arXiv:2305.02894v1 [cs.LG] 4 May 2023

FedCBO：通过基于共识的优化，在聚类联邦学习中达成群体共识

JOS´E A. CARRILLO、NICOL´AS GARC´IA TRILLOS、SIXU LI 和 YUHUA ZHU

Abstract. 联邦学习是现代机器学习中的一个重要框架，旨在整合多个用户的学习模型训练；每个用户都有自己的本地数据集，同时兼顾数据隐私和通信丢失约束。在聚类联邦学习中，我们假设用户之间还存在未知的群体结构，目标是训练出适用于各个群体的模型，而不是仅为所有用户训练一个全局模型。本文提出了一种受基于共识的优化（CBO）思想启发的聚类联邦学习新方法。我们提出的 CBO 型方法基于一个不依赖群体成员关系的相互作用粒子系统。该模型有严谨的数学推导作为依据，包括对粒子数量趋于无穷时粒子系统的描述所进行的平均场分析，以及在平均场状态下对一般非凸目标函数（对应于各用户群体的损失函数）进行同步全局优化的收敛保证。实验结果表明，与其他最先进的方法相比，我们的 FedCBO 算法效果显著，并为本文的方法和理论工作提供了验证。

1. Introduction

在物联网设备广泛应用于家庭自动化、个人健康监测和车联网通信等各种场景的背景下，众多用户产生了海量数据。然而，出于对数据隐私和安全的顾虑，加之通信成本和带宽方面的限制，单个用户难以利用这些大量存储的信息。这促使人们设计和开发联邦学习（FL）策略，旨在汇集在本地设备上训练的学习模型中的信息，以构建模型，而无需收集本地数据 [19, 28]。

标准的联邦学习方法旨在为所有本地客户端/用户训练一个全局模型 [21, 23, 28, 30]。然而，数据异质性（也称为非独立同分布数据）在联邦学习应用中自然存在，因为数据通常由用户的个人设备生成。因此，可以预期，不存在一个能够在所有客户端上都表现良好的全局模型 [34]。另一方面，可以合理地认为，背景相似的用户可能会做出相似的决策，从而生成符合相似分布的数据。本文研究一种针对非独立同分布数据的联邦学习方案，即聚类联邦学习（CFL） [15, 25, 26, 33, 34]。在聚类联邦学习中，用户被划分为不同的簇，目标是为每个用户簇训练模型。例如，这些簇可以代表偏好不同类别电影和电视连续剧的用户群体。本文着重对聚类联邦学习方法进行数学建模与分析，并探究聚类联邦学习在处理非独立同分布数据时提升联邦学习性能的效果。具体而言，我们研究聚类联邦学习如何为偏好相似的用户簇创建个性化模型。我们的研究受到此前聚类联邦学习研究的启发；这些研究表明，在非独立同分布数据设置下，聚类联邦学习有望提升联邦学习的性能。

为了更精确地描述我们的设定，我们来考虑一个包含一个全局服务器和 N 个不同智能体的聚类联邦学习场景。我们假设每个智能体都属于 K 个互不重叠的群组之一。

致谢： 作者姓名按照姓氏的字母顺序排列。作者感谢 Huang Hui 和 Jinniao Qiu 就共识型贝叶斯优化（CBO）的平均场极限问题所进行的富有启发性的讨论。本研究始于 2021 年秋季，当时作者们访问了 Simons 研究所，并参加了题为“优化与采样中的几何方法”的研究项目。作者谨向该研究所提供的热情接待与支持表示感谢。NGT 获得了 NSF-DMS 项目资助，项目编号为 2005797 和 2236447。此外，NGT 与 SL 共同感谢威斯康星大学麦迪逊分校的 IFDS 以及 NSF 通过 TRIPODS 项目（资助编号 2023239）提供的支持。JAC 获得了欧洲研究委员会执行局（ERC）在欧盟“地平线 2020”研究与创新计划框架下提供的高级资助项目 Nonlocal-CPD（复杂粒子动力学的非局部偏微分方程：相变、模式与同步，资助协议编号 883363）的支持。JAC 还获得了 EPSRC 项目资助，项目编号分别为 EP/T022132/1 和 EP/V051121/1。

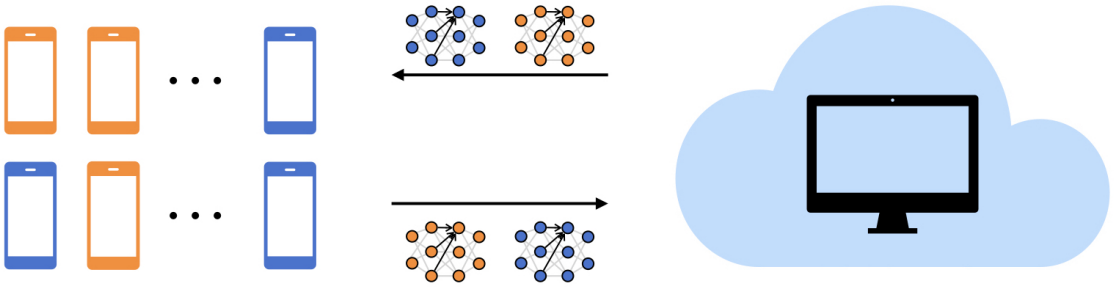

FIGURE 1. 聚类联邦学习问题的简单示例。每位手机用户都有一个潜在的簇身份，此处以橙色和蓝色表示。我们的目标是在通过与云端全局服务器通信模型参数、为每个簇训练模型的同时，识别用户的群体成员关系。

记作 S_1^*、……、S_K^*。我们进一步假设，每个属于组 S_k^* 的代理都拥有从分布 $\mathcal{D}_k$ 生成的数据点，可用于训练其自身的学习模型。理想情况下，代理会寻求与组内其他代理通信，以加速自身模型的训练过程。然而，在数据隐私约束下（参见备注 4中关于隐私的讨论），即代理不会与全局服务器和其他代理共享其本地数据，因此底层划分 S_1^*、……、S_K^* 永远不会被学习算法获知。令 $l(\theta; z) : \Theta \to \mathbb{R}$ 为与数据点 z相关联的损失函数，其中 $\Theta \subset \mathbb{R}^d$ 是学习模型的参数空间。我们的目标是最小化总体损失函数

$$L_i(\theta) := \mathbb{E}_{z\sim\mathcal{D}_i}[l(\theta; z)] \tag{1}$$

对所有 $i \in [K]$ 同时。换言之，目标是为所有损失函数找到极小值点 θ_i^* s

$$\theta_i^* \in \arg\min_{\theta\in\Theta} L_i(\theta), \quad i \in [K]. \tag{2}$$

图 1 展示了一个聚类联邦学习框架的简单示例1。

如上文所述，聚类联邦学习的主要难点在于用户的簇身份未知。因此，聚类联邦学习算法必须能够对用户进行聚类，并在分布式环境中同时训练模型，而不依赖于本地数据收集。为了提出一种能够实现这一目标的算法，本文借鉴基于共识的优化（CBO）的思路 [6, 8, 36]，对聚类联邦学习问题进行抽象并建立数学模型。CBO 是一类全局优化方法，它基于相互作用粒子系统，使粒子围绕目标函数的全局极小值点达成共识。具体而言，考虑

$$\min_{\theta\in\mathbb{R}^d} L(\theta),$$

其中，目标函数 L 可能是非凸函数，是一个具有唯一全局极小值点 θ^* 的连续函数。对于每个 $i \in [N]$，令 $\theta^i \in \mathbb{R}^d$ 表示粒子 i 的位置，并考虑以下方程组

$$d\theta_t^i = -\lambda(\theta_t^i - m_t)dt + \sigma|\theta_t^i - m_t|dB_t^i, \qquad \text{for } i = 1, 2, \ldots, N, \tag{3}$$

其中 t he $\{B^i\}_i$ 是相互独立的布朗运动， m_t 是定义为以下形式的加权平均值

$$m_t := \frac{\sum_{i=1}^N \theta_t^i \exp\big(-\alpha L(\theta_t^i)\big)}{\sum_{i=1}^N \exp\big(-\alpha L(\theta_t^i)\big)}.$$

也可以考虑用于（3）的其他类型噪声（参见 [7, 8]）。例如，可以用3）中的扩散项替换为 [8]所示的几何分量布朗运动。这使得通过令某些假设与维度无关 d，可以提升共识型贝叶斯优化算法的性能，从而适用于高维机器学习应用。还可以通过采用类似于 m_t 的协方差矩阵，以更一般且具有各向异性的方式调整噪声，如 [7], 中所述；该方法称为优化模式下的基于共识的采样（CBS）。为简洁起见，本文将采用基本的共识型贝叶斯优化方法，关于其他现有的共识型贝叶斯优化变体的更多细节，请读者参阅 [7, 8] 。

我们注意到，$-\alpha L(\theta)$中的 exp(m_t 公式里的项是与目标函数 $L(\theta)$和温度 $\frac{1}{\alpha}$ 相对应的吉布斯分布。以这种方式赋予权重的动机来自拉普拉斯原理 [10,29], ，该原理指出，对于任意概率测度 $\rho \in \mathcal{P}(\mathbb{R}^d)$，若其支撑集紧且 $\theta^* \in \text{supp}(\rho)$，则有

$$\lim_{\alpha\to+\infty}\left(-\frac{1}{\alpha}\log\left(\int_{\mathbb{R}^d}\exp\big(-\alpha L(\theta)\big)d\rho(\theta)\right)\right) = L(\theta^*).$$

因此，如果 θ^* 是 L的唯一极小值点，那么经常数因子归一化后的测度 exp($-\alpha L(\theta)$)$\rho(\theta$会将其大部分质量分配到 θ^*的一个小邻域内；而且，如果 α 足够大，该测度就会近似于 θ^*处的狄拉克 δ 分布。因此，加权平均 m_t，即测度 exp($-\alpha L(\theta)$)$\rho(\theta)$ 的归一化版本的经验一阶矩，是一个合理的目标，可供粒子按照方程（3）跟随。

尽管当已知簇身份时，CBO 可以很容易地适配到分布式环境中（例如，只需在每个簇上运行 CBO），但如果目标是提出不依赖于智能体身份的动力学，它就不能直接应用于聚类联邦学习。我们的问题也不同于标准的多目标优化，后者已通过 CBO 实现适配；见 [3, 4]。事实上，在标准多目标优化中，目标是找到一个对于 K 个不同目标函数而言 Pareto 最优的点；而我们的目标是同时找到 K 个不同目标函数的全局极小值点。另一方面，CBO 方法本质上不依赖梯度，因此当目标函数不够平滑，或其导数计算成本高昂时，更适合采用该方法。不过，如果通信成本像真实联邦学习应用中那样高昂，可以考虑在每个用户的动力学中引入额外的局部梯度项，以便即使用户之间没有通信，训练也能继续进行。

1.1. **贡献与相关工作**。

1.1.1. 贡献。 受上述讨论的启发，我们提出了一种新的、受 CBO 启发的相互作用粒4子系统（见下文的（，以及其后的讨论），适用于聚类联邦学习场景。在我们的系统中，每个粒子的演化完全由其自身的损失函数、自身身份以及其他粒子的位置决定，但不需要了解它们的身份。更准确地说，我们的主要贡献可概括如下：

(1) 我们提出了一种新颖的 CBO 类型框架，使得在不知道任何粒子所属簇身份的情况下，也能在 CFL 场景中最小化 K 目标函数。为此，我们引入了一种机制，暗中迫使属于同一簇的粒子达成共识。此外，我们在粒子动力学中加入了梯度项，从而大幅减少 CBO 算法取得良好性能所需的通信轮数。

(2) 我们为所提出的框架提供了严格的理论论证。具体而言，我们证明了所提出粒子系统的适定性及其向平均场极限方程的收敛性，并研究了平均场动力学的共识行为，以及其围绕各个底层损失函数的全局极小值点进行集中的能力。

(3) 我们以合理的方式对连续动力学进行离散化，并将其融入传统的联邦训练协议，从而得到一种新的联邦学习算法，称为 FedCBO。我们开展了大量数值实验，以验证该算法的效率，并证明其在非独立同分布数据设置下优于其他最先进的方法。

1.1.2. 聚类联邦学习中的相关工作。 在聚类联邦学习（CFL）设置中，假设用户之间存在底层聚类结构，目标是识别各簇的身份，并在每个群组内进行联邦学习。IFCA [15] 和 HypCluster [27] 均交替进行用户簇身份识别，以及通过本地梯度下降更新用户簇的模型。这些方法通过在每个本地数据集上寻找损失最低的模型来识别簇身份。FedSEM [25] 在每个联邦步骤中，根据模型参数和准确率衡量用户之间的距离，再运行简单的 K-均值算法来对用户分组。所有这些方法都需要预先了解或估计底层簇的数量，而在实际中这可能难以做到。相比之下，正如下文所述，我们的方法不需要任何关于聚类结构的先验知识，同一簇中的客户端之间将通过粒子动力学自动形成共识。

在 [34], 中，簇以层次化方式划分。具体而言，根据客户端模型梯度或权重更新的余弦相似度，将客户端递归地划分为两个集合。WeCFL [26] 形式化了

聚类联邦学习问题，将其纳入统一的双层优化框架。不同于上述两种框架 [26, 34]，在凸性假设下进行收敛分析，我们的论文考虑非凸目标函数，并给出了平均场情形下的渐近收敛结果。

1.1.3. 共识型贝叶斯优化中的相关工作。 利用包含促共识项的相互作用粒子动力学来求解全局优化问题的想法，最早由 [31] 提出。此后，这种方法在理论和应用领域都引起了广泛关注。

在理论方面， [6] 首次对一个平均场 CBO 方程进行了局部收敛分析，但对系统的初始化提出了相当严格的假设。具体而言，首先证明在无限时间范围内，平均场层面会形成共识，然后利用拉普拉斯原理调整共识点。随后， [13] 放宽了其中一些严格假设，并证明平均场动力学能够在有限时间区间内达到任意程度的集中，使其聚集于全局极小值点周围；不过，这一时间范围可能难以事先估计。通过证明有限粒子 CBO 系统收敛到平均场极限， [16] 进一步奠定了 CBO 框架的理论基础。在本文中，我们采用与 [13, 32] 和 [16] 类似的策略，研究平均场系统的行为（第4节），并证明所提出的粒子动力学收敛至一个适当的平均场极限（第5节）。在我们的设定中，由于粒子可能具有不同的动力学，而这些动力学取决于它们试图优化的损失函数，因此我们需要应对新的挑战。例如，与 [13, 32]不同，我们需要采用不同的方法来估计平均场层面达到给定准确率所需的时间范围。同样，有限粒子系统向适当平均场极限的收敛也涉及额外的技术难题，因为在我们的设定中，存在多种类型的粒子彼此相互作用。

在算法方面，受多种应用的启发，研究人员扩展并改造了原始 CBO 模型，以涵盖新的情形，例如球面等紧流形上的全局优化 [12]，一般约束 [9]，高维机器学习问题 [8]，具有多个极小值的目标函数的全局优化 [5]，从分布中采样 [5, 7]，以及鞍点优化问题 [17]。在 [32]，中，作者在 CBO 系统中引入了梯度项，并通过数值结果表明，该项应用于压缩感知问题时效果良好。在本文中，我们也在粒子动力学中纳入了梯度信息，但我们的动机不同于 [32]，：我们旨在降低用户之间的通信成本，这是联邦学习中的一项重要实际约束。若想更全面地了解 CBO 类方法的发展，感兴趣的读者可参阅近期综述 [36]。

1.2. 记号。我们用 $|\cdot|$ 表示欧几里得空间中向量的绝对值或 ℓ_2-范数，并用 $B_r(\theta)$表示以 $\theta \in \mathbb{R}^d$为中心、半径为 r 的开球。我们用 $\mathcal{C}(X, Y)$表示从 $X \subset \mathbb{R}^n$ 到给定拓扑空间 Y的连续函数 $f\colon X \to Y$所构成的空间。按照惯例，空间 $\mathcal{C}(X, Y)$赋予上确界范数。当 $Y = \mathbb{R}$时，我们简记为 $\mathcal{C}(X)$。我们还用 $\mathcal{C}_c^k(X)$和 $\mathcal{C}_b^k(X)$分别表示具有紧支集的 k 次连续可微实值函数空间，以及 k 次连续可微有界函数空间。令 $\mathcal{P}(\mathbb{R}^d)$为 $\mathbb{R}^d$ 上所有博雷尔概率测度构成的空间，并赋予其列维-普罗霍罗夫度量；该度量可度量弱收敛拓扑。给定 $p \geq 1$，令 $\mathcal{P}_p(\mathbb{R}^d) \subseteq \mathcal{P}(\mathbb{R}^d)$表示具有有限 p 阶矩的概率测度 $\rho \in \mathcal{P}(\mathbb{R}^d)$的集合，即 $\int_{\mathbb{R}^d} |\theta|^p d\rho(\theta) < \infty$。空间 $\mathcal{P}_p(\mathbb{R}^d)$赋予 p-Wasserstein 距离 W_p $(1 \leq p < \infty)$，其定义如下

$$W_p(\rho, \hat{\rho}) := \inf_{\pi \in \Gamma(\rho, \hat{\rho})} \int_{\mathbb{R}^d \times \mathbb{R}^d} |\theta - \hat{\theta}|^p \pi(d\theta, d\hat{\theta}), \qquad \rho, \hat{\rho} \in \mathcal{P}_p(\mathbb{R}^d),$$

其中， $\Gamma(\rho, \hat{\rho})$ 表示所有以 $\mathbb{R}^d \times \mathbb{R}^d$ 为定义域、边缘分布分别为 ρ 和 $\hat{\rho}$ 的联合概率测度的集合。

研究随机过程 $\rho \in \mathcal{C}([0, T], \mathcal{P}(\mathbb{R}^d))$ 的分布时，我们将时刻 t 的分布记为 $\rho_t \in \mathcal{P}(\mathbb{R}^d)$。给定连续函数 $f \in \mathcal{C}(\mathbb{R}^d)$ 和固定的概率测度 $\rho \in \mathcal{P}(\mathbb{R}^d)$，我们将 f 关于测度 ρ 的 L^1-范数记为 $\|f\|_{\mathbb{L}^1(\rho)} := \int_{\mathbb{R}^d} |f(\theta)| d\rho(\theta)$。

1.3. **论文结构**。本文其余部分安排如下。在第2.1节中，我们介绍了启发 FedCBO 算法的相互作用粒子系统。在第2.2节中，我们陈述主要理论结果，包括所提出的相互作用粒子动力学相关的平均场系统的适定性

（定理2.1）、随着系统中粒子数量增加，相互作用粒子系统向平均场收敛（定理2.2），以及平均场随时间的演化行为，及其对每个目标函数的全局最优解的集中能力（定理2.3）。受粒子系统的启发，我们在第2.3节介绍 FedCBO 算法。在第2.4节中，我们介绍一系列数值实验，以验证所提出的算法。第3 节用于证明定理2.1，第4 节用于证明定理2.3，第5 节用于证明定理2.2。最后，我们在第6节总结本文的贡献，并讨论未来的研究方向。

