boldsymbol{\eta}_0\langle\zeta_0\rangle]$24 对于 $s\leftarrow 1$ 到 t， 执行 25 抽取 $(\boldsymbol{x}_s,1,\boldsymbol{x}_s,2)\sim\mathrm{M}$ $[\boldsymbol{M}_{s-1,\theta_1}\langle\zeta_s\rangle(\boldsymbol{x}_{s-1},1,\cdot)\,\boldsymbol{M}_{s-1,\theta_2}\langle\zeta_s\rangle(\boldsymbol{x}_{s-1},2,\cdot)]$

└

首先，我们来证明一步选择–变异核是利普希茨的。

引理 12. 对于所有 $t\in\mathbb{N}$, $\boldsymbol{x}_{t-1}\in\mathbf{X}_{t-1}$ 以及 $(\theta_1,\theta_2)\in\Theta^2$,

$$\int\mathbb{1}_{\{x_1=x_2\}}\mathbb{M}\left[\Phi_{t-1,\theta_1}(\mu(\boldsymbol{x}_{t-1})),\Phi_{t-1,\theta_2}(\mu(\boldsymbol{x}_{t-1}))\right](\mathrm{d}(x_1,x_2))\geq 1-\frac{\sum_{i=1}^N\lambda_t\left(L_{t-1}^q(x_{t-1}^i,\cdot)\right)}{N\bar{\tau}_n}\|\theta_1-\theta_2\|. \tag{C.83}$$

证明。 由 **A3.1(i)** 和 **A4.1(iii)**,

$$
\begin{aligned}
&\int \mathbb{1}_{\{x_1=x_2\}}\mathbb{M}\left[\Phi_{t-1,\theta_1}(\mu(\boldsymbol{x}_{t-1})),\Phi_{t-1,\theta_2}(\mu(\boldsymbol{x}_{t-1}))\right](\mathrm{d}(x_1,x_2))\\
&\quad=\int \min\left(\sum_{i=1}^{N}\frac{q_{t-1,\theta_1}(x_{t-1}^i,x)}{\sum_{j=1}^{N} g_{t-1,\theta_1}(x_{t-1}^j)},\sum_{i=1}^{N}\frac{q_{t-1,\theta_2}(x_{t-1}^i,x)}{\sum_{j=1}^{N} g_{t-1,\theta_2}(x_{t-1}^j)}\right)\lambda_t(\mathrm{d}x)\\
&\quad\geq\sum_{j=1}^{N}\int \min\left(\frac{q_{t-1,\theta_1}(x_{t-1}^i,x)}{\sum_{j=1}^{N} g_{t-1,\theta_1}(x_{t-1}^j)},\frac{q_{t-1,\theta_2}(x_{t-1}^i,x)}{\sum_{j=1}^{N} g_{t-1,\theta_2}(x_{t-1}^j)}\right)\lambda_t(\mathrm{d}x)\\
&\quad\geq\frac{1}{\sum_{j=1}^{N}\max\left(g_{t-1,\theta_1}(x_{t-1}^j),g_{t-1,\theta_2}(x_{t-1}^j)\right)}\sum_{j=1}^{N}\int \min\left(q_{t-1,\theta_1}(x_{t-1}^j,x),q_{t-1,\theta_2}(x_{t-1}^j,x)\right)\lambda_t(\mathrm{d}x)\\
&\quad\geq\frac{\sum_{j=1}^{N}\max\left(g_{t-1,\theta_1}(x_{t-1}^j),g_{t-1,\theta_2}(x_{t-1}^j)\right)-\sum_{i=1}^{N}\lambda_t\left(L_{t-1}^{q}(x_{t-1}^i,\cdot)\right)\|\theta_1-\theta_2\|}{\sum_{j=1}^{N}\max\left(g_{t-1,\theta_1}(x_{t-1}^j),g_{t-1,\theta_2}(x_{t-1}^j)\right)}\\
&\quad\geq 1-\frac{\sum_{i=1}^{N}\lambda_t\big(L_{t-1}^{q}(x_{t-1}^i,\cdot)\big)}{N\bar{\tau}_n}\|\theta_1-\theta_2\|,
\end{aligned}
$$

这里我们利用了

$$
\begin{aligned}
\int \max(q_{t-1,\theta_1}(x_{t-1}^i,x),q_{t-1,\theta_2}(x_{t-1}^i,x))\lambda_t(\mathrm{d}x)&\geq\max\left(\int q_{t-1,\theta_1}(x_{t-1}^i,x)\lambda_t(\mathrm{d}x),\int q_{t-1,\theta_2}(x_{t-1}^i,x)\lambda_t(\mathrm{d}x)\right)\\
&\geq\max(g_{t-1,\theta_1}(x_{t-1}^i),g_{t-1,\theta_2}(x_{t-1}^i)).
\end{aligned}
$$

□

引理13。 对于所有 $t\in\mathbb{N}$, $\boldsymbol{x}_{t-1}\in\mathbf{X}_{t-1}$, $z\in\mathsf{X}_t$ 以及 $(\theta_1,\theta_2)\in\Theta^2$,

$$
\|\boldsymbol{M}_{t-1,\theta_1}\langle z\rangle(\boldsymbol{x}_{t-1},\cdot)-\boldsymbol{M}_{t-1,\theta_2}\langle z\rangle(\boldsymbol{x}_{t-1},\cdot)\|_{\mathrm{TV}}\leq L_{t-1}^{\boldsymbol{M}}(\boldsymbol{x}_{t-1})\|\theta_1-\theta_2\|
$$

其中 $L_{t-1}^{\boldsymbol{M}}(\boldsymbol{x}_{t-1})=(1-N^{-1})\bar{\tau}_{t-1}^{-1}\sum_{i=1}^{N}\lambda_t\left(L_{t-1}^{q}(x_{t-1}^i,\cdot)\right)$。

证明。 记 $\mathrm{U}[\![1,n]\!]$ 为 $[\![1,n]\!]$上的均匀分布。根据核$\boldsymbol{M}_{t-1,\theta}\langle z\rangle$的定义，我们有

$$
\boldsymbol{M}_{t-1,\theta}\langle z\rangle(\boldsymbol{x}_{t-1},\mathrm{d}\boldsymbol{x}_t)=\int \mathrm{U}[\![1,n]\!](\mathrm{d}j)\left\{\Phi_{t-1}(\mu(\boldsymbol{x}_{t-1}))^{\otimes j}\otimes\delta_z\otimes\Phi_{t-1}(\mu(\boldsymbol{x}_{t-1}))^{\otimes(N-j-1)}\right\}(\mathrm{d}\boldsymbol{x}_t)
$$

