参数隐私保护数据共享：粒子信念MDP表述 ⋆

Haokun Yu [a]、Jingyuan Zhou [b]、Kaidi Yang [b]

[a] *Institute of Operations Research and Analytics, National University of Singapore*

[b] *Civil and Environmental Engineering, National University of Singapore*

arXiv:2602.04262v1 [eess.SY] 4 Feb 2026

摘要

本文研究连续状态动力系统中的参数隐私保护数据共享问题。在这一问题中，数据所有者设计数据共享策略，以支持下游估计与控制，同时防止对手推断敏感参数。该数据共享问题被表述为一个优化问题：在满足数据可用性约束的前提下，权衡隐私泄露与数据共享对数据所有者效用的影响。我们证明，该问题可等价地表述为信念马尔可夫决策过程（MDP），从而简化最优策略的表示。为有效刻画连续状态与动作空间中的信息论隐私泄露，我们提出粒子信念MDP表述，通过序贯蒙特卡洛跟踪参数后验分布，得到一种易于处理的信念状态近似；随着粒子数量增加，该近似渐近收敛。我们进一步利用高斯混合近似，推导出基于粒子的互信息的一个便于处理的闭式上界，从而能够高效优化粒子信念MDP。在混合自动化车队上的实验表明，所学习到的连续策略在保持数据可用性和系统性能的同时，显著阻碍了针对人类驾驶行为参数的推断攻击。

Key words: 数据隐私；参数隐私；互信息；粒子滤波器。

1 引言

在智能交通和智能建筑等网络化控制系统中，主体之间对数据共享的需求日益增长。例如，Uber 等出行平台发布的轨迹数据已被主管部门广泛用于提升交通效率和安全性 [1,2]。然而，共享运行数据可能会暴露有关系统状态与参数的敏感信息。尤其是，轨迹数据可能泄露用户的出行模式（例如目的地）或运营方的商业机密（例如算法参数）。

现有的隐私保护数据共享机制主要通过密码技术 [3–7] 和差分隐私 [8–12]来保护个体级记录（如状态）。然而，针对控制增益和运行约束等敏感系统参数 [13–16],的保护却鲜少受到关注。这些参数对于出行平台等主体尤为关键，因为一旦这些参数泄露，便可能削弱其商业竞争力，有时甚至会使系统漏洞暴露于对手的利用之下。值得注意的是，差分隐私等个体级技术旨在隐藏群体中的数据记录，因此并不能自然地扩展到参数隐私，因为参数并不存在可供其隐藏其中的群体。

多项研究通过信息论表述探索了保护参数隐私的数据共享方法：有意扰动数据，以限制对手能够推断出的敏感参数信息，并以互信息（MI）或熵等量度衡量泄漏。代表性方法包括：能够维持与私有参数之间统计依赖关系的随机核 [17],；用于掩盖线性高斯系统动态特性的基于高斯的滤波器 [16], ；计算效率更高的随机参数生成方法 [18], ；以及实现最优效用-隐私权衡的动态规划表述 [15]。

尽管取得了显著进展，当前的参数隐私保护机制仍受两项根本性问题的限制。

⋆ 通讯作者 Kaidi Yang。*Email addresses:* yuhaokun@u.nus.edu (Haokun Yu), jingyuanzhou@u.nus.edu (Jingyuan Zhou), kaidi.yang@nus.edu.sg (Kaidi Yang)。

首先，大多数现有机制（例如 [15,19]）仅在状态离散化的低维场景中得到实现，且往往依赖对过去观测的穷举枚举。这类建模在实践中并不适用，因为：（i）速度、交通流量和电压等状态变量通常在连续域中变化；（ii）离散化会导致计算互信息时的复杂度呈指数级增长，使这些方法在更高维度下的计算开销难以承受。其次，许多研究假设数据提供方的运营不受其数据共享机制影响，实际上忽略了使用共享数据所引发的反馈回路 [20]。然而在实践中，发布数据会被其他主体使用，而这些主体的决策会改变数据提供方的运行环境。例如，发布的轨迹数据可能为交通管理部门制定拥堵定价策略提供依据，进而影响出行平台的运营。

贡献概述。本文通过构建一种用于参数隐私保护数据共享的粒子信念框架，解决了上述局限。该框架旨在求解一个优化问题，在以下两者之间进行权衡：（i）参数隐私，以参数与共享数据之间的互信息（MI）衡量；（ii）共享数据对系统效用的影响，该影响通过使用共享数据所引发的反馈回路来刻画。我们的贡献主要有三点。首先，我们推导出一个信念马尔可夫决策过程（MDP），为数据共享优化问题提供了一种简化的信念空间表述，其中最优策略与历史无关。我们从理论上证明了该表述与原始优化问题等价。其次，为克服连续状态空间中信念更新难以处理的问题，我们提出一种粒子信念表述，利用一组粒子近似精确的信念分布。我们证明，随着粒子数量增加，由粒子信念诱导的贝尔曼值将收敛于原始信念MDP的价值函数。第三，为应对连续数据共享策略带来的计算负担，我们设计了一种高效的近似框架，通过高斯混合模型近似为互信息提供一个易于处理的上界。

本文的初步版本曾在第7届年度动力学与控制学习会议 [21]上发表。本期刊版本拓宽了问题设定，对信念马尔可夫决策过程进行了严格分析，通过粒子滤波和高斯分布混合近似将框架扩展至连续数据共享，为粒子信念表述建立了理论保证，并给出了更全面的仿真研究。

2 问题陈述

考虑一个建模为离散时间马尔可夫过程的数据提供方，其状态为 $X_t \in \mathbf{X} \subset \mathbb{R}^N$，控制动作为 $U_t \in \mathbb{R}^M$，时域为 $\mathcal{T} = \{0, 1, \cdots, T\}$，采样间隔为 ΔT。系统演化由转移核 $p(X_{t+1} \mid \Theta,\ X_t,\ W_t)$ 描述，该核结合了系统动力学 $p(X_{t+1} \mid X_t,\ U_t)$ 和控制策略 $p(U_t \mid \Theta,\ X_t,\ W_t)$，其中敏感参数 $\Theta \in \mathcal{S}_\Theta \subset \mathbb{R}^P$被视为时不变机密（例如控制增益），而 W_t 是来自环境的外部输入。为在保护 Θ的同时实现数据共享，数据提供方依据数据共享策略 $\pi_t \in \Pi$发布失真输出 Y_t，而非 X_t。

为清晰起见，我们对外部输入和对手的知识施加以下假设。

假设1（马尔可夫外部输入） 外部输入 W_t 遵循转移为 $p(W_{t+1} \mid W_t, Y_t)$的马尔可夫过程，该过程可能不为数据提供方所知。

假设1允许转移依赖于共享数据 Y_t。这是合理的，因为 Y_t 可能被生态系统中的其他参与方利用（例如，交通管理部门利用出行数据来优化信号灯时序）。

假设2（对手的知识） 假设对手知道*(i)*由 $p(\cdot \mid \cdot)$表示的全部条件概率，*(ii)*敏感参数 Θ 的先验分布$p(\Theta)$ 但不知道其实现，以及*(iii)*任意时刻 t的数据共享策略 π_t 。

我们提出两点说明。首先，(i) 和 (iii) 体现了保守的最坏情况立场，即隐私保障不应依赖于隐瞒条件模型 [22]。放宽这一假设只会增强隐私保护。其次，先验知识 Θ 是合理的，因为对手通常知道 Θ 的可能范围和常用值。

我们设计一种数据共享策略 π_t，使其在任意时刻 t 均满足以下标准。首先，该策略必须保护敏感参数 Θ，防止他人根据随时间发布的数据准确推断其真实值。其次，该策略应保持控制性能，确保失真数据 Y_t不会显著降低系统运行表现。请注意，Y_t 可能通过外部输入 W_t对未来状态 $X_{>t}$ 产生负面影响，从而影响系统性能（见假设 1）。第三，共享数据应保留数据可用性，这要求限制失真数据 Y_t 与真实状态 X_t之间的偏差。

为满足这些标准，数据共享策略的设计可表述为

$$\min_{\boldsymbol{\pi}} \rho\, I^{\boldsymbol{\pi}}(\Theta; Y_{1:T}, W_{1:T}) + \sum_{t=1}^{T} \mathbb{E}^{\boldsymbol{\pi}}[r(\theta^*, X_{t:t+1}, Y_t)], \tag{1a}$$

$$\text{s.t. } \mathbb{E}^{\pi_t}[d(X_t, Y_t)] \le \hat{D}, \quad t = 1, 2, \ldots, T, \tag{1b}$$

其中 $\boldsymbol{\pi} = \{\pi_t\}_{t=1}^T$。(1a) 中的目标函数以权重系数 ρ 将隐私度量与系统成本结合起来。隐私通过 Θ与已发布数据序列和外部输入对（$Y_{1:T}$，$W_{1:T}$）之间的互信息来量化，从而刻画关于 Θ 的信息泄露。系统成本 $r(\theta^*, X_{t:t+1}, Y_t)$ 是关于所有参数的连续函数，用于刻画在真实参数 θ^* 给定的情况下，从状态 X_t 转移到 X_{t+1}时发布Y_t 所产生的影响。约束 (1b) 要求策略 π_t 下的期望失真 $d(X_t, Y_t)$ 不超过规定的上界 $\hat{D}$，其中 $d(\cdot,\cdot)$ 也是连续函数。

我们利用互信息链式法则（见附录A）简化 $I(\Theta; Y_{1:T}, W_{1:T})$ 的计算，得到

$$I^{\boldsymbol{\pi}}(\Theta; Y_{1:T}, W_{1:T}) = \sum_{t=1}^{T} I^{\boldsymbol{\pi}}(\Theta; Y_t \mid Y_{1:t-1}, W_{1:t}). \quad (2)$$

3 信念MDP表述

本节将推导一种与数据共享优化问题(1)等价的信念马尔可夫决策过程表述。在最一般的形式 $\pi_t(Y_t \mid \Theta, X_{1:t}, Y_{1:t-1}, W_{1:t})$ 中，策略 π_t 根据参数 Θ、状态序列 $X_{1:t}$、外部输入序列 $W_{1:t}$ 和共享数据历史 $Y_{1:t-1}$ 生成失真输出 Y_t。尽管这种表示在条件化于截至时间 t 的所有可用信息这一意义上是最优的，但随着 t 增大，由于维度不断增长，采用该策略求解 (1) 在计算上变得不可行。为解决这一问题，我们引入一种等价的简化表示来描述最优策略。

我们首先在定理1中证明，问题 (1) 的最优解可以限制为形式为 $\pi_t^s(Y_t \mid \Theta, X_t, W_{1:t}, Y_{1:t-1}) \in \boldsymbol{\Pi}^s$ 的策略，这些策略不再需要以完整状态历史 $X_{1:t-1}$ 为条件。证明见附录B.1。

定理1（简化策略的最优性）存在 $ist\ of\ simplified\ policies\ \boldsymbol{\pi}^{s,*} = \{\pi_t^{s,*}\}_{t=1} \in \boldsymbol{\Pi}^s$ 的一个序列，达到了问题(1)的最优性。特别地，$\boldsymbol{\pi}^{s,*}$ 满足约束(1b)，且相应的目标值 $L(\boldsymbol{\pi}^{s,*})$与最优策略序列 $\boldsymbol{\pi}^*$的目标值一致，即 $L(\boldsymbol{\pi}^{s,*}) = L(\boldsymbol{\pi}^*)$。

根据定理1，最优策略的搜索可限制在简化类 $\boldsymbol{\Pi}^s$中。因此，我们将(1)重新表述为

$$\min_{\boldsymbol{\pi}^s \in \boldsymbol{\Pi}^s} \sum_{t=1}^{T} \Big(\rho I^{\boldsymbol{\pi}^s}(\Theta; Y_t \mid Y_{1:t-1}, W_{1:t}) + \mathbb{E}^{\boldsymbol{\pi}^s}[r(\theta^*, X_{t:t+1}, Y_t)] + \lambda\Big(\mathbb{E}^{\boldsymbol{\pi}^s}[d(X_t, Y_t)] - \hat{D}\Big)\Big) \quad (3)$$

其中，拉格朗日乘子 λ与约束(1b)相关联。我们采用动态规划求解(3)：对于任意历史 $h_t = (w_{1:t}, y_{1:t-1})$，代价到达函数 $V_t(h_t)$定义为从时间 t 起的最小期望累积代价：

$$V_t^*(h_t) = \min_{\{\pi_k^s\}_{k=t}^T} \Bigg[\sum_{k=t}^{T} \Big(\rho I^{\boldsymbol{\pi}^s}(\Theta; Y_k \mid Y_{1:k-1}, W_{1:k}) + \mathbb{E}^{\boldsymbol{\pi}^s}[r(X_{k:k+1}, Y_k, \theta^*)] + \lambda\Big(\mathbb{E}^{\boldsymbol{\pi}^s}[d(X_k, Y_k)] - \hat{D}\Big)\Big) \,\Big|\, h_t \Bigg] \quad (4)$$

终端条件为 $V_{T+1}^*(\cdot) = 0$。因此，我们可以将贝尔曼最优性方程写为

$$V_t^*(h_t) = \min_{\pi_t^s(Y_t \mid \Theta, X_t, h_t)} [C_t(h_t, \pi_t^s) + \mathbb{E}(V_{t+1}^*(h_{t+1} \mid h_t))], \quad (5)$$

其中 $C_t(h_t, \pi_t^s)$表示策略 π_t^s下的阶段代价。其具体表达式见附录 B.3。

然而，简化策略 $\boldsymbol{\pi}^s$ 仍依赖于历史 h_t，导致计算复杂度随着 T 的增加而迅速增长。为解决这一问题，我们采用信念状态表示法重新表述该问题。具体而言，我们将历史 h_t 映射为信念分布 $\beta_t(\Theta, X_t) = p(\Theta, X_t \mid h_t)$，该分布以当前状态和敏感参数上的概率分布形式，紧凑地概括了所有过去的信息。观测到新信息（即失真数据 y_t 和外部输入 W_t）后，信念状态根据贝叶斯定理进行更新，写作

$$\beta_{t+1}(\Theta, X_{t+1}) \propto \int_{x_t} p(X_{t+1} \mid \Theta, x_t, W_t) a(y_t \mid \Theta, x_t) \cdot \beta_t(\Theta, x_t) dx_t \quad (6)$$

其中，$a(y_t \mid \Theta, x_t) = \pi_t^s(Y_t \mid \Theta, x_t, W_{1:t}, Y_{1:t-1})$表示策略。详细推导见附录 B.2。