2. CBO for Clustered Federated Learning

2.1. **模型形式化**。 在本文其余部分中，不失一般性，我们假设聚类联邦学习问题中只有两个簇（2），即 $K = 2$。事实上，从下文的讨论可以清楚地看出，将所提出的模型及其相应的理论分析扩展到 $K > 2$ 的情形十分直接。我们还假设类别1中的所有代理使用同一个损失函数 L_1，类别2中的所有代理使用同一个损失函数 L_2。[1] 为了同时优化 L_1 和 L_2，我们考虑由 $N \in \mathbb{N}$ 个相互作用粒子组成的集合，其位置分别为 $\{\theta_t^{1,i}\}_{i=1}^{N_1} \in \mathbb{R}^d$（类别1粒子）和 $\{\theta_t^{2,j}\}_{j=1}^{N_2} \in \mathbb{R}^d$（类别2粒子），并由以下随机微分方程组描述：

$$d\theta_t^{1,i} = -\lambda_1(\theta_t^{1,i} - m_t^1)dt - \lambda_2 \nabla L_1(\theta_t^{1,i})dt + \sigma_1|\theta_t^{1,i} - m_t^1|dB_t^{1,i} + \sigma_2|\nabla L_1(\theta_t^{1,i})|d\widetilde{B}_t^{1,i} \tag{4a}$$

$$d\theta_t^{2,j} = -\lambda_1(\theta_t^{2,j} - m_t^2)dt - \lambda_2 \nabla L_2(\theta_t^{2,j})dt + \sigma_1|\theta_t^{2,j} - m_t^2|dB_t^{2,j} + \sigma_2|\nabla L_2(\theta_t^{2,i})|d\widetilde{B}_t^{2,j} \tag{4b}$$

$$m_t^1 := \frac{\sum_{i=1}^{N_1} \theta_t^{1,i} w_{L_1}^\alpha(\theta_t^{1,i}) + \sum_{j=1}^{N_2} \theta_t^{2,j} w_{L_1}^\alpha(\theta_t^{2,j})}{\sum_{i=1}^{N_1} w_{L_1}^\alpha(\theta_t^{1,i}) + \sum_{j=1}^{N_2} w_{L_1}^\alpha(\theta_t^{2,j})} \tag{4c}$$

$$m_t^2 := \frac{\sum_{i=1}^{N_1} \theta_t^{1,i} w_{L_2}^\alpha(\theta_t^{1,i}) + \sum_{j=1}^{N_2} \theta_t^{2,j} w_{L_2}^\alpha(\theta_t^{2,j})}{\sum_{i=1}^{N_1} w_{L_2}^\alpha(\theta_t^{1,i}) + \sum_{j=1}^{N_2} w_{L_2}^\alpha(\theta_t^{2,j})}, \tag{4d}$$

其中 λ_1，λ_2，σ_1，$\sigma_2 > 0$，$w_{L_k}^\alpha(\theta) := \exp(-\alpha L_k(\theta))$ 对于 $k = 1$，2，以及 $\alpha > 0$。下文中，我们可能会交替使用“粒子”和“用户”这两个术语来指代代理。

我们逐项讨论上述系统。方程（4a）描述了类别1中代理 i 的模型参数随时间的演变，而方程（4b）则对类别2中的代理 j 作出相同描述。在（4c）中定义的 m_t^1是相对于损失函数 L_1对所有粒子位置 $\{\theta_t^{1,i}\}_{i=1}^{N_1}$，$\{\theta_t^{2,j}\}_{j=1}^{N_2}$ 进行加权平均的结果。特别是，类别1中的代理可以在不知道其他任何代理的类别身份的情况下计算出 m_t^1，这对我们的研究目的而言是一项关键特性。暂且设想类别1的粒子集中在 L_1 较小的区域，那么可以预期 $\{\theta_t^{1,i}\}_{i=1}^{N_1}$ 的 L_1损失会低于类别2粒子 $\{\theta_t^{2,j}\}_{j=1}^{N_2}$；后者理应集中在损失函数 L_2较小的区域。由此直观来看，在 m_t^1 的表达式中，类别1粒子 $\{\theta_t^{1,i}\}_{i=1}^{N_1}$ 将获得比类别2粒子更高的权重，因此 m_t^1 应当接近仅对粒子 $\{\theta_t^{1,i}\}_{i=1}^{N_1}$ 进行加权平均的结果。因此，m_t^1 可以被视为一个随时间演变的共识点，且仅对应类别1粒子。对于 m_t^2也有类似的直观解释，它是仅对应类别2粒子的随时间演变的共识点。于是，（4a）和（4b）中漂移项的第一部分可以被视为针对各类别的共识诱导项。另一方面，（4a）和（4b）中漂移项的第二部分引入了局部梯度信息，可解释为通过梯度下降进行的本地模型更新。在联邦学习中，利用局部梯度信息可确保即使模型之间没有通信，所有模型仍会继续更新；正如 [28], 所述，在联邦环境中，通信通常成本高昂。最后，动力学中的扩散项确保每个粒子持续探索其损失函数对应的优化景观，直至到达临界点并与其所属类别的共识点对齐。假设 $\{B^{k,i}\}_{k,\ i}$和 $\{\widetilde{B}^{k,i}\}_{k,\ i}$ 相互独立，这些是 d维布朗运动。

[1]在实践中，每个代理只能访问有限的数据样本，因此其经验损失函数实际上可能与同一集群中其他代理的经验损失函数不同。我们将这一更符合现实且更具挑战性的情形的建模与研究留待未来工作。

我们将 $\boldsymbol{\theta}^{1,N}_t := \{\theta^{1,1}_t, \dots, \theta^{1,N_1}_t\}$，$\boldsymbol{\theta}^{2,N}_t := \{\theta^{2,1}_t, \dots, \theta^{2,N_2}_t\}$ 在 $N = N_1 + N_2 \in \mathbb{N}$ 时记作粒子系统（4）的解。考虑其经验测度如下：

$$\rho^{1,N}_t := \frac{1}{N_1}\sum_{i=1}^{N_1}\delta_{\theta^{1,i}_t}, \qquad \rho^{2,N}_t := \frac{1}{N_2}\sum_{j=1}^{N_2}\delta_{\theta^{2,j}_t}, \qquad \rho^N_t := \frac{N_1}{N}\rho^{1,N}_t + \frac{N_2}{N}\rho^{2,N}_t, \tag{5}$$

其中，我们用 δ_θ 表示位于 $\theta \in \mathbb{R}^d$ 处的狄拉克δ测度。注意，m^1_t，m^2_t可改写为关于 ρ^N_t 的形式，如下所示：

$$m^1_t = \frac{1}{\|w^\alpha_{L_1}\|_{\mathbb{L}^1(\rho^N_t)}}\int \theta w^\alpha_{L_1} d\rho^N_t =: m^\alpha_{L_1}[\rho^N_t], \qquad m^2_t = \frac{1}{\|w^\alpha_{L_2}\|_{\mathbb{L}^1(\rho^N_t)}}\int \theta w^\alpha_{L_2} d\rho^N_t =: m^\alpha_{L_2}[\rho^N_t].$$

进而，系统（4）可改写为

$$\begin{aligned} d\theta^{1,i}_t &= -\lambda_1(\theta^{1,i}_t - m^\alpha_{L_1}[\rho^N_t])dt - \lambda_2\nabla L_1(\theta^{1,i}_t) + \sigma_1|\theta^{1,i}_t - m^\alpha_{L_1}[\rho^N_t]|dB^{1,i}_t + \sigma_2|\nabla L_1(\theta^{1,i}_t)|d\widetilde{B}^{1,i}_t, \\ d\theta^{2,j}_t &= -\lambda_1(\theta^{2,j}_t - m^\alpha_{L_2}[\rho^N_t])dt - \lambda_2\nabla L_2(\theta^{2,j}_t) + \sigma_1|\theta^{2,j}_t - m^\alpha_{L_2}[\rho^N_t]|dB^{2,j}_t + \sigma_2|\nabla L_2(\theta^{2,j}_t)|d\widetilde{B}^{2,j}_t. \end{aligned}$$

根据上述表达式，我们可以形式化地假设一个平均场随机微分方程系统，用以描述粒子随时间的演化 N_1, $N_2 \to \infty$。具体而言，我们考虑以下两个随机微分方程组成的系统：

$$d\bar{\theta}^1_t = -\lambda_1(\bar{\theta}^1_t - m^\alpha_{L_1}[\rho_t])dt - \lambda_2\nabla L_1(\bar{\theta}^1_t)dt + \sigma_1|\bar{\theta}^1_t - m^\alpha_{L_1}[\rho_t]|dB^1_t + \sigma_2|\nabla L_1(\bar{\theta}^1_t)|d\widetilde{B}^1_t, \tag{7a}$$

$$d\bar{\theta}^2_t = -\lambda_1(\bar{\theta}^2_t - m^\alpha_{L_2}[\rho_t])dt - \lambda_2\nabla L_2(\bar{\theta}^2_t)dt + \sigma_1|\bar{\theta}^2_t - m^\alpha_{L_2}[\rho_t]|dB^2_t + \sigma_2|\nabla L_2(\bar{\theta}^2_t)|d\widetilde{B}^2_t, \tag{7b}$$

其中

$$m^\alpha_{L_k}[\rho_t] = \int \theta d\eta^\alpha_{k,t}, \quad \eta^\alpha_{k,t} = w^\alpha_{L_k}\rho_t/\|w^\alpha_{L_k}\|_{\mathbb{L}^1(\rho_t)}, \quad \rho^k_t = Law(\bar{\theta}^k_t), \quad \rho_t = w_1\rho^1_t + w_2\rho^2_t,$$

满足独立初始条件 $\theta^1_0 \sim \rho^1_0$ 和 $\theta^2_0 \sim \rho^2_0$；在上述表达式中，B^1、$\widetilde{B}^1$、B^2、$\widetilde{B}^2$ 是相互独立的 d 维布朗运动。方程（7a）和（7b）分别描述了在粒子数量较大的情形下，类别1和类别2中单个粒子的有效时间演化。权重 w_k 表示系统中类型 k 粒子的渐近比例。最后，ρ^1_t 和 ρ^2_t 分别表示时刻 t 类别1和类别2粒子的分布。请注意，系统（7）通过两种类型代理的分布 ρ_t 相互耦合。在第 5 节中，我们将以精确的数学方式讨论有限相互作用粒子系统（4）与（7）中所述平均场极限系统之间的关系；另请参见下文的定理 2.2。

与（7）相对应的福克-普朗克方程组为：

$$\partial_t\rho^1_t := \Delta(\kappa^1_t\rho^1_t) + \nabla\cdot(\mu^1_t\rho^1_t), \qquad \lim_{t\to 0}\rho^1_t = \rho^1_0 \tag{8a}$$

$$\partial_t\rho^2_t := \Delta(\kappa^2_t\rho^2_t) + \nabla\cdot(\mu^2_t\rho^2_t), \qquad \lim_{t\to 0}\rho^2_t = \rho^2_0, \tag{8b}$$

其中

$$\mu^k_t := \lambda_1(\theta - m^\alpha_{L_k}[\rho_t]) + \lambda_2\nabla L_k(\theta), \quad \kappa^k_t := \frac{\sigma_1^2}{2}|\theta - m^\alpha_{L_k}[\rho_t]|^2 + \frac{\sigma_2^2}{2}|\nabla L_k(\theta)|^2, \quad \text{for } k = 1, 2.$$

这是一个非线性方程组，用于描述平均场极限下各类别代理的分布随时间的演化。值得注意的是，由于过程 θ^1 和 θ^2相互独立，我们确实可以仅通过两个边缘分布的分布来描述平均场极限下代理联合系统的分布，这一点我们将在注释 7中讨论。然而，在第 5.2 节中，显式地研究联合过程（θ^1_t，θ^2_t）的分布将更加方便。这样做是为了便于研究有限粒子系统向平均场极限的收敛性。相关细节将在适当的时候给出。

在下文中，我们将以弱意义解释福克-普朗克系统（8）。

定义2.1。对于 $k = 1, 2$，设 $\rho^k_0 \in \mathcal{P}(\mathbb{R}^d)$。设 $T > 0$ 为给定的时间范围。若 ρ^1，$\rho^2 \in \mathcal{C}([0,T], \mathcal{P}(\mathbb{R}^d))$ 在弱意义下满足福克-普朗克方程 (8) 在时间区间 $[0,T]$ 且

具有初始条件 (ρ_0^1, ρ_0^2)，若对于 $\forall \phi \in \mathcal{C}_c^\infty(\mathbb{R}^d)$，所有 $\forall t \in (0,T)$ 以及 $k=1,2$，有

$$\frac{d}{dt}\int \phi(\theta) d\rho_t^k(\theta) = \int \left(-\lambda_1(\theta - m_{L_k}^\alpha[\rho_t]) - \lambda_2 \nabla L_k(\theta)\right)\cdot \nabla\phi(\theta) d\rho_t^k(\theta) + \int \left(\frac{\sigma_1^2}{2}|\theta - m_{L_k}^\alpha[\rho_t]|^2 + \frac{\sigma_2^2}{2}|\nabla L_k(\theta)|^2\right)\Delta\phi(\theta) d\rho_t^k(\theta),$$

$and\ \lim_{t\to 0}\rho_t^k = \rho_0^k$（按概率测度弱收敛的意义）对于 $k=1,2$.

我们的主要理论结果将在下一节介绍，分为三个关键定理。首先，我们讨论平均场随机微分方程系统（7）及其对应的福克-普朗克系统（8）的适定性。其次，我们给出一个结果，建立有限粒子系统（4）的经验测度演化与福克-普朗克方程（8）之间的联系。最后，我们讨论平均场偏微分方程的长期行为性质，并表明，在对初始化作出一些温和假设且正确调整参数的情况下，在一定时间区间内，分布 ρ_t^1 和 ρ_t^2 分别集中在 L_1 和 L_2的全局极小值点周围。这些理论结果结合起来的实际意义如下：考虑具有足够大的 N_1 和 N_2 的系统（4），并假设初始化适当，则类别1的粒子将集中在损失函数 L_1的全局极小值点周围，而类别2的粒子也将集中在 L_2的全局极小值点周围。在第2.3节，我们利用数学模型为一种新的基于聚类的联邦学习算法提供理论动机。在第2.4节，我们通过数值实验表明，所提出的算法确实能够为具有相似数据集的用户群体生成性能优异的学习模型。

2.2. **主要理论结果**。 在所有理论分析中，我们对损失函数 L_1, L_2作出以下假设。

假设 1。 对于 $k=1,2$，损失函数 $L_k\colon \mathbb{R}^d \to \mathbb{R}$ 有下界 $\underline{L_k} := \inf L_k$。此外，存在常数 $M_{L_k}, C_{L_k}, M_{\nabla L_k}, C_{\nabla L_k} > 0$，使得对所有 $\theta, \hat{\theta} \in \mathbb{R}^d$,

$$|L_k(\theta) - L_k(\hat{\theta})| \le M_{L_k}(|\theta| + |\hat{\theta}|)|\theta - \hat{\theta}| \tag{9a}$$

$$L_k(\theta) - \underline{L_k} \le C_{L_k}(1 + |\theta|^2) \tag{9b}$$

$$|\nabla L_k(\theta) - \nabla L_k(\hat{\theta})| \le M_{\nabla L_k}|\theta - \hat{\theta}| \tag{9c}$$

$$|\nabla L_k(\theta)| \le C_{\nabla L_k} \tag{9d}$$

简单来说，在（9a）和（9b）中，我们假设损失函数 L_k 是局部利普希茨的，且有二次函数上界。我们还假设 L_k 的梯度在（9c）和（9d）中是利普希茨的且有界。除了假设 1之外，我们还考虑满足以下条件之一的损失函数 L_k：（1）有上界。特别地，L_k 的上界为 $\underline{L_k} := \sup L_k$；或（2）在无穷远处二次增长，即存在常数 $M>0$和 $c_{q_k}>0$，使得 $L_k(\theta) - \underline{L_k} \ge c_{q_k}|\theta|^2$对所有 $|\theta| \ge M$成立。

定理 2.1 (平均场方程的适定性)。 设 L_1, L_2 满足假设 *1* 且有界或在无穷远处具有二次增长，并设 $\rho_0^1, \rho_0^2 \in \mathcal{P}_4(\mathbb{R}^d)$。则存在唯一的非线性过程 $\theta^1, \theta^2 \in \mathcal{C}([0,T],\mathbb{R}^d)$ $T>0$，满足

$$\begin{aligned} d\overline{\theta}_t^1 &= -\lambda_1(\overline{\theta}_t^1 - m_{L_1}^\alpha[\rho_t])dt - \lambda_2\nabla L_1(\overline{\theta}_t^1)dt + \sigma_1|\overline{\theta}_t^1 - m_{L_1}^\alpha[\rho_t]|dB_t^1 + \sigma_2|\nabla L_1(\overline{\theta}_t^1)|d\widetilde{B}_t^1, \\ d\overline{\theta}_t^2 &= -\lambda_1(\overline{\theta}_t^2 - m_{L_2}^\alpha[\rho_t])dt - \lambda_2\nabla L_2(\overline{\theta}_t^2)dt + \sigma_1|\overline{\theta}_t^2 - m_{L_2}^\alpha[\rho_t]|dB_t^2 + \sigma_2|\nabla L_2(\overline{\theta}_t^2)|d\widetilde{B}_t^2, \\ \rho_t &= w_1\rho_t^1 + w_2\rho_t^2, \qquad w_1 + w_2 = 1 \end{aligned}$$

在强意义下，且 $\rho_t^1 =$分布(θ_t^1) $\rho_t^2 =$分布(θ_t^2) $\rho_t \in \mathcal{C}([0,T],\mathcal{P}_2(\mathbb{R}^d))$ 在弱意义下满足相应的福克-普朗克方程（*8a*）和（*8b*），且 $\lim_{t\to 0}\rho_t^k = \rho_0^k \in \mathcal{P}_2(\mathbb{R}^d)$ 对于 $k=1,2$。

已知平均场随机微分方程系统（7）及其对应的福克-普朗克系统均适定，现在我们可以精确地说明有限粒子系统（4）与平均场系统之间的联系。我们先引入一些记号。给定 ρ_0^1，$\rho_0^2 \in \mathcal{P}(\mathbb{R}^d)$，并给定 N_1，$N_2 \in \mathbb{N}$，我们用 $\rho_0^{1\otimes N_1}$表示 ρ_0^1 自身的 N_1 重乘积测度；同样地，用 $\rho_0^{2\otimes N_2}$表示 ρ_0^2 自身的 N_2 重乘积测度。

定理 *2.2*（收敛至平均场极限）。设 L_1, L_2 满足假设 *1*，且在无穷远处具有二次增长，并假设 $\rho_0^1, \rho_0^2 \in \mathcal{P}_4(\mathbb{R}^d)$。设 $w_1, w_2 > 0$ 使得 $w_1 + w_2 = 1$，并设 $N_1, N_2 \geq 1, N = N_1 + N_2$ 使得 $\frac{N_1}{N} \to w_1$，$\frac{N_2}{N} \to w_2$，当 $N_1, N_2 \to \infty$ 时。假设 $\{\boldsymbol{\theta}_t^{1,N}\}, \{\boldsymbol{\theta}_t^{2,N}\}$ 是粒子系统 *(4)* 的唯一解，其初始数据 $\{\boldsymbol{\theta}_0^{1,N}\}, \{\boldsymbol{\theta}_0^{2,N}\}$ 服从 $\rho_0^{1\otimes N_1} \otimes \rho_0^{2\otimes N_2}$ 分布。

那么，对于 $k = 1,2$ 以及任意 $\zeta > 0$，我们有

$$\lim_{N\to\infty} \mathbb{P}\left(\sup_{t\in[0,T]} d_{LP}(\rho_t^{k,N}, \rho_t^k) \geq \zeta \right) = 0,$$

其中 d_{LP}是定义在 $\mathbb{R}^d$上的概率测度之间的列维-普罗霍罗夫度量，且 $t \in [0,T] \mapsto (\rho_t^1, \rho_t^2)$ 是满足初始条件 ρ_0^1 和 ρ_0^2的方程(8) 的唯一解。