因此，将引理 11 的两项结论与以下事实结合起来：$\mathbb{M}[\mu,\mu]\left(\mathrm{d}(x_1,x_2)\right)=$

对于任意概率测度 μ，$(\mathrm{d}x_1)\delta_{x_1}(\mathrm{d}x_2)$，我们得到

$$\begin{aligned}
&\int \mathbb{1}_{\{\boldsymbol{x}_{t,1}=\boldsymbol{x}_{t,2}\}}\mathbb{M}\left[\boldsymbol{M}_{t-1,\theta_1}\langle z\rangle(\boldsymbol{x}_{t-1},\cdot),\boldsymbol{M}_{t-1,\theta_2}\langle z\rangle(\boldsymbol{x}_{t-1},\cdot)\right]\mathrm{d}(\boldsymbol{x}_{t,1},\boldsymbol{x}_{t,2})\\
&\qquad\geq\int \mathbb{1}_{\boldsymbol{x}_{t,1}=\boldsymbol{x}_{t,2},i_1=i_2}\mathbb{M}\left[\mathrm{U}[\![1,n]\!],\mathrm{U}[\![1,n]\!]\right]\left(\mathrm{d}(i_1,i_2)\right)\\
&\qquad\qquad\times\mathbb{M}\left[\Phi_{t-1,\theta_1}(\mu(\boldsymbol{x}_{t-1})),\Phi_{t-1,\theta_2}(\mu(\boldsymbol{x}_{t-1}))\right]^{\otimes i_1}\otimes\mathbb{M}\left[\delta_z,\delta_z\right]\\
&\qquad\qquad\otimes\mathbb{M}\left[\Phi_{t-1,\theta_1}(\mu(\boldsymbol{x}_{t-1})),\Phi_{t-1,\theta_2}(\mu(\boldsymbol{x}_{t-1}))\right]^{\otimes N-i_1-1}\mathrm{d}(\boldsymbol{x}_{t,1},\boldsymbol{x}_{t,2})\\
&\qquad=\frac{1}{N}\sum_{i=1}^{N}\int\prod_{k=1,k\neq i}^{n}\mathbb{1}_{x^i_{t,1}=x^i_{t,2}}\mathbb{M}\left[\Phi_{t-1,\theta_1}(\mu(\boldsymbol{x}_{t-1})),\Phi_{t-1,\theta_2}(\mu(\boldsymbol{x}_{t-1}))\right]\left(\mathrm{d}(x^i_{t,1},x^i_{t,2})\right)\\
&\qquad\geq\left(1-\frac{\sum_{i=1}^{N}\lambda_t\big(L^q_{t-1}(x^i_{t-1},\cdot)\big)}{N\bar{\tau}_{t-1}}\|\theta_1-\theta_2\|\right)^{N-1}\\
&\qquad\geq 1-\frac{N-1}{\bar{\tau}_{t-1}N}\sum_{i=1}^{N}\lambda_t\big(L^q_{t-1}(x^i_{t-1},\cdot)\big)\|\theta_1-\theta_2\|\,.
\end{aligned}$$

wh这里我们在倒数第二行应用了引理 12，在最后一行应用了引理 10。 □

引理14。 对于每个 $t\in\mathbb{N}^*$，存在 $L^{\mathbb{C}}_t\in\mathsf{M}(\mathcal{X}_{0:t})$ 使得

$$\left\|\mathbb{C}_{t,\theta_1}(z_{0:t})-\mathbb{C}_{t,\theta_2}(z_{0:t})\right\|_{\mathrm{TV}}\leq L^{\mathbb{C}}_t(z_{0:t})\|\theta_1-\theta_2\|\,, \tag{C.84}$$

其中 $L^{\mathbb{C}}_t(z_{0:t})=\sup_\theta\mathbb{C}_{t,\theta}\left[\sum_{i=0}^{t-1}L^{\boldsymbol{M}}_i\right](z_{0:t})$。在 ***AB.9(i)*** 条件下，我们得到 $\|L^{\mathbb{C}}_t\|_\infty\leq(N-1)\sum_{\ell=0}^{t-1}\bar{\tau}_\ell\|L^q_\ell\|_\infty$。

证明。 这直接应用了引理 20。

□

C.1.2 $\theta\mapsto\mathbb{B}_{t,\theta}(\boldsymbol{x}_{0:t},\cdot)$ 是利普希茨连续的

我们先回顾一下 $\mathbb{B}_m$的定义

$$\mathbb{B}_{t,\theta}:\mathsf{X}_{0:t}\times\mathcal{X}_{0:t}\ni(\boldsymbol{x}_{0:t},A)\mapsto\int\cdots\int\mathbb{1}_A(x_{0:t})\left(\prod_{s=0}^{t-1}\overleftarrow{Q}_{s,\mu(\boldsymbol{x}_s)}(x_{s+1},\mathrm{d}x_s)\right)\mu(\boldsymbol{x}_t)(\mathrm{d}x_t)\,. \tag{C.85}$$

引理 15。 对于所有 $s\in[\![0,t]\!]$，$x_{t+1}\in\mathsf{X}_{t+1}$，$\boldsymbol{x}_t\in\mathsf{X}_t$ 以及 $(\theta_1,\theta_2)\in\Theta^2$

$$\left\|\overleftarrow{Q}_{s,\mu(\boldsymbol{x}_s),\theta_1}(x_{s+1},\cdot)-\overleftarrow{Q}_{s,\mu(\boldsymbol{x}_s),\theta_2}(x_{s+1},\cdot)\right\|_{\mathrm{TV}}\leq L^{\overleftarrow{Q}}_s(x_{s+1},\boldsymbol{x}_s)\|\theta_1-\theta_2\|\,. \tag{C.86}$$

与 $L^{\overleftarrow{Q}}_s(x_{s+1},\boldsymbol{x}_s)=(N\bar{\tau}_t\bar{\sigma}_s)^{-1}\sum_{i=1}^{N}L^q_s(x^i_s,x_{s+1})$。根据 ***AB.9(i)***，我们有 $\|L^{\overleftarrow{Q}}_m\|_\infty=(\bar{\tau}_m\bar{\sigma}_m)^{-1}\|L^q_m\|_\infty$。