由式（6）可知，β_t 仅取决于当前观测 y_t、策略 π_t^s和先前的信念状态 β_{t-1}。基于此，引理 2 表明，以信念状态 β_t 表示的最优代价到达函数与基于完整历史 h_t 得到的结果等价（证明见附录 B.3）。需要注意的是，这一结论基于 [15] 中的思路，并将其扩展为明确纳入数据共享对预期控制性能 $\mathbb{E}^{\boldsymbol{\pi}^s}[r(\theta^*, X_{t:t+1}, Y_t)]$的影响。

引理 2（信念状态上的贝尔曼等价性）在任意时刻 t，给定历史 h_t，最优剩余代价函数 $V_t^*(h_t)$ 可以仅用信念状态 β_t 表示，即 $V_t^*(h_t) = V_t^*(\beta_t)$

由引理 2 可知，给定 h_t，数据共享策略 $\pi_t^s(Y_t \mid \Theta, X_t, h_t)$ 诱导出一个马尔可夫核 $\mathcal{K}_t$，该核将状态对 (Θ, X_t) 映射为 Y_t 上的概率分布。对于 (Θ, X_t) 与Y_t的任意可测空间，该核 $\mathcal{K}_t$ 给出了策略 π_t^s 下Y_t 的条件分布。基于这一表示，我们可以在 (5) 中用信念状态 β_t 替换 h_t，并用其诱导的马尔可夫核 $\mathcal{K}_t$ 替换策略 π_t^s，从而将贝尔曼最优性方程转化为

$$V_t^*(\beta_t) = \min_{\mathcal{K}_t} \big[C_t(\beta_t, \mathcal{K}_t) + \mathbb{E}[V_{t+1}^*(\beta_{t+1}) \mid h_t]\big]. \quad (7)$$

根据式（7），可知最优马尔可夫核 $\mathcal{K}_t^* = \arg\min V_t^*(\beta_t)$ 是一个可测映射 $\beta \mapsto \mathcal{K}_t^*(\cdot \mid \cdot, \cdot, \beta)$；给定实现的信念 β_t，我们使用 $\mathcal{K}_t^*(\cdot \mid \cdot, \cdot, \beta_t)$。因此，我们在定理 3 中证明，数据共享策略由 β_t、Θ 和 X_t 唯一确定。

定理 3（最优策略等价性）对于任意历史 h_t 及其信念 $\beta_t(\cdot \mid h_t)$，存在一个最优马尔可夫核 $\mathcal{K}_t^*(\cdot \mid \theta, x, \beta)$，使得最优数据共享策略满足

$$\pi_t^{s,*}(Y_t \mid \Theta, X_t, h_t) = \mathcal{K}_t^*(Y_t \mid \Theta, X_t, \beta_t), \quad (8)$$

对于所有 $(\theta, x) \in \mathcal{S}_\Theta \times \mathsf{X}$。此外，对于每个固定的信念 β，核 $\mathcal{K}_t^*(\cdot \mid \cdot, \cdot, \beta)$ 在信念状态贝尔曼方程 （7）中达到最小值。

定理 3 表明，最优数据共享策略完全由当前状态（Θ, X_t）和信念状态 β_t 决定。因此，为避免混淆，在本文其余部分，我们用 $\mathcal{K}_t^*(Y_t \mid \Theta, X_t, \beta_t)$ 代替 $\pi^{s,*}_t(Y_t \mid \Theta, X_t, h_t)$。

在现有研究中，状态 X_t 与策略 $\mathcal{K}_t^*(Y_t \mid \Theta, X_t, \beta_t)$ 在离散空间中建模，从而将状态表示和共享数据限制在预定义的离散集合内。将该框架扩展到连续状态与动作空间，能够更准确地表示系统动力学，并实现更平滑、更具适应性的数据共享策略。

4 粒子-信念形式化与连续空间中互信息的上界

求解最优数据共享策略可能颇具挑战，尤其是在状态空间和动作空间连续的情形下。为处理连续状态空间，我们采用粒子滤波跟踪信念状态β_t 的演化，并据此重新表述贝尔曼最优性方程（7）。为处理连续动作空间，我们采用高斯数据共享策略生成共享数据。由于在连续空间中精确计算互信息通常难以实现，我们通过结合基于粒子的信念与高斯策略，开发了一种高效方法来估计互信息的上界。我们证明，由粒子信念引起的贝尔曼值误差依概率一致收敛于零。此外，我们还推导出一种易于计算的互信息估计器，其误差间隙由簇相似性／分离参数（KL界 κ、巴塔恰里亚 γ）限定。

4.1 信念状态的粒子形式化

信念状态 β_t 在 (6) 中定义为 (Θ, X_t) 的后验分布，在连续空间中通常难以计算。为应对这一挑战，我们采用粒子滤波，用支撑于加权粒子有限集 $\{(\theta_i, x_{i,t})\}_{i=1}^N$ 上的经验概率测度 $\hat{\beta}_t^{(N)}$ 来近似 β_t，其关联的归一化非负权重 $\{\omega_{i,t}\}_{i=1}^N$ 满足 $\sum_{i=1}^N \omega_{i,t} = 1$:

$$\hat{\beta}_t^{(N)}(d\theta, dx) = \sum_{i=1}^N \omega_{i,t} \delta_{(\theta_i, x_{i,t})}(d\theta, dx), \quad (9)$$

其中，$\delta_{(\theta,x)}$ 表示位于（θ，x）处的狄拉克测度（单位点质量）。注意，在 $\hat{\beta}_t^{(N)}$ 下的所有期望都可化为加权求和，即对于任意有界可测测试函数 φ,

$$\int \varphi(\theta, x)\, \hat{\beta}_t^{(N)}(d\theta, dx) \;=\; \sum_{i=1}^N \omega_{i,t}\, \varphi(\theta_i, x_{i,t}). \quad (10)$$

与固定网格上的基于网格的离散化不同，PF 在连续空间中使用自适应加权样本，聚焦于高后验区域，从而避免维数爆炸和离散化偏差。粒子根据系统动力学传播。其权重根据与最新观测的一致性进行更新，具体形式见引理 4（证明见附录 C.1）。

引理 4（递归权重更新）给定策略 $\mathcal{K}_t(y_t \mid \theta_i, x_{i,t}, \beta_t)$ 和共享数据 y_t，粒子 i 的权重 $\omega_{i,t}$ 按递归方式更新为 $\omega_{i,t} \propto \omega_{i,t-1} \mathcal{K}(y_t \mid \theta_i, x_{i,t}, \beta_t)$，其中 $\sum_{i=1}^N \omega_{i,t} = 1$。

在粒子权重更新之后，采用条件无偏的重采样方法（例如 [23]中的序贯重要性重采样），以缓解仅有少数粒子占主导且多样性丧失的退化现象。当权重退化评分$D(\tilde{\omega}_t)$降至阈值以下时触发重采样（例如有效样本量 $\hat{N}_{\text{eff}} = \frac{1}{\sum_{i=1}^N (\tilde{\omega}_{i,t})^2}$）。权重更新之后，粒子再根据系统动力学进行传播。

引理5概述了粒子信念的弱一致性。证明见 [24,定理 1]（另见 [25,26]）。此处，$\Rightarrow$ 表示概率测度的弱收敛，即对于任意有界连续函数 φ,

$\int \varphi(\theta,x)\,\hat{\beta}_t^{(N)}(d\theta,dx) \to \int \varphi(\theta,x)\,\beta_t(d\theta,dx)$。为方便记号起见，我们用 β_t 同时表示信念状态及其对应的概率测度。

引理5（粒子信念一致性）固定 t。在序贯蒙特卡洛（*SMC*）方法的标准正则性条件下（即费勒状态转移，以及严格为正、有界且连续的数据共享策略），粒子信念 $\hat{\beta}_t^{(N)}$ 弱一致，即

$$\hat{\beta}_t^{(N)} \Rightarrow \beta_t \qquad a.s. \tag{11}$$

将真实信念 β_t 替换为粒子信念 $\hat{\beta}_t^{(N)}$，并使用粒子信念更新映射，可得到以下粒子（替代）贝尔曼最优性方程：

$$V_t^{*,(N)}(\hat{\beta}_t^{(N)}) = \min_{\mathcal{K}_t\in\mathcal{A}_t} \Big\{ C_t^{(N)}(\hat{\beta}_t^{(N)},\mathcal{K}_t) + \mathbb{E}\big[V_{t+1}^{*,(N)}(\hat{\beta}_{t+1}^{(N)}) \mid h_t,\mathcal{K}_t\big] \Big\}. \tag{12}$$

$\hat{}$ 必须确保在贝尔曼方程中用 β_t 替代真实信念 β_t 不会扭曲最优性原理。因此，接下来我们将正式证明，随着粒子数量趋于无穷，基于粒子信念的贝尔曼最优性方程渐近收敛于其真实形式。

4.2 贝尔曼方程的渐近收敛

现在，我们分析基于粒子信念的贝尔曼最优性方程（12）的渐近收敛性。不失一般性，假设 Θ 取值于有限离散集合 $\mathcal{S}_\Theta = \{\theta^1,\dots,\theta^K\}$。对于任意 θ，令 $\mathcal{I}_\theta = \{i \in \{1,\cdots,N\} \mid \theta_i = \theta\}$ 表示满足 $\theta_i = \theta$的粒子集合。根据（10），边缘质量为 $\hat{q}_t^{(N)}(\theta) = \sum_{i\in\mathcal{I}_\theta} \omega_{i,t}$，条件经验测度为 $\hat{b}_t^{(N)}(dx \mid \theta) = \sum_{i\in\mathcal{I}_\theta} \omega_{i,t}\,\hat{q}_t^{(N)}(\theta)^{\delta_{x_{i,t}}}(dx)$。令 $q_0(\theta^j) > 0$表示对所有 $j = 1,\dots,K$ 均成立的先验，并满足 $\sum_{j=1}^K q_0(\theta^j) = 1$。

在每个时间步 t，接下来我们将证明，随着粒子数量 N 趋于无穷大，粒子形式化 (12) 中的最优价值函数 $V^{*,(N)}_t(\cdot)$ 收敛于原始形式化 (7) 中的最优价值函数 $V_t^*(\cdot)$。在本小节中，只要不会引起歧义，我们始终固定$t \in \{1,\dots,T\}$，并省略核 $\mathcal{K}_t$ 参数中的下标 t 以简化记号，同时保留 $\mathcal{K}_t$上的时间索引。我们还将采用以下正则性假设。

假设6（正则性条件）状态空间和数据共享策略满足以下条件：

(i) 状态空间 S 与观测空间 Y 是 *Polish* 空间，即可分且可完备度量化的拓扑空间；动作空间 Θ 是 *compact*。

(ii) 对于每个 t，可行数据共享策略$\mathcal{K}_t(dy \mid \theta,x,\beta)$ 均对应严格为正的概率密度函数 $k_t(y \mid \theta,x,\beta;u)$，该函数由决策变量 $u \in \mathcal{U}_t$ 参数化，其中 $\mathcal{U}_t \subset \mathbb{R}^{d_u}$ 是紧集，即 $\mathcal{K}_t(dy \mid \theta,x,\beta) = k_t(y \mid \theta,x,\beta;u)dy$。此外，对于任意固定的 (θ,y,β,u)，映射 $x \mapsto k_t(y \mid \theta,x,\beta;u)$ 在 X 上连续。

(iii) *There exist measurable functions* $\underline{k}_t,\overline{k}_t : \mathbb{R}^N \to (0,\infty)$ *such that* $0 < \underline{k}_t(y) \le k_t(y \mid \theta,x,\beta;u) \le \overline{k}_t(y)$, $\forall(\theta,x,\beta,u,y)$. *These functions satisfy the integrability conditions* $\int \overline{k}_t(y)\,dy < \infty$, $\int \overline{k}_t(y)\Big(1+\big|\log \underline{k}_t(y)\big|+\big|\log \overline{k}_t(y)\big|\Big)\,dy < \infty$, *and* $\int (1+\|y\|_2)\overline{k}_t(y)\,dy < \infty$.

(iv) 存在 $L_k < \infty$，使得对于所有 u、$v \in \mathcal{U}_t$ 以及所有 (θ、x、β、y)

$$|k_t(y \mid \theta,x,\beta;u) - k_t(y \mid \theta,x,\beta;v)| \le L_k\|u-v\|\,\overline{k}_t(y).$$

(v) $\pi(\theta^*, x_{1:t-1}, y_{1:t-1})$ 连续且有界。

(vi) *The transition* $p(X_{t+1} \mid \Theta,X_t,W_t)$ *is Feller, i.e., for any bounded continuous function* g, *the mapping* $x_t \mapsto \int g(x_{t+1})p(dx_{t+1} \mid \Theta,x_t,w_t)$ *is continuous.*

为证明假设 6 下贝尔曼最优性方程的渐近收敛性，我们首先在引理7中证明逐步成本的渐近收敛性（证明见附录C.2）。

引理7（逐步成本的渐近收敛）在假设 *6* 下，$\lim_{N\to\infty} C_t^{(N)}\big(\hat{\beta}_t^{(N)},\mathcal{K}_t\big) = C_t\big(\beta_t,\mathcal{K}_t\big)$ 几乎必然。

随后陈述贝尔曼最优性方程的渐近收敛如下。

定理8（贝尔曼方程的渐近收敛）在假设 *6* 下，$\lim_{N\to\infty} V^{*,(N)}_t\big(\hat{\beta}_t^{(N)}\big) = V_t^*\big(\beta_t\big)$ 几乎必然。

定理8的证明概要。我们采用反向归纳法。在终止时刻，由于 $V^{*,(N)}_{T+1} \equiv V^*_{T+1} \equiv 0$，收敛显然成立。在归纳步骤中，假设 $V^{*,(N)}_{t+1}(\hat{\beta}_{t+1}^{(N)}) \to V^*_{t+1}(\beta_{t+1})$ 几乎必然成立。定义 $Q_t^{(N)}(\beta_t^{(N)},\mathcal{K}_t) = C_t^{(N)}(\beta_t^{(N)},\mathcal{K}_t) + \mathbb{E}[V^{*,N}_{t+1}(\beta_{t+1}^{(N)}) \mid h_t,\mathcal{K}_t]$ 以及 $Q_t(\beta_t,\mathcal{K}_t) = C_t(\beta_t,\mathcal{K}_t) + \mathbb{E}[V^*_{t+1}(\beta_{t+1}) \mid h_t = w_{1:t},y_{1:t-1}]$，

回忆一下，h_t 是历史。于是，

$$V_t^{*,(N)}(\hat{\beta}_t^{(N)}) = \min_{\mathcal{K}_t\in\mathcal{A}_t} Q_t^{(N)}(\hat{\beta}_t^{(N)},\mathcal{K}_t)$$
$$V_t^*(\beta_t) = \min_{\mathcal{K}_t\in\mathcal{A}_t} Q_t(\beta_t,\mathcal{K}_t),$$