注 1. 定理*2.2* 的推论是，有限粒子系统中类型为k 的粒子位置的经验测度，随时间一致地依概率收敛于平均场中类型为 k 的粒子位置的分布。

对于下一个结果，我们对损失函数 L_1, L_2施加额外假设。

假设 2. 对于 $k = 1,2$，我们假设函数 $L_k \in \mathcal{C}(\mathbb{R}^d)$ 满足

(I) 存在 $\theta_k^* \in \mathbb{R}^d$，使得 $L_k(\theta_k^*) = \inf_{\theta\in\mathbb{R}^d} L_k(\theta) =: \underline{L_k}$；

(II) 存在 L_∞^k、R_0^k、$\eta_k > 0$ 和 $\nu_k \in (0, \frac{1}{2}]$，使得

$$|\theta - \theta_k^*| \leq \frac{1}{\eta_k}\big(L_k(\theta) - \underline{L_k}\big)^{\nu_k} \qquad \textit{for all } \theta \in B_{R_0^k}(\theta_k^*), \tag{11}$$

$$L_\infty^k < L_k(\theta) - \underline{L_k} \qquad \textit{for all } \theta \in \big(B_{R_0^k}(\theta_k^*)\big)^c. \tag{12}$$

假设 2 与 [13]中使用的假设类似。假设的第一部分，即假设(I)，指出目标函数的最小值 L_k 在 θ_k^*处取得。第二部分，即假设 (II)，规定了我们的理论成立所需的目标函数景观的某些性质。具体而言，不等式（11）对 L_k 在全局极小值点θ_k^*附近的局部增长设定了下界，而条件（12）排除了以下可能性：对于某个 $L_k(\theta) \approx L_k$，若 θ 位于 θ_k^*的某个邻域之外，则

定理 *2.3*（平均场向全局极小值点集中）。对于 $k = 1,2$，假设 $L_k \in \mathcal{C}(\mathbb{R}^d)$ 满足假设 *1* 和 *2*。此外，设 $\rho_0^k \in \mathcal{P}_4(\mathbb{R}^d)$ 满足 $\rho_0^k(B_r(\theta_k^*)) > 0$ 对所有 $r > 0$ 均成立。定义$\mathcal{V}(\rho_t^k) := \frac{1}{2}\int |\theta - \theta_k^*|^2 d\rho_t^k(\theta)$。对于任意 $\varepsilon \in (0, \mathcal{V}(\rho_0^1) + \mathcal{V}(\rho_0^2))$、$\tau \in (0,1)$ 以及满足 $2\lambda_1 > 2\lambda_2 M + d\sigma_1^2 + d\sigma_2^2 M^2$ 的参数 $\lambda_1, \lambda_2, \sigma_1, \sigma_2 > 0$，其中 $M := \max\{M_{\nabla L_1}, M_{\nabla L_2}\}$，以及时间范围

$$T^* := \frac{1}{(1-\tau)(2\lambda_1 - 2\lambda_2 M - d\sigma_1^2 - d\sigma_2^2 M^2)} \log\left(\frac{\mathcal{V}(\rho_0^1) + \mathcal{V}(\rho_0^2)}{\varepsilon}\right), \tag{13}$$

存在 $\alpha_0 > 0$，它仅取决于 ε和 τ，使得对于所有 $\alpha > \alpha_0$，若 $\rho^1, \rho^2 \in \mathcal{C}([0,T], \mathcal{P}_4(\mathbb{R}^d))$ 分别是在时间区间 $[0,T^*]$上、具有初始条件 ρ_0^1, ρ_0^2 的福克-普朗克方程 *(8a)* 和 *(8b)* 的弱解，则有 $\min_{t\in[0,T^*]} \big(\mathcal{V}(\rho_t^1) + \mathcal{V}(\rho_t^2)\big) \leq \varepsilon$。此外，在 $\mathcal{V}(\rho_t^1) + \mathcal{V}(\rho_t^2)$ 首次达到规定的准确率 ε 之前，我们有指数衰减

$$\mathcal{V}(\rho_t^1) + \mathcal{V}(\rho_t^2) \leq \big(\mathcal{V}(\rho_0^1) + \mathcal{V}(\rho_0^2)\big) \exp\big(-(1-\tau)(2\lambda_1 - 2\lambda_2 M - d\sigma_1^2 - d\sigma_2^2 M^2)t\big). \tag{14}$$

注释 2。参数 λ_1 在(4) 决定了驱动粒子趋向各自共识点的力的强度。同样，λ_2 和 σ_1，σ_2 分别表征梯度项和噪声项的强度。在定理*2.3*中，我们要求参数 λ_1，λ_2，σ_1，$\sigma_2 > 0$ 满足 $2\lambda_1 > 2\lambda_2 M + d\sigma_1^2 + d\sigma_2^2 M^2$。这一要求确保动力学中的共识诱导项占主导地位，这对于系统在损失函数的全局最小值附近达成共识至关重要。不过，这只是出于理论考虑而施加的假设；事实上，在应用中，更强的共识漂移意味着代理之间需要更多通信轮次。尽管如此，正如我们将在第*2.4*节的数值实验中看到的，即使施加合理的通信约束，我们在下一节介绍的 *FedCBO* 算法仍能促使集群成员达成共识。

2.3. **FedCBO 算法**。要将系统（4）应用于联邦学习的实际算法中，我们需要进行一系列调整。首先，我们对所提出的连续时间系统进行离散化，这可以通过对（4）进行欧拉-丸山离散化来实现。其次，必须调整所得的离散化方案，使其适应常规联邦训练协议，其中用户之间的通信轮次受到限制。我们将这些调整结合起来，得到名为 FedCBO 的算法。接下来将详细介绍。

设 $\{\theta_0^{1,i}\}_{i=1}^{N_1}, \{\theta_0^{2,j}\}_{j=1}^{N_2}$ 分别从给定的分布 $\rho_0^1, \rho_0^2 \in \mathcal{P}(\mathbb{R}^d)$ 中采样。由于未给出类别成员关系，合理的假设是 ρ_0^1 和 ρ_0^2 相同，但并不要求这一假设成立。考虑以下迭代序列

$$\theta_{n+1}^{1,i} \leftarrow \theta_n^{1,i} - \lambda_1\gamma(\theta_n^{1,i} - m_n^1) - \lambda_2\gamma\nabla L_1(\theta_n^{1,i}) + \sigma_1\sqrt{\gamma}|\theta_n^{1,i} - m_n^1|z_n^{1,i} + \sigma_2\sqrt{\gamma}|\nabla L_1(\theta_n^{1,i})|\tilde{z}_n^{1,i}, \tag{15a}$$

$$\theta_{n+1}^{2,j} \leftarrow \theta_n^{2,j} - \lambda_1\gamma(\theta_n^{2,j} - m_n^2) - \lambda_2\gamma\nabla L_2(\theta_n^{2,j}) + \sigma_1\sqrt{\gamma}|\theta_n^{2,j} - m_n^2|z_n^{2,j} + \sigma_2\sqrt{\gamma}|\nabla L_2(\theta_n^{2,j})|\tilde{z}_n^{2,j}, \tag{15b}$$

对于 $n = 0, 1, 2, \ldots$。其中，γ 是离散化步长；对于 $k = 1, 2$，$z^{k,i}_n, \tilde{z}^{k,i}_n$ 是相互独立的正态随机向量 $N(0, I_{d\times d})$；$m_n^k, k = 1, 2$ 是 $\{\theta_n^{1,i}\}_{i=1}^{N_1}, \{\theta_n^{2,j}\}_{j=1}^{N_2}$ 的加权平均值，定义如下

$$m_n^1\left[\{\theta_n^{1,i}\}, \{\theta_n^{2,j}\}\right] = \frac{\sum_{i=1}^{N_1}\theta_n^{1,i}w_{L_1}^\alpha(\theta_n^{1,i}) + \sum_{j=1}^{N_2}\theta_n^{2,j}w_{L_1}^\alpha(\theta_n^{2,j})}{\sum_{i=1}^{N_1}w_{L_1}^\alpha(\theta_n^{1,i}) + \sum_{j=1}^{N_2}w_{L_1}^\alpha(\theta_n^{2,j})}, \tag{16a}$$

$$m_n^2\left[\{\theta_n^{1,i}\}, \{\theta_n^{2,j}\}\right] = \frac{\sum_{i=1}^{N_1}\theta_n^{1,i}w_{L_2}^\alpha(\theta_n^{1,i}) + \sum_{j=1}^{N_2}\theta_n^{2,j}w_{L_2}^\alpha(\theta_n^{2,j})}{\sum_{i=1}^{N_1}w_{L_2}^\alpha(\theta_n^{1,i}) + \sum_{j=1}^{N_2}w_{L_2}^\alpha(\theta_n^{2,j})}, \tag{16b}$$

其中 $w_{L_k}^\alpha(\theta) = \exp(-\alpha L_k(\theta))$。给定一个固定整数 $\tau > 0$，将（15a）和（15b）求和 τ 次后，可以将（15a）和（15b）改写为（为简洁起见，省略噪声项[2]）：

$$\theta_{(n+1)\tau}^{1,i} \leftarrow \theta_{n\tau}^{1,i} - \lambda_1\gamma\sum_{q=0}^{\tau-1}(\theta_{n\tau+q}^{1,i} - m_{n\tau+q}^1) - \lambda_2\gamma\sum_{q=0}^{\tau-1}\nabla L_1(\theta_{n\tau+q}^{1,i}), \tag{17a}$$

$$\theta_{(n+1)\tau}^{2,j} \leftarrow \theta_{n\tau}^{2,j} - \lambda_1\gamma\sum_{q=0}^{\tau-1}(\theta_{n\tau+q}^{2,j} - m_{n\tau+q}^2) - \lambda_2\gamma\sum_{q=0}^{\tau-1}\nabla L_1(\theta_{n\tau+q}^{2,j}), \tag{17b}$$

然而，上述更新规则要求在每次迭代时计算共识点 $m_{n\tau+q}^1$ 和 $m_{n\tau+q}^2$。这将导致用户与服务器之间产生过量的通信，而在实际场景中必须避免这种情况。事实上，请注意，代理需要下载所有参与训练的其他用户的参数，才能计算其相应的 $m_{n\tau+q}^k$。如果通信过于频繁，成本可能会迅速变得难以承受。为适应这一实际限制，我们考虑采用一种分裂方案，以如下方式近似更新公式（17a）和（17b）：对于 $n = 0$、1、2，$\cdots$，

$$\widehat{\theta}_{n\tau}^{1,i} \leftarrow \theta_{n\tau}^{1,i}, \quad \widehat{\theta}_{n\tau}^{2,j} \leftarrow \theta_{n\tau}^{2,j}, \tag{18a}$$

$$\widehat{\theta}_{n\tau+q+1}^{1,i} \leftarrow \widehat{\theta}_{n\tau+q}^{1,i} - \lambda_2\gamma\nabla L_1(\widehat{\theta}_{n\tau+q}^{1,i}), \quad \widehat{\theta}_{n\tau+q+1}^{2,j} \leftarrow \widehat{\theta}_{n\tau+q}^{2,j} - \lambda_2\gamma\nabla L_1(\widehat{\theta}_{n\tau+q}^{2,j}) \qquad \text{for } q = 0, \ldots, \tau-1, \tag{18b}$$

$$\theta_{(n+1)\tau}^{1,i} \leftarrow \widehat{\theta}_{(n+1)\tau}^{1,i} - \lambda_1\gamma\big(\widehat{\theta}_{(n+1)\tau}^{1,i} - m_{(n+1)\tau}^1\big), \quad \theta_{(n+1)\tau}^{2,j} \leftarrow \widehat{\theta}_{(n+1)\tau}^{2,j} - \lambda_1\gamma\big(\widehat{\theta}_{(n+1)\tau}^{2,j} - m_{(n+1)\tau}^2\big), \tag{18c}$$

其中，共识点 $m_{(n+1)\tau}^k := m_{(n+1)\tau}^k\big[\{\widehat{\theta}^{1,i}_{(n+1)\tau}\}, \{\widehat{\theta}^{2,j}_{(n+1)\tau}\}\big]$ 对于 $k = 1, 2$ 而言，是按 (16a) 和 (16b) 中的定义对 $\{\widehat{\theta}^{1,i}_{(n+1)\tau}\}_{i=1}^{N_1}, \{\widehat{\theta}^{2,j}_{(n+1)\tau}\}_{j=1}^{N_2}$ 取加权平均值所得。简而言之，在每个通信轮次中，我们首先通过梯度下降更新模型 τ 次（18b），然后计算一次共识点（18c）。为了使上述方案尽可能接近 (17a)，我们将 λ_1 的值设得比 λ_2 更大。按照联邦训练的标准术语，（18b）可解释为每个用户的模型参数经过 τ 个本地梯度下降训练轮次后进行的本地更新，而（18c）则可视为聚合步骤。我们的更新规则有一个有趣的特点：模型聚合并不在全局服务器上进行。相反，智能体可以下载其他用户的模型，并通过（18c）在本地进行聚合。因此，可以假设服务器完全无法获知类别成员关系，甚至也无法获知所有用户参数的实际

[2]如 [8], 所述，在 MNIST 数据集上训练神经网络时，噪声项对算法性能的影响并不显著。

数值。这一特点使我们的 FedCBO 方法成为一种相当去中心化的联邦学习方法。

算法1 FedCBO

输入： 初始化模型 $\theta_0^j \in \mathbb{R}^d$，$j \in [N]$；迭代次数 T；本地梯度步数 τ；下载的模型数量 M；CBO系统超参数 λ_1，λ_2，α；离散化步长 γ；初始化采样概率 $P_0 \in \mathbb{R}^{N\times(N-1)}$;

1: **对于** $n = 0, \dots, T-1$ **执行**
2: $G_n \leftarrow$ 随机选择一组智能体（参与设备）；
3: **本地更新**(θ_n^j，τ，λ_2，γ），针对 $j \in G_n$;
4: **本地聚合**(代理 j)，用于 $j \in G_n$;
5: **结束循环**
输出： θ_T^j 对于 $j \in [N]$. ^ **本地更新**(θ_0，τ，λ_2，γ) 在第 j 个代理6: **对于** $q = 0, \cdots, \tau - 1$ **执行**
^ ^ ^7: （随机）梯度下降 $\theta_{q+1} \leftarrow \theta_q - \lambda_2\gamma\nabla L_j(\theta_q)$;8: **结束循环**^9: **返回** θ_τ;

算法 2本地聚合(代理 j)

输入： 代理 j 的模型 $\theta_n^j \in \mathbb{R}^d$；第 n 次迭代时的参与设备 G_n；采样概率 $P_n^j \in \mathbb{R}^{N-1}$；CBO系统超参数 λ_1，α；离散化步长 γ；随机采样比例 $\varepsilon \in (0, 1)$；下载的模型数量 M;

1: $A_n \leftarrow$ ε-**贪心采样**(P_n^j，G_n，M);
2: 代理 j 下载用于 $i \in A_n$ 的模型 θ_n^i ;
3: 分别在代理 j的数据集上评估模型 θ_n^i ，并将相应的损失记为 L_j^i;
4: 计算共识点 m b

$$m_j \leftarrow {}^{j}\ {}^{\mathrm{y}} \frac{1}{\sum_{i\in A_n}\mu_j^i} \sum_{i\in A_n} \theta_n^i \mu_j^i, \qquad \text{with } \mu_j^i = \exp(-\alpha L_j^i) \tag{19}$$

5: 通过以下方式更新代理 j 的模型

$$\theta_{n+1}^j \leftarrow \theta_n^j - \lambda_1\gamma(\theta_n^j - m_j), \tag{20}$$

6: 通过以下方式更新采样概率 P_n^j

$$P_{n+1}^{j,i} \leftarrow P_n^{j,i} + (L_j^j - L_j^i), \qquad \text{for } i \in A_n \tag{21}$$

输出： θ_{n+1}^j，P_{n+1}^j

$\underline{\varepsilon\text{-}\textbf{贪心采样}(P_n^j, G_n, M)}$

7: 从 $\varepsilon * M$ 中随机采样数量为 G_n 的代理，记为 A_n^1；8: 选择 $1-\varepsilon$ 个 $* M$ 中的代理，其数值为 $G_n \backslash A_n^1$ 且具有最高值 $P^{j,}{}_j{}^i$，$i \in G_n \backslash A_n^1$记为 A_n^2；9: **返回** $A_n = A_n^1 \cup A_n^2$

现在可以正式介绍 FedCBO 算法（算法1）。在 FedCBO 的第 n 次迭代中，中央服务器选择一组参与代理 $G_n \subseteq [N]$；实际上，服务器可以从当前在线且可用的代理中选择这一组代理。每个选定的代理 $j \in G_n$ 使用其个人数据集 θ_n^j 对其模型执行本地 SGD 更新。本地更新完成后，每个参与代理 $j \in G_n$ 开始本地聚合（算法 2）。具体而言，代理 j 首先使用例如 ε-贪心采样策略 [38] 选择一组代理$A_n \subseteq G_n$ （详情请参见备注 3），然后下载这些代理的模型（关于这一做法及其他联邦学习方案中的数据隐私漏洞，请参见备注 4）。随后，代理 j 在其本地数据集上评估所有下载的模型 $\theta_j^i i \in A_n$，并获得相应的损失 L_j^i。利用这些损失 L_j^i，代理 j 根据式（19）计算共识点 m_j，并根据公式（20）更新其自身模型 θ_n^j。最后，代理 j 根据式（21）更新其采样概率向量P_n^j，以用于未来的通信轮次。正如上文所述，FedCBO 中的模型聚合步骤不同于大多数传统联邦学习算法中的模型聚合步骤。在 FedCBO 中，模型在每台设备上进行本地聚合，而传统联邦学习算法则在全局服务器上对模型进行平均。

注释 *3*（ε-贪心采样策略）。对于每个代理 *j*，我们使用 ε-贪心采样方案，如 *[38]* 所示，以选择从其他代理下载哪些模型。具体而言，我们维护一个矩阵 P，该矩阵由行向量 $p^j = (p^{j,1}, \ldots, p^{j,N})$ 组成，其中 $p^{j,i}$ 衡量代理 j 下载模型 θ^i 的概率。初始时，我们将 P 设为零矩阵，即每个模型被任何其他代理选中的机会均等。在每次联邦迭代期间，我们根据公式 *(21)* 更新 P。由于允许下载的模型数量 M 远小于代理总数 N，我们可以通过额外的探索获益：随机选择 ε 比例的代理，然后根据 P，按照采样概率最高的顺序选择剩余比例的代理。经过若干次迭代后，概率矩阵 P 应能更准确地识别相似的代理。因此，我们逐渐降低 ε 的值，以控制随机探索率。

注释 4。鉴于联邦学习方法所面临的数据隐私约束，各个代理不得与全局服务器或其他代理共享其本地数据。在标准联邦训练协议中，代理通常会交换基于各自本地数据集训练的模型的梯度或参数。不过，需要注意的是，共享梯度并非完全安全，因为攻击者有可能从公开共享的梯度中恢复私有训练数据 *[14, 18, 24, 37, 39, 40]*。相比之下，从模型参数中重建训练数据信息比从梯度中重建更困难，尤其是在查询次数有限时。因此，在我们的 *FedCBO* 算法中，即使存在潜在的训练数据反演攻击，允许代理下载其他代理的模型也相对安全。

2.4. **实验**。本节将对我们提出的 FedCBO 算法进行实证研究，并评估其在聚类联邦学习场景下相较于其他最先进学习方法的表现[3]。

数据集设置：我们采用 [15] 中使用的方法，构建基于标准 MNIST 数据集 [22]的簇式联邦学习设置。具体而言，我们从原始 MNIST 数据集开始，其中包含 60,000 张训练图像和 10,000 张测试图像。我们通过将每张图像分别旋转 0、90、180 和 270 度来扩充该数据集，从而生成 $k = 4$ 个簇，每个簇对应一个旋转角度。训练时，我们将全部训练图像 $60000k$ 随机划分到 N台客户端机器上，使每个代理持有 $n = \frac{60000k}{N}$ 张图像，且这些图像均来自同一旋转角度。推理时，我们不对测试数据进行划分。因此，每个本地代理的模型都将在 10,000 张旋转后的测试图像上进行评估，这些图像对应于该代理所属的簇。图2展示了旋转 MNIST 数据集的几个示例。