证明。 注意到 $\overleftarrow{Q}_{t,\mu(\boldsymbol{x}_t)}(x_{t+1},\cdot)=\sum_{\ell=1}^{N}\frac{q_t(x^\ell_t,x_{t+1})}{\sum_{\ell'=1}^{N}q_t(x^{\ell'}_t,x_{t+1})}\delta_{x^\ell_t}$。因此，与引理 12 的证明类似，

$$\begin{aligned}
&\int\mathbb{1}_{\{x_{t,1}=x_{t,2}\}}\mathbb{M}\left[\overleftarrow{Q}_{t,\mu(\boldsymbol{x}_t),\theta_1}(x_{t+1},\cdot),\overleftarrow{Q}_{t,\mu(\boldsymbol{x}_t),\theta_2}(x_{t+1},\cdot)\right]\mathrm{d}(x_{t,1},x_{t,2})\\
&\qquad\geq\frac{\sum_{\ell=1}^{N}\max(q_{t,\theta_1}(x^\ell_t,x_{t+1}),q_{t,\theta_2}(x^\ell_t,x_{t+1}))-L^q_t(x^\ell_t,x_{t+1})\|\theta_1-\theta_2\|}{\sum_{\ell=1}^{N}\max(q_{t,\theta_1}(x^\ell_t,x_{t+1}),q_{t,\theta_2}(x^\ell_t,x_{t+1}))}\\
&\qquad\geq 1-\frac{\sum_{\ell=1}^{N}L^q_t(x^\ell_t,x_{t+1})}{N\bar{\tau}_t\bar{\sigma}_t}\|\theta_1-\theta_2\|\,.
\end{aligned}$$

□

引理 16。 对于所有 $t \in \mathbb{N}$, $\boldsymbol{x}_{0:t} \in \mathbf{X}_{0:t}$ 以及 $(\theta_1, \theta_2) \in \Theta^2$

$$\|\mathbb{B}_{t,\theta_1}(\boldsymbol{x}_{0:t}, \cdot) - \mathbb{B}_{t,\theta_2}(\boldsymbol{x}_{0:t}, \cdot)\|_{\mathrm{TV}} \leq L_t^{\mathbb{B}}(\boldsymbol{x}_{0:t})\|\theta_1 - \theta_2\| \tag{C.87}$$

其中 $L_t^{\mathbb{B}}(\boldsymbol{x}_{0:t}) = \text{上确界}_\theta \, \mathbb{B}_t \left[\sum_{i=0}^{t-1} L_i^{\overleftarrow{Q}}\right](\boldsymbol{x}_{0:t})$。根据 ***AB.9(i)***，我们有 $\|L_t^{\mathbb{B}}\|_\infty = \sum_{i=0}^{t-1} (\bar{\tau}_i \bar{\sigma}_i)^{-1} \|L_i^q\|_\infty$。

证明。应用引理 19 和引理 15。 □

C.1.3 $\theta \mapsto \int \mathbb{S}_{t,\theta}(\boldsymbol{x}_{0:t}, \mathrm{d}\boldsymbol{b}_t)\mu(\boldsymbol{b}_t)(\mathrm{id})$ 是利普希茨连续的

定义后向祖先核

$$\mathcal{B}_{\theta,t} : \mathsf{X}_{t+1} \times \mathbf{X}_t \times \sigma([\![1, N]\!]) \mapsto \int \mathbb{1}_A(\tilde{j}) \left(\sum_{\ell=1}^{N} \frac{q_t(x_t^\ell, x_{t+1})}{\sum_{\ell'=1}^{N} q_t(x_t^{\ell'}, x_{t+1})} \delta_\ell(\mathrm{d}\tilde{j}) \right) .$$

引理 17. *($\mathcal{B}_{\theta,t}$ 是利普希茨连续的) 对于任意 $m \in [\![0, t]\!]$，存在 $L_m^{BK} \in \mathsf{M}(\boldsymbol{\mathcal{X}}_{m:m+1})$ 使得 t*

$$\|\mathcal{B}_{\theta_1,m}(x_{m+1}, \boldsymbol{x}_m) - \mathcal{B}_{\theta_2,m}(x_{m+1}, \boldsymbol{x}_m)\|_{\mathrm{TV}} \leq L_m^{\overleftarrow{Q}}(x_{m+1}, \boldsymbol{x}_m)\|\theta_1 - \theta_2\| , \tag{C.88}$$

其中 $L_s^{\overleftarrow{Q}}$在引理 *15* 中定义

证明。 $\mathcal{B}_{\theta,\ s}$ 是核 (C.85) 的指标版本，因此具有相同的利普希茨常数。**命题 14**。 对于任意 $m \in [\![0,\ t]\!]$，我们有 □

$$\left| \int \mathbb{C}_m \mathbb{S}_{m,\theta}(z_{0:m}, \mathrm{d}\boldsymbol{b}_m)\mu(\boldsymbol{b}_m)(\mathrm{Id}) \right| \leq \sum_{\ell=0}^{m-1} s_\ell^\infty \tag{C.89}$$

and

$$\left| \int \mathbb{S}_{m,\theta_1}(\boldsymbol{x}_{0:m}, \mathrm{d}\boldsymbol{b}_m)\mu(\boldsymbol{b}_m)(\mathrm{Id}) - \int \mathbb{S}_{m,\theta_2}(\boldsymbol{x}_{0:m}, \mathrm{d}\boldsymbol{b}_m)\mu(\boldsymbol{b}_m)(\mathrm{Id}) \right| \leq L_m^{\mathbb{S}\mu}(\boldsymbol{x}_{0:m})\|\theta_1 - \theta_2\| . \tag{C.90}$$

其中 $L_m^{\mathbb{S}\mu}(\boldsymbol{x}_{0:m}) = N^{-1} \sum_{i=1}^{N} L_m^B(x_m^k, \boldsymbol{x}_{0:m})$，并且 L_m^B 递归定义为

$$L_{m+1}^B(x_{m+1}^k, \boldsymbol{x}_{0:m}) = L_m^{\overleftarrow{Q}}(x_{m+1}^k, \boldsymbol{x}_m) \sum_{\ell=0}^{m} s_\ell^\infty + \int \mathcal{B}_{\theta,m}(x_{m+1}^k, \boldsymbol{x}_m, \mathrm{dJ}) \left\{ L_m^s(x_m^{\mathrm{J}}, x_{m+1}^k) + L_m^B(x_m^{\mathrm{J}}, \boldsymbol{x}_{0:m-1}) \right\} .$$

特别地，在 ***AB.9*** 条件下，我们有 $L_m^B \leq \sum_{j=1}^{m} \|L_j^{\overleftarrow{Q}}\|_\infty \left[\sum_{\ell=0}^{m-1} s_\ell^\infty\right] + \sum_{j=1}^{m} \|L_j^s\|_\infty$ (C.91)

In .