根据假设 (ii)，我们可以将 $\min_{\mathcal{K}_t \in \mathcal{A}_t}(\cdot)$ 替换为 $\min_{u_t \in \mathcal{U}_t}(\cdot)$，并在 Q_t 和 $Q_t^{(N)}$ 中都将 $\mathcal{K}_t$ 替换为 u_t。然后，根据三角不等式，

$$|V_t^{*,(N)}(\hat{\beta}_t^{(N)}) - V_t^*(\beta_t)| \leq \sup_{u_t \in \mathcal{U}_t} |Q_t^{(N)}(\hat{\beta}_t^{(N)}, u_t) - Q_t(\hat{\beta}_t, u_t)|. \tag{13}$$

定义 $\Delta Q_N := \sup_{u_t \in \mathcal{U}_t} |Q_t^{(N)}(\hat{\beta}_t^{(N)}, u_t) - Q_t(\beta_t, u_t)|$。只需证明 $\Delta_N \to 0$ 几乎必然成立。根据三角不等式，我们将 $\Delta_N \leq \Delta Q_N^{(C)} + \Delta Q_N^{(V)}$ 分解为两部分，其中 $\Delta Q_N^{(C)} = \sup |C_t^{(N)}(\hat{\beta}_t^{(N)}, u_t) - C_t(\beta_t, u_t)|$ 是当前成本差距，而 $\Delta Q^{(V)}{}_N = \sup_{u_t \in \mathcal{U}_t} |\mathbb{E}[V^*{}_{(N)}{}_{t+1}(\hat{\beta}_{t+1}^{(N)}) \mid h_t, u_t] - \mathbb{E}[V_{t+1}^*(\beta_{t+1}) \mid h_t, u_t]|$ 是未来价值差距。随后，我们证明这些差距的逐点收敛和一致收敛（详见附录 C.3）。

4.3 Upper bound on particle MI under Gaussian policy

先前研究 [15,19]通常会将共享数据离散化，从而导致表示粗糙且不准确，并使高维空间中的可能数据共享结果呈组合式增长。为解决这一问题，我们通过设计一种由当前状态、私有参数和信念状态决定的高斯策略，实现连续数据的共享，即$Y_t \sim \mathcal{N}(\mu_t(X_t, \Theta, \hat{\beta}_t^{(N)}; u)$ $\Sigma_t(X_t, \Theta, \hat{\beta}_t^{(N)}; u))$，其中均值 μ_t 和协方差矩阵 Σ_t 是确定性映射（例如神经网络）的输出，$u \in \mathcal{U}$ 为其参数，$X_t, \Theta, \hat{\beta}_t^{(N)}$ 为输入。

接下来，我们量化此类策略下的粒子互信息。逐步粒子互信息表示为 $I_u^{(N)}(\Theta; Y_t \| h_t) = H^{(N)}(Y_t \mid h_t; u) - H^{(N)}(Y_t \mid \Theta, h_t; u)$，其中 $H^{(N)}$表示相应的条件微分熵。

我们计算 Y_t 给定h_t 时诱导出的边缘密度，如下所示的高斯混合:

$$p^{(N)}(y \mid h_t; u) = \int k_t(y \mid \theta, x, \hat{\beta}_t^{(N)}; u)\, \hat{\beta}_t^{(N)}(d\theta, dx) = \sum_{i=1}^{N} \omega_{i,t} \mathcal{N}(y; \mu_{i,t}(u), \Sigma_{i,t}(u)), \tag{14}$$

其中，$\mu_{i,t}(u) := \mu_t(\theta_i, x_{i,t}, \hat{\beta}_t^{(N)}; u)$ 以及 $\Sigma_{i,t}(u) := \Sigma_t(\theta_i, x_{i,t}, \hat{\beta}_t^{(N)}; u)$ 表示通过在粒子 i 处评估策略而得到的均值和协方差。

回顾边缘质量 $\hat{q}_t^{(N)}(\theta) = \sum_{i \in \mathcal{I}_\theta} \omega_{i,t}$ 以及条件经验测度 $\hat{b}_t^{(N)}(dx \mid \theta) = \sum_{i \in \mathcal{I}_\theta} \frac{\omega_{i,t}}{\hat{q}_t^{(N)}(\theta)} \delta_{x_{i,t}}(dx)$。给定 θ，Y_t 的条件密度因此是一个高斯混合:

$$p^{(N)}(y \mid \theta, h_t; u) = \int k_t(y \mid \theta, x, \hat{\beta}_t^{(N)}; u)\, \hat{b}_t^{(N)}(dx \mid \theta) = \sum_{i \in \mathcal{I}_\theta} \frac{\omega_{i,t}}{\hat{q}_t^{(N)}(\theta)} \mathcal{N}\big(y; \mu_{i,t}(u), \Sigma_{i,t}(u)\big). \tag{15}$$

因此，$H^{(N)}(Y_t \mid h_t)$ 和 $H^{(N)}(Y_t \mid \Theta = \theta, h_t)$ 都是高斯混合模型（GMM）的微分熵。由于一般 GMM 的微分熵没有闭式表达式，我们推导出互信息 $I^{(N)}(\Theta; Y_t \| h_t)$ 的一个可计算上界，以支持可处理规划和学习。

定理 9（KL/Chernoff 混合界） 记 $\mathcal{N}_{i,t}(\cdot; u) := \mathcal{N}\big(\cdot; \mu_{i,t}(u), \Sigma_{i,t}(u)\big)$ 以及 $\bar{\omega}_{i,t}(\theta) := \omega_{i,t} / \hat{q}_t^{(N)}(\theta)$。则 $\widehat{I}_u^{(N)}(\Theta; Y_t \mid h_t) := \widehat{H}^{(N)}{}_{\mathrm{KL}}(Y_t \mid h_t; u) - \sum_{\theta \in \mathcal{S}_\Theta} \hat{q}_t^{(N)}(\theta) \cdot \widehat{H}_{C_\alpha}^{(N)}\big(Y_t \mid \theta, h_t; u\big)$ 是逐步粒子互信息 $I_u^{(N)}(\Theta; Y_t \mid h_t)$ 的一个上界，其中

$$\widehat{H}_{\mathrm{KL}}^{(N)}(Y_t \mid h_t; u) = \sum_{i=1}^{N} \omega_{i,t} H(\mathcal{N}_{i,t}(\cdot; u)) - \sum_{i=1}^{N} \omega_{i,t} \log \left(\sum_{j=1}^{N} \omega_{j,t} e^{-\mathrm{KL}\left(\mathcal{N}_{i,t}(\cdot;u) \| \mathcal{N}_{j,t}(\cdot;u)\right)} \right),$$

$$\widehat{H}_{C_\alpha}^{(N)}\big(Y_t \mid \theta, h_t; u\big) = \sum_{i \in \mathcal{I}_\theta} \bar{\omega}_{i,t}(\theta) H(\mathcal{N}_{i,t}(\cdot; u)) - \sum_{i \in \mathcal{I}_\theta} \bar{\omega}_{i,t}(\theta) \log \left(\sum_{j \in \mathcal{I}_\theta} \bar{\omega}_{i,t}(\theta) e^{-C_\alpha\left(\mathcal{N}_{i,t}(\cdot;u) \| \mathcal{N}_{j,t}(\cdot;u)\right)} \right).$$

定理 9 的证明结合了 (i) 基于 KL 散度的上界 $H^{(N)}(Y_t \mid h_t; u) \leq \widehat{H}^{(N)}{}_{\mathrm{KL}}(Y_t \mid h_t; u)$，其中 $\mathrm{KL}(\cdot\|\cdot)$ 表示 KL 散度；以及 (ii) 基于 Chernoffα 散度的下界 $H^{(N)}\big(Y_t \mid \Theta, h_t; u\big) \geq \sum_{\theta \in \mathcal{S}_\Theta} \hat{q}_t^{(N)}(\theta) \cdot \widehat{H}_{C_\alpha}^{(N)}\big(Y_t \mid \theta, h_t; u\big)$，其中 $C_\alpha(\cdot\|\cdot)$ 表示 Chernoff α 散度 [27]。由于所有组成部分均为高斯分布，KL 散度和 C_α 都有闭式表达式，因此该界完全可计算。

接下来，我们分析粒子互信息与已有上界之间的差距，记为 $\Delta I_u^{(N)} = \widehat{I}_u^{(N)}(\Theta; Y_t \mid h_t) - I_u^{(N)}(\Theta; Y_t \mid h_t)$。沿用 [27], 中的紧致性分析，我们进一步为 $\Delta I_u^{(N)}$建立一个上界。我们根据以下两项准则，使用聚类函数 $g(\cdot)$ 将粒子划分为多个簇：(1) 对某个较小的 κ，当 $g(i) = g(j)$ 时，$\mathrm{KL}(\mathcal{N}_{i,t}(\cdot; u) \| \mathcal{N}_{j,t}(\cdot; u)) \leq \kappa$；(2) 对某个较大的 κ，当 $g(i) \neq g(j)$ 时，$\mathrm{BD}(\mathcal{N}_{i,t}(\cdot; u) \| \mathcal{N}_{j,t}(\cdot; u)) \geq \gamma$

其中，BD表示巴塔查里亚距离 $\mathrm{BD}(\cdot\|\cdot) = C_{0.5}(\cdot\|\cdot)$ [28]。类似地，对于每个固定的 $\theta \in \mathcal{S}_\Theta$，我们为条件子集合 I_θ 定义聚类函数 g_θ ：(1) 对某个较小的 κ_θ，当 $g_\theta(i) = g_\theta(j)$ 时，$\mathrm{KL}(\mathcal{N}_{i,t}(\cdot;u)\|\mathcal{N}_{j,t}(\cdot;u)) \le \kappa_\theta$；(2) 对某个较大的 γ_θ，当 $g_\theta(i) \neq g_\theta(j)$ 时，$\mathrm{BD}(\mathcal{N}_{i,t}(\cdot;u)\|\mathcal{N}_{j,t}(\cdot;u)) \ge \gamma_\theta$。

在上述聚类条件下，$\widehat{I}_u^{(N)}(\Theta; Y_t \mid h_t)$ 与粒子互信息 $I_u^{(N)}(\Theta; Y_t \mid h_t)$ 之间的差距 $\Delta I_u^{(N)}$在形式上如推论 10 所示被上界约束。证明可参见附录C.4。

推论 10（上界的紧致性） *Let $|G_\theta|$ be the number of clusters within I_θ, and let $\alpha, \alpha_\theta \in [0, 1]$ be Chernoff parameters.*

$$\Delta\tilde{I}_u^{(N)} \le \kappa + (|G| - 1)e^{\left(-(1-|1-2\alpha|)\gamma\right)} + \sum_{\theta\in\mathcal{S}_\Theta} \hat{q}_t^{(N)}(\theta)\Big[\kappa_\theta + (|G_\theta| - 1)e^{\left(-(1-|1-2\alpha_\theta|)\gamma_\theta\right)}\Big] \quad (16)$$

推论 10 表明，差距 $\Delta I_u^{(N)}$受簇内相似性（通过 KL 项，κ，κ_θ）和簇间分离度（通过 BD 项，γ，γ_θ）控制。当簇内相似性较小、簇间分离度较大时，我们得到的上界$\widehat{I}_u^{(N)}$可以接近粒子互信息 $I_u^{(N)}$。

然而，在强隐私保护约束下，KL/Chernoff 混合界可能较松。具体而言，为了降低 Θ 与发布数据之间的可区分性，策略往往会使不同粒子的均值和协方差趋于相似，即使 i 和 j 属于不同的簇（或不同的 θ-组），也会有 $\mu_{i,t}(u) \approx \mu_{j,t}(u)$ 和 $\Sigma_{i,t}(u) \approx \Sigma_{j,t}(u)$。相应地，簇内相似性参数 κ和 κ_θ 可能仍然较小，但簇间分离度量γ 和 γ_θ 会大幅下降。注意，(16) 包含形如 $(|G| - 1)e^{-(1-|1-2\alpha|)\gamma}$ 和 $(|G_\theta| - 1)e^{-(1-|1-2\alpha_\theta|)\gamma_\theta}$ 的项；当 γ 和 γ_θ 较小时，这些项可能很大。在极端情况下，当 $\gamma, \gamma_\theta \approx 0$ 时，(16) 化简为 $\Delta I_u^{(N)} \lesssim \kappa + (|G| - 1) + \sum_{\theta\in\mathcal{S}_\Theta} \hat{q}_t^{(N)}(\theta)\Big[\kappa_\theta + (|G_\theta| - 1)\Big]$；只要存在多个簇，该式就可能很大。因此，即使真实的粒子互信息已经很小，KL/Chernoff 混合界仍可能偏高，导致其在高隐私区域过于保守。

为缓解高隐私区域中 KL/切尔诺夫混合上界的松弛性，我们利用一个通用的最大熵上界。具体而言，对于式 (14) 中的边缘高斯混合模型 $p^{(N)}(y \mid h_t; u)$，令 $\bar{\mu}_t = \sum_{i=1}^N \omega_{i,t}\mu_{i,t}(u)$ 和 $\Sigma_{\mathrm{glob},t} = \sum_{i=1}^N \omega_{i,t}\Big(\Sigma_{i,t}(u) + (\mu_{i,t}(u) - \bar{\mu}_t)(\mu_{i,t}(u) - \bar{\mu}_t)^\top\Big)$。由于在给定协方差 [29], 时，高斯分布的熵最大，因此有 $H^{(N)}(Y_t \mid h_t; u) \le \widehat{H}^{(N)}\mathrm{G}(Y_t \mid h_t; u)$，其中 $H^{(N)}\mathrm{G}(Y_t \mid h_t; u) = \frac{1}{2}\log((2\pi e)^d |\Sigma_{\mathrm{glob},t}|)$。将其与定理 9 中关于 $H^{(N)}(Y_t \mid \Theta, h_t; u)$ 的基于切尔诺夫的下界相结合，我们提出以下粒子互信息的自适应区域上界。在附录 C.5 中，我们证明，当诱导的混合分量有大量重叠时（例如，对于许多 (i, j)，$\mu_{i,t}(u) \approx \mu_{j,t}(u)$ 且 $\Sigma_{i,t}(u) \approx \Sigma_{j,t}(u)$），高斯项$\widehat{H}^{(N)}\mathrm{G}(Y_t \mid h_t; u)$ 通常会成为起作用的上界（即使表达式最小），从而得到更紧的估计。

定理 11（自适应机制互信息上界） 逐步粒子互信息的上界为：

$$I_u^{(N)}(\Theta; Y_t \mid h_t) \le \widehat{I}_{u,\mathrm{RA}}^{(N)}(\Theta; Y_t \mid h_t), \quad (17)$$

where $\widehat{I}_{u,\mathrm{RA}}^{(N)}(\Theta; Y_t \mid h_t) = \min\Big\{\widehat{H}_{\mathrm{KL}}^{(N)}(Y_t \mid h_t; u), \widehat{H}_{\mathrm{G}}^{(N)}(Y_t \mid h_t; u)\Big\} - \sum_{\theta\in\mathcal{S}_\Theta} \hat{q}_t^{(N)}(\theta)\widehat{H}_{C_\alpha}^{(N)}\big(Y_t \mid \theta, h_t; u\big)$.