基线方法与实现：我们将 FedCBO 算法与三种基线算法进行比较：IFCA [15], FedAvg [28], 以及一种本地模型训练 方案。我们使用具有一个大小为 200 的隐藏层并采用 ReLU 激活函数的全连接神经网络作为基础模型。我们设置代理总数$N = 1200$ 和通信轮次 $T = 100$。在每个通信轮次中，所有代理均参与训练，即对于所有 n，$|G_n| = N$。当一个代理在其本地数据集上训练自己的模型（每轮中的本地更新步骤）时，我们运行 $\tau = 10$ 轮随机梯度下降（SGD）训练，学习率为

[3]我们的实验实现已开源，代码位于 https://github.com/SixuLi/FedCBO。11

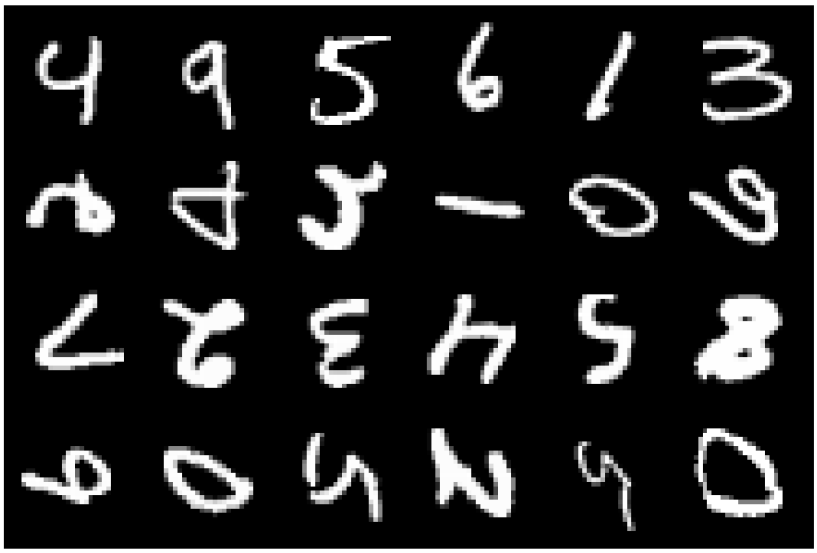

FIGURE 2. 旋转 MNIST 数据集示例。每一行包含来自某一特定旋转的一组样本。

$\gamma = 0.1$，动量为 0.9。下面我们介绍各基线算法的具体实现细节：

- **FedCBO：** 我们将模型下载预算设为 $M = 200$。我们选择超参数 $\lambda_1 = 10$、$\lambda_2 = 1$ 和 $\alpha = 10$。对于 ε-贪心采样，我们采用衰减方案 $\varepsilon(n) = \max\{0.5 - 0.01n, 0.1\}$，即初始随机采样比例为 $\varepsilon = 0.5$。该参数在每个通信轮次减少 0.01，直到达到阈值 0.1。
- **IFCA [15]：** 我们将全局服务器上初始化的模型数量设置为等于底层簇的数量（$k = 4$），如 [15] 所建议的，以进行公平比较。这应当为 IFCA 算法提供最佳结果。
- **FedAvg [28]：** 该算法试图训练一个适用于所有本地分布的单一全局模型。因此，在模型聚合步骤中，对代理训练的本地模型进行平均，以获得更新后的全局模型。
- **本地模型训练：** 各智能体仅使用各自的本地数据训练自己的模型，不与全局服务器或其他任何智能体进行通信。为确保比较的公平性，每个智能体均对其模型进行总计 $T * \tau = 1000$ 个训练轮次的训练。

备注 5. 请注意，在 *FedCBO* 中，我们无需输入底层簇的数量 k；而在 *IFCA* 中，我们需要输入该值或其估计值。

对于 FedCBO 和本地模型训练方案，我们在与训练数据具有相同分布的测试数据上测试训练好的模型，以进行推理（即具有相同旋转的数据点）。对于 IFCA 和 FedAvg，按照 [15], ，我们在每种数据分布上对所有学习到的模型（IFCA 有 k 个模型，FedAvg 有一个模型）进行推理，并计算损失值最小的模型的准确率。我们对所有算法使用 5 个不同的随机种子进行实验，并报告平均准确率和标准差。

实验结果： 测试结果汇总于表1。我们观察到，FedCBO 算法优于三种基线方法。尽管 IFCA 算法能够逐步准确估计用户的簇身份，并对估计为属于同一簇的用户模型进行平均，但它在模型聚合步骤中为各模型赋予相同的权重，因此算法无法进一步利用不同模型之间的相对相似性。相比之下，在运行 FedCBO 算法时，我们观察到，在（本地）模型聚合步骤中，智能体能够成功选取来自其他数据分布相同用户的模型（如备注6所述），并在聚合步骤中使用（19）为下载的模型赋予不同的重要性（权重）。这样，每位用户都能更好地利用其他用户提供的最有益模型。我们将这一平均方案（19）归因于

TABLE 1. 旋转 MNIST 上的测试准确率 ± 标准差%。

FEDCBO	IFCA	FEDAVG	LOCAL
96.51 ± 0.04	94.44 ± 0.01	85.50 ± 0.19	81.27 ± 0.02

这是 FedCBO 方法优于 IFCA 算法的原因之一。正如 [15], 所指出的，FedAvg 基线的表现不如 FedCBO 和 IFCA，因为它试图用一个模型拟合异质数据，因此无法提供簇级预测。由于每个智能体仅存储少量数据，本地模型训练方案很容易对本地数据集过拟合。这解释了为什么它在所有其他方法中表现最差。

备注 6. 为验证 *FedCBO* 中采样方案的正确性，我们将代理 j 在迭代 n 时的成功选择率（*SR*）定义如下：

$$SR_n^j := \frac{Number\ of\ selected\ agents\ in\ the\ same\ cluster\ as\ agent\ j}{Total\ number\ of\ selected\ agents}, \tag{22}$$

其中，选中的智能体总数等于模型下载预算 M。在 *FedCBO* 算法运行过程中，我们计算每个通信轮次 n ：$= \frac{1}{N}\sum_{j=1}^{N} SR_n^j$ 的平均成功选择率 n，对应图 3中的蓝色曲线。同时，在实施 ε-贪心采样时，我们将随机探索比例设为 $\varepsilon\,0。5$，并在 $n = 0$ 采用 $\varepsilon(n)$ $= \max\{0.5 - 0.01n, 0.1\}$衰减方案。因此，我们可以计算每轮的预言机期望成功选择率，如图 3中的橙色曲线所示。我们注意到，经验平均成功 *SR*（蓝色曲线）与最佳预期成功 *SR*（橙色曲线）非常接近。这表明我们的 *FedCBO* 算法能够成功识别具有相同数据分布的智能体。未来我们将研究设计更好的采样策略，以缩小经验成功 *SR* 与预言机成功 *SR* 之间的差距。

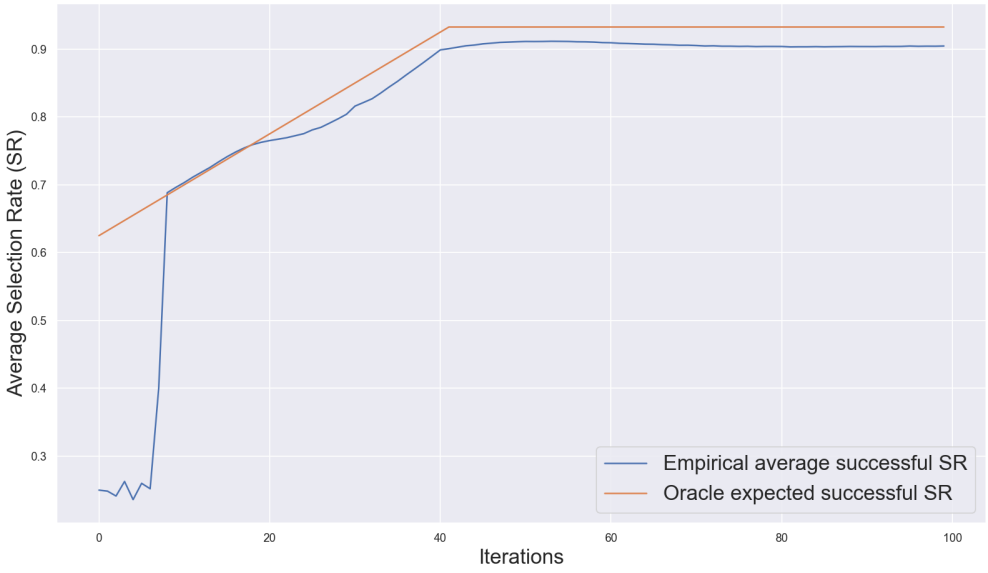

FIGURE 3. 每个通信轮次的平均成功选择率（SR）。

3. WELL-POSEDNESS OF MEAN-FIELD EQUATIONS

本节中，我们证明定理 2.1。我们将详细阐述假设损失函数有界的情形。二次增长情形的证明与此类似，更多详情请参阅附录。

引理 *3.1.* 设 L_1, L_2 满足假设 *1*，且 $\nu^1, \nu^2 \in \mathcal{C}([0,T], \mathcal{P}_2(\mathbb{R}^d))$ 满足 $\sup_{t\in[0,T]} \int |\theta|^4 d\nu_t^1 \le K$、$\sup_{t\in[0,T]} \int |\theta|^4 d\nu_t^2 \le K$。记 L_1, L_2 为各损失函数的上确界。设 $w_1, w_2 > 0$ 满足 $w_1 + w_2 = 1$，且令 $\nu := w_1\nu^1 + w_2\nu^2$。那么，对于所有 $s, t \in (0, T)$，以下稳定性估计成立。

$$|m_{L_k}^\alpha[\nu_t] - m_{L_k}^\alpha[\nu_s]| \le C\Big(w_1 W_2(\nu_t^1, \nu_s^1) + w_2 W_2(\nu_t^2, \nu_s^2)\Big), \tag{23}$$

对于 $k = 1, 2$ 且对于仅取决于 α、M_{L_1}、M_{L_2} 和 K 的常数 C。

证明。对于 $k = 1, 2$ 和 $s, t \in [0, T]$，可以计算其差值

$$\begin{aligned}
|m_{L_k}^\alpha[\nu_t] - m_{L_k}^\alpha[\nu_s]| &= |\frac{\int \theta w_{L_k}^\alpha d\nu_t}{\|w_{L_k}^\alpha\|_{\mathbb{L}^1(\nu_t)}} - \frac{\int \theta w_{L_k}^\alpha d\nu_s}{\|w_{L_k}^\alpha\|_{\mathbb{L}^1(\nu_s)}}| \\
&\le \frac{1}{\|w_{L_k}^\alpha\|_{\mathbb{L}^1(\nu_t)}} |\int \theta w_{L_k}^\alpha d\nu_t - \int \theta w_{L_k}^\alpha d\nu_s| + |\int \theta w_{L_k}^\alpha d\nu_s| |\frac{1}{\|w_{L_k}^\alpha\|_{\mathbb{L}^1(\nu_t)}} - \frac{1}{\|w_{L_k}^\alpha\|_{\mathbb{L}^1(\nu_s)}}| \\
&=: A_1 + A_2.
\end{aligned}$$

为估计 A_1，请注意

$$|\int \theta w_{L_k}^\alpha d\nu_t - \int \theta w_{L_k}^\alpha d\nu_s| \le w_1 |\int \theta w_{L_k}^\alpha d\nu_t^1 - \int \theta w_{L_k}^\alpha d\nu_s^1| + w_2 |\int \theta w_{L_k}^\alpha d\nu_t^2 - \int \theta w_{L_k}^\alpha d\nu_s^2|. \tag{24}$$

让我们来看一下右侧的第一项。

$$|\int \theta w_{L_k}^\alpha(\theta) d\nu_t^1(\theta) - \int \hat\theta w_{L_k}^\alpha(\hat\theta) d\nu_s^1(\hat\theta)| = |\int\int \big[\theta w_{L_k}^\alpha(\theta) - \hat\theta w_{L_k}^\alpha(\hat\theta)\big] d\pi(\theta, \hat\theta)|,$$

其中 π 是 ν_t^1 与 ν_s^1 之间的任意耦合。记 $h(\theta) - h(\hat\theta)$ 为：$= \theta w_{L_k}^\alpha(\theta) - \hat\theta w_{L_k}^\alpha(\hat\theta)$，则根据关于 1 的假设 L_k，我们有

$$\begin{aligned}
|h(\theta) - h(\hat\theta)| &= |w_{L_k}^\alpha(\theta)(\theta - \hat\theta) + \hat\theta\big[w_{L_k}^\alpha(\theta) - w_{L_k}^\alpha(\hat\theta)\big]| \\
&\le \exp(-\alpha \underline{L_k})|\theta - \hat\theta| + \alpha \exp(-\alpha \underline{L_k}) M_{L_k} |\hat\theta|(|\theta| + |\hat\theta|)|\theta - \hat\theta|.
\end{aligned}$$

因此，c 取 π 为 ν_t^1 与 ν_s^1 之间关于 W_2-距离的最优耦合，我们得到 n

$$\begin{aligned}
|\int \theta w_{L_k}^\alpha(\theta) d\nu_t^1(\theta) - \int \hat\theta w_{L_k}^\alpha(\hat\theta) d\nu_s^1(\hat\theta)| &\le \iint \exp(-\alpha \underline{L_k})\big(1 + \alpha M_{L_k} |\hat\theta|(|\theta| + |\hat\theta|)\big)|\theta - \hat\theta|) d\pi \\
&\le \Big(\iint \exp(-2\alpha \underline{L_k})\big(1 + \alpha M_{L_k} |\hat\theta|(|\theta| + |\hat\theta|)\big)^2 d\pi\Big)^{\frac{1}{2}} \Big(\iint |\theta - \hat\theta|^2 d\pi\Big)^{\frac{1}{2}} \\
&\le \exp(-\alpha \underline{L_k}) \alpha M_{L_k} p_K \Big(\iint |\theta - \hat\theta|^2 d\pi\Big)^{\frac{1}{2}} \\
&= \exp(-\alpha \underline{L_k}) \alpha M_{L_k} p_K W_2(\nu_t^1, \nu_s^1)\,.
\end{aligned}$$

这里，p_K 是关于 K的多项式。同理，我们推得

$$|\int \theta w_{L_k}^\alpha d\nu_t^2 - \int \theta w_{L_k}^\alpha d\nu_s^2| \le \exp(-\alpha \underline{L_k}) \alpha M_{L_k} p_K W_2(\nu_t^2, \nu_s^2)\,.$$

因此，我们得出结论

$$\begin{aligned}
|\int \theta w_{L_k}^\alpha d\nu_t - \int \theta w_{L_k}^\alpha d\nu_s| &\le w_1 |\int \theta w_{L_k}^\alpha d\nu_t^1 - \int \theta w_{L_k}^\alpha d\nu_s^1| + w_2 |\int \theta w_{L_k}^\alpha d\nu_t^2 - \int \theta w_{L_k}^\alpha d\nu_s^2| \\
&\le \exp(-\alpha \underline{L_k}) \alpha M_{L_k} p_K \big(w_1 W_2(\nu_t^1, \nu_s^1) + w_2 W_2(\nu_t^2, \nu_s^2)\big)\,,
\end{aligned}$$

and

$$A_1 := \frac{1}{\|w_{L_k}^\alpha\|_{\mathbb{L}^1(\nu_t)}} |\int \theta w_{L_k}^\alpha d\nu_t - \int \theta w_{L_k}^\alpha d\nu_s| \le \exp(\alpha(\overline{L_k} - \underline{L_k})) \alpha M_{L_k} p_K \big(w_1 W_2(\nu_t^1, \nu_s^1) + w_2 W_2(\nu_t^2, \nu_s^2)\big)\,.$$

对于第二项 A_2，我们有

$$A_2 := |\int \theta w_{L_k}^\alpha d\nu_s| |\frac{1}{\|w_{L_k}^\alpha\|_{\mathbb{L}^1(\nu_t)}} - \frac{1}{\|w_{L_k}^\alpha\|_{\mathbb{L}^1(\nu_s)}}| = \frac{\int \theta w_{L_k}^\alpha d\nu_s}{\|w_{L_k}^\alpha\|_{\mathbb{L}^1(\nu_t)} \|w_{L_k}^\alpha\|_{\mathbb{L}^1(\nu_s)}} |\int w_{L_k}^\alpha d\nu_t - \int w_{L_k}^\alpha d\nu_s|.$$

与（24）中的估计类似，我们有

$$|\int w_{L_k}^{\alpha} d\nu_t - \int w_{L_k}^{\alpha} d\nu_s| \leq \exp(-\alpha \underline{L_k}) \alpha M_{L_k} q_K \Big(w_1 W_2(\nu_t^1, \nu_s^1) + w_2 W_2(\nu_t^2, \nu_s^2) \Big),$$

其中， q_K 是 K上的多项式。于是，

$$\begin{aligned} A_2 &= \frac{\int \theta w_{L_k}^{\alpha} d\nu_s}{\|w_{L_k}^{\alpha}\|_{\mathbb{L}^1(\nu_t)} \|w_{L_k}^{\alpha}\|_{\mathbb{L}^1(\nu_s)}} |\int w_{L_k}^{\alpha} d\nu_t - \int w_{L_k}^{\alpha} d\nu_s| \\ &\leq C_{K_k} \exp(\alpha \overline{L_k}) \exp(-\alpha \underline{L_k}) \alpha M_{L_k} q_K \Big(w_1 W_2(\nu_t^1, \nu_s^1) + w_2 W_2(\nu_t^2, \nu_s^2) \Big) \\ &= C_{K_k} \exp(\alpha(\overline{L_k} - \underline{L_k})) \alpha M_{L_k} q_K \Big(w_1 W_2(\nu_t^1, \nu_s^1) + w_2 W_2(\nu_t^2, \nu_s^2) \Big). \end{aligned}$$

结合对 A_1 和 A_2的估计，可得

$$|m_{L_k}^{\alpha}[\nu_t] - m_{L_k}^{\alpha}[\nu_s]| \leq C \Big(w_1 W_2(\nu_t^1, \nu_s^1) + w_2 W_2(\nu_t^2, \nu_s^2) \Big),$$

where $C := \max \Big\{ \exp(\alpha(\overline{L_1} - \underline{L_1})) \alpha M_{L_1}(p_K + C_{K_1} q_K), \exp(\alpha(\overline{L_2} - \underline{L_2})) \alpha M_{L_2}(p_K + C_{K_2} q_K) \Big\}$. □

定理*2.1*的证明。证明分为几个步骤。

步骤 1： 给定一对 $u^1, u^2 \in \mathcal{C}([0,T], \mathbb{R}^d)$，我们可以应用随机微分方程的标准理论，得出该随机微分方程存在唯一的强解

$$dY_t^1 = -\lambda_1(Y_t^1 - u_t^1)dt - \lambda_2 \nabla L_1(Y_t^1)dt + \sigma_1 |Y_t^1 - u_t^1| dB_t^1 + \sigma_2 |\nabla L_1(Y_t^1)| d\widetilde{B}_t^1 \tag{25a}$$

$$dY_t^2 = -\lambda_1(Y_t^2 - u_t^2)dt - \lambda_2 \nabla L_2(Y_t^2)dt + \sigma_1 |Y_t^2 - u_t^2| dB_t^2 + \sigma_2 |\nabla L_2(Y_t^2)| d\widetilde{B}_t^2, \tag{25b}$$