证明。 考虑以下核,

$$\widetilde{\mathbb{S}}_{m,\theta}(\boldsymbol{x}_{0:m+1}, \mathrm{d}(\mathrm{J}_0^{i,j}, \ldots, \mathrm{J}_m^{i,j})_{i=1,j=1}^{N,M}) \coloneqq \prod_{\ell=0}^{m} \prod_{k=1}^{N} \widetilde{\boldsymbol{S}}_{\ell,\theta}(x_{\ell+1}^k, \boldsymbol{x}_\ell, \mathrm{d}(\mathrm{J}_\ell^{k,j})_{j=1}^{M}) , \tag{C.92}$$

$$\widetilde{\boldsymbol{S}}_{\ell,\theta}(x_{\ell+1}^k, \boldsymbol{x}_\ell, \mathrm{d}(\mathrm{J}_\ell^{k,j})_{j=1}^{M}) \coloneqq \prod_{j=1}^{M} \mathcal{B}_{\theta,\ell}(x_{\ell+1}^k, \boldsymbol{x}_\ell, \mathrm{dJ}_\ell^{k,j}) . \tag{C.93}$$

对所有 $k \in [1$ 定义： $N]$, $m \in \mathbb{N}_{>0}$,

$$B_{m+1,k} : \theta \mapsto \int \widetilde{\mathbb{S}}_{m,\theta}(\boldsymbol{x}_{0:m+1}, \mathrm{d}(\mathrm{J}_0^{i,j}, \ldots, \mathrm{J}_m^{i,j})_{i=1,j=1}^{N,M}) b_{m+1}^k(\boldsymbol{x}_{0:m+1}, (\mathrm{J}_0^{i,j}, \ldots, \mathrm{J}_m^{i,j})_{i=1,j=1}^{N,M}) ,$$

其中 $b^k_{m+1}\big(\boldsymbol{x}_{0:m+1}, \big(\mathrm{J}^{i,j}_0, \ldots, \mathrm{J}^{i,j}_m\big)^{N,M}_{i=1,j=1}\big)$ 递归定义为

$$b^k_{m+1}\big(\boldsymbol{x}_{0:m+1}, \big(\mathrm{J}^{i,j}_0, \ldots, \mathrm{J}^{i,j}_m\big)^{N,M}_{i=1,j=1}\big) = M^{-1}\sum_{\ell=1}^{M} b^{\mathrm{J}^{k,\ell}_m}_m\big(\boldsymbol{x}_{0:m}, \big(\mathrm{J}^{i,j}_0, \ldots, \mathrm{J}^{i,j}_{m-1}\big)^{N,M}_{i=1,j=1}\big) + s_{m,\theta}(x^{\mathrm{J}^{k,\ell}_m}_m, x^k_{m+1}).$$

为方便记号，以下省略参数，直接写作 b^k_{m+1}。

下面证明$B_{m+1,\,k}$是 Lipschitz 的，其 Lipschitz 常数为 $L^B_m(x^k_{m+1},\ \boldsymbol{x}_m)$，且有界于 $\sum_{\ell=0}^{m-1} s^\infty_\ell$。对于 $m>2$ 和 $k\in[1:\ N]$,

$$\begin{aligned}
B_{m+1,k}(\theta) &= \int \widetilde{\mathbb{S}}_{m,\theta}(\boldsymbol{x}_{0:m+1}, \mathrm{d}(\mathrm{J}^{i,j}_0, \ldots, \mathrm{J}^{i,j}_m)^{N,M}_{i=1,j=1}) b^k_{m+1} \\
&= \int\cdots\int \widetilde{\mathbb{S}}_{m-1,\theta}(\boldsymbol{x}_{0:m}, \mathrm{d}(\mathrm{J}^{i,j}_0, \ldots, \mathrm{J}^{i,j}_{m-1})^{N,M}_{i=1,j=1}) \widetilde{\boldsymbol{S}}_{m,\theta}(x^k_{m+1}, \boldsymbol{x}_m, \mathrm{d}(\mathrm{J}^{k,j}_m)^M_{j=1}) \\
&\qquad\qquad \times \left\{ M^{-1}\sum_{\ell=1}^{M} b^{\mathrm{J}^{k,\ell}_m}_m + s_{m,\theta}(x^{\mathrm{J}^{k,\ell}_m}_m, x^k_{m+1}) \right\} \\
&= \int\cdots\int \widetilde{\boldsymbol{S}}_{m,\theta}(x^k_{m+1}, \boldsymbol{x}_m, \mathrm{d}\{\mathrm{J}^{k,j}_m\}^M_{j=1}) \Bigg[M^{-1}\sum_{\ell=1}^{M} \Bigg\{ s_{m,\theta}(x^{\mathrm{J}^{k,\ell}_m}_m, x^k_{m+1}) \\
&\qquad\qquad + \int \widetilde{\mathbb{S}}_{m-1,\theta}(\boldsymbol{x}_{0:m}, \mathrm{d}(\mathrm{J}^{i,j}_0, \ldots, \mathrm{J}^{i,j}_{m-1})^{N,M}_{i=1,j=1}) b^{\mathrm{J}^{k,\ell}_m}_m \Bigg\} \Bigg] \\
&= \int\cdots\int \widetilde{\boldsymbol{S}}_{m,\theta}(x^k_{m+1}, \boldsymbol{x}_m, \mathrm{d}(\mathrm{J}^{k,j}_m)^M_{j=1}) \left[M^{-1}\sum_{\ell=1}^{M} \Big\{ s_{m,\theta}(x^{\mathrm{J}^{k,\ell}_m}_m, x^k_{m+1}) + B_{m,\mathrm{J}^{k,\ell}_m}(\theta) \Big\} \right] \\
&= \int \mathcal{B}_{\theta,m}(x^k_{m+1}, \boldsymbol{x}_m, \mathrm{dJ}) \Big\{ s_{m,\theta}(x^{\mathrm{J}}_m, x^k_{m+1}) + B_{m,\mathrm{J}}(\theta) \Big\}
\end{aligned}$$