对于所有 $u \in \mathcal{U}$，自适应机制上界始终不比定理 9 中的 KL/Chernoff 混合界宽松程度更低。

注 1（其他成本项） 在高斯数据共享策略下，偏差项 $\mathbb{E}[d(X_t, Y_t)]$ 和预期系统影响 $\mathbb{E}[r(\theta^*, X_t, X_{t+1}, Y_t)]$ 可能具有解析形式。例如，当偏差定义为欧几里得距离时，根据非中心χ^2 分布的性质，其期望可表示为闭式表达式。一般而言，当无法获得解析形式时，可以通过蒙特卡洛抽样对相应的期望进行数值计算。

我们将在定理 12 中证明（证明见附录 C.6），可以为高斯混合模型策略构造类似的粒子互信息闭式上界。

定理 12 (高斯混合模型策略下粒子互信息的上界) 如果每个粒子对应的数据共享策略都是有限高斯混合，则定理 *11* 中的粒子互信息闭式上界仍然适用。

因此，假设对于所有 θ，$H(Y_t \mid h_t; u)$ 和 $H(Y_t \mid \theta, h_t; u)$ 均为有限值，根据有限高斯混合模型（例如 [30,31]）的通用逼近性质，我们提出的基于粒子的互信息上界可推广至任意连续策略。证明见附录 C.7。

推论13 (通用闭式上界) 对于任意连续策略核密度$k_t(\cdot \mid \theta, x, \hat{\beta}_t^{(N)}; u)$ 以及任意 $\varepsilon > 0$，存在一个有限高斯混合策略核 $\mathcal{K}_{J(\delta)}$，使得

诱导的逐步基于粒子的互信息满足

$$\left|I^{(N)}(\Theta;Y_t \mid h_t;\mathcal{K}_t) - I^{(N)}(\Theta;Y_t \mid h_t;\hat{\mathcal{K}}_{J(\delta)})\right| \le \varepsilon.$$

此外，将定理 *11* 应用于 $\mathcal{K}_{J(\delta)}$，可得可计算的上界

$$I^{(N)}(\Theta;Y_t \mid h_t;\mathcal{K}_t) \le \widehat{I}^{(N)}_{\hat{\mathcal{K}}_{J(\delta)},\mathrm{RA}}(\Theta;Y_t \mid h_t) + \varepsilon.$$

因此，借助这些易于处理的表达式，原始的隐私-效用优化问题可重新表述为一个粒子信念状态马尔可夫决策过程，其最优数据共享参数 u^* 满足粒子贝尔曼最优方程（12）。接下来，我们通过强化学习（RL）学习参数化的数据共享策略。

4.4 基于粒子滤波的强化学习求解方法

利用信念状态的粒子表示，我们采用演员–评论家算法学习最优参数 u^*，该参数诱导出最优策略 $\mathcal{K}_t^*(Y_t \mid \theta,\ x_t,\ \hat{\beta}_t^{(N)}) = \mathcal{K}_t(Y_t \mid \theta,\ x_t,\ \hat{\beta}_t^{(N)};\ u^*)$。我们提出的算法包含两个主要部分：一个MGF编码器和一个深度确定性策略梯度学习算法。

（*1*）矩母函数（*MGF*）编码器。我们采用MGF编码器 [32] 来应对基于粒子的信念表示带来的两项挑战。首先，在高维状态空间中，通常需要大量粒子才能实现高精度的信念估计，这可能导致训练不稳定。其次，粒子滤波中的重采样步骤会引入粒子置换，进一步增加收敛难度。MGF编码器通过捕获高阶矩并在温和的正则性条件下刻画分布来应对这些挑战 [33]。具体而言，我们近似粒子信念 $\hat{\beta}_t^{(N)}$ 的矩母函数，得到一个低维向量 $s_t = \left[\bar{s}_t,\ M^{1:m}{}_{\hat{\beta}_t^{(N)}}\right]$，其中 $\bar{s}_t = \left[\sum_{i=1}^N \omega_{i,t}\theta_i, \sum_{i=1}^N \omega_{i,t}x_{i,t}\right]$ 是 $\hat{\beta}_t^{(N)}$ 的一阶矩。通过在学习得到的评估点 $\{v^i\}_{i=1}^m$ 处计算矩母函数特征来捕获 $\hat{\beta}_t^{(N)}$ 的高阶信息，得到 $M^{1:}{}_{\hat{\beta}_t^{(N)}}^{m} = \left[M^{1}{}_{\hat{\beta}_t^{(N)}},\ M^{2}{}_{\hat{\beta}_t^{(N)}},\ \ldots\ldots,\ M^{m}{}_{\hat{\beta}_t^{(N)}}\right]$，其中 $M^{i}{}_{\hat{\beta}_t^{(N)}}$ 是 v^i 处计算得到的矩母函数特征。MFG编码器 $S_t = \mathrm{Enc}_\phi(\hat{\beta}_t^{(N)})$ 由带有ReLU激活函数的全连接层表示，参数为 ϕ。

(2) 基于 *DDPG* 的学习算法。我们采用一种包含演员网络和评论家网络 [34] 的 DDPG 算法。演员网络 $(\hat{\mu}_t, \Sigma_t) = f_u(\theta, x_t, s_t)$ 通过生成其均值 $\hat{\mu}_t$ 和协方差矩阵 Σ_t，根据当前状态 x_t、私有参数 θ 和均场博弈特征 s_t 实例化高斯策略 $\mathcal{K}_t\big(\cdot \mid \theta, x_t, \hat{\beta}_t^{(N)}; u\big)$。随后按此对共享数据 y_t 进行采样。由参数 w 表示的评论家网络估计动作价值函数 $Q(\theta^*, s_t, x_t, \hat{\mu}_t, \Sigma_t \mid w)$。目标网络 Q' 和 $f_{u'}$（以及目标编码器 $\mathrm{Enc}_{\phi'}$）以速率 $\tau \in (0,1)$ 通过 Polyak 平均进行更新。评论家通过最小化时序差分（TD）损失进行训练，而演员参数 u 则使用确定性策略梯度更新。MGF 编码器与评论家联合训练，通过 $s_t = \mathrm{Enc}_\phi(\hat{\beta}_t^{(N)})$ 反向传播 TD 损失，以鼓励获得能够预测价值（符合贝尔曼一致性）的表示。这种联合训练产生的学习信号比仅更新演员更密集、方差更低。

5 实验

本节通过数值仿真，评估所提出的隐私保护数据共享框架在混合自主车队控制场景中的表现 [21,35]。

5.1 实验设置

我们考虑一个由三辆车组成的三车混合自动化车队，包括一辆领航车辆（编号为 0）、一辆位于中间的网联自动驾驶车辆（CAV，编号为 1），以及一辆位于车队末尾的人工驾驶车辆（HDV，编号为 2）。HDV 作为数据提供方，通过车对车通信向 CAV 分享其实时车间距和速度。CAV 利用这些信息提升车队性能。尽管 HDV 可能受益于车队运行状况的改善，但它也担心隐私问题：对手可能利用共享数据推断其驾驶行为参数，从而可能降低安全性，并带来经济风险，例如保险费上涨 [35]。

仿真设置。与 [35,36], 类似，CAV 和 HDV 的系统动态定义如下：

$$\begin{aligned}
\dot{s}_i(t) &= v_{i-1}(t) - v_i(t) + \zeta_{s,i}(t), i = 1, 2 \\
\dot{v}_i(t) &= \begin{cases} u(t), & i = 1, \\ \mathbf{F}_{\theta^*}\left(s_i(t), v_i(t), v_{i-1}(t)\right) + \zeta_{a,i}(t), & i = 2. \end{cases}
\end{aligned}$$

其中，扰动 $\zeta_a(t)$ 和 $\zeta_s(t)$ 分别表示离散化误差和人类建模误差。网联自动驾驶车辆的控制动作 $u(t)$ 是通过分布式线性控制器 $u(t) = \sum_{i=1}^3 \left(\mu_i\left(s_i - s^\star\right) + \eta_i\left(v_i - v^\star\right)\right)$ 确定的加速度，其中 $(s^\star, v^\star)$ 表示平衡状态，μ_i 和 η_i 是基于串稳定性条件的反馈增益。人工驾驶车辆的加速度率由以下模型决定：

跟驰模型$\mathbf{F}_{\theta^*}$，该模型由敏感参数 θ^* 参数化，此参数是人类驾驶员的固有属性。这里，我们采用全速度差模型（FVD） [37] ，其形式为$\mathbf{F}_{\theta^*}(\cdot) = \theta_1^*(V_{\mathrm{FVD}}(s_3(t)) - v_3(t)) + \theta_2^*(v_2(t) - v_3(t))$，其中 $V_{\mathrm{FVD}}(s)$ 是间距相关期望速度（详见 [37]）。

式 (1) 中的偏差项选为共享数据 $Y_t = (\tilde{v}_t,\ \tilde{s}_t)$ 与真实数据 $X_t = (v_t,\ s_t)$ 之间的加权欧氏距离，权重为 $\omega_1 \geq \omega_2$，即$d(X_t,\ Y_t) = \sqrt{\omega_1(v_t - \tilde{v}_t)^2 + \omega_2(s_t - \tilde{s}_t)^2}$。在我们的实验中，设定 $\omega_1 = 3$ 和 $\omega_2 = 1$。系统成本$r(\theta^*,\ X_{t:t+1},\ Y_t)$ 定义为人工驾驶车辆的实际燃油消耗率 f_i (mL/s) [38]:

$$f_i = \begin{cases} 0.444 + 0.090 R_i v_i + \left[0.054 a_i^2 v_i\right]_{a_i > 0}, & R_i > 0 \\ 0.444, & R_i \leq 0 \end{cases}$$

其中 $R_i = 0.333 + 0.00108 v_i^2 + 1.200 a_i$，$v_i$ 表示车辆速度，a_i 表示其加速度。

系统动力学以 $\Delta t = 0.2$s 的时间步长进行离散化。仿真环境的参数如下：平衡间距和速度 $(s^*, v^*) = (20\ \mathrm{m},\ 15\ \mathrm{m/s})$，间距阈值 $(s_{st}, s_{go}) = (5\ \mathrm{m},\ 35\ \mathrm{m})$，加速度限值 $(a_{\min}, a_{\max}) = (-5,\ 5)\ \mathrm{m/s^2}$，最大速度$v_{\max} = 30\ \mathrm{m/s}$，以及自动驾驶车辆速度的标准差 $\mathrm{std}(\zeta_{a,i}) = 1\ \mathrm{m/s^2}$。对于自动驾驶车辆，线性控制器的反馈增益为 $(\mu_1,\ \eta_1) = (0.1,\ 0)$；对于网联自动驾驶车辆，反馈增益为 $(\mu_2,\ \eta_2) = (-0.5,\ 0.5)$；对于人工驾驶车辆，反馈增益为 $(\mu_3,\ \eta_3) = (-0.2,\ 0.2)$。

训练配置。每个回合的时域为$L = 50$，期望总失真的上界为$\hat{D} = 200$，拉格朗日乘子为 $\lambda = 0.2$，隐私权重为 $\rho = 80$，学习得到的位置向量维度为$D^v = 10$。演员和评论家的学习率分别为 $lr_a = 10^{-3}$ 和 $lr_c = 10^{-3}$，两个网络均包含2层，MGF网络包含1层。

基准方法。将所提出的数据共享策略与以下方法进行评估：(i) 真实数据，其中人工驾驶车辆将其真实数据共享给网联自动驾驶车辆；(ii) 离散策略[15,19,21]，其中人工驾驶车辆采用基于优势演员-评论家的数据共享策略，将真实数据和共享数据均离散化为网格。对于所提出的方法和离散策略基准，网络架构包括一个1层的MGF编码器多层感知机，以及用于演员和评论家网络的2层多层感知机。

5.2 性能比较

本节将从隐私保护、数据可用性和系统级成本（即燃料消耗和控制平滑性）等方面评估训练策略。敏感参数 θ^* 从 $\theta^1 = (0.4, 0.5)$、$\theta^2 = (0.7, 0.8)$、 $\theta^3 = (1.0, 1.1)$ 和 $\theta^4 = (1.3, 1.4)$ 中选取。所有实验均在 50 个时间步的时域内进行。为准确评估隐私泄露，我们使用 5,184 个粒子来跟踪信念状态。

参数隐私。图1展示了对于每个真实参数 θ^*，攻击者后验概率质量的变化（左图），以及最大似然占用估计器的检测结果（右图）。

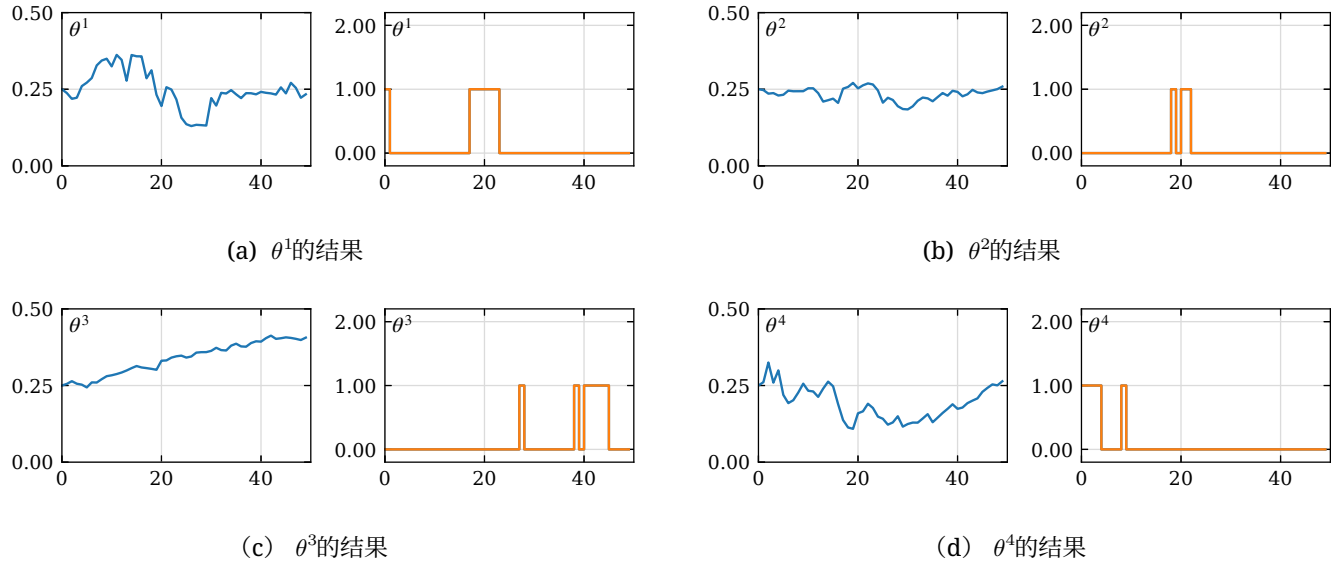
可以看出， θ^* 的后验概率质量在所有情况下都保持较低，表明隐私泄漏有限，这与估计器输出中频繁出现的漏检相一致。