对于初始条件 $Y_0^1 \sim \rho_0^1$ 和 $Y_0^2 \sim \rho_2$（彼此独立），其中 $\rho_0^1, \rho_0^2 \in \mathcal{P}_4(\mathbb{R}^d)$。令$\nu_t^1 = \mathsf{Law}(Y_t^1)$和 $\nu_t^2 = \mathsf{Law}(Y_t^2)$分别表示 Y_t^1和 Y_t^2的分布。由于变量 Y^1, Y^2 取值于 $\mathcal{C}([0,T], \mathbb{R}^d)$，因此 $\nu^1, \nu^2 \in \mathcal{C}([0,T], \mathcal{P}(\mathbb{R}^d))$。此外， ν_t^1, ν_t^2满足以下福克-普朗克方程组（弱形式）：

$$\frac{d}{dt} \int \varphi d\nu_t^1 = \int \Big[\big(-\lambda_1(\theta - u_t^1) - \lambda_2 \nabla L_1(\theta) \big) \cdot \nabla \varphi + \big(\frac{\sigma_1^2}{2} |\theta - u_t^1|^2 + \frac{\sigma_2^2}{2} |\nabla L_1(\theta)|^2 \big) \Delta \varphi \Big] d\nu_t^1, \tag{26a}$$

$$\frac{d}{dt} \int \varphi d\nu_t^2 = \int \Big[\big(-\lambda_1(\theta - u_t^2) - \lambda_2 \nabla L_2(\theta) \big) \cdot \nabla \varphi + \big(\frac{\sigma_1^2}{2} |\theta - u_t^2|^2 + \frac{\sigma_2^2}{2} |\nabla L_2(\theta)|^2 \big) \Delta \varphi \Big] d\nu_t^2, \tag{26b}$$

对所有 $\varphi \in \mathcal{C}_c^2(\mathbb{R}^d)$。考虑乘积空间 $\mathcal{C}([0,T], \mathbb{R}^d) \times \mathcal{C}([0,T], \mathbb{R}^d)$，其赋予的范数$\|\cdot\|$ 定义为

$$\|(f,g)\| := \|f\|_{\infty} + \|g\|_{\infty},$$

其中 f、$g \in \mathcal{C}([0,T], \mathbb{R}^d)$ 且 $\|f\|_{\infty} := \sup_{t \in [0,T]} |f_t|$。令 P_k: $\mathcal{C}([0,T], \mathbb{R}^d) \times \mathcal{C}([0,T], \mathbb{R}^d) \to \mathcal{C}([0,T], \mathbb{R}^d)$ 表示对 $k = 1$、2 的第 k 个坐标的投影。也就是说，对于任意 $(f,g) \in \mathcal{C}([0,T], \mathbb{R}^d) \times \mathcal{C}([0,T], \mathbb{R}^d)$，$P_1(f,g) = f$ 且 $P_2(f,g) = g$。注意，对于 $k = 1$、2，有 $m_{L_k}[\nu] \in \mathcal{C}([0,T], \mathbb{R}^d)$，因此我们可以定义映射

$$\begin{aligned} \mathcal{T} : \mathcal{C}([0,T], \mathbb{R}^d) \times \mathcal{C}([0,T], \mathbb{R}^d) &\longrightarrow \mathcal{C}([0,T], \mathbb{R}^d) \times \mathcal{C}([0,T], \mathbb{R}^d), \\ (u^1, u^2) &\longrightarrow \mathcal{T}(u^1, u^2) = (m_{L_1}^{\alpha}[\nu], m_{L_2}^{\alpha}[\nu]), \end{aligned}$$

其中 $\nu = w_1 \nu^1 + w_2 \nu^2$。

接下来，我们证明 $\mathcal{T}$ 存在唯一的不动点。

步骤 2： 首先，我们证明 $\mathcal{T}$是紧的，也就是说， $\mathcal{C}([0,T], \mathbb{R}^d)$ 中的任意有界序列 $\{(f_n, g_n)\}_{n \in \mathbb{N}} \times \mathcal{C}([0,T], \mathbb{R}^d)$ 都是预紧的。令 $(\varphi_n, \psi_n) := \mathcal{T}(f_n, g_n)$。只需证明序列 $\{\varphi_n\}_{n \in \mathbb{N}}$ 和 $\{\psi_n\}_{n \in \mathbb{N}}$ 各自都是预紧的。下面给出序列 $\{\varphi_n\}_{n \in \mathbb{N}}$的详细证明。由于 $\rho_0^1, \rho_0^2 \in \mathcal{P}_4(\mathbb{R}^d)$,

随机微分方程的标准理论（例如，参见 [1]第 7 章）给出了方程 (25a) 和 (25b) 解的四阶矩估计，其形式为

$$\mathbb{E}|Y_t^1|^4 \le (1+\mathbb{E}|Y_0^1|^4)\exp(ct) \qquad \text{and} \qquad \mathbb{E}|Y_t^2|^4 \le (1+\mathbb{E}|Y_0^2|^4)\exp(ct),$$

其中常数 $c>0$ 仅取决于 $\|u^1\|_\infty$ 和 $\|u^2\|_\infty$。特别地，

$$\sup_{t\in[0,T]}\int |x|^4 d\nu_t^1, \ \sup_{t\in[0,T]}\int |x|^4 d\nu_t^2 \le K$$

其中某个 $K<\infty$。另一方面，对于任意 $t>s$， t， $s\in(0,\ T)$，Itô 等距性和柯西–施瓦茨不等式给出

$$\begin{aligned}
\mathbb{E}|Y_t^1 - Y_s^1|^2 &= \mathbb{E}\bigg[\int_s^t \big(-\lambda_1(Y_\tau^1-u_\tau^1)-\lambda_2\nabla L_1(Y_\tau^1)\big)d\tau + \int_s^t \sigma_1|Y_\tau^1-u_\tau^1|dB_\tau^1 + \int_s^t \sigma_2|\nabla L_1(Y_\tau^1)|d\widetilde{B}_\tau^1\bigg]^2 \\
&\le 2\mathbb{E}\bigg[\int_s^t \big(-\lambda_1(Y_\tau^1-u_\tau^1)-\lambda_2\nabla L_1(Y_\tau^1)\big)d\tau\bigg]^2 + 2\mathbb{E}\bigg[\int_s^t \big(\sigma^2|Y_\tau^1-u_\tau^1|^2+\sigma_2^2|\nabla L_1(Y_\tau^1)|^2\big)d\tau\bigg] \\
&\le 2\lambda_1^2|t-s|\mathbb{E}\big[\int_s^t |Y_\tau^1-u_\tau^1|^2 d\tau\big] + 2\lambda_2^2|t-s|\mathbb{E}\big[\int_s^t |\nabla L_1(Y_\tau^1)|^2 d\tau\big] \\
&\quad + \mathbb{E}\bigg[\int_s^t \big(\sigma^2|Y_\tau^1-u_\tau^1|^2+\sigma_2^2|\nabla L_1(Y_\tau^1)|^2\big)d\tau\bigg] \\
&\le 4\lambda_1^2(K+\|u^1\|_\infty^2)T|t-s| + 2\lambda_2^2 C_{\nabla L_1}^2 T|t-s| + 2\sigma_1^2(K+\|u^1\|_\infty^2)|t-s| + \sigma_2^2 C_{\nabla L_1}^2|t-s| \\
&=: C_1|t-s|.
\end{aligned}$$

类似地， $\mathbb{E}|Y_t^2-Y_s^2|^2 \le C_2|t-s|$ 对于常数 $C_2>0$。因此， $W_2(\nu_t^1,\ \nu_s^1)\le c_1|t-s|^{\frac{1}{2}}$和 $W_2(\nu_t^2,\ \nu_s^2)\le c_2|t-s|^{\frac{1}{2}}$，其中常数 c_1， $c_2>0$仅取决于 $\|u^1\|_\infty$和 $\|u^2\|_\infty$。应用引理3.1，可得

$$|m_{L_1}^\alpha[\nu_t]-m_{L_1}^\alpha[\nu_s]| \le C(w_1c_1+w_2c_2)|t-s|^{\frac{1}{2}},$$

这证明了 $t\to m_{L_1}^\alpha[\nu_t]$ 具有指数为 $\frac{1}{2}$ 的 Hölder 连续性。由此，结合紧嵌入 $\mathcal{C}^{0,\ 1/2}([0,T],\ \mathbb{R}^d)\hookrightarrow\mathcal{C}([0,T],\ \mathbb{R}^d)$，我们可以得出 $\{\varphi_n\}_{n\in\mathbb{N}}$ 是预紧的，其中 $\mathcal{C}^{0,\ 1/2}([0,T],\ \mathbb{R}^d)$ 是从 $[0,T]$ 到 $\mathbb{R}^d$ 的 $\frac{1}{2}$-Hölder 连续函数空间。

步骤 3：接下来，我们验证 Leray-Schauder不动点定理中的条件。为此，假设对某个 $\tau\in[0,1]$，有序对 $(u^1,u^2)\in\mathcal{C}([0,T],\mathbb{R}^d)\times\mathcal{C}([0,T],\mathbb{R}^d)$ 满足 $(u^1,u^2)=\tau\mathcal{T}(u^1,u^2)$。特别地，存在满足式 (26a) 和 (26b) 的 ν_1，$\nu_2\in\mathcal{C}([0,T],\mathcal{P}_2(\mathbb{R}^d))$，使得 $(u^1,u^2)=\tau(m_{L_1}^\alpha[\nu],m_{L_2}^\alpha[\nu])$，其中 $\nu=w_1\nu_1+w_2\nu_2$。根据关于 L_1 的有界性假设，对于所有 $t\in(0,T)$，我们有

(27)

$$|u_t^1|^2=\tau^2|m_{L_1}[\nu]|^2\le\tau^2\exp(\alpha(\overline{L_1}-\underline{L_1}))\int|x|^2d\nu_t=\tau^2\exp(\alpha(\overline{L_1}-\underline{L_1}))\bigg(w_1\int|x|^2d\nu_t^1+w_2\int|x|^2d\nu_t^2\bigg).$$

对 ν_t^1 的二阶矩进行计算可得

$$\begin{aligned}
\frac{d}{dt}\int|\theta|^2d\nu_t^1 &= \int\bigg[\big(-\lambda_1(\theta-u_t^1)-\lambda_2\nabla L_1(\theta)\big)\cdot 2\theta + 2d\big(\frac{\sigma_1^2}{2}|\theta-u_t^1|^2+\frac{\sigma_2^2}{2}|\nabla L_1(\theta)|^2\big)\bigg]d\nu_t^1 \\
&= \int\bigg[-2\lambda_1|\theta|^2+2\lambda_1\theta\cdot u_t^1-2\lambda_2\theta\cdot\nabla L_1(\theta)+d\sigma_1^2(|\theta|^2-2\theta\cdot u_t^1+|u_t^1|^2)+d\sigma_2^2|\nabla L_1(\theta)|^2\bigg]d\nu_t^1 \\
&\le \int\bigg[(d\sigma_1^2-2\lambda_1+|\gamma|+\lambda_2)|\theta|^2+(d\sigma_1^2+|\gamma|)|u_t^1|^2+(\lambda_2+d\sigma_2^2)|\nabla L_1(\theta)|^2\bigg]d\nu_t^1 \\
&\le (d\sigma_1^2-2\lambda_1+|\gamma|+\lambda_2)\int|\theta|^2d\nu_t^1+(d\sigma_1^2+|\gamma|)|u_t^1|^2+(\lambda_2+d\sigma_2^2)C_{\nabla L_1}^2 \\
&\le (d\sigma_1^2+|\gamma|+\lambda_2)\big(1+\exp(\alpha(\overline{L_1}-\underline{L_1}))\big)\bigg(\int|\theta|^2d\nu_t^1+\int|\theta|^2d\nu_t^2\bigg)+(\lambda_2+d\sigma_2^2)C_{\nabla L_1}^2,
\end{aligned}$$

其中 $\gamma := \lambda_1 - d\sigma_1^2$。类似地，

$$\frac{d}{dt}\int |\theta|^2 d\nu_t^2 \le (d\sigma_1^2 + |\gamma| + \lambda_2)\big(1+\exp(\alpha(\overline{L_2} - \underline{L_2}))\big)\bigg(\int |\theta|^2 d\nu_t^1 + \int |\theta|^2 d\nu_t^2\bigg) + (\lambda_2 + d\sigma_2^2) C_{\nabla L_2}^2.$$

将上述两个不等式相加，我们得到

$$\frac{d}{dt}\bigg(\int |\theta|^2 d\nu_t^1 + \int |\theta|^2 d\nu_t^2\bigg) \le C_1\bigg(\int |\theta|^2 d\nu_t^1 + \int |\theta|^2 d\nu_t^2\bigg) + C_2,$$

其中某些常数为 C_1， $C_2 > 0$。利用 Gr¨onwall 不等式，我们得到

$$\int |\theta|^2 d\nu_t^1 + \int |\theta|^2 d\nu_t^2 \le \bigg(\int |\theta|^2 d\nu_0^1 + \int |\theta|^2 d\nu_0^2\bigg)\exp(C_1 t) + \frac{C_2}{C_1}\big(\exp(C_1 t) - 1\big).$$

然后，由（27），我们得出存在一个常数 $q_1 > 0$ 使得 $\|u^1\|_\infty < q_1$。对于$\|u^2\|_\infty$，也有类似的界，即存在一个常数 $q_2 > 0$ 使得 $\|u^2\|_\infty < q_2$。因此， $\|(u^1,\ u^2)\| = \|u^1\|_\infty + \|u^2\|_\infty < q_1 + q_2$。现在我们可以应用 Leray-Schauder 不动点定理（参见 [11]第 9.2 节），从而得出映射 u^1， u^2存在一个不动点（ $\mathcal{T}$，由此得到（25a）和（25b）的一个解。

步骤 4： 关于唯一性，我们首先注意到， $\mathcal{T}$ 的一个不动点（u^1， u^2）必须满足 $\|(u^1,\ u^2)\| < q$。因此，**步骤 2** 中给出的四阶矩估计成立，并且对于 $k = 1$， 2，有 $\sup_{t\in[0,T]}\int |x|^4 d\nu_t^k \le K < \infty$。现在假设有两个不动点（$u^1$， u^2）和（$\hat{}u^1$，$\hat{}u^2$），且

$$\|(u^1,u^2)\|, \|(\hat{u}^1,\hat{u}^2)\| < q, \qquad \sup_{t\in[0,T]}\int |x|^4 d\nu_t^k, \sup_{t\in[0,T]}\int |x|^4 d\hat{\nu}_t^k \le K \quad \text{for } k = 1,2,$$

并考虑它们对应的过程（Y_t^1， Y_t^2），（$\hat{}Y_t^1$，$\hat{}Y_t^2$），它们满足（25a）和（25b），且两者使用相同的布朗运动，对于 $k = 1$， 2。取差值 z_t^k: $= Y_t^k - \hat{}Y_t^k$，对于 $k = 1$， 2，我们得到

$$\begin{aligned} z_t^k &= z_0^k + \int_0^t \big(-\lambda_1 z_s^k + \lambda_1(u_s^k - \hat{u}_s^k) - \lambda_2(\nabla L_k(Y_s^k) - \nabla L_k(\hat{Y}_s^k))\big)ds \\ &\quad + \sigma_1\int_0^t \big(|Y_s^k - u_s^k| - |\hat{Y}_s^k - \hat{u}_s^k|\big)dB_s^k + \sigma_2\int_0^t \big(|\nabla L_k(Y_s^k)| - |\nabla L_k(\hat{Y}_s^k)|\big)d\widetilde{B}_s^k. \end{aligned}$$

两边平方后取期望，并利用 Itô 等距性，可得

$$\begin{aligned} \mathbb{E}|z_t^k|^2 &= \mathbb{E}\bigg[z_0^k + \int_0^t \big(-\lambda_1 z_s^k + \lambda_1(u_s^k - \hat{u}_s^k) - \lambda_2(\nabla L_k(Y_s^k) - \nabla L_k(\hat{Y}_s^k))\big)ds \\ &\quad + \sigma_1\int_0^t \big(|Y_s^k - u_s^k| - |\hat{Y}_s^k - \hat{u}_s^k|\big)dB_s^k + \sigma_2\int_0^t \big(|\nabla L_k(Y_s^k)| - |\nabla L_k(\hat{Y}_s^k)|\big)d\widetilde{B}_s^k\bigg]^2 \\ &\le 2\mathbb{E}|z_0^k|^2 + 2t\mathbb{E}\bigg[\int_0^t \big(-\lambda_1 z_s^k + \lambda_1(u_s^k - \hat{u}_s^k) - \lambda_2(\nabla L_k(Y_s^k) - \nabla L_k(\hat{Y}_s^k))\big)^2 ds\bigg] \\ &\quad + 2\sigma_1^2\mathbb{E}\bigg[\int_0^t \big(|Y_s^k - u_s^k| - |\hat{Y}_s^k - \hat{u}_s^k|\big)^2 ds\bigg] + 2\sigma_2^2\mathbb{E}\bigg[\int_0^t \big(|\nabla L_k(Y_s^k)| - |\nabla L_k(\hat{Y}_s^k)|\big)^2 ds\bigg] \\ &\le 2\mathbb{E}|z_0^k|^2 + 6t\lambda_1^2\int_0^t \mathbb{E}|z_s^k|^2 ds + 6t\lambda_1^2\int_0^t |u_s^k - \hat{u}_s^k|^2 ds + 6t\lambda_2^2\int_0^t \mathbb{E}\big[\nabla L_k(Y_s^k) - \nabla L_k(\hat{Y}_s^k)\big]^2 ds \\ &\quad + 2\sigma_1^2\mathbb{E}\bigg[\int_0^t |(Y_s^k - \hat{Y}_s^k) - (u_s^k - \hat{u}_s^k)|^2 ds\bigg] + 2\sigma_2^2\mathbb{E}\bigg[\int_0^t |\nabla L_k(Y_s^k) - \nabla L_k(\hat{Y}_s^k)|^2 ds\bigg] \\ &\le 2\mathbb{E}|z_0^k|^2 + (6\lambda_1^2 t + 4\sigma_1^2)\int_0^t \mathbb{E}|z_s^k|^2 ds + (6\lambda_1^2 t + 4\sigma_1^2)\int_0^t |u_s^k - \hat{u}_s^k|^2 ds \\ &\quad + 6\lambda_2^2 M_{\nabla L_1}^2\int_0^t \mathbb{E}|z_s^k|^2 ds + 2\sigma_2^2 M_{\nabla L_1}^2\int_0^t \mathbb{E}|z_s^k|^2 ds \\ &= 2\mathbb{E}|z_0^k|^2 + \big[6\lambda_1^2 t + 4\sigma_1^2 + M_{\nabla L_1}^2(6\lambda_2^2 + 2\sigma_2^2)\big]\int_0^t \mathbb{E}|z_s^k|^2 ds + (6\lambda_1^2 t + 4\sigma_1^2)\int_0^t |u_s^k - \hat{u}_s^k|^2 ds. \end{aligned}$$

根据引理3.1，对于 $k=1,\ 2$，可得

$$
\begin{aligned}
|u_s^k-\hat{u}_s^k|^2=|m_{L_k}[\nu_s]-m_{L_k}[\hat{\nu}_s]|^2 &\le C^2\big(w_1W_2(\nu_s^1,\hat{\nu}_s^1)+w_2W_2(\nu_s^2,\hat{\nu}_s^2)\big)^2\\
&\le C^2\big(w_1\sqrt{\mathbb{E}|z_s^1|^2}+w_2\sqrt{\mathbb{E}|z_s^2|^2}\big)^2\\
&\le 2C^2\big(\mathbb{E}|z_s^1|^2+\mathbb{E}|z_s^2|^2\big).
\end{aligned}
\tag{28}
$$