根据归纳假设，在给定 $\mathrm{J}^{k,\ell}_m$、$B_{m,\,\mathrm{J}^{k,\ell}_m}$的条件下，$\mathrm{J}^{k,\ell}_m$是 Lipschitz 的，其 Lipschitz 常数为 $L^B_m(x^{\mathrm{J}^{k,\ell}_m}_m,\ \boldsymbol{x}_{0:m-1})$，因此 $B_{m+1,\,k}$ 的 Lipschitz 常数为

$$L^B_{m+1}(x^k_{m+1}, \boldsymbol{x}_{0:m}) = \overleftarrow{Q}^m_m(x^k_{m+1}, \boldsymbol{x}_m)\sum_{\ell=0}^{m} s^\infty_\ell + \int \mathcal{B}_{\theta,m}(x^k_{m+1}, \boldsymbol{x}_m, \mathrm{dJ}) \Big\{ L^s_m(x^{\mathrm{J}}_m, x^k_{m+1}) + L^B_m(x^{\mathrm{J}}_m, \boldsymbol{x}_{0:m-1}) \Big\}\ . \tag{C.94}$$

这里用到了$\mathcal{B}_{\theta,\,m}$ 和 $s_{m,\,\theta}$ 也都是 Lipschitz 这一事实。再次根据归纳法，$B_{m+1,\,k}$ 一致有界于 $\sum_{\ell=0}^{m} s^\infty_\ell$。
注意到在基例中，$m=0$,$\beta^k_m=0$ 对所有 $k\in\mathbb{N}$ 均成立，因此结论成立，归纳过程至此完成。

现在只需验证，对于所有 $\theta\in\Theta$、$m\in[\![0,t]\!]$ 和 $k\in[1:\ N]$,

$$B_{m,k}(\theta) = \int \mathbb{S}_m(\boldsymbol{x}_{0:m}, \mathrm{d}\boldsymbol{b}_m) b^k_m\ .$$

同样，我们采用归纳法。

$$
\begin{aligned}
&\int \mathbb{S}_m(\boldsymbol{x}_{0:m}, \mathrm{d}\boldsymbol{b}_m) b_m^k \\
&= \int \cdots \int \mathbb{S}_{m-1}(\boldsymbol{x}_{0:m-1}, \mathrm{d}\boldsymbol{b}_{m-1}) \boldsymbol{S}_m(\boldsymbol{b}_{m-1}, \boldsymbol{x}_{m-1:m}, \mathrm{d}\boldsymbol{b}_m) b_m^k \\
&= \int \cdots \int \mathbb{S}_{m-1}(\boldsymbol{x}_{0:m-1}, \mathrm{d}\boldsymbol{b}_{m-1}) \\
&\qquad \times \prod_{j=1}^{M} \left(\sum_{p=1}^{N} \frac{q_{m-1}(x_{m-1}^p, x_m^k)}{\sum_{\ell=1}^{N} q_{m-1}(x_{m-1}^\ell, x_m^k)} \delta_{x_{m-1}^p, b_{m-1}^p}\left(\mathrm{d}(\tilde{x}_{m-1}^{k,j}, \tilde{b}_{m-1}^{k,j})\right) \right) \\
&\qquad \times \left[M^{-1} \sum_{n=1}^{M} \left\{ \tilde{b}_{m-1}^{k,n} + s_{m,\theta}(\tilde{x}_{m-1}^{k,n}, x_m^k) \right\} \right] \\
&= \int \cdots \int \mathbb{S}_{m-1}(\boldsymbol{x}_{0:m-1}, \mathrm{d}\boldsymbol{b}_{m-1}) \\
&\qquad \times \prod_{j=1}^{M} \left(\sum_{p=1}^{N} \frac{q_{m-1}(x_{m-1}^p, x_m^k)}{\sum_{\ell=1}^{N} q_{m-1}(x_{m-1}^\ell, x_m^k)} \delta_p(\mathrm{d}\mathrm{J}_{m-1}^{k,j}) \right) \left[M^{-1} \sum_{n=1}^{M} \left\{ b_{m-1}^{\mathrm{J}_{m-1}^{k,n}} + s_{m,\theta}(x_{m-1}^{\mathrm{J}_{m-1}^{k,n}}, x_m^k) \right\} \right] \\
&= \int \cdots \int \widetilde{\boldsymbol{S}}_{m,\theta}(x_{m-1}^k, \boldsymbol{x}_{\ell-1}, \mathrm{d}(\mathrm{J}_{\ell-1}^{k,j})_{j=1}^{M}) \\
&\quad \times \left[M^{-1} \sum_{\ell=1}^{M} \left\{ s_{m,\theta}(x_{m-1}^{\mathrm{J}_{m-1}^{k,\ell}}, x_m^k) + \mathbb{S}_{m-1}(\boldsymbol{x}_{0:m-1}, \mathrm{d}\boldsymbol{b}_{m-1}) b_{m-1}^{\mathrm{J}_{m-1}^{k,\ell}} \right\} \right] \\
&= \int \cdots \int \widetilde{\boldsymbol{S}}_{m,\theta}(x_{m-1}^k, \boldsymbol{x}_{\ell-1}, \mathrm{d}(\mathrm{J}_{\ell-1}^{k,j})_{j=1}^{M}) \\
&\quad \times \left[M^{-1} \sum_{\ell=1}^{M} \left\{ s_{m,\theta}(x_{m-1}^{\mathrm{J}_{m-1}^{k,\ell}}, x_m^k) + \int \mathbb{S}_{m-1}(\boldsymbol{x}_{0:m-1}, \mathrm{d}\boldsymbol{b}_{m-1}) b_{m-1}^{\mathrm{J}_{m-1}^{k,\ell}} \right\} \right] \\
&= \int \cdots \int \widetilde{\boldsymbol{S}}_{m,\theta}(x_{m-1}^k, \boldsymbol{x}_{\ell-1}, \mathrm{d}(\mathrm{J}_{\ell-1}^{k,j})_{j=1}^{M}) \left[M^{-1} \sum_{\ell=1}^{M} \left\{ s_{m,\theta}(x_{m-1}^{\mathrm{J}_{m-1}^{k,\ell}}, x_m^k) + B_{m-1,\mathrm{J}_{m-1}^{k,\ell}}(\theta) \right\} \right] \\
&= B_{m,k}(\theta)
\end{aligned}
$$