数据可用性。为评估数据可用性，我们对共享数据运行粒子滤波器，以跟踪x_t 的后验分布，并报告后验均值估计 $\mathbb{E}[x_t \mid y_{1:t}]$。图2（出于篇幅考虑，仅展示 $\theta^* = \theta^1$ ）表明，后验均值与真实速度和间距高度吻合，只有短暂的初始瞬态，这是由先验初始化和早期观测有限所致。对于 θ^*的其他取值，我们也观察到类似的跟踪性能，这证实了所提出的数据共享机制能够保留状态估计效用。

系统成本。我们比较应用离散策略和连续策略前后，HDV 的燃料消耗性能（mL/s）以及由此产生的运动平滑度。燃料消耗结果汇总于表1。结果表明，对于不同的驾驶行为 Θ，连续策略始终比离散策略具有更低的预期燃料消耗，并且非常接近“真实数据”基线。表2报告了两种平滑性度量：平均加速度平方 $\mathbb{E}[a^2]$ 和平均加速度变化平方 $\mathbb{E}[(\Delta a)^2]$。可以看出，对于所有真实参数，与离散策略相比，连续策略始终能降低驱动强度。具体而言，连续策略下的 $\mathbb{E}[a^2]$ 比离散策略下低约 $1.15\times$–$1.55\times$，并且始终更接近真实数据基线。此外，连续策略带来的控制也更加平滑，因为对于所有 θ^*，其 $\mathbb{E}[(\Delta a)^2]$ 均低于离散策略下的对应值，并且同样更接近真实数据基线。总体而言，与离散策略相比，连续策略具有更低的加速度幅值和更小的加速度变化，表明其在驱动效率和平滑度方面具有明显的系统成本优势。

6 结论

本文研究了连续状态动力系统中的参数隐私保护数据共享。我们通过构建粒子信念马尔可夫决策过程，实现隐私保护



(a) θ^1的结果　(b) θ^2的结果

(c) θ^3的结果　(d) θ^4的结果

图 1. 隐私性能。对于每个 θ^*，左图展示后验概率的演变，右图则报告随时间变化的误检实例。

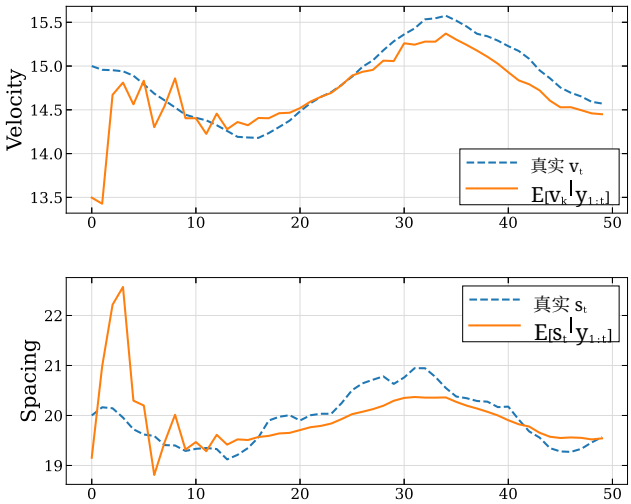


图 2. 数据可用性：真实状态与后验均值估计。

表 1 预期燃料消耗（毫升/秒）

θ^*	真实数据	离散策略	连续策略
θ^1	1.24	1.29	1.25
θ^2	1.24	1.30	1.25
θ^3	1.25	1.34	1.26
θ^4	1.26	1.35	1.27

数据共享，并通过序贯蒙特卡洛近似难以处理的信念动态。由此得到的形式化模型易于处理，且具有渐近一致性

表2 预期加速度指标

	真实数据		离散策略		连续策略	
θ^*	$\mathbb{E}[a^2]$	$\mathbb{E}[(\Delta a)^2]$	$\mathbb{E}[a^2]$	$\mathbb{E}[(\Delta a)^2]$	$\mathbb{E}[a^2]$	$\mathbb{E}[(\Delta a)^2]$
θ^1	0.04	0.03	0.11	0.04	0.10	0.03
θ^2	0.05	0.04	0.13	0.06	0.11	0.05
θ^3	0.06	0.06	0.16	0.09	0.12	0.07
θ^4	0.08	0.09	0.19	0.15	0.12	0.10

随着粒子数量的增加。我们进一步利用高斯混合近似推导出基于粒子的互信息的状态自适应闭式上界，从而能够高效优化连续数据共享策略。在混合自动化车队上的实验表明，所学习的策略在保持数据可用性的同时，显著阻碍了对人类驾驶行为参数的推断，并且仅造成可忽略的性能下降。未来的工作将研究隐私感知的高维数据共享以及多利益相关方的数据共享机制。

致谢

本研究得到新加坡教育部（MOE）学术研究基金第二级（A-8003064-00-00）的支持。

参考文献

[1] Qiqing Wang 和 Kaidi Yang。用于交通状态估计的隐私保护数据融合：一种纵向联邦学习

方法。 *Transportation Research Part C: Emerging Technologies* 168 104743, 2024

Qiqing Wang、Haokun Yu 和 Kaidi Yang。促进数字交通市场中的数据协作：隐私保护机制的作用，2026年。[3] Yang Lu 和 Minghui Zhu。使用同态加密的隐私保护分布式优化。 *Automatica*, 96 314: –325, 2018 年。[4] Jun Shang 、 Jing Zhou 和 Tongwen Chen。针对远程状态估计中基于创新的隐蔽攻击的单维加密。 *Automatica* 136 110015, , : 。2022

[5] Le Wang、Xianghui Cao、Heng Zhang、Changyin Sun 和 Wei Xing Zheng。针对远程状态估计中的窃听攻击的隐私最优加密传输调度。 *Automatica*, 137:110145, 2022年。[6] Tongxiang Li、Zidong Wang、Lei Zou、Bo Chen 和 Li Yu。用于信息物理系统重放攻击检测的动态加解密方案。 *Automatica*, 151:110926, 2023年。[7] Lei Zou、Zidong Wang、Bo Shen 和 Hongli Dong。中继信道中的递归状态估计及增强的窃听防护：一种创新的加解密框架。 *Automatica*, 174:112159, 2025年。[8] Bo Chen、Kevin Leahy、Austin Jones 和 Matthew Hale。符号系统的差分隐私及其在马尔可夫链中的应用。 *Automatica*, 152:110908, 2023年。[9] Chenxi Chen、Xianbiao Hu、Yang Li 和 Qing Tang。面向智慧出行应用的、基于差分隐私的公共交通轨迹数据发布中的隐私预算分配优化。 *IEEE Transactions on Intelligent Transportation Systems* 2023 , 。

[10] Lei Wang、Ian R Manchester、Jochen Trumpf 和 Guodong Shi。线性动态系统的差分初值隐私与可观测性。 *Automatica*, 148:110722, 2023年。[11]Lei Wang、Weijia Liu、Fanghong Guo、Zixin Qiao 和 Zhengguang Wu。具有改进精度—隐私权衡的差分隐私平均一致性。*Automatica*, 167 111769, 2024年：。[12]Jing Zhao 、 Taeho Jung 、 Yu Wang 和 Xiangyang Li。实现智能电网数据披露的差分隐私，载于 *IEEE INFOCOM 2014-IEEE conference on computer communications* 504 512 2014 , 第 – 页。IEEE, 。[13] Ehsan Nekouei 、 Mohammad Pirani 、 Henrik Sandberg 和 Karl H Johansson。针对主动转向控制系统推断攻击的随机滤波策略。 *IEEE Transactions on Information Forensics and Security* 17 16 , : –27, 2021年。[14] Farhad Farokhi 和 Henrik Sandberg。通过最小化费舍尔信息，利用受约束加性噪声确保隐私。*Automatica*, 99:275–288, 2019年。[15] Chuanghong Weng、Ehsan Nekouei 和 Karl H Johansson。最优隐私感知状态估计。 *IEEE Transactions on Automatic Control* 2025 , 。

[16]Ingvar Ziemann 和 Henrik Sandberg。参数隐私与控制性能：费舍尔信息正则化控制。载于 *2020 American Control Conference (ACC)*, 第1259–1265页。IEEE, 2020年。[17]Germ´an Bassi、Mikael Skoglund 和 Pablo Piantanida。统计参数隐私约束下的有损通信。载于 *2018 IEEE international symposium on information theory (ISIT)* 1031 1035 2018 , 第 – 页。IEEE, 。

[18]Ehsan Nekouei、Henrik Sandberg、Mikael Skoglund 和 Karl Henrik Johansson。统计参数隐私的模型随机化方法。 *IEEE Transactions on Automatic Control* 68 2 839 , ( ): –850, 2022年。[19]Ecenaz Erdemir、Pier Luigi Dragotti 和 Deniz G¨und¨uz。基于深度强化学习的隐私感知时间序列数据共享。 *IEEE Transactions on Information Forensics and Security* 16 389 401, 2020 , : – .[20]Yu Kawano、Kenji Kashima 和 Ming Cao。隐私保护下的模块化控制：基本权衡。*Automatica*, 127:109518, 2021年。[21]Haokun Yu、Jingyuan Zhou 和 Kaidi Yang。耦合系统中基于粒子滤波强化学习的交互感知参数隐私保护数据共享。载于 *7thAnnual Learning for Dynamics\& Control Conference*, 第1525–1536页。PMLR, 2025年。[22]Claude E Shannon。保密系统的通信理论。 *The Bell system technical journal* 28 4 656 715, , ( ): – .[23]Donald B Rubin。贝叶斯自助法。 *The annals of statistics* 130 134, 1981 , 第 – 页。1949

[24]Dan Crisan 和 Arnaud Doucet。面向实践者的粒子滤波方法收敛结果综述。*Transactions on signal processing* 50 3 736 746, 2002 , ( ): – .*IEEE*

[25] Nicolas Chopin。序贯蒙特卡洛方法的中心极限定理及其在贝叶斯推断中的应用。*Annals of Statistics* 32 6 2385 2411, 2004 , ( ): – .[26] Pierre Moral。 *Feynman-Kac formulae: genealogical and interacting particle systems with applications* 2004 。Springer, 。*The*

[27] Artemy Kolchinsky 和 Brendan D Tracey。利用成对距离估计混合熵。 *Entropy* 19 7 361, , ( ): .[28] Keinosuke Fukunaga。 *Introduction to statistical pattern recognition* 2013 。Elsevier, 。2017

[29]Thomas M Cover 和 Joy A Thomas。 *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)* 2006 。Wiley-Interscience, 美国, 。[30]T Tin Nguyen、Hien D Nguyen、Faicel Chamroukhi 和 Geoffrey J McLachlan。由在无穷远处趋于零的连续密度函数有限混合进行逼近。 *Cogent Mathematics & Statistics* 7 1 1750861, 2020 , ( ): .[31] Andriy NORETS。通过平滑回归混合逼近条件密度。 *Annals of statistics*, 38 3 1733 1766, 2010 ( ): – .[32]Xiao Ma 、 Peter Karkus 、 David Hsu 、 Wee Sun Lee 和 Nan Ye。面向复杂部分观测的判别式粒子滤波强化学习。 *arXiv preprint arXiv:2002.09884* 2020 , .[33] Michael George Bulmer。 *Principles of statistics*。Courier Corporation, 2012年。[34]Timothy P Lillicrap、Jonathan J Hunt、Alexander Pritzel、Nicolas Heess、Tom Erez、Yuval Tassa、David Silver 和 Daan Wierstra。基于深度强化学习的连续控制。 *arXiv preprint arXiv:1509.02971*, 2015年。[35]Jingyuan Zhou 和 Kaidi Yang。混合自主车队控制的参数隐私保护策略。 *Transportation Research Part C: Emerging Technologies*, 169:104885, 2024年。[36]Jiawei Wang、Yang Zheng、Chaoyi Chen、Qing Xu 和 Keqiang Li。混合交通流中的领航巡航控制：系统建模、可控性与串稳定性。*IEEE Transactions on Intelligent Transportation Systems*, 23 8 12861( ): –12876, 2021 年。

[37] 姜锐、吴清松和朱作金。跟驰理论的全速度差模型。*Physical Review E* 64 1 017101, 2001 , ( ): .