我们进一步得到

$$
\begin{aligned}
\mathbb{E}|z_t^1|^2 &\le 2\mathbb{E}|z_0^1|^2+\widetilde{C}_1\int_0^t\mathbb{E}|z_s^1|^2ds+\widetilde{C}_2\int\mathbb{E}|z_s^2|^2ds\\
\mathbb{E}|z_t^2|^2 &\le 2\mathbb{E}|z_0^2|^2+\widetilde{C}_1\int_0^t\mathbb{E}|z_s^2|^2ds+\widetilde{C}_2\int\mathbb{E}|z_s^1|^2ds,
\end{aligned}
$$

其中 $\widetilde{C}_1=(1+2C^2)(6\lambda_1^2t+4\sigma_1^2)+(6\lambda_2^2+2\sigma_2^2)\max\{M_{\nabla L_1}^2,M_{\nabla L_2}^2\}$ 且 $\widetilde{C}_2=2C^2(6\lambda_1^2t+4\sigma_1^2)$。结合上述两个不等式，我们推得

$$
\mathbb{E}|z_t^1|^2+\mathbb{E}|z_t^2|^2\le 2\big(\mathbb{E}|z_0^1|^2+\mathbb{E}|z_0^2|^2\big)+(\widetilde{C}_1+\widetilde{C}_2)\int_0^t\big(\mathbb{E}|z_s^1|^2+\mathbb{E}|z_s^2|^2\big)ds.
$$

接着，根据 Gr¨onwall 不等式以及 $\mathbb{E}|z_0^1|^2=\mathbb{E}|z_0^2|^2=0$这一事实，我们推知 $\mathbb{E}|z_t^1|^2+\mathbb{E}|z_t^2|^2=0$ 对所有 $t\in[0,T]$成立。由不等式 (28，可得 $\|u^1-\hat{}u^1\|_\infty=\|u^2-\hat{}u^2\|_\infty=0$，即 $(u^1,\ u^2)\equiv(\hat{}u^1,\ \hat{}u^2)$，由此证明了唯一性。□

注 7. 注意，对于任意输入函数 $(u^1,\ u^2)$，*(25a)* 和 *(25b)* 中的随机过程 Y^1 和 Y^2 彼此独立。进而，由于 *(7a)* 和 *(7b)* 在特定的 $(u^1,\ u^2)$ 选择下（即在 $\mathcal{T}$ 的一个不动点处）实现为 $(Y^1,\ Y^2)$，我们得出结论：*(7a)* 和 *(7b)* 中的过程 θ^1、θ^2 作为随机过程彼此独立。不过请注意，这两个随机微分方程共享参数，例如，同时出现在方程漂移项和扩散项中的分布ρ 。

4. Large time behavior of mean-field equation and consensus formation

定理的证明 *2.3*。令 $M:=$ 最大值$\{M_{\nabla L_1},M_{\nabla L_2}\}$。利用引理 C.1 可得

$$
\begin{aligned}
\frac{d}{dt}\big(\mathcal{V}(\rho_t^1)+\mathcal{V}(\rho_t^2)\big) &\le -(2\lambda_1-2\lambda_2M-d\sigma_1^2-d\sigma_2^2M^2)\big(\mathcal{V}(\rho_t^1)+\mathcal{V}(\rho_t^2)\big)\\
&\quad+\sqrt{2}(\lambda_1+d\sigma_1^2)\Big(\sqrt{\mathcal{V}(\rho_t^1)}|m_{L_1}^\alpha[\rho_t]-\theta_1^*|+\sqrt{\mathcal{V}(\rho_t^2)}|m_{L_2}^\alpha[\rho_t]-\theta_2^*|\Big)\\
&\quad+\frac{d\sigma_1^2}{2}\big(|m_{L_1}^\alpha[\rho_t]-\theta_1^*|^2+|m_{L_2}^\alpha[\rho_t]-\theta_2^*|^2\big)\\
&\le -(2\lambda_1-2\lambda_2M-d\sigma_1^2-d\sigma_2^2M^2)\big(\mathcal{V}(\rho_t^1)+\mathcal{V}(\rho_t^2)\big)\\
&\quad+\sqrt{2}(\lambda_1+d\sigma_1^2)\big(|m_{L_1}^\alpha[\rho_t]-\theta_1^*|+|m_{L_2}^\alpha[\rho_t]-\theta_2^*|\big)\sqrt{\mathcal{V}(\rho_t^1)+\mathcal{V}(\rho_t^2)}\\
&\quad+\frac{d\sigma_1^2}{2}\big(|m_{L_1}^\alpha[\rho_t]-\theta_1^*|^2+|m_{L_2}^\alpha[\rho_t]-\theta_2^*|^2\big).
\end{aligned}
\tag{29}
$$

设 $T_\alpha\ge 0$ 由下式给出

$$
T_\alpha:=\sup\big\{t\ge 0:\mathcal{V}(\rho_{t'}^1)+\mathcal{V}(\rho_{t'}^2)>\varepsilon,|m_{L_1}^\alpha[\rho_{t'}]-\theta_1^*|+|m_{L_2}^\alpha[\rho_{t'}]-\theta_2^*|<C(t')\ \forall t'\in[0,t]\big\}
\tag{30}
$$

其中

$$
C(t):=C\sqrt{\mathcal{V}(\rho_t^1)+\mathcal{V}(\rho_t^2)}
\tag{31}
$$

with

$$
C:=\min\left\{\frac{\tau}{2}\frac{(2\lambda_1-2\lambda_2M-d\sigma_1^2-d\sigma_2^2M^2)}{\sqrt{2}(\lambda_1+d\sigma_1^2)},\sqrt{\tau\frac{(2\lambda_1-2\lambda_2M-d\sigma_1^2-d\sigma_2^2M^2)}{d\sigma_1^2}}\right\}.
\tag{32}
$$

然后，将式（29）与式（30）结合起来，对于所有 $t\in[0,T_\alpha]$，我们有

$$
\frac{d}{dt}\big(\mathcal{V}(\rho_t^1)+\mathcal{V}(\rho_t^2)\big)\le -(1-\tau)(2\lambda_1-2\lambda_2M-d\sigma_1^2-d\sigma_2^2M^2)\big(\mathcal{V}(\rho_t^1)+\mathcal{V}(\rho_t^2)\big)<0.
\tag{33}
$$

这意味着总和 $\mathcal{V}(\rho_t^1) + \mathcal{V}(\rho_t^2)$ 随时间递减。此外，Gr¨onwall 不等式给出了上界

$$\mathcal{V}(\rho_t^1) + \mathcal{V}(\rho_t^2) \le \big(\mathcal{V}(\rho_0^1) + \mathcal{V}(\rho_0^2)\big) \exp\big(-(1-\tau)(2\lambda_1 - 2\lambda_2 M - d\sigma_1^2 - d\sigma_2^2 M^2)t\big), \quad \text{对于 } t \in [0, T_\alpha]. \tag{34}$$

因此，和式 $\mathcal{V}(\rho_t^1) + \mathcal{V}(\rho_t^2)$ 随时间衰减，意味着辅助函数 $C(t)$ 也随之减小。因此，对于 $k = 1, 2$,

$$\max_{t\in[0,T_\alpha]} |m_{L_k}^\alpha[\rho_t] - \theta_k^*| \le \max_{t\in[0,T_\alpha]} |m_{L_1}^\alpha[\rho_t] - \theta_1^*| + |m_{L_2}^\alpha[\rho_t] - \theta_2^*| \le \max_{t\in[0,T_\alpha]} C(t) \le C\sqrt{\mathcal{V}(\rho_0^1) + \mathcal{V}(\rho_0^2)}. \tag{35}$$

此外，请注意,

$$\begin{aligned}
\int |\theta - \theta_1^*| d\rho_{T_\alpha}(\theta) &= w_1 \int |\theta - \theta_1^*| d\rho_{T_\alpha}^1(\theta) + w_2 \int |\theta - \theta_1^*| d\rho_{T_\alpha}^2(\theta) \\
&\le w_1 \sqrt{2\mathcal{V}(\rho_{T_\alpha}^1)} + w_2 \sqrt{\int |\theta - \theta_1^*|^2 d\rho_{T_\alpha}^2} \\
&\le w_1 \sqrt{2\mathcal{V}(\rho_{T_\alpha}^1)} + w_2 \sqrt{4\mathcal{V}(\rho_{T_\alpha}^2) + 2|\theta_1^* - \theta_2^*|^2} \\
&\le 2\sqrt{\mathcal{V}(\rho_{T_\alpha}^1)} + 2\sqrt{\mathcal{V}(\rho_{T_\alpha}^2)} + \sqrt{2}|\theta_1^* - \theta_2^*| \\
&\le 2\sqrt{2}\sqrt{\mathcal{V}(\rho_{T_\alpha}^1) + \mathcal{V}(\rho_{T_\alpha}^2)} + \sqrt{2}|\theta_1^* - \theta_2^*| \\
&\le 2\sqrt{2}\sqrt{\mathcal{V}(\rho_0^1) + \mathcal{V}(\rho_0^2)} + \sqrt{2}|\theta_1^* - \theta_2^*|,
\end{aligned} \tag{36}$$

同理,

$$\int |\theta - \theta_2^*| d\rho_{T_\alpha}(\theta) \le 2\sqrt{2}\sqrt{\mathcal{V}(\rho_0^1) + \mathcal{V}(\rho_0^2)} + \sqrt{2}|\theta_1^* - \theta_2^*|. \tag{37}$$

要得出 $\mathcal{V}(\rho_{T_\alpha}^1) + \mathcal{V}(\rho_{T_\alpha}^2) \le \varepsilon$，还需分析以下三种不同的情况。

情形 $T_\alpha \ge T^*$: 若 $T_\alpha \ge T^*$，我们可以利用（13）中 T^* 的定义以及（34）中 $\mathcal{V}(\rho_t^1)+\mathcal{V}(\rho_t^2)$ 的界，得出 $\mathcal{V}(\rho_{T^*}^1)+\mathcal{V}(\rho_{T^*}^2) \le \varepsilon$。因此，根据（30）中 T_α 的定义，我们发现 $\mathcal{V}(\rho_{T_\alpha}^1)+\mathcal{V}(\rho_{T_\alpha}^2) = \varepsilon$ 且 $T_\alpha = T^*$。

情形 $T_\alpha < T^*$ **且** $\mathcal{V}(\rho_{T_\alpha}^1) + \mathcal{V}(\rho_{T_\alpha}^2) = \varepsilon$**：** 此情形无需讨论。

情形 $T_\alpha < T^*$**，** $\mathcal{V}(\rho_{T_\alpha}^1) + \mathcal{V}(\rho_{T_\alpha}^2) > \varepsilon$**，且** $|m_{L_1}^\alpha[\rho_{T_\alpha}] - \theta_1^*| + |m_{L_2}^\alpha[\rho_{T_\alpha}] - \theta_2^*| \ge C(T_\alpha)$**：** 我们将证明存在 $\alpha_0 > 0$，使得对于任意 $\alpha > \alpha_0$，都有

$$|m_{L_1}^\alpha[\rho_{T_\alpha}] - \theta_1^*| + |m_{L_2}^\alpha[\rho_{T_\alpha}] - \theta_2^*| < C(T_\alpha), \tag{38}$$

这将与 $|m_{L_1}^\alpha[\rho_{T_\alpha}] - \theta_1^*| + |m_{L_2}^\alpha[\rho_{T_\alpha}] - \theta_2^*| \ge C(T_\alpha)$矛盾。换句话说，我们证明，只要选择 α 足够大，最后一种情况就不会发生。为证明（38），我们定义

$$q_1 := \frac{1}{2}\min\left\{ \left(\frac{\eta_1}{4} C\sqrt{\varepsilon}\right)^{\frac{1}{\nu_1}}, L_\infty^1 \right\} \qquad \text{and} \qquad r_1 := \max_{s\in[0,R_0^1]} \left\{ \max_{\theta\in B_s(\theta_1^*)} L_1(\theta) - \underline{L_1} \le q_1 \right\},$$

其中，$\underline{L_1} := \inf_{\theta\in\mathbb{R}^d} L_1(\theta)$，$\eta_1$、$\nu_1$、$L_\infty^1$ 来自假设（II），而 C 定义于（32）式中。根据构造，这些选择满足 $r_1 \le \overline{R_0^1}$ 和 $q_1 + \sup_{\theta\in B_{r_1}(\theta_1^*)} L_1(\theta) - \underline{L_1} \le 2q_1 \le L_\infty^1$。此外，我们注意到$q_1 > 0$，根据 L_1的连续性，存在 $s_{q_1} > 0$，使得对于所有$\theta \in B_{s_{q_1}}(\theta_1^*)$，均有 $L_1(\theta) - \underline{L_1} \le q_1$，从而得到 $r_1 > 0$。因此，我们可以将上述 q_1 和 r_1 代入引理 C.2，得到

$$\begin{aligned}
|m_{L_1}^\alpha[\rho_{T_\alpha}] - \theta_1^*| &\le \frac{\big(q_1 + \sup_{\theta\in B_{r_1}(\theta_1^*)} L_1(\theta) - \underline{L_1}\big)^{\nu_1}}{\eta_1} + \frac{\exp(-\alpha q_1)}{\rho_{T_\alpha}(B_{r_1}(\theta_1^*))} \int |\theta - \theta_1^*| d\rho_{T_\alpha}(\theta) \\
&\le \frac{(2q_1)^{\nu_1}}{\eta_1} + \frac{\exp(-\alpha q_1)}{\rho_{T_\alpha}(B_{r_1}(\theta_1^*))} \int |\theta - \theta_1^*| d\rho_{T_\alpha}(\theta) \\
&\le \frac{\big[\big(\frac{\eta_1}{4} C\sqrt{\varepsilon}\big)^{\frac{1}{\nu_1}}\big]^{\nu_1}}{\eta_1} + \frac{\exp(-\alpha q_1)}{\rho_{T_\alpha}(B_{r_1}(\theta_1^*))} \int |\theta - \theta_1^*| d\rho_{T_\alpha}(\theta) \\
&= \frac{C}{4}\sqrt{\varepsilon} + \frac{\exp(-\alpha q_1)}{\rho_{T_\alpha}(B_{r_1}(\theta_1^*))} \int |\theta - \theta_1^*| d\rho_{T_\alpha}(\theta).
\end{aligned} \tag{39}$$

同样，通过选择

$$q_2 := \frac{1}{2}\min\left\{(\frac{\eta_2}{4}C\sqrt{\varepsilon})^{\frac{1}{\nu_2}}, L_\infty^2\right\} \qquad \text{and} \qquad r_2 := \max_{s\in[0,R_0^2]}\left\{\max_{\theta\in B_s(\theta_1^*)} L_2(\theta) - \underline{L_2} \le q_2\right\},$$

我们有

$$|m_{L_2}^\alpha[\rho_{T_\alpha}] - \theta_2^*| \le \frac{C}{4}\sqrt{\varepsilon} + \frac{\exp(-\alpha q_2)}{\rho_{T_\alpha}(B_{r_2}(\theta_2^*))}\int |\theta-\theta_2^*| d\rho_{T_\alpha}(\theta).$$

结合不等式（36）和（37），进一步得到

$$\begin{aligned}
|m_{L_1}^\alpha[\rho_{T_\alpha}] - \theta_1^*| + |m_{L_2}^\alpha[\rho_{T_\alpha}] - \theta_2^*| &\le \frac{C}{2}\sqrt{\varepsilon} + \frac{\exp(-\alpha q_1)}{\rho_{T_\alpha}(B_{r_1}(\theta_1^*))}\int |\theta-\theta_1^*| d\rho_{T_\alpha}(\theta) \\
&\quad + \frac{\exp(-\alpha q_2)}{\rho_{T_\alpha}(B_{r_2}(\theta_2^*))}\int |\theta-\theta_2^*| d\rho_{T_\alpha}(\theta) \\
&\le \frac{C}{2}\sqrt{\varepsilon} + \left(\frac{\exp(-\alpha q_1)}{\rho_{T_\alpha}(B_{r_1}(\theta_1^*))} + \frac{\exp(-\alpha q_2)}{\rho_{T_\alpha}(B_{r_2}(\theta_2^*))}\right)\left(2\sqrt{2}\sqrt{\mathcal{V}(\rho_0^1)+\mathcal{V}(\rho_0^2)} + \sqrt{2}|\theta_1^*-\theta_2^*|\right) \\
&\le \frac{C}{2}\sqrt{\varepsilon} + \exp(-\alpha q)\left(\frac{1}{\rho_{T_\alpha}(B_{r_1}(\theta_1^*))} + \frac{1}{\rho_{T_\alpha}(B_{r_2}(\theta_2^*))}\right)\left(2\sqrt{2}\sqrt{\mathcal{V}(\rho_0^1)+\mathcal{V}(\rho_0^2)} + \sqrt{2}|\theta_1^*-\theta_2^*|\right),
\end{aligned} \tag{40}$$

其中 $q :=$ 最小值 $\{q_1, q_2\}$. 根据 (35)，我们有界 $G_{\alpha,k} :=$ 最大值$_{t\in[0,T_\alpha]}|m_{L_k}^\alpha[\rho_t]-\theta_k^*| \le C\sqrt{\mathcal{V}(\rho_0^1)+\mathcal{V}(\rho_0^2)} =: G$，这意味着引理 C.3 的所有假设均已满足。因此，根据引理 C.3，对于 $k = 1$、2 以及在 (61) 中定义的磨光函数 $\phi_{r_k}^k$，存在 $a_k := a_k^l + a_k^g > 0$ 使得

$$\rho_{T_\alpha}^k\big(B_{r_k}(\theta_k^*)\big) \ge \int \phi_{r_k}^k(\theta) d\rho_0^k(\theta)\exp(-a_k T_\alpha),$$

其中

$$a_k^l := \max\left\{h_1^l + h_2^l\frac{G_{\alpha,k}}{r_k} + h_3^l\frac{G_{\alpha,k}^2}{r_k^2}, h_4^l\right\} \qquad \text{and} \qquad a_k^g := \max\left\{h_1^g M_{\nabla L_k} + h_2^g M_{\nabla L_k}^2, h_3^g\right\},$$

其中 h_1^l、h_2^l、h_3^l、h_4^l 并且 h_1^g、h_2^g、h_3^g 仅取决于 λ_1、λ_2、σ_1、σ_2 和 d。现在我们令 $\widetilde{a} := \widetilde{a}^l + \widetilde{a}^g$，其中

$$\widetilde{a}^l := \max\left\{h_1^l + h_2^l\frac{G}{r} + h_3^l\frac{G^2}{r^2}, h_4^l\right\} \qquad \text{and} \qquad \widetilde{a}^g := \max\left\{h_1^g M + h_2^g M^2, h_3^g\right\},$$

其中 $r := \min\{r_1, r_2\}$ 和 $M := \max\{M_{\nabla L_1}, M_{\nabla L_2}\}$。然后

$$\begin{aligned}
\rho_{T_\alpha}(B_{r_1}(\theta_1^*)) &= w_1\rho_{T_\alpha}^1(B_{r_1}(\theta_1^*)) + w_2\rho_{T_\alpha^2}(B_{r_1}(\theta_1^*)) \\
&\ge w_1\rho_{T_\alpha}^1(B_{r_1}(\theta_1^*)) \\
&\ge w_1\int \phi_{r_1}^1(\theta)d\rho_0^1(\theta)\exp(-a_1 T_\alpha) \\
&\ge w_1\int \phi_{r_1}^1(\theta)d\rho_0^1(\theta)\exp(-\widetilde{a}T^*) > 0,
\end{aligned}$$