证明最后指出

$$
\int \mathbb{S}_m(\boldsymbol{x}_{0:m}, \mathrm{d}\boldsymbol{b}_m) \mu(\boldsymbol{b}_m)(\mathrm{Id}) = N^{-1} \sum_{k=1}^{N} B_{m,k}(\theta)
$$

因此，它是 Lipschitz 的，常数为 $L_m^{\mathbb{S}\mu}(\boldsymbol{x}_{0:m}) = N^{-1} \sum_{i=1}^{N} L_m^B(x_m^k, \boldsymbol{x}_{m-1})$。 □

C.2 马尔可夫核的 Lipschitz 性质

引理18 （遍历利普希茨核的复合是利普希茨的）。 设 P_θ 是定义在$X \times \mathcal{Y}$ 上的马尔可夫核，对于任意 θ，它都是一致的 π-几何遍历的，收缩常数为 ρ，且与θ 无关；并且存在 $L_p > 0$，使得对于每个 $x \in$X

$$
\|P_{\theta_0}(x, \cdot) - P_{\theta_1}(x, \cdot)\|_{\mathrm{TV}} \leq L_P \|\theta_0 - \theta_1\|.
$$

那么，对于所有 $k > 0$

$$
\left\|P_{\theta_0}^k(x, \cdot) - P_{\theta_1}^k(x, \cdot)\right\|_{\mathrm{TV}} \leq \frac{L_P}{1-\rho} \|\theta_0 - \theta_1\|.
$$

证明。我们采用借鉴自 [Fort 等人 2011] 的以下分解。对于任意 $k \geq 1$,

$$P_{\theta_0}^k f - P_{\theta_1}^k f = \sum_{j=0}^{k-1} P_{\theta_0}^j (P_{\theta_0} - P_{\theta_1}) \big(P_{\theta_1}^{k-j-1} f - \pi f\big).$$

那么，对于任意满足 $\|f\|_\infty \leq 1$ 且 $x \in \mathsf{X}$ 的 f，

$$\begin{aligned}
|P_{\theta_0}^k f(x) - P_{\theta_1}^k f(x)| &\leq \sum_{j=0}^{k-1} \left| \int P_{\theta_0}^j (x, \mathrm{d}y) \sup_{z \in \mathsf{X}} |P_{\theta_1}^{k-j-1} f(z) - \pi f| \right| L_P \|\theta_0 - \theta_1\| \\
&\leq L_P \left(\sum_{j=0}^{k-1} \rho^{k-j-1} \right) \|\theta_0 - \theta_1\| \\
&\leq \frac{L_P}{1-\rho} \|\theta_0 - \theta_1\|.
\end{aligned}$$

□

引理 19（Lipschitz 核的复合仍为利普希茨）。设 P_θ、Q_θ 是定义在 $X \times \mathcal{Y}$ 和 $Y \times \mathcal{Z}$ 上的两个核，使得对于每个 $x \in X$、$y \in Y$，都存在 $L_p \in \mathsf{M}(X)$、$L_q \in \mathsf{M}(Y)$，满足

$$\|P_{\theta_0}(x, \cdot) - P_{\theta_1}(x, \cdot)\|_{\mathrm{TV}} \leq L_p(x) \|\theta_0 - \theta_1\|$$

and

$$\|Q_{\theta_0}(y, \cdot) - Q_{\theta_1}(y, \cdot)\|_{\mathrm{TV}} \leq L_q(y) \|\theta_0 - \theta_1\| .$$

Then

$$\|P_{\theta_0} Q_{\theta_0}(x, \cdot) - P_{\theta_1} Q_{\theta_1}(x, \cdot)\|_{\mathrm{TV}} \leq L_{pq} \|\theta_0 - \theta_1\| ,$$

其中 $L_{pq} = (\sup_\theta P_\theta L_q(x) + L_p(x) \sup_y \sup_\theta Q_\theta(y, Z))$。

证明。令 $f \in \mathsf{M}$，使得 $\|f\|_\infty \leq 1$。

$$\begin{aligned}
\|P_{\theta_1} Q_{\theta_1} f - P_{\theta_2} Q_{\theta_2} f\| &\leq \|P_{\theta_1} [Q_{\theta_1} f - Q_{\theta_2} f]\| + \|(P_{\theta_1} - P_{\theta_2}) Q_{\theta_2} f\| \\
&\leq (P_{\theta_1} L_q(x) + L_p(x) \|Q_{\theta_2} f\|_\infty) \|\theta_1 - \theta_2\| .
\end{aligned}$$

□

推论 4。设 P_θ、Q_θ 是定义在 $X \times \mathcal{Y}$ 和 $Y \times \mathcal{Z}$ 上的两个马尔可夫核，使得对于每个$x \in X$、$y \in Y$，都存在 $L_p \in \mathsf{M}(X)$、$L_q \in \mathsf{M}(Y)$，满足

$$\|P_{\theta_0}(x, \cdot) - P_{\theta_1}(x, \cdot)\|_{\mathrm{TV}} \leq L_p(x) \|\theta_0 - \theta_1\|$$

and

$$\|Q_{\theta_0}(y, \cdot) - Q_{\theta_1}(y, \cdot)\|_{\mathrm{TV}} \leq L_q(y) \|\theta_0 - \theta_1\| .$$

Then

$$\|P_{\theta_0} Q_{\theta_0}(x, \cdot) - P_{\theta_1} Q_{\theta_1}(x, \cdot)\|_{\mathrm{TV}} \leq L_{pq} \|\theta_0 - \theta_1\| ,$$

其中 $L_{pq} = (\sup_\theta P_\theta L_q(x) + L_p(x))$。

引理 20（Lipschitz 核的乘积仍为 Lipschitz）。设 P_θ、Q_θ 是两个常数分别为 L_P、L_Q 的一致 *Lipschitz* 马尔可夫核。那么 $P_\theta \otimes Q_\theta$ 是常数为 $L_P + L_Q$ 的一致 *Lipschitz* 核。

证明。令 $h_\theta \colon y \mapsto \int Q_\theta(y, \mathrm{d}z) f(y, z)$。则 $(P_{\theta_i} \otimes Q_{\theta_i})(f) = P_{\theta_i}(h_{\theta_i})$，证明与前一个引理类似，因为 h_θ 是利普希茨连续的，其常数为 L_Q 和 $\|h_\theta\|_\infty \leq 1$。