[38] D.P. Bowyer、D.C. Biggs 和 R. Ackelik。*Guide to Fuel Consumption Analyses for Urban Traffic Management* 。ARRB内部报告。澳大利亚道路研究局，1985。

第2节中互信息化简的证明

根据条件互信息的链式法则，我们有：

$$I^{\boldsymbol{\pi}}(\Theta; Y_{1:T}, W_{1:T}) = \sum_{t=1}^{T} I^{\boldsymbol{\pi}}(\Theta; Y_t, W_t \mid Y_{1:t-1}, W_{1:t-1})$$
$$= \sum_{t=1}^{T} \Big(I^{\boldsymbol{\pi}}(\Theta; Y_t \mid Y_{1:t-1}, W_{1:t}) + I^{\boldsymbol{\pi}}(\Theta; W_t \mid Y_{1:t-1}, W_{1:t-1}) \Big) \tag{A.1}$$

我们只需证明 $I^{\boldsymbol{\pi}}(\Theta; W_t \mid Y_{1:t-1}, W_{t-1}) = 0$。这是因为给定 $(W_{1:t-1}, Y_{1:t-1})$ 时，W_t 与 Θ 相互独立；根据自动驾驶车辆的动态特性，Θ 对 W_t 的影响是通过 $(W_{1:t-1}, Y_{1:t-1})$ 传递的。

附录 B 第3节中的证明

B.1 定理*1*的证明

为证明定理1，我们首先证明

$$p^{\pi}(\Theta, X_{t:t+1}, W_{1:t}, Y_{1:t}) = p^{\pi_s}(\Theta, X_{t:t+1}, W_{1:t}, Y_{1:t}). \tag{B.1}$$

为此，我们将 $p^{\pi}(\Theta, X_{t:t+1}, W_{1:t}, Y_{1:t})$ 分解为

$$\begin{aligned}
&p^{\pi}(\Theta, X_{t:t+1}, W_{1:t}, Y_{1:t}) \\
&= p(X_{t+1} \mid \Theta, X_t, W_t, Y_{1:t})\; p^{\pi}(\Theta, X_t, W_{1:t}, Y_{1:t}) \\
&= p(X_{t+1} \mid \Theta, X_t, W_t)\; \pi_t(Y_t \mid \Theta, X_t, W_{1:t}, Y_{1:t-1}) \\
&\quad \cdot p^{\pi}(\Theta, X_t, W_{1:t}, Y_{1:t-1}) \\
&= p(X_{t+1} \mid \Theta, X_t, W_t)\; \pi_t(Y_t \mid \Theta, X_t, W_{1:t}, Y_{1:t-1}) \\
&\quad \cdot p(W_t \mid W_{t-1}, Y_{t-1})\; p^{\pi}(\Theta, X_t, W_{1:t-1}, Y_{1:t-1}) \\
&= p(X_{t+1} \mid \Theta, X_t, W_t)\; \pi_t(Y_t \mid \Theta, X_t, W_{1:t}, Y_{1:t-1}) \\
&\quad \cdot p(W_t \mid W_{t-1}, Y_{t-1}) \\
&\quad \cdot \int_{x_{t-1}} p^{\pi}(\Theta, X_{t-1}, x_t, W_{1:t-1}, Y_{1:t-1})\, dx_{t-1}.
\end{aligned} \tag{B.2}$$

其中，第二个等式由数据提供方系统动力学的马尔可夫性质得出。

对于任意 $\pi_t \in \Pi$，我们都可以选择一个简化策略 $\pi_t^s \in \Pi_s$，使得：

$$\begin{aligned}
&\pi_t^s(Y_t \mid \Theta, X_t, W_{1:t}, Y_{1:t-1}) \\
&= \pi_{t, Y_t \mid \Theta, X_t, W_{1:t}, Y_{1:t-1}}(Y_t \mid \Theta, X_t, W_{1:t}, Y_{1:t-1}).
\end{aligned} \tag{B.3}$$

我们用归纳法证明 (B.1)。对于 $t = 1$，$\pi_1(Y_1 \mid \Theta, X_1, W_1) = \pi_1^s(Y_1 \mid \Theta, X_1, W_1)$ 成立，因为我们有 $p^{\pi}(\Theta, X_1, X_2, W_1, Y_1) = p(X_2 \mid \Theta, X_1, W_1)\pi_1(Y_1 \mid \Theta, X_1, W_1)p(\Theta, X_1, W_1)$。因此，归纳法可以从 $p^{\pi}(\Theta, X_1, X_2, W_1, Y_1) = p^{\pi^s}(\Theta, X_1, X_2, W_1, Y_1)$ 开始。在归纳步骤中，假设 $p^{\pi}(\Theta, X_{t-1:t}, W_{1:t-1}, Y_{1:t-1}) = p^{\pi_s}(\Theta, X_{t-1:t}, W_{1:t-1}, Y_{1:t-1})$，则由 (B.2) 可自然推出 (B.1)。

结合式 (B.1)，我们证明互信息的等价性。我们有

$$\begin{aligned}
p^{\pi}(\Theta, W_{1:t}, Y_{1:t}) &= \int_{x_{t:t+1}} p^{\pi}(\Theta, x_{t:t+1}, W_{1:t}, Y_{1:t})\, dx_{t:t+1} \\
&= \int_{x_{t:t+1}} p^{\pi^s}(\Theta, x_{t:t+1}, W_{1:t}, Y_{1:t})\, dx_{t:t+1} \\
&= p^{\pi^s}(\Theta, W_{1:t}, Y_{1:t}),
\end{aligned} \tag{B.4}$$

根据互信息的定义，可得 $I^{\pi}(\theta; Y_{1:t}, W_{1:t}) = I^{\pi_s}(\theta; Y_{1:t}, W_{1:t})$。

接下来，我们证明策略 π 和 π_s 下控制性能的等价性，即 $\mathbb{E}^{\boldsymbol{\pi}}[r(\theta^*, X_t, X_{t+1}, Y_t)]$。注意

$$\begin{aligned}
&p^{\pi}(\theta^*, X_t, X_{t+1}, Y_t) \\
&= \int_{w_{1:t}, y_{1:t-1}} p^{\pi}(\theta^*, X_{t:t+1}, w_{1:t}, Y_t, y_{1:t-1})\, dy_{1:t-1}\, dw_{1:t} \\
&= \int_{w_{1:t}, y_{1:t-1}} p^{\pi_s}(\theta^*, X_{t:t+1}, w_{1:t}, Y_t, y_{1:t-1})\, dy_{1:t-1}\, dw_{1:t} \\
&= p^{\pi^s}(\theta^*, X_t, X_{t+1}, Y_t),
\end{aligned} \tag{B.5}$$

其中，第二个等式由（B.1）得出。随后，根据期望的定义即可证明控制性能等价。

同样，要证明策略 π 和 π_s 下数据偏差约束的等价性，只需证明 $p^{\pi}(X_t, Y_t) = p^{\pi^s}(X_t, Y_t)$。通过计算等式两边的边缘概率，即可由（B.4）得出这一结论。

B.2 信念状态更新的推导

信念状态更新使用贝叶斯定理推导。以其后验概率 $\beta_{t+1}(\Theta, X_{t+1})$ 为起点，可得到（B.8）中给出的递归更新方程。

注意，第三个等式（B.6）由以下关系得出

$$p(W_{t+1} \mid \Theta, X_{t:t+1}, W_{1:t}, Y_{1:t}) = p(W_{t+1} \mid W_t, Y_t),$$

这是因为在给定 (W_t, Y_t) 的条件下，W_{t+1} 与 $(\Theta, X_{t:t+1}, Y_{1:t-1}, W_{1:t-1})$ 条件独立。因此，可以进一步分解 (B.8) 中的 $\beta_{t+1}(\Theta, X_{t+1})$。

$$\beta_{t+1}(\Theta, X_{t+1}) = p(\Theta, X_{t+1} \mid W_{1:t+1}, Y_{1:t}) = \frac{p(\Theta, X_{t+1}, W_{t+1}, Y_t \mid W_{1:t}, Y_{1:t-1})}{p(W_{t+1}, Y_t \mid W_{1:t}, Y_{1:t-1})}$$
$$= \frac{p(W_{t+1} \mid W_t, Y_t)p(\Theta, X_{t+1}, Y_t \mid W_{1:t}, Y_{1:t-1})}{p(W_{t+1} \mid W_t, Y_t)p(Y_t \mid W_{1:t}, Y_{1:t-1})} = \frac{p(\Theta, X_{t+1}, Y_t \mid W_{1:t}, Y_{1:t-1})}{p(Y_t \mid W_{1:t}, Y_{1:t-1})} \tag{B.6}$$
$$= \frac{\int_{x_t} p(\Theta, x_t, X_{t+1}, Y_t \mid W_{1:t}, Y_{1:t-1})dx_t}{\int_{x_{t:t+1}} \int_\theta p(\theta, x_{t:t+1}, Y_t \mid W_{1:t}, Y_{1:t-1})d\theta dx_{t:t+1}} = \frac{\int_{x_t} p(X_{t+1} \mid \Theta, x_t, W_t)p(\Theta, x_t, Y_t \mid W_{1:t}, Y_{1:t-1})dx_t}{\int_{x_{t:t+1}} \int_\theta p(x_{t+1} \mid \theta, x_t, W_t)p(\theta, x_t, Y_t \mid W_{1:t}, Y_{1:t-1})d\theta dx_{t:t+1}}$$
$$= \frac{\int_{x_t} p(X_{t+1} \mid \Theta, x_t, W_t)p(Y_t \mid \Theta, x_t, W_{1:t}, Y_{1:t-1})p(\Theta, x_t \mid W_{1:t}, Y_{1:t-1})dx_t}{\int_{x_{t:t+1}} \int_\theta p(x_{t+1} \mid \theta, x_t, W_t)p(Y_t \mid \theta, x_t, W_{1:t}, Y_{1:t-1})p(\theta, x_t \mid W_{1:t}, Y_{1:t-1})d\theta dx_{t:t+1}} \tag{B.7}$$
$$= \frac{\int_{x_t} p(X_{t+1} \mid \Theta, x_t, W_t)\pi_t^s(Y_t \mid \Theta, x_t, W_{1:t}, Y_{1:t-1})\beta_t(\Theta, x_t)dx_t}{\int_{x_{t:t+1}} \int_\theta p(x_{t+1} \mid \theta, x_{t+1}, W_t)\pi_t^s(Y_t \mid \theta, x_t, W_{1:t}, Y_{1:t-1})\beta_t(\theta, x_t)d\theta dx_{t:t+1}}. \tag{B.8}$$

B.3 引理 *2* 的证明

根据策略 π_t^s 和信念状态 β_t，贝尔曼最优性方程 (B.11) 可分解为即时成本项和期望未来价值。由于已知 h_t，我们将记号 $\pi_t^s(Y_t \mid \Theta, X_t, h_t)$ 简化为 $\pi_t^s(Y_t \mid \Theta, X_t)$。贝尔曼方程为：

$$V_t^*(h_t) = \min_{\pi_t^s} \left\{ C_t(h_t, \pi_t^s) + \mathbb{E}\left[V_{t+1}^*(h_{t+1}) \mid h_t, \pi_t^s \right] \right\}, \tag{B.9}$$

其中期望是相对于观测 Y_t 和外生噪声 W_{t+1}取的。我们将即时成本 $C_t(h_t, \pi_t^s)$定义为隐私损失（互信息）、系统成本以及数据可用性惩罚之和：

$$\begin{aligned} C_t(h_t, \pi_s^t) &= \int_{y_t, \theta} \int_{x_t} \pi_t^s(Y_t \mid \theta, x_t)\beta_t(\theta, x_t)dx_t \\ &\cdot \log \frac{\int_{x_t} \pi_t^s(Y_t \mid \theta, x_t)\beta_t(\theta, x_t)dx_t}{\beta_t(\theta) \int_\theta \int_{x_t} \pi_t^s(Y_t \mid \theta, x_t)\beta_t(\theta, x_t)dx_t d\theta} d\theta dy_t \\ &+ \int_{x_{t:t+1}} \int_{y_t} r(\theta^*, x_{t:t+1}, y_t)p(x_{t+1} \mid \theta^*, x_t, w_t) \\ &\cdot \pi_t^s(Y_t \mid \theta^*, x_t)\beta_t(\theta^*, x_t)dy_t dx_{t:t+1} \\ &+ \lambda \int_{x_t} \int_{\theta, y_t} \pi_t^s(Y_t \mid \theta, x_t)\beta_t(\theta, x_t)d(x_t, y_t)dy_t d\theta dx_t \\ &- \lambda \hat{D} \end{aligned} \tag{B.10}$$

由 C_t的每一部分可以看出，给定历史 h_t，C_t 由信念状态 β_t 和策略 π_t^s决定。因此，我们用 $C_t(h_t, \pi_s^t)$表示在历史 h_t下采取策略集 π_t^s 时的即时成本 C_t。

基于这一依赖关系，我们可以用归纳法假设 $V_{t+1}^*(h_{t+1})$ 也依赖于信念状态 β_{t+1}。

然后，我们将贝尔曼最优性方程改写为：

$$V_t^*(h_t) = \min_{\pi_t^s} \{ C_t(h_t, \pi_t^s) + \mathbb{E}(V_{t+1}^*(\beta_{t+1} \mid h_t)) \}. \tag{B.11}$$

使用 $\beta_{t+1} = \Phi(\beta_t, \pi_t^s, y_t)$来紧凑地表示该递推，我们可以将第二项$\mathbb{E}(V_{t+1}^*(\beta_{t+1} \mid h_t))$分解为：

$$\begin{aligned} &\mathbb{E}(V_{t+1}^*(\beta_{t+1} \mid h_t)) = \int_{y_t} p(y_t \mid h_t)V_{t+1}^*(\beta_{t+1})dy_t \\ &= \int_{\theta, x_t, y_t} \pi_t^s(y_t \mid \theta, x_t)\beta_t(\theta, x_t)V_{t+1}^*(\Phi(\beta_t, \pi_t^s, y_t))dx_t d\theta dy_t, \end{aligned}$$

该项同样依赖于 β_t。因此，将该项与 $C_t(h_t, \pi_s^t)$相结合并使用归纳法，我们可以证明$V_t^*(h_t)$依赖于信念状态 β_t。

C 第 4 节的证明

C.1 引理 *4* 的证明

我们用式 (9) 中定义的粒子测度 $\hat{}\beta_t^{(N)}$ 近似信念状态 β_t，即 $\hat{}\beta_t^{(N)}(d\theta, dx) = \sum_{i=1}^N \omega_{i,t}\delta_{(\theta_i, x_{i,t})}(d\theta, dx)$，它为 $p(\Theta, X_t \mid Y_{1:t}, W_{1:t})$ 提供了一种离散加权近似。

权重 $\{\omega_{i,t}\}_{i=1}^N$ 根据重要性采样原理选取。假设 $p(\Theta, X_t \mid Y_{1:t-1}, W_{1:t})$ 是一种难以从中抽取样本的概率密度。我们定义 θ_i 和 X_t 的易于采样的提议联合分布（即重要性密度）为 $q(\theta_i, X_{i,t} \mid Y_{1:t-1}, W_{1:t})$。我们从新分布中获取样本 $(\theta_i, X_{i,t})$，并将权重 $\omega_{i,t}$ 更新为 $\omega_{i,t} \propto {}^{p}(\theta_i, X_{i,t} \mid Y_{1:t-1}, W_{1:t})\, q(\theta_i, X_{i,t} \mid Y_{1:t-1}, W_{1:t})$。提议联合分布 $q(\Theta, X_t \mid W_{1:t}, Y_{1:t-1})$ 的选取方式是使其具有一个

遵循贝叶斯规则的因式分解形式：

$$q(\Theta, X_t \mid W_{1:t}, Y_{1:t-1}) = q(X_t \mid \Theta, X_{t-1}, W_{1:t}, Y_{1:t-1}) \cdot q(\Theta, X_{t-1} \mid W_{1:t-1}, Y_{1:t-2}) \tag{C.1}$$