由于 $\widetilde{a} \ge a_1$ 和 $T^* \ge T_\alpha$，类似地，

$$\rho_{T_\alpha}(B_{r_2}(\theta_2^*)) \ge w_2\int \phi_{r_2}^2(\theta)d\rho_0^2(\theta)\exp(-\widetilde{a}T^*) > 0.$$

记 $K := \min\left\{w_1\int\phi_{r_1}^1(\theta)d\rho_0^1(\theta), w_2\int\phi_{r_2}^2(\theta)d\rho_0^2(\theta)\right\}$。然后使用 $\alpha > \alpha_0$，

$$\alpha_0 := \frac{\widetilde{a}T^* - \log\left(\frac{CK\sqrt{\varepsilon}}{2\sqrt{2}\sqrt{\mathcal{V}(\rho_0^1)+\mathcal{V}(\rho_0^2)}+\sqrt{2}|\theta_1^*-\theta_2^*|}\right)}{q},$$

式（40）中的第二项严格小于 $\frac{C}{2}\sqrt{\varepsilon}$。也就是说，

$$\begin{aligned}
&\exp(-\alpha q)\left(\frac{1}{\rho_{T_\alpha}(B_{r_1}(\theta_1^*))}+\frac{1}{\rho_{T_\alpha}(B_{r_2}(\theta_2^*))}\right)\left(2\sqrt{2}\sqrt{\mathcal{V}(\rho_0^1)+\mathcal{V}(\rho_0^2)}+\sqrt{2}|\theta_1^*-\theta_2^*|\right)\\
&<\exp(-\alpha_0 q)\frac{2}{K\exp(-\widetilde{a}T^*)}\left(2\sqrt{2}\sqrt{\mathcal{V}(\rho_0^1)+\mathcal{V}(\rho_0^2)}+\sqrt{2}|\theta_1^*-\theta_2^*|\right)\\
&=\frac{C}{2}\sqrt{\varepsilon}.
\end{aligned}$$

由（40）可得

$$|m_{L_1}^\alpha[\rho_{T_\alpha}]-\theta_1^*|+|m_{L_2}^\alpha[\rho_{T_\alpha}]-\theta_2^*|<C\sqrt{\varepsilon}<C\sqrt{\mathcal{V}(\rho_{T_\alpha}^1)+\mathcal{V}(\rho_{T_\alpha}^2)},$$

以此方式与（38）矛盾。 □

5. Mean-field Limit of CBO for Clustered Federated Learning

5.1. **经验测度的紧性**。首先，我们给出粒子系统（4）的一些一致矩估计。这些估计直接由引理 A.1 得出。

引理 5.1。设 L_1、L_2 满足假设 *1*，并且在无穷远处具有二次增长。对于 N_1、$N_2\geq 1$ 和 $N=N_1+N_2$，令 $\{\boldsymbol{\theta}_t^{1,N}\}$、$\{\boldsymbol{\theta}_t^{2,N}\}$ 为粒子系统（4）的唯一解，其中 $\rho_0^{\otimes N}$-分布的初始数据为 $\{\boldsymbol{\theta}_0^{1,N}\}$、$\{\boldsymbol{\theta}_0^{2,N}\}$。则存在一个与 $K>0$ 无关的常数 N 使得

$$\sup_{i=1,2,\dots,N_1}\left\{\sup_{t\in[0,T]}\mathbb{E}\left[|\theta_t^{1,(i,N)}|^2+|\theta_t^{1,(i,N)}|^4\right]+\sup_{t\in[0,T]}\mathbb{E}\left[|m_{L_1}^\alpha[\rho_t^N]|^2+|m_{L_1}^\alpha[\rho_t^N]|^4\right]\right\}\leq K$$

and

$$\sup_{j=1,2,\dots,N_2}\left\{\sup_{t\in[0,T]}\mathbb{E}\left[|\theta_t^{2,(j,N)}|^2+|\theta_t^{2,(j,N)}|^4\right]+\sup_{t\in[0,T]}\mathbb{E}\left[|m_{L_2}^\alpha[\rho_t^N]|^2+|m_{L_2}^\alpha[\rho_t^N]|^4\right]\right\}\leq K.$$

在下文中，我们将 $\theta^{1,(i,N)}$、$\theta^{2,(j,N)}$ 视为随机元素，其取值于 $\mathcal{C}([0,T],\mathbb{R}^d)$。这些随机元素定义在一个足够丰富的概率空间（$\Omega$ $\mathcal{F}$、$\mathbb{P}$）上。具体而言，$\theta^{1,(i,N)}$、$\theta^{2,(j,N)}:\Omega\to\mathcal{C}([0,T],\mathbb{R}^d)$ 是可测映射。因此，我们能够将 $\rho^{1,N}=\frac{1}{N_1}\sum_{i=1}^{N_1}\delta_{\theta^{1,(i,N)}}$、$\rho^{2,N}=\frac{1}{N_2}\sum_{j=1}^{N_2}\delta_{\theta^{2,(j,N)}}$、$\rho^N=\frac{N_1}{N}\rho^{1,N}+\frac{N_2}{N}\rho^{2,N}:\Omega\to\mathcal{P}(\mathcal{C}([0,T],\mathbb{R}^d))$ 解释为随机测度。 — —

$\mathcal{L}(\rho^N):=Law(\rho^N)\in\mathcal{P}(\mathcal{P}(\mathcal{C}([0,T];\mathbb{R}^d)))$ 表示随机变量 ρ^N 的分布，并类似地定义 $\mathcal{L}(\rho^{1,N})$ 和 $\mathcal{L}(\rho^{2,N})$。接下来我们证明 $\{\mathcal{L}(\rho^{1,N})\}_{N\geq 2}$，$\{\mathcal{L}(\rho^{2,N})\}_{N\geq 2}$ 和 $\{\mathcal{L}(\rho^N)\}_{N\geq 2}$ 是紧的。正如概率论中的常见做法，我们将简单地说随机变量序列 $\{\rho^{1,N}\}_{N\geq 2}$，$\{\rho^{2,N}\}_{N\geq 2}$ 和 $\{\rho^N\}_{N\geq 2}$ 是紧的。

定理 5.2. 在与引理 *5.1* 相同的假设下，序列 $\{\mathcal{L}(\rho^{1,N})\}_{N\geq 2}$、$\{\mathcal{L}(\rho^{2,N})\}_{N\geq 2}$ 在 $\mathcal{P}(\mathcal{P}(\mathcal{C}([0,T];\mathbb{R}^d)))$ 中是紧的。

定理 *5.2* 的证明。根据 [35, 命题 2.2 (ii)], ，由于粒子系统的可交换性，只需证明 $\{\mathcal{L}(\theta^{1,(1,N)})\}$ 和 $\{\mathcal{L}(\theta^{2,(1,N)})\}$ 在 $\mathcal{P}(\mathcal{C}([0,T];\mathbb{R}^d))$ 中是紧的。换言之，只需证明以 N 为指标的系统中单个粒子的轨迹分布族是紧的。我们将通过验证 $\{\mathcal{L}(\theta^{1,(1,N)})\}$ 满足 Aldous 判据（引理 D.1）中的两个条件来完成这一证明。对于 $\{\mathcal{L}(\theta^{2,(1,N)})\}$ 的证明类似，因此省略细节。

验证（Con1）：给定 $\varepsilon>0$，考虑紧子集 $U_\varepsilon:=\{\theta\in\mathbb{R}^d:\ |\theta|^2\leq\frac{K}{\varepsilon}\}$。那么，根据马尔可夫不等式，

$$\mathcal{L}(\theta_t^{1,(1,N)})\big((U_\varepsilon)^c\big)=\mathbb{P}\left(|\theta_t^{1,(1,N)}|^2>\frac{K}{\varepsilon}\right)\leq\frac{\varepsilon\mathbb{E}[|\theta_t^{1,(1,N)}|^2]}{K}\leq\varepsilon\quad\forall N\geq 2,$$

其中，最后一个不等式用到了引理 5.1。这意味着，对于每个固定的 $t\in[0,T]$，序列$\{\mathcal{L}(\theta^{1,(1,N)}_t)\}_{N\geq 2}$ 都是紧的，从而验证了引理（*Con1*）中的条件 D.1。

验证（条件（Con2））： 令 β为一个取离散值的 $\sigma(\theta^{1,(1,N)}_s;\ s\in[0,T])$-停时，且满足$\beta+\delta_0\le T$，其中 δ_0将在下文选取。令 $\delta\in[0,\delta_0]$。回顾（4a），可得

$$\begin{aligned}\theta^{1,(1,N)}_{\beta+\delta}-\theta^{1,(1,N)}_{\beta}=&-\int_\beta^{\beta+\delta}\lambda_1(\theta^{1,(1,N)}_s-m^\alpha_{L_1}[\rho^N_s])ds-\int_\beta^{\beta+\delta}\lambda_2\nabla L_1(\theta^{1,(1,N)}_s)ds\\&+\int_\beta^{\beta+\delta}\sigma_1|\theta^{1,(1,N)}_s-m^\alpha_{L_1}[\rho^N_s]|dB^{1,1}_s+\int_\beta^{\beta+\delta}\sigma_2|\nabla L_1(\theta^{1,(1,N)}_s)|d\widetilde{B}^{1,1}_s.\end{aligned}$$

由柯西–施瓦茨不等式和引理5.1，我们有

$$\begin{aligned}\mathbb{E}\Big[|\int_\beta^{\beta+\delta}\lambda_1\big(\theta^{1,(1,N)}_s-m^\alpha_{L_1}[\rho^N_s]\big)ds|^2\Big]&\le\lambda_1^2\delta\int_0^T\mathbb{E}\Big[|\theta^{1,(1,N)}_s-m^\alpha_{L_1}[\rho^N_s]|^2\Big]ds\\&\le 2\lambda_1^2\delta T\Big(\sup_{t\in[0,T]}\mathbb{E}[|\theta^{1,(1,N)}_t|^2]+\sup_{t\in[0,T]}\mathbb{E}[|m^\alpha_{L_1}[\rho^N_s]|^2]\Big)\\&\le 2\lambda_1^2KT\delta.\end{aligned}$$

利用假设 1，我们得到

$$\mathbb{E}\Big[|\int_\beta^{\beta+\delta}\lambda_2\nabla L_1(\theta^{1,(1,N)}_s)ds|^2\Big]\le\lambda_2^2\delta\int_0^T|\nabla L_1(\theta^{1,(1,N)}_s)|^2ds\le\lambda_2^2C^2_{\nabla L_1}T\delta.$$

此外，我们应用 Itô 等距性，得到

$$\begin{aligned}\mathbb{E}\Big[|\sigma_1\int_\beta^{\beta+\delta}|\theta^{1,(1,N)}_s-m^\alpha_{L_1}[\rho^N_s]|dB^{1,1}_s|^2\Big]&=\sigma_1^2\mathbb{E}\Big[\int_\beta^{\beta+\delta}|\theta^{1,(1,N)}_s-m^\alpha_{L_1}[\rho^N_s]|^2ds\Big]\\&\le\sigma_1^2\delta^{\frac12}\mathbb{E}\Big[\big(\int_0^T|\theta^{1,(1,N)}_s-m^\alpha_{L_1}[\rho^N_s]|^4ds\big)^{\frac12}\Big]\\&\le\sigma_1^2\delta^{\frac12}\Big(\int_0^T\mathbb{E}[|\theta^{1,(1,N)}_s-m^\alpha_{L_1}[\rho^N_s]|^4]ds\Big)^{\frac12}\\&\le\sigma_1^2(8K)^{\frac12}T^{\frac12}\delta^{\frac12}\end{aligned}$$

and

$$\mathbb{E}\Big[|\sigma_2\int_\beta^{\beta+\delta}|\nabla L_1(\theta^{1,(1,N)}_s)|d\widetilde{B}^{1,1}_s|^2\Big]=\sigma_2^2\mathbb{E}\Big[\int_\beta^{\beta+\delta}|\nabla L_1(\theta^{1,(1,N)}_s)|^2ds\Big]\le\sigma_2^2C^2_{\nabla L_1}\delta.$$

结合上述四个估计式，我们得到

$$\mathbb{E}|\theta^{1,(1,N)}_{\beta+\delta}-\theta^{1,(1,N)}_{\beta}|^2\le C(\lambda_1,\lambda_2,\sigma_1,\sigma_2,T,K)(\delta^{\frac12}+\delta).$$

因此，对于任意 $\varepsilon>0$ 和 $\eta>0$，都存在某个 $\delta_0>0$，使得对于所有 $N\ge2$，均有】【。

$$\sup_{\delta\in[0,\delta_0]}\mathbb{P}\big(|\theta^{1,(1,N)}_{\beta+\delta}-\theta^{1,(1,N)}_{\beta}|^2\ge\eta\big)\le\sup_{\delta\in[0,\delta_0]}\frac{\mathbb{E}|\theta^{1,(1,N)}_{\beta+\delta}-\theta^{1,(1,N)}_{\beta}|^2}{\eta}\le\varepsilon.$$

<escape>这证明了引理 条件（*Con2*） 中的条件，并且 D.1 完成了证明。</escape> □

引理 5.3。 在与定理 *5.2* 相同的假设下，以下结论成立：

(1) 存在 $\{(\rho^{1,N},\rho^{2,N})\}_{N\ge2}$ 的一个子序列（为简洁起见，不再重新标记）以及一个随机变量$(\tilde\rho^1,\tilde\rho^2):\Omega\to\mathcal{P}(\mathcal{C}([0,T];\mathbb{R}^d))\times\mathcal{P}(\mathcal{C}([0,T];\mathbb{R}^d))$，使得

$$(\rho^{1,N},\rho^{2,N})\rightharpoonup(\tilde\rho^1,\tilde\rho^2)\quad in\ law\ as\ N\to+\infty.$$

（*2*）设 $\rho^N=\frac{N_1}{N}\rho^{1,N}+\frac{N_2}{N}\rho^{2,N}$，其中 N_1 和 N_2 满足 $\frac{N_1}{N_2}\to\frac{w_1}{w_2}$，且 $N=N_1+N_2\to+\infty$ 和 w_1、$w_2>0$、$w_1+w_2=1$。那么

$$\rho^N\rightharpoonup w_1\tilde\rho^1+w_2\tilde\rho^2\quad in\ law\ as\ N\to+\infty.$$

证明。我们知道 $\{(\rho^{1,N},\rho^{2,N})\}_{N\geq 2}$是紧的，因为根据定理 5.2，$\{\rho^{1,N}\}_{N\geq 2}$和 $\{\rho^{2,N}\}_{N\geq 2}$均为紧的。因此，断言（1）可直接由普罗霍罗夫定理得出。

关于断言 (2)，容易验证，$(\frac{N_1}{N}\rho^{1,N}, \frac{N_2}{N}\rho^{2,N})$ 依分布收敛到 $(w_1\tilde{\rho}^1, w_2\tilde{\rho}^2)$——注意，这些随机变量取值于 $\mathcal{M}_+(\mathcal{C}([0,T],\mathbb{R}^d)) \times \mathcal{M}_+(\mathcal{C}([0,T],\mathbb{R}^d))$，也就是说，每个坐标都是 $\mathcal{C}([0,T],\mathbb{R}^d)$ 上的一个正测度——前提是 $(\rho^{1,N},\rho^{2,N}) \rightharpoonup (\tilde{\rho}^1,\tilde{\rho}^2)$，且 $\frac{N_1}{N_2}\to\frac{w_1}{w_2}$、$N=N_1+N_2$。断言 (2) 由连续映射定理得出。□

5.2. **通过偏微分方程识别极限测度。** 本节将平均场随机微分方程（7）改写为以下形式：

$$
\begin{aligned}
d\begin{pmatrix}\bar{\theta}_t^1\\ \bar{\theta}_t^2\end{pmatrix} &= -\begin{pmatrix}\lambda_1(\bar{\theta}_t^1 - m_{L_1}^{\alpha}[\rho_t]) + \lambda_2\nabla L_1(\bar{\theta}_t^1)\\ \lambda_1(\bar{\theta}_t^1 - m_{L_2}^{\alpha}[\rho_t]) + \lambda_2\nabla L_2(\bar{\theta}_t^2)\end{pmatrix}dt\\
&\quad + \sigma_1\begin{pmatrix}\mathrm{diag}(|\bar{\theta}_t^1 - m_{L_1}^{\alpha}[\rho_t]|) & 0\\ 0 & \mathrm{diag}(|\bar{\theta}_t^2 - m_{L_2}^{\alpha}[\rho_t]|)\end{pmatrix}\begin{pmatrix}dB_t^1\\ dB_t^2\end{pmatrix}\\
&\quad + \sigma_2\begin{pmatrix}\mathrm{diag}(|\nabla L_1(\bar{\theta}_t^1)|) & 0\\ 0 & \mathrm{diag}(|\nabla L_2(\bar{\theta}_t^2)|)\end{pmatrix}\begin{pmatrix}d\widetilde{B}_t^1\\ d\widetilde{B}_t^2\end{pmatrix},
\end{aligned}
\tag{41}
$$

其中，回顾一下，θ_0^1和 θ_0^2是独立初始条件。还要回顾的是，$\rho_t^1 = $分布$(\theta_t^1)$ $\rho_t^2 = $分布$(\theta_t^2)$，$\rho_t = w_1\rho_t^1 + w_2\rho_t^2$。在上述表达式中，我们用 $\mathrm{diag}(v)$表示 $\mathbb{R}^{d\times d}$中的对角矩阵，其对角元素等于向量 v的各个坐标。对于每个 $t\in[0,T]$，令 $\mu_t = $分布$((\theta_t^1,\theta_t^2))$。我们将 μ视为 $\mathcal{C}([0,T],\mathcal{P}(\mathbb{R}^d\times\mathbb{R}^d))$中的一个元素。我们注意到，它必须满足以下联合福克–普朗克方程：

$$
\begin{aligned}
\partial_t\mu_t(\theta) &= \nabla_{(\theta_1,\theta_2)}\cdot\Big(\big(\lambda_1(\theta_1 - m_{L_1}^{\alpha}[\rho_t]) + \lambda_2\nabla L_1(\theta_1), \lambda_1(\theta_2 - m_{L_2}^{\alpha}[\rho_t]) + \lambda_2\nabla L_2(\theta_2)\big)\mu_t\Big)\\
&\quad + \Delta_{\theta_1}\Big(\big(\frac{\sigma_1^2}{2}|\theta_1 - m_{L_1}^{\alpha}[\rho_t]|^2 + \frac{\sigma_2^2}{2}|\nabla L_1(\theta_1)|^2\big)\mu_t\Big) + \Delta_{\theta_2}\Big(\big(\frac{\sigma_1^2}{2}|\theta_2 - m_{L_2}^{\alpha}[\rho_t]|^2 + \frac{\sigma_2^2}{2}|\nabla L_2(\theta_2)|^2\big)\mu_t\Big),
\end{aligned}
\tag{42}
$$

其中 $\theta=(\theta_1,\theta_2)\in\mathbb{R}^{2d}$，$\Delta_{\theta_1} := \sum_{i=1}^{d}\frac{\partial^2}{\partial\theta_i^2}$且 $\Delta_{\theta_2} := \sum_{i=d+1}^{2d}\frac{\partial^2}{\partial\theta_i^2}$，并且 $\nabla_{(\theta_1,\theta_2)}\cdot$表示在两个坐标方向上的散度 θ_1，θ_2。方程（42）必须按照下文定义 5.1 进行解释。

注 8。 需要注意的是，由于 θ^1 和 θ^2 根据注 7相互独立，分布 μ_t满足 $\mu_t = \rho_t^1\otimes\rho_t^2\in\mathcal{P}(\mathbb{R}^d\times\mathbb{R}^d)$ 对于 ρ_t^1和 ρ_t^2如（8）所示。虽然以这种形式引入 μ 似乎有些多余，但这样做是为了便于我们在下文中进行分析。

定义 5.1。 我们称 $\tilde{\mu}\in\mathcal{C}([0,T],\mathcal{P}(\mathbb{R}^d\times\mathbb{R}^d))$ 是偏微分方程（42）的弱解，当且仅当

(i) 对于所有 $\phi\in\mathcal{C}_b(\mathbb{R}^d\times\mathbb{R}^d)$ 和 $t_n\to t$，均有

$$
\int_{\mathbb{R}^d\times\mathbb{R}^d}\phi(\theta_1,\theta_2)d\tilde{\mu}_{t_n}(\theta_1,\theta_2) \to \int_{\mathbb{R}^d\times\mathbb{R}^d}\phi(\theta_1,\theta_2)d\tilde{\mu}_t(\theta_1,\theta_2);
\tag{43}
$$

(ii) 以下结论成立

$$
\begin{aligned}
&\langle\varphi(\theta_1,\theta_2),\tilde{\mu}_t(d\theta_1,d\theta_2)\rangle - \langle\varphi(\theta_1,\theta_2),\tilde{\mu}_0(d\theta_1,d\theta_2)\rangle\\
&\quad + \int_0^t\langle\big(\lambda_1(\theta_1 - m_{L_1}^{\alpha}[\rho_s]) + \lambda_2\nabla L_1(\theta_1)\big)\cdot\nabla_{\theta_1}\varphi\\
&\qquad\qquad + \big(\lambda_1(\theta_2 - m_{L_2}^{\alpha}[\rho_s]) + \lambda_2\nabla L_2(\theta_2)\big)\cdot\nabla_{\theta_2}\varphi, \tilde{\mu}_s(d\theta_1,d\theta_2)\rangle ds\\
&\quad - \int_0^t\langle\big(\frac{\sigma_1^2}{2}|\theta_1 - m_{L_1}^{\alpha}[\rho_s]|^2 + \frac{\sigma_2^2}{2}|\nabla L_1(\theta_1)|^2\big)\Delta_{\theta_1}\varphi\\
&\qquad\qquad + \big(\frac{\sigma_1^2}{2}|\theta_2 - m_{L_2}^{\alpha}[\rho_s]|^2 + \frac{\sigma_2^2}{2}|\nabla L_2(\theta_2)|^2\big)\Delta_{\theta_2}\varphi, \tilde{\mu}_s(d\theta_1,d\theta_2)\rangle ds = 0,
\end{aligned}
$$

对于所有 $\varphi\in\mathcal{C}_c^2(\mathbb{R}^d\times\mathbb{R}^d)$ 以及所有 $t\in[0,T]$。这里，$\rho_t := w_1\rho_t^1 + w_2\rho_t^2$ 和 $\rho_t^1 := \pi_{1\sharp}\tilde{\mu}_t$、$\rho_t^2 := \pi_{2\sharp}\tilde{\mu}_t$ 是 $\tilde{\mu}_t$ 的边缘分布。在上文及下文中，我们用 $\langle\cdot,\cdot\rangle$ 表示测试函数 φ 与测度之间的标准对偶配对。

对于每个 $\varphi \in \mathcal{C}_c^2(\mathbb{R}^d \times \mathbb{R}^d)$ 和固定的 $t \in (0,T)$，我们定义 F_φ 上的泛函 $\mathcal{C}([0,T], \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)) \times [0,1]$，其定义为

$$
\begin{aligned}
F_\varphi(\nu, \tilde{w}) := {} & |\tilde{w} - w_1| + \langle \varphi(\theta_1, \theta_2), \nu_t(d\theta_1, d\theta_2) \rangle - \langle \varphi(\theta_1, \theta_2), \nu_0(d\theta_1, d\theta_2) \rangle \\
& + \int_0^t \langle \big(\lambda_1(\theta_1 - m_{L_1}^\alpha[\rho_{s,\nu,\tilde{w}}]) + \lambda_2 \nabla L_1(\theta_1) \big) \cdot \nabla_{\theta_1} \varphi \\
& \qquad + \big(\lambda_1(\theta_2 - m_{L_2}^\alpha[\rho_{s,\nu,\tilde{w}}]) + \lambda_2 \nabla L_2(\theta_2) \big) \cdot \nabla_{\theta_2} \varphi, \nu_s(d\theta_1, d\theta_2) \rangle ds \\
& - \int_0^t \langle \big(\frac{\sigma_1^2}{2} |\theta_1 - m_{L_1}^\alpha[\rho_{s,\nu,\tilde{w}}]|^2 + \frac{\sigma_2^2}{2} |\nabla L_1(\theta_1)|^2 \big) \Delta_{\theta_1} \varphi \\
& \qquad + \big(\frac{\sigma_1^2}{2} |\theta_2 - m_{L_2}^\alpha[\rho_{s,\nu,\tilde{w}}]|^2 + \frac{\sigma_2^2}{2} |\nabla L_2(\theta_2)|^2 \big) \Delta_{\theta_2} \varphi, \nu_s(d\theta_1, d\theta_2) \rangle ds,
\end{aligned}
$$

对于 $\nu \in \mathcal{C}([0,T], \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d))$ 和 $\tilde{w} \in [0,1]$。在上述内容中，$\rho^1_{s,\nu,\tilde{w}} := \pi_{1\sharp}\nu_s$, $\rho^2_{s,\nu,\tilde{w}} := \pi_{2\sharp}\nu_s$，以及 $\rho_{s,\nu,\tilde{w}} := \tilde{w}\rho^1_{s,\nu,\tilde{w}} + (1-\tilde{w})\rho^2_{s,\nu,\tilde{w}}$。

我们有如下估计。

引理 5.4. 设 L_1, L_2 满足假设 1，且在无穷远处具有二次增长；设 $\mu_0^1, \mu_0^2 \in \mathcal{P}_4(\mathbb{R}^d)$。假设对于每个 $N \geq 2$，$\{\boldsymbol{\theta}_t^{1,N}\}, \{\boldsymbol{\theta}_t^{2,N}\}$ 都是粒子系统 (4) 的唯一解，其初始数据 $\{\boldsymbol{\theta}_0^{1,N}\}, \{\boldsymbol{\theta}_0^{2,N}\}$ 服从 $(\mu_0^1)^{\otimes N_1} \otimes (\mu_0^2)^{\otimes N_2}$ 分布。那么存在一个常数 $C > 0$，它仅取决于 σ_1, σ_2, K, T 和 $\|\nabla\varphi\|_\infty$，使得

$$
\mathbb{E}\big[|F_\varphi(\mu^N, \frac{N_1}{N})|^2 \big] \leq \frac{C}{N_1 N_2} + 2|\frac{N_1}{N} - w_1|^2,
$$

其中 $\mu_s^N := \rho^{1,N}_s \otimes \rho^{2,N}_s \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$，且经验测度 $\rho^{1,N}_s$, $\rho^{2,N}_s$ 如 (5) 所示。μ^N 因此可以解释为一个取值于 $\mathcal{C}([0,T], \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d))$ 。

证明。根据 F_φ 的定义可得

$$
\begin{aligned}
F_\varphi(\mu^N, \frac{N_1}{N}) - |\frac{N_1}{N} - w_1| = {} & \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \varphi(\theta_t^{1,i}, \theta_t^{2,j}) - \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \varphi(\theta_0^{1,i}, \theta_0^{2,j}) \\
& + \int_0^j \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \Big[\big(\lambda_1(\theta_s^{1,i} - m_{L_1}^\alpha[\rho_s^N]) + \lambda_2 \nabla L_1(\theta_s^{1,i}) \big) \cdot \nabla_{\theta_1} \varphi(\theta_s^{1,i}, \theta_s^{2,j}) \\
& \qquad + \big(\lambda_1(\theta_s^{2,j} - m_{L_2}^\alpha[\rho_s^N]) + \lambda_2 \nabla L_2(\theta_s^{2,j}) \big) \cdot \nabla_{\theta_2} \varphi(\theta_s^{1,i}, \theta_s^{2,j}) \Big] ds \\
& - \int_0^t \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \Big[\big(\frac{\sigma_1^2}{2} |\theta_s^{1,i} - m_{L_1}^\alpha[\rho_s^N]|^2 + \frac{\sigma_2^2}{2} |\nabla L_1(\theta_s^{1,i})|^2 \big) \Delta_{\theta_1} \varphi(\theta_s^{1,i}, \theta_s^{2,j}) \\
& \qquad + \big(\frac{\sigma_1^2}{2} |\theta_s^{2,j} - m_{L_2}^\alpha[\rho_s^N]|^2 + \frac{\sigma_2^2}{2} |\nabla L_2(\theta_s^{2,j})|^2 \big) \Delta_{\theta_2} \varphi(\theta_s^{1,i}, \theta_s^{2,j}) \Big] ds \\
= {} & \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \Big\{ \varphi(\theta_t^{1,i}, \theta_t^{2,j}) - \varphi(\theta_0^{1,i}, \theta_0^{2,j}) \\
& \qquad + \int_0^t \Big[\big(\lambda_1(\theta_s^{1,i} - m_{L_1}^\alpha[\rho_s^N]) + \lambda_2 \nabla L_1(\theta_s^{1,i}) \big) \cdot \nabla_{\theta_1} \varphi(\theta_s^{1,i}, \theta_s^{2,j}) \\
& \qquad\qquad + \big(\lambda_1(\theta_s^{2,j} - m_{L_2}^\alpha[\rho_s^N]) + \lambda_2 \nabla L_2(\theta_s^{2,j}) \big) \cdot \nabla_{\theta_2} \varphi(\theta_s^{1,i}, \theta_s^{2,j}) \Big] ds \\
& \qquad - \int_0^t \Big[\big(\frac{\sigma_1^2}{2} |\theta_s^{1,i} - m_{L_1}^\alpha[\rho_s^N]|^2 + \frac{\sigma_2^2}{2} |\nabla L_1(\theta_s^{1,i})|^2 \big) \Delta_{\theta_1} \varphi(\theta_s^{1,i}, \theta_s^{2,j}) \\
& \qquad\qquad + \big(\frac{\sigma_1^2}{2} |\theta_s^{2,j} - m_{L_2}^\alpha[\rho_s^N]|^2 + \frac{\sigma_2^2}{2} |\nabla L_2(\theta_s^{2,j})|^2 \big) \Delta_{\theta_2} \varphi(\theta_s^{1,i}, \theta_s^{2,j}) \Big] ds \Big\}.
\end{aligned}
$$

另一方面，对于每个 $i\in[N_1]$，$j\in[N_2]$，由伊藤公式可得

$$\begin{aligned}
\varphi(\theta_s^{1,i},\theta_t^{2,j})-\varphi(\theta_0^{1,i},\theta_0^{2,j}) =& -\int_0^t\Big[\big(\lambda_1(\theta_s^{1,i}-m_{L_1}^{\alpha}[\rho_s^N])+\lambda_2\nabla L_1(\theta_s^{1,i})\big)\cdot\nabla_{\theta_1}\varphi(\theta_s^{1,i},\theta_s^{2,j})\\
&\qquad+\big(\lambda_1(\theta_s^{2,j}-m_{L_2}^{\alpha}[\rho_s^N])+\lambda_2\nabla L_2(\theta_s^{2,j})\big)\cdot\nabla_{\theta_2}\varphi(\theta_s^{1,i},\theta_s^{2,j})\Big]ds\\
&+\int_0^t\Big[\big(\frac{\sigma_1^2}{2}|\theta_s^{1,i}-m_{L_1}^{\alpha}[\rho_s^N]|^2+\frac{\sigma_2^2}{2}|\nabla L_1(\theta_s^{1,i})|^2\big)\Delta_{\theta_1}\varphi(\theta_s^{1,i},\theta_s^{2,j})\\
&\qquad+\big(\frac{\sigma_1^2}{2}|\theta_s^{2,j}-m_{L_2}^{\alpha}[\rho_s^N]|^2+\frac{\sigma_2^2}{2}|\nabla L_2(\theta_s^{2,j})|^2\big)\Delta_{\theta_2}\varphi(\theta_s^{1,i},\theta_s^{2,j})\Big]ds\\
&+\sigma_1\int_0^t\begin{pmatrix}\mathrm{diag}(|\theta_s^{1,i}-m_{L_1}^{\alpha}[\rho_s^N]|) & 0\\ 0 & \mathrm{diag}(|\theta_s^{2,j}-m_{L_2}^{\alpha}[\rho_s^N]|)\end{pmatrix}\nabla\varphi(\theta_s^{1,i},\theta_s^{2,j})\cdot\begin{pmatrix}dB_s^{1,i}\\ dB_s^{2,j}\end{pmatrix}\\
&+\sigma_2\int_0^t\begin{pmatrix}\mathrm{diag}(|\nabla L_1(\theta_s^{1,i})|) & 0\\ 0 & \mathrm{diag}(|\nabla L_2(\theta_s^{2,j})|)\end{pmatrix}\nabla\varphi(\theta_s^{1,i},\theta_s^{2,j})\cdot\begin{pmatrix}d\widetilde{B}_s^{1,i}\\ d\widetilde{B}_s^{2,j}\end{pmatrix}.
\end{aligned}$$

这意味着

$$\begin{aligned}
&F_\varphi(\mu^N,N_1/N)-|\frac{N_1}{N}-w_1|\\
&=\frac{1}{N_1N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\Big[\sigma_1\int_0^t\begin{pmatrix}\mathrm{diag}(|\theta_s^{1,i}-m_{L_1}^{\alpha}[\rho_s^N]|) & 0\\ 0 & \mathrm{diag}(|\theta_s^{2,j}-m_{L_2}^{\alpha}[\rho_s^N]|)\end{pmatrix}\nabla\varphi(\theta_s^{1,i},\theta_s^{2,j})\cdot\begin{pmatrix}dB_s^{1,i}\\ dB_s^{2,j}\end{pmatrix}\\
&\qquad+\sigma_2\int_0^t\begin{pmatrix}\mathrm{diag}(|\nabla L_1(\theta_s^{1,i})|) & 0\\ 0 & \mathrm{diag}(|\nabla L_2(\theta_s^{2,j})|)\end{pmatrix}\nabla\varphi(\theta_s^{1,i},\theta_s^{2,j})\cdot\begin{pmatrix}d\widetilde{B}_s^{1,i}\\ d\widetilde{B}_s^{2,j}\end{pmatrix}\Big].
\end{aligned}$$

然后，根据伊藤等距性，我们得出

$$\begin{aligned}
&\mathbb{E}[\big|F_\varphi(\mu^N,N_1/N)-|N_1/N-w_1|\big|^2]\\
&=\frac{1}{N_1^2N_2^2}\mathbb{E}\Big[\Big|\sum_{i,j}\Big[\sigma_1\int_0^t\begin{pmatrix}\mathrm{diag}(|\theta_s^{1,i}-m_{L_1}^{\alpha}[\rho_s^N]|) & 0\\ 0 & \mathrm{diag}(|\theta_s^{2,j}-m_{L_2}^{\alpha}[\rho_s^N]|)\end{pmatrix}\nabla\varphi(\theta_s^{1,i},\theta_s^{2,j})\cdot\begin{pmatrix}dB_s^{1,i}\\ dB_s^{2,j}\end{pmatrix}\\
&\qquad+\sigma_2\int_0^t\begin{pmatrix}\mathrm{diag}(|\nabla L_1(\theta_s^{1,i})|) & 0\\ 0 & \mathrm{diag}(|\nabla L_2(\theta_s^{2,j})|)\end{pmatrix}\nabla\varphi(\theta_s^{1,i},\theta_s^{2,j})\cdot\begin{pmatrix}d\widetilde{B}_s^{1,i}\\ d\widetilde{B}_s^{2,j}\end{pmatrix}\Big]\Big|^2\Big]\\
&=\frac{1}{N_1^2N_2^2}\sum_{i,j}\Big\{\sigma_1^2\mathbb{E}\Big[\Big|\int_0^t\begin{pmatrix}\mathrm{diag}(|\theta_s^{1,i}-m_{L_1}^{\alpha}[\rho_s^N]|) & 0\\ 0 & \mathrm{diag}(|\theta_s^{2,j}-m_{L_2}^{\alpha}[\rho_s^N]|)\end{pmatrix}\nabla\varphi(\theta_s^{1,i},\theta_s^{2,j})\cdot\begin{pmatrix}dB_s^{1,i}\\ dB_s^{2,j}\end{pmatrix}\Big|^2\Big]\\
&\qquad+\sigma_2^2\mathbb{E}\Big[\Big|\int_0^t\begin{pmatrix}\mathrm{diag}(|\nabla L_1(\theta_s^{1,i})|) & 0\\ 0 & \mathrm{diag}(|\nabla L_2(\theta_s^{2,j})|)\end{pmatrix}\nabla\varphi(\theta_s^{1,i},\theta_s^{2,j})\cdot\begin{pmatrix}d\widetilde{B}_s^{1,i}\\ d\widetilde{B}_s^{2,j}\end{pmatrix}\Big|^2\Big]\Big\}\\
&=\frac{1}{N_1^2N_2^2}\sum_{i,j}\Big\{\sigma_1^2\mathbb{E}\Big[\int_0^t\Big\|\begin{pmatrix}\mathrm{diag}(|\theta_s^{1,i}-m_{L_1}^{\alpha}[\rho_s^N]|) & 0\\ 0 & \mathrm{diag}(|\theta_s^{2,j}-m_{L_2}^{\alpha}[\rho_s^N]|)\end{pmatrix}\nabla\varphi(\theta_s^{1,i},\theta_s^{2,j})\Big\|_F^2ds\Big]\\
&\qquad+\sigma_2^2\mathbb{E}\Big[\int_0^t\Big\|\begin{pmatrix}\mathrm{diag}(|\nabla L_1(\theta_s^{1,i})|) & 0\\ 0 & \mathrm{diag}(|\nabla L_2(\theta_s^{2,j})|)\end{pmatrix}\nabla\varphi(\theta_s^{1,i},\theta_s^{2,j})\Big\|_F^2ds\Big]\Big\}\\
&=\frac{1}{N_1^2N_2^2}\sum_{i,j}\mathbb{E}\Big[\int_0^t\Big(\big(\sigma_1^2|\theta_s^{1,i}-m_{L_1}^{\alpha}[\rho_s^N]|^2+\sigma_2^2|\nabla L_1(\theta_s^{1,i})|^2\big)|\nabla_{\theta_1}\varphi(\theta_s^{1,i},\theta_s^{2,j})|^2\\
&\qquad+\big(\sigma_1^2|\theta_s^{2,j}-m_{L_2}^{\alpha}[\rho_s^N]|^2+\sigma_2^2|\nabla L_2(\theta_s^{2,j})|^2\big)|\nabla_{\theta_2}\varphi(\theta_s^{1,i},\theta_s^{2,j})|^2\Big)ds\Big]\\
&\le C(\sigma_1,\sigma_2,K,T,C_{\nabla L_1},C_{\nabla L_2},\|\nabla\varphi\|_\infty)\frac{1}{N_1N_2},
\end{aligned}$$

其中我们在最后一个不等式中用到了引理 5.1 和假设 1 。由上可得所需的估计。 □

根据 Skorokhod 引理（参见 [2, 定理 6.7]）和引理 5.3，我们可以改变概率空间 $(\Omega, \mathcal{F}, \mathbb{P})$，并在这个新的概率空间中假设随机变量序列 $\{(\rho^{1,N}, \rho^{2,N})\}_{N\geq 2}$ 几乎处处收敛到某个随机变量 $(\tilde{\rho}^1, \tilde{\rho}^2)$；注意，这些随机变量的取值属于 $\mathcal{P}(\mathcal{C}([0,T], \mathbb{R}^d)) \times \mathcal{P}(\mathcal{C}([0,T], \mathbb{R}^d))$（详见附录中的注 10）。特别地，若令 $\tilde{\mu}$ 为