□

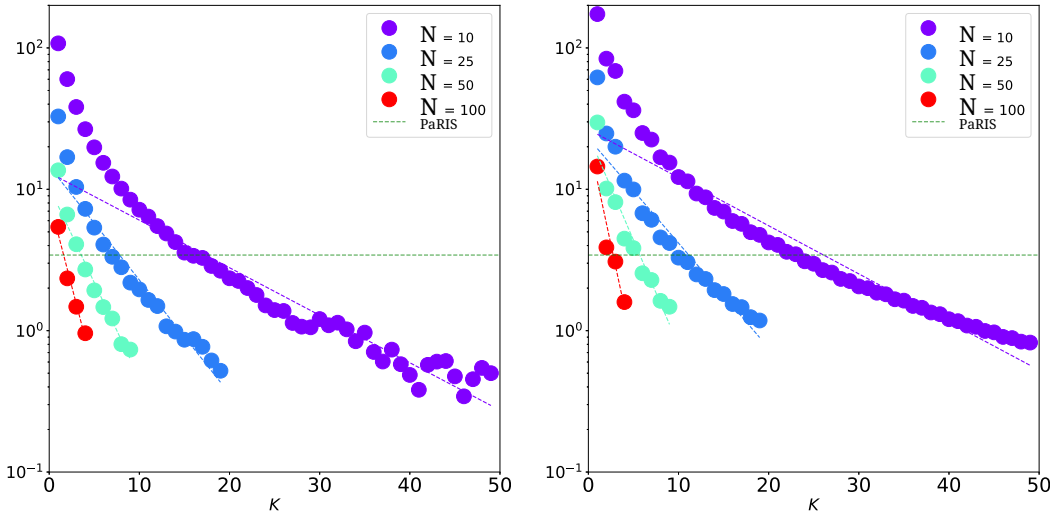

图 3：LGSSM 的 PPG 展开估计器输出。曲线描述了随着 k 增加，在不同粒子样本数 N 下偏差的变化。左图和右图分别对应 $k_0 = k - 1$ 和 $k_0 = \lfloor k/2 \rfloor$。

D. 补充数值结果 D.1 PPG D.2 学习

在两个实验中，所有参数均通过从均值为零、协方差矩阵为 $0.01I$ 的多元高斯分布中采样进行初始化。我们使用了 ADAM 优化器[Kingma 和 Ba， 2014]，其学习率衰减为 $1/\sqrt{\ell}$，其中 ℓ 是迭代索引，初始学习率为 0。2。我们将梯度按 T 进行缩放。

线性高斯状态空间模型对于线性高斯状态空间模型，我们在粒子数固定为 ($N = 64$)、吉布斯迭代次数固定为 ($k = 8$) 的情况下，评估了预热阶段 (k_0) 对最终所得结果与最大似然估计量之间距离的影响。表3表明， k_0 较小的配置表现更好。对此现象的一种可能解释是，由于条件路径会在两次梯度上升迭代之间传递，从某一时刻起，这条条件路径会使估计值的偏差减小，因此，为了降低偏差而令 k_0 较大的重要性便不复存在；但通过 k_0 增大方差的影响仍然存在，因为拥有一个来自正确边缘分布的条件粒子并不会影响估计量的方差，只会影响其偏差。

参考文献

[Anderson 和 Qiu， 1997] Anderson, G. D. 和 Qiu, S.-L. (1997)。伽马函数的一个单调性性质。美国数学会会刊，125(11):3355–3362。

[Andrieu 和 Doucet， 2003] Andrieu, C. 和 Doucet, A. (2003). 用于一般状态空间模型参数估计的在线期望最大化类算法。载于*IEEE*国际声学、语音与信号处理会议论文集，第6卷，第69–72页。

表3：线性高斯状态空间模型案例中每种配置与 θ_{MLE} 的距离。

算法	N	k_0	k	D_{mle}
PPG	64	0	8	0.205 ± 0.013
PPG	64	1	8	0.213 ± 0.016
PPG	64	2	8	0.201 ± 0.010
PPG	64	3	8	0.201 ± 0.010
PPG	64	4	8	0.207 ± 0.012
PPG	64	5	8	0.212 ± 0.015
PPG	64	6	8	0.210 ± 0.017
PPG	64	7	8	0.211 ± 0.018

[Andrieu 等，2010a] Andrieu, C., Doucet, A. 和 Holenstein, R. (2010a). 粒子马尔可夫链蒙特卡洛方法。《皇家统计学会杂志：B辑》，72(3):269–342。[Andrieu 等，2010b] Andrieu, C., Doucet, A. 和 Holenstein, R. (2010b). 粒子马尔可夫链蒙特卡洛方法（附讨论）。《皇家统计学会杂志：B辑》，72:269–342。[Andrieu 等， 2018] Andrieu, C., Lee, A. 和 Vihola, M. (2018). 迭代条件序贯蒙特卡洛的一致遍历性与粒子 Gibbs 采样器的几何遍历性。Bernoulli，24(2):842–872。[Capp´e， 2001] Capp´e, O. (2001). 使用粒子近似递归计算隐马尔可夫过程的平滑泛函。《蒙特卡洛方法及应用》，7(1–2):81–92。[Capp´e， 2011] Capp´e, O. (2011). 隐马尔可夫模型的在线期望最大化算法。《计算与图形统计学杂志》，20(3):728–749。[Capp´e 等人， 2007] Capp´e, O., Godsill, S. J. 和 Moulines, E. (2007). 序贯蒙特卡洛现有方法与最新进展综述。IEEE会刊，95(5):899–924。[Capp´e 和 Moulines， 2005] Capp´e, O. 和 Moulines, E. (2005). 粒子滤波在最大似然参数估计中的应用。载于欧洲信号处理会议（EUSIPCO），土耳其安塔利亚。[Capp´e 等人， 2005] Capp´e, O., Moulines, E. 和 Ryd´en, T. (2005). 《隐马尔可夫模型中的推断》。施普林格。[Capp´e 等人， 2009] Capp´e, O., Moulines, E. 和 Ryd´en, T. (2009). 隐马尔可夫模型中的推断。载于EUSFLAT会议论文集，第14–16页。[Chopin 和 Papaspiliopoulos， 2020] Chopin, N. 和 Papaspiliopoulos, O. (2020). 《序贯蒙特卡洛导论》。施普林格。[Chopin 和 Singh，2015a] Chopin, N. 和 Singh, S. S. (2015a). 关于粒子吉布斯采样。Bernoulli，21(3):1855–1883。[Chopin 和 Singh，2015b] Chopin, N. 和 Singh, S. S. (2015b). 关于粒子吉布斯采样。Bernoulli，21(3):1855–1883。[Del Moral， 2004] Del Moral, P. (2004). 《费曼-卡茨公式：谱系与相互作用粒子系统及其应用》。施普林格。[Del Moral， 2013] Del Moral, P. (2013). 《用于蒙特卡洛积分的平均场模拟》。CRC出版社。[Del Moral 等人， 2010] Del Moral, P., Doucet, A. 和 Singh, S. S. (2010). 费曼-卡茨公式的反向解释。ESAIM：数学建模与数值分析，44:947–975。

[Del Moral 和 Jasra， 2018] Del Moral, P. 和 Jasra, A. (2018). Feynman–Kac 粒子模型的一阶精细分析，第二部分：粒子吉布斯采样器。《随机过程及其应用》，128(1):354– 371.[Del Moral 等人， 2016] Del Moral, P., Kohn, R. 和 Patras, F. (2016). 关于粒子吉布斯采样器。《H. Poincaré 概率与统计研究所年刊》，52(4):1687–1733.[Douc 等人， 2011] Douc, R., Garivier, A., Moulines, E. 和 Olsson, J. (2011). 一般状态空间隐马尔可夫模型的序贯蒙特卡洛平滑。《应用概率年鉴》，21(6):1201–2145.[Douc 和 Moulines， 2008] Douc, R. 和 Moulines, E. (2008). 加权样本的极限定理及其在序贯蒙特卡洛方法中的应用。《统计年鉴》，36(5):2344–2376.[Douc 等人， 2018] Douc, R., Moulines, E., Priouret, P. 和 Soulier, P. (2018). 马尔可夫链。施普林格。[Douc 等人， 2014] Douc, R., Moulines, E. 和 Stoffer, D. (2014). 非线性时间序列：理论、方法及 R 示例应用。CRC出版社。[Dubarry 和 Le Corff， 2013] Dubarry, C. 和 Le Corff, S. (2013). 非线性状态空间模型中平滑加性泛函的非渐近偏差不等式。Bernoulli，19(5B):2222–2249.[Fort 等人， 2011] Fort, G., Moulines, E. 和 Priouret, P. (2011). 自适应和交互式马尔可夫链蒙特卡洛算法的收敛。《统计年鉴》，39(6).[Gloaguen 等人， 2022] Gloaguen, P., Le Corff, S. 和 Olsson, J. (2022). 伪边际序贯蒙特卡洛在线平滑算法。Bernoulli，28(4):2606–2633.[Glynn 和 Rhee， 2014] Glynn, P. W. 和 Rhee, C.-H. (2014). 马尔可夫链平衡期望的精确估计。《应用概率杂志》，51(A):377–389.[Godsill 等人， 2004] Godsill, S. J., Doucet, A. 和 West, M. (2004). 非线性时间序列的蒙特卡洛平滑。《美国统计协会杂志》，50:438–449.[Jacob 等人， 2020] Jacob, P. E., O'Leary, J. 和 Atchadé, Y. F. (2020). 通过耦合实现的无偏马尔可夫链蒙特卡洛方法。《皇家统计学会杂志：B辑（统计方法学）》，82(3):543–600.[Karimi 等人， 2019] Karimi, B., Miasojedow, B., Moulines, E. 和 Wai, H.-T. (2019). 有偏随机逼近方案的非渐近分析。载于 Beygelzimer, A. 和 Hsu, D.（编），《第三十二届学习理论会议论文集》，第 99 卷《机器学习研究论文集》，第 1944–1974 页。PMLR。[Kingma 和 Ba， 2014] Kingma, D. P. 和 Ba, J. (2014). Adam：一种随机优化方法。[Lee 等人， 2020] Lee, A., Singh, S. S. 和 Vihola, M. (2020). 耦合条件后向采样粒子滤波器。《统计年鉴》，48(5):3066–3089.[Lindholm 和 Lindsten， 2018] Lindholm, A. 和 Lindsten, F. (2018). 利用粒子随机逼近期望最大化学习动力系统。[Lindsten 等，2014a] Lindsten, F., Jordan, M. I. 和 Schön, T. B. (2014a). 带祖先采样的粒子吉布斯。《机器学习研究杂志》，15(1):2145–2184.[Lindsten 等，2014b] Lindsten, F., Jordan, M. I. 和 Schön, T. B. (2014b). 带祖先采样的粒子吉布斯。《机器学习研究杂志》，15(63):2145–2184.

[Naesseth 等人， 2020] Naesseth, C. A., Lindsten, F. 和 Blei, D. (2020). 马尔可夫得分攀升：使用 kl(p——q) 的变分推断。[Olsson 和 Westerborn， 2017] Olsson, J. 和 Westerborn, J. (2017). 一般隐马尔可夫模型中高效的基于粒子的在线平滑：PaRIS 算法。*Bernoulli*，23(3):1951–1996.[Poyiadjis 等人， 2005] Poyiadjis, G., Doucet, A. 和 Singh, S. S. (2005). 用于最优滤波器导数的粒子方法：在参数估计中的应用。载于*IEEE* 国际声学、语音与信号处理会议论文集，第 v/925–v/928 页。[Poyiadjis 等人， 2011] Poyiadjis, G., Doucet, A. 和 Singh, S. S. (2011). 状态空间模型中得分和观测信息矩阵的粒子近似及其在参数估计中的应用。*Biometrika*，98(1):65–80.[S¨ärkk¨ä， 2013] S¨ärkk¨ä, S. (2013). 贝叶斯滤波与平滑。剑桥大学出版社。[Singh 等人， 2017] Singh, S. S., Lindsten, F. 和 Moulines, E. (2017). 粒子吉布斯采样器的分块策略与稳定性。*Biometrika*，104(4):953–969.[Whiteley， 2010] Whiteley, N. (2010). 关于粒子马尔可夫链蒙特卡洛方法的讨论。第 306–307 页。[Zhao 等人， 2021] Zhao, Y., Nassar, J., Jordan, I., Bugallo, M. 和 Park, I. M. (2021). 流式变分蒙特卡洛。