同理，联合密度 $p(\Theta,\ X_t \mid Y_{1:t},\ W_{1:t})$ 可分解为：

$$p(\Theta, X_t \mid Y_{1:t}, W_{1:t}) \propto p(\Theta, X_{t-1} \mid W_{1:t-1}, Y_{1:t-2}) \cdot p(W_t \mid W_{t-1}, Y_{t-1}) p(X_t \mid \Theta, X_{t-1}, W_{t-1}) \cdot p(Y_{t-1} \mid \Theta, X_{t-1}, W_{1:t-1}, Y_{1:t-2}) \tag{C.2}$$

利用式(C.1)和(C.2)中的因式分解形式，我们可以推导出如(C.3)所示的递归权重更新方案。

假设密度 $q(X_{i,t} \mid \theta_i, X_{i,t-1}, W_{1:t}, Y_{1:t-1}) = p(X_{i,t} \mid \theta_i, X_{i,t-1}, W_{1:t-1})$，则可以得出权重更新的最终过程

$$\omega_{i,t} \propto \omega_{i,t-1} p(W_t \mid W_{t-1}, Y_{t-1}) \cdot p(Y_{t-1} \mid \theta_i, X_{i,t-1}, W_{1:t-1}, Y_{1:t-2}). \tag{C.4}$$

C.2 引理7的证明

我们首先用粒子信念分量显式表示 $C_t^{(N)}(\hat{\beta}_t^{(N)}, \mathcal{K}_t)$。由于 Θ 在有限集中取值，给定信念 $\beta(\theta^j, dx) = q(\theta^j) b(dx \mid \theta^j)$，定义诱导密度

$$f_{Y|\Theta}(y \mid \theta^j; \beta, u) = \int k_t(y \mid \theta^j, x, \beta; u) b(dx \mid \theta^j),$$
$$p_Y(y; \beta, u) = \sum_{j=1}^{K} q(\theta^j) f_{Y|\Theta}(y \mid \theta^j; \beta, u).$$

应用粒子信念$\hat{\beta}_t^{(N)}$，我们得到

$$f_{Y|\Theta}^{(N)}(y \mid \theta^j; u) = \int k_t(y \mid \theta^j, x, \hat{\beta}_t^{(N)}; u) \hat{b}_t^{(N)}(dx \mid \theta^j),$$
$$p_Y^{(N)}(y; u) = \sum_{j=1}^{K} \hat{q}_t^{(N)}(\theta^j) f_{Y|\Theta}^{(N)}(y \mid \theta^j; u).$$

因此，在粒子信念β_t^N 和参数 u下的逐步互信息为

$$I^{(N)}(\Theta; Y_t \mid h_t, u) = \sum_{j=1}^{K} \hat{q}_t^{(N)}(\theta^j) \int f_{Y|\Theta}^{(N)}(y \mid \theta^j; u) \log \frac{f_{Y|\Theta}^{(N)}(y \mid \theta^j; u)}{p_Y^{(N)}(y; u)} dy. \tag{C.5}$$

类似地，系统代价项 $C_{t\,1}^{(N)}$,和数据失真项 $C_{t\,2}^{(N)}$,的粒子近似通过将 β_t 替换为$\hat{\beta}_t^{(N)}$ 得到。令 x' 为 x_{t+1}的简写。则

$$C_{t,1}^{(N)}(\hat{\beta}_t^{(N)}, u) = \iiint r(\theta^*, x, x', y) p(dx' \mid \theta^*, x, w_t) \cdot k_t(y \mid \theta^*, x, \hat{\beta}_t^{(N)}; u) dy \hat{\beta}_t^{(N)}(\theta^*, dx)$$

$$C_{t,2}^{(N)}(\hat{\beta}_t^{(N)}, u) = \sum_{\theta \in \mathcal{S}_\Theta} \iint d(x, y) k_t(y \mid \theta, x, \hat{\beta}_t^{(N)}; u) \cdot dy \hat{\beta}_t^{(N)}(\{\theta\}, dx) - \lambda \hat{D}$$

为简洁起见，我们只展示 $I^{(N)}(\Theta; Y_t \mid h_t,\ u)$的收敛性。其他项 $C_{t\,1}^{(N)}$, $(\hat{\beta}_t^{(N)},\ u)$和$C_{t\,2}^{(N)}$, $(\hat{\beta}_t^{(N)},\ u)$的证明类似。

由于 $\mathcal{S}_\Theta = \{\theta^1, \ldots, \theta^K\}$ 是有限的，引理5中的联合弱收敛立即推出边缘信念分量和条件信念分量的弱收敛。

对每个 j，在X上定义边缘测度为$\hat{\beta}_{t,\ j}^{(N)}(A) := \hat{\beta}_t^{(N)}(\{\theta^j\} \times A)$和$\beta_{t,\ j}(A) := \beta_t(\{\theta^j\} \times A)$。任取X上的有界连续函数 g，并定义 $h(\theta,\ x) = g(x)\mathbf{1}_{\{\theta=\theta^j\}}$。示性函数 $\mathbf{1}_{\{\theta=\theta^j\}}$ 在 Θ上连续，因此 h也是有界连续函数。由$\hat{\beta}_t^{(N)} \Rightarrow \beta_t$ 几乎必然地，有

$$\int h(\theta, x)\, \hat{\beta}_t^{(N)}(d\theta, dx) \longrightarrow \int h(\theta, x)\, \beta_t(d\theta, dx).$$

几乎必然。取 $g \equiv 1$ 可得 $\hat{q}_t^{(N)}(\theta^j) = \hat{\beta}_{t,\ j}^{(N)}(\mathbf{X}) \to \beta_{t,\ j}(\mathbf{X}) = q_t(\theta^j)$ 对所有 j 几乎必然成立。此外，如果$q_t(\theta^j) > 0$

$$\int g(x)\, \hat{b}_t^{(N)}(dx \mid \theta^j) = \frac{\int g\, d\hat{\beta}_{t,j}^{(N)}}{\hat{q}_t^{(N)}(\theta^j)} \longrightarrow \frac{\int g\, d\beta_{t,j}}{q_t(\theta^j)} = \int g(x)\, b_t(dx \mid \theta^j), \qquad \text{a.s.}$$

这表明 $\hat{b}_t^{(N)}(\cdot \mid \theta^j) \Rightarrow b_t(\cdot \mid \theta^j)$ 几乎必然成立。

接下来，固定 θ^j 和 y。根据假设 6(ii)–(iii)，映射 $x \mapsto k_t(y \mid \theta^j,\ x,\ \beta;\ u)$ 在 X 上连续，且对所有 x 均满足$k_t(y) \le k_t(\underline{y} \mid \theta^j,\ x,\ \beta;\ u) \le k_t(y)$。因此，利用上述条件弱收敛，我们得到

$$f_{Y|\Theta}^{(N)}(y \mid \theta^j; u) = \int k_t(y \mid \theta^j, x, \hat{\beta}_t^{(N)}; u)\, \hat{b}_t^{(N)}(dx \mid \theta^j) \longrightarrow \int k_t(y \mid \theta^j, x, \beta_t; u)\, b_t(dx \mid \theta^j) = f_{Y|\Theta}(y \mid \theta^j; u)$$

$$\omega_{i,t} \propto \frac{\omega_{i,t-1} p(W_t \mid W_{t-1}, Y_{t-1}) p(Y_{t-1} \mid \theta_i, X_{i,t-1}, W_{1:t-1}, Y_{1:t-2}) p(X_{i,t} \mid \theta_i, X_{i,t-1}, W_{t-1})}{q(X_{i,t} \mid \theta_i, X_{i,t-1}, W_{1:t}, Y_{1:t-1})} \tag{C.3}$$

几乎必然。将此结果与 $\hat{q}_t^{(N)}(\theta^j) \to q_t(\theta^j)$ 结合，可得边缘密度收敛

$$p_Y^{(N)}(y;u) = \sum_{j=1}^{K} \hat{q}_t^{(N)}(\theta^j)\, f_{Y|\Theta}^{(N)}(y \mid \theta^j; u)$$
$$\longrightarrow \sum_{j=1}^{K} q_t(\theta^j)\, f_{Y|\Theta}(y \mid \theta^j; u) = p_Y(y;u), \qquad \text{a.s.}$$

最后，包络界通过取平均继承：

$$\underline{k}_t(y) \le f_{Y|\Theta}^{(N)}(y \mid \theta^j; u) \le \overline{k}_t(y), \tag{C.7a}$$

$$\underline{k}_t(y) \le p_Y^{(N)}(y;u) \le \overline{k}_t(y), \tag{C.7b}$$

且极限也满足相同的界。接着定义逐点对数比 $\ell_N(\theta^j, y; u) = \log f_{Y|\Theta}^{(N)}(y \mid \theta^j; u) - \log p_Y^{(N)}(y; u)$，并

$\ell(\theta^j, y; u) = \log f_{Y|\Theta}(y \mid \theta^j; u) - \log p_Y(y;u)$。对于每个 θ^j 和每个 y，$\ell_N(\theta^j, y; u) \to \ell(\theta^j, y; u)$ 几乎必然成立。此外，由一致界 (C.7)，

$$|\ell_N(\theta^j, y; u)| = \left| \log \frac{f_{Y|\Theta;u}^{(N)}(y \mid \theta^j; u)}{p_Y^{(N)}(y;u)} \right|$$
$$\le |\log \overline{k}_t(y)| + |\log \underline{k}_t(y)|$$

因此，(C.5) 中的被积函数由 $\left|\hat{q}_t^{(N)}(\theta^j)\, f_{Y|\Theta}^{(N)}(y \mid \theta^j)\, \ell_N(\theta^j, y)\right| \le k_t(y)\Big(|\log k_t(y)| + |\log k_t(y)|\Big)$ 控制。定义 $g_t = k_t(y)\Big(|\log k_t(y)| + |\log k_t(y)|\Big)$，且根据假设 6(ii)，$g_t$ 在 $\mathbb{R}^N$ 上可积。由于 $\ell_N(\theta^j, y) \to \ell(\theta^j, y)$ 逐点几乎必然成立，将控制收敛定理 (DCT) 应用于 (C.5) 可得 $I^{(N)}(\Theta; Y_t \mid h_t) \to I(\Theta; Y_t \mid h_t)$ 几乎必然成立。

定理*8*的证明

根据证明概要，我们只需证明 $\Delta_N^{(C)}$ 和 $\Delta_N^{(V)}$ 一致收敛于 0。根据引理7，对于每个 $u \in \mathcal{U}_t$，$C_t^{(N)}(\hat{\beta}_t^{(N)}, u) \to C_t(\beta_t, u)$ 几乎必然成立。下面的引理可以将这一逐点收敛加强为在 $\mathcal{U}_t$ 上的一致收敛。

引理14（逐点到一致提升）令$\mathcal{U}$ 为紧集。假设 $f_N: \mathcal{U} \to \mathbb{R}$ 是等度连续的，并且 $f_N(u) \to f(u)$ 在 $\mathcal{U}$上逐点成立，其中 f是连续的。那么$\sup_{u\in\mathcal{U}} |f_N(u) - f(u)| \to 0$。

证明。假设该收敛并非一致收敛。那么存在 $\varepsilon > 0$ 以及一个子列 $\{f_{N_k}\}$，使得

$$\sup_{u\in\mathcal{U}} |f_{N_k}(u) - f(u)| \ge \varepsilon, \qquad \forall k.$$

由于 $\mathcal{U}$ 是紧的且 $\{f_N\}$ 是等度连续的，Arzel`a–Ascoli 定理给出了一个进一步的子序列 $\{f_{N_{k_j}}\}$，其一致收敛于某个 $g \in C(\mathcal{U})$。特别地，对于每个 $u \in \mathcal{U}$，$f_{N_{k_j}}(u) \to g(u)$。但是，根据假设，$f_{N_{k_j}}(u) \to f(u)$ 逐点成立，因此 $g \equiv f$。所以 $\sup_{u\in\mathcal{U}} |f_{N_{k_j}}(u) - f(u)| \to 0$，这与 ε 的选取相矛盾。

注意到假设 6(iv) 意味着关于 u 的利普希茨连续性：存在 $L_C < \infty$，使得对于所有 N 以及所有 $u,\ v \in \mathcal{U}_t$，

$$\left|C_t^{(N)}(\hat{\beta}_t^{(N)}, u) - C_t^{(N)}(\hat{\beta}_t^{(N)}, v)\right| \le L_C \|u - v\|, \tag{C.8a}$$
$$\left|C_t(\beta_t, u) - C_t(\beta_t, v)\right| \le L_C \|u - v\|. \tag{C.8b}$$

由于 $\mathcal{U}_t$ 是紧的，族 $\{u \mapsto C_t^{(N)}(\hat{\beta}_t^{(N)}, u)\}_N$ 在 $\mathcal{U}_t$ 上等度连续。此外，根据引理7，对于每个固定的 $u \in \mathcal{U}_t$，$C_t^{(N)}(\hat{\beta}_t^{(N)}, u) \to C_t(\beta_t, u)$，且 $u \mapsto C_t(\beta_t, u)$ 连续。因此，由引理14可得

$$\Delta_N^{(C)} = \sup_{u\in\mathcal{U}_t} \left|C_t^{(N)}(\hat{\beta}_t^{(N)}, u) - C_t(\beta_t, u)\right| \to 0 \qquad \text{a.s..}$$

接下来证明 $\Delta_N^{(V)} \to 0$。根据三角不等式

$$\begin{aligned}
\Delta_N^{(V)} &\le \sup_{u_t\in\mathcal{U}_t} \left|\mathbb{E}\big[V_{t+1}^{*,(N)}(\hat{\beta}_{t+1}^{(N)}) - V_{t+1}^{*}(\hat{\beta}_{t+1}^{(N)}) \mid h_t, u_t\big]\right| \\
&\quad + \sup_{u_t\in\mathcal{U}_t} \left|\mathbb{E}\big[V_{t+1}^{*}(\hat{\beta}_{t+1}^{(N)}) - V_{t+1}^{*}(\beta_{t+1}) \mid h_t, u_t\big]\right| \\
&\le \sup_{u_t\in\mathcal{U}_t} \Big(\mathbb{E}\Big[\left|V_{t+1}^{*,(N)}(\hat{\beta}_{t+1}^{(N)}) - V_{t+1}^{*}(\beta_{t+1})\right| \mid h_t, u_t\Big] \\
&\quad + \mathbb{E}\Big[\left|V_{t+1}^{*}(\beta_{t+1}) - V_{t+1}^{*}(\hat{\beta}_{t+1}^{(N)})\right| \mid h_t, u_t\Big]\Big) \\
&\quad + \sup_{u_t\in\mathcal{U}_t} \left|\mathbb{E}\big[V_{t+1}^{*}(\hat{\beta}_{t+1}^{(N)}) - V_{t+1}^{*}(\beta_{t+1}) \mid h_t, u_t\big]\right|.
\end{aligned}$$

对于第一项，我们有 (i) $V_{t+1}^{*,(N)} \to V_{t+1}^{*}$ 在 β_{t+1} 上逐点成立；(ii) $\hat{\beta}_{t+1}^{(N)} \to \beta_{t+1}$ 几乎必然成立；以及 (iii) 对于固定的 u_t，$V_{t+1}^{*}(\cdot, u_t)$ 是等度连续的（根据归纳法假设，对这个固定的 V_{t+1}^{*} 应用），几乎必然成立。由于粒子信念更新的一致性，几乎必然成立。此外，有限时域值函数一致有界。因此，根据

支配收敛定理，第一项中的条件期望对每个固定的 u_t 都收敛到 0。为将这种逐点收敛提升为关于 $u_t \in \mathcal{U}_t$ 的一致收敛，注意到假设 6(iv) 表明映射 $u_t \mapsto \mathbb{E}[V^{*,(N)}_{t+1}(\hat{\beta}^{(N)}_{t+1}) - V^{*}_{t+1}(\hat{\beta}^{(N)}_{t+1}) \mid h_t,\ u_t]$ 在 $\mathcal{U}_t$ 上是利普希茨连续的（因而等度连续），且其常数与 N 无关。由于 $\mathcal{U}_t$ 是紧集，引理14表明第一项关于 $u_t \in \mathcal{U}_t$ 一致收敛到 0。

对于第二项，同样的论证也适用。对于每个固定的 u_t，粒子信念更新的一致性意味着 $V^{*}_{t+1}(\hat{\beta}^{(N)}_{t+1}) \to V^{*}_{t+1}(\beta_{t+1})$ 几乎必然成立，而支配收敛定理给出相应条件期望逐点收敛至 0。此外，假设 6(iv) 意味着关于 u_t 的利普希茨性质在 N 上一致，因此根据引理14，该收敛在 $u_t \in \mathcal{U}_t$ 上一致。

结合上述结果可得 $\Delta_N \to 0$ 几乎必然成立。因此，根据 (13)，我们得到 $|V^{*,(N)}_t(\hat{\beta}^{(N)}_t) - V^{*}_t(\beta_t)| \to 0$ 几乎必然成立。由此完成归纳步骤及证明。□

C.4 推论 *10* 的证明

基于粒子的互信息与其易于处理的上界之间的差距为

$$
\begin{aligned}
&\tilde{I}^{(N)}(u) = \widehat{H}^{(N)}_{\mathrm{KL}}(Y_t \mid h_t; u) \\
&\quad - \sum_{\theta \in \mathcal{S}_\Theta} \hat{q}^{(N)}_t(\theta) \widehat{H}^{(N)}_{C_\alpha}\big(Y_t \mid \theta, h_t; u\big) - I^{(N)}(\Theta; Y_t \mid h_t; u) \\
&= \Big(\widehat{H}^{(N)}_{\mathrm{KL}}(Y_t \mid h_t; u) - H^{(N)}(Y_t \mid h_t; u)\Big) \\
&\quad - \sum_{\theta \in \mathcal{S}_\Theta} \hat{q}^{(N)}_t(\theta) \Big(\widehat{H}^{(N)}_{C_\alpha}(Y_t \mid \theta, h_t; u) - H^{(N)}(Y_t \mid \theta, h_t; u)\Big) \\
&\le \Big(\widehat{H}^{(N)}_{\mathrm{KL}}(Y_t \mid h_t; u) - H^{(N)}_{C_\alpha}(Y_t \mid h_t; u)\Big) \\
&\quad - \sum_{\theta \in \mathcal{S}_\Theta} \hat{q}^{(N)}_t(\theta) \Big(\widehat{H}^{(N)}_{C_\alpha}(Y_t \mid \theta, h_t; u) - H^{(N)}_{\mathrm{KL}}(Y_t \mid \theta, h_t; u)\Big).
\end{aligned}
$$

在簇条件下，[27] 已经证明 $\widehat{H}^{(N)}_{\text{KL散度}}(Y_t \mid h_t; u) - H^{(N)}_{C_\alpha}(Y_t \mid h_t; u) \le \kappa + (|G| - 1)e^{\big(-(1-|1-2\alpha|)\gamma\big)}$。当以 θ 为条件时，我们有 $\widehat{H}^{(N)}_{C_\alpha}(Y_t \mid \theta, h_t; u) - H^{(N)}_{\text{KL散度}}(Y_t \mid \theta, h_t; u) \le \kappa_\theta + (|G_\theta| - 1)e^{\big(-(1-|1-2\alpha_\theta|)\gamma_\theta\big)}$。由此可得

$$
\begin{aligned}
&\tilde{I}^{(N)}(u) \le \kappa + (|G| - 1)e^{\big(-(1-|1-2\alpha|)\gamma\big)} + \\
&\sum_{\theta \in \mathcal{S}_\Theta} \hat{q}^{(N)}_t(\theta) \Big[\kappa_\theta + (|G_\theta| - 1)e^{\big(-(1-|1-2\alpha_\theta|)\gamma_\theta\big)}\Big].
\end{aligned}
$$

□

C.5 定理 *11* 的证明

考虑一个具有共同协方差 Σ、权重 ω_i和均值 μ_i的同方差高斯混合。令 $S = \sum_{i=1}^{N} \omega_i(\mu_i - \bar{\mu})(\mu_i - \bar{\mu})^\top$ 为类间散度矩阵。我们定义分离矩阵 $A = \Sigma^{-1}S$，并考虑 $\|A\|_F \to 0$ 的情形（这意味着 $\mu_i \approx \mu_j$）。该混合的总协方差为 $\Sigma_{\mathrm{glob}} = \Sigma + S$。则 $\widehat{H}_G$ 可由 $\widehat{H}_{\mathrm{G}} = \frac{1}{2}\log\det(2\pi e(\Sigma + S)) = \frac{1}{2}\log\det(2\pi e\Sigma) + \frac{1}{2}\log\det(I + \Sigma^{-1}S) = H_0 + \frac{1}{2}\log\det(I + A)$ 计算得到，其中 H_0 是单个分量的熵。利用泰勒展开 $\log\det(I + A) = \mathrm{tr}(A) - \frac{1}{2}\mathrm{tr}(A^2) + O(\|A\|^3)$，我们得到：

$$
\widehat{H}_{\mathrm{G}} = H_0 + \frac{1}{2}\mathrm{tr}(A) + O(|A|^2). \tag{C.9}
$$

基于 KL 的界为 $H_{\mathrm{KL}} = H_0 - \sum_{i=1}^{N} \omega_i \log\Big(\sum_{j=1}^{N} \omega_j e^{-D_{ij}}\Big)$，其中 $D_{ij} = \mathrm{KL}(\mathcal{N}_i \| \mathcal{N}_j) = \frac{1}{2}(\mu_i - \mu_j)^\top \Sigma^{-1}(\mu_i - \mu_j)$。注意 $D_{ij} = O(\|A\|)$。利用展开式 $e^{-x} \approx 1 - x$，内层求和变为 $\sum_j \omega_j e^{-D_{ij}} = 1 - \sum_j \omega_j D_{ij} + O(\|A\|^2)$。应用 $\log(1 - x) \approx x$，H_{KL} 可以近似为 $\widehat{H}_{\mathrm{KL}} \approx H_0 + \sum_i \omega_i \Big(\sum_j \omega_j D_{ij}\Big) = H_0 + \sum_{i,j} \omega_i \omega_j \frac{1}{2}(\mu_i - \mu_j)^\top \Sigma^{-1}(\mu_i - \mu_j)$。在迹中使用 $\sum_{i,j} \omega_i \omega_j (\mu_i - \mu_j)(\mu_i - \mu_j)^\top = 2S$：

$$
\begin{aligned}
\widehat{H}_{\mathrm{KL}} &= H_0 + \frac{1}{2}\mathrm{tr}(\Sigma^{-1}(2S)) + O(|A|^2) \\
&= H_0 + \mathrm{tr}(A) + O(|A|^2).
\end{aligned} \tag{C.10}
$$

比较 (C.9) 和 (C.10)，两者之差为：$H_{\mathrm{KL}} - \widehat{H}_{\mathrm{G}} = (H_0 + \mathrm{tr}(A)) - \big(H_0 + \frac{1}{2}\mathrm{tr}(A)\big) = \frac{1}{2}\mathrm{tr}(A) + O(\|A\|^2)$。由于 Σ 和 S 均为半正定矩阵，因此 $\mathrm{tr}(A) = \mathrm{tr}(\Sigma^{-1}S) \ge 0$。因此，对于足够小的分离（此时一阶项占主导），我们有 $\widehat{H}_{\mathrm{G}} \lesssim H_{\mathrm{KL}}$。□

C.6 定理*12*的证明

对于粒子 i，假设数据共享策略是一个有限高斯混合模型，即 $k_t(y \mid \theta_i, x_{i,t}, \hat{\beta}^{(N)}_t; u) = \sum_{k=1}^{K_i} a_{i,k,t}(u)\mathcal{N}\big(y; \mu_{i,k,t}(u), \Sigma_{i,k,t}(u)\big)$，其中非负混合权重满足 $\sum_{k=1}^{K_i} a_{i,k,t}(u) = 1$，且高斯参数 $\big(\mu_{i,k,t}(u), \Sigma_{i,k,t}(u)\big)$ 可能依赖于

关于 u。Y_t 的历史条件边缘密度为

$$
\begin{aligned}
p^{(N)}(y \mid h_t; u) &= \iint k_t(y \mid \theta, x, \hat{\beta}_t^{(N)}; u)\, \hat{\beta}_t^{(N)}(d\theta, dx) \\
&= \sum_{i=1}^{N} \omega_{i,t}\, k_t\big(y \mid \theta_i, x_{i,t}, \hat{\beta}_t^{(N)}; u\big) \\
&= \sum_{i=1}^{N} \sum_{k=1}^{K_i} \tilde{\omega}_{i,k,t}(u)\, \mathcal{N}_{i,k,t}(y; u),
\end{aligned}
$$

其中 $\mathcal{N}_{i,k,t}(\cdot;\, u) = \mathcal{N}\big(\cdot; \mu_{i,k,t}(u), \Sigma_{i,k,t}(u)\big)$ 是粒子 i 的分量密度，而 $\tilde{\omega}_{i,k,t}(u) = \omega_{i,t}\, a_{i,k,t}(u)$ 表示其（全局）混合权重，满足 $\sum_{i=1}^{N} \sum_{k=1}^{K_i} \tilde{\omega}_{i,k,t}(u) = 1$。因此，$p^{(N)}(y \mid h_t; u)$ 本身是一个单一的有限高斯混合模型，其高斯分量为 $\{\mathcal{N}_{i,k,t}(\cdot; u)\}_{i,k}$，权重为 $\{\tilde{\omega}_{i,k,t}(u)\}_{i,k}$。

类似地，给定 $\Theta = \theta$ 时 Y_t 的条件密度为

$$
\begin{aligned}
p^{(N)}(y \mid \theta, h_t; u) &= \int k_t(y \mid \theta, x, \hat{\beta}_t^{(N)}; u)\, \hat{b}_t^{(N)}(dx \mid \theta) \\
&= \sum_{i \in \mathcal{I}_\theta} \bar{\omega}_{i,t}(\theta)\, k_t\big(y \mid \theta_i, x_{i,t}, \hat{\beta}_t^{(N)}; u\big) \\
&= \sum_{i \in \mathcal{I}_\theta} \sum_{k=1}^{K_i} \tilde{\omega}_{i,k,t}(\theta; u)\, \mathcal{N}_{i,k,t}(y; u),
\end{aligned}
$$

其中，$\tilde{\omega}_{i,\,k,\,t}(\theta; u) = \bar{\omega}_{i,\,t}(\theta) a_{i,\,k,\,t}(u)$ 是粒子 i 的条件混合权重，满足 $\sum_{i \in \mathcal{I}_\theta} \sum_{k=1}^{K_i} \tilde{\omega}_{i,\,k,\,t}(\theta; u) = 1$。因此，$p^{(N)}(y \mid \theta, h_t; u)$ 也是一个有限高斯混合模型（限制于索引 $i \in \mathcal{I}_\theta$）。将定理 11 应用于这些混合模型，可得到 $I^{(N)}(\Theta; Y_t \mid h_t)$ 的显式可计算上界。这证明了引理 12。□

C.7 推论 *13* 的证明

令 $\varepsilon > 0$ 任意。根据条件密度的 L^1/总变差稠密性，有限高斯混合模型 [30,31]，存在一个具有密度 $\hat{k}_{t,\,J(\delta)}(\cdot \mid \theta, x, \hat{\beta}_t^{(N)})$ 的有限高斯混合策略核 $\mathcal{K}_{J(\delta)}$，使得对于某个 $\delta > 0$，

$$
\sup_{(\theta, x)} \Big\| k_t(\cdot \mid \theta, x, \hat{\beta}_t^{(N)}) - \hat{k}_{t,J(\delta)}(\cdot \mid \theta, x, \hat{\beta}_t^{(N)}) \Big\|_1 \le \delta.
$$

在假设 6 下，条件熵和互信息是 L^1/TV-连续的。因此，存在连续性模 $\omega(\cdot)$，使得 $\omega(\delta) \to 0$ 当 $\delta \to 0$ 时

$$
\Big| I^{(N)}(\Theta; Y_t \mid h_t; \mathcal{K}_t) - I^{(N)}(\Theta; Y_t \mid h_t; \hat{\mathcal{K}}_{J(\delta)}) \Big| \le \omega(\delta). \tag{C.11}
$$

将定理 11 应用于有限高斯混合模型策略 $\mathcal{K}_{J(\delta)}$，可得可计算上界

$$
I^{(N)}(\Theta; Y_t \mid h_t; \hat{\mathcal{K}}_{J(\delta)}) \le \widehat{I}^{(N)}_{\hat{\mathcal{K}}_{J(\delta)}, \mathrm{RA}}(\Theta; Y_t \mid h_t). \tag{C.12}
$$

结合式 (C.11) 和式 (C.12) 可得

$$
I^{(N)}(\Theta; Y_t \mid h_t; \mathcal{K}_t) \le \widehat{I}^{(N)}_{\hat{\mathcal{K}}_{J(\delta)}, \mathrm{RA}}(\Theta; Y_t \mid h_t) + \omega(\delta).
$$

最后，选择 δ 使得 $\omega(\delta) \le \varepsilon$。证明完毕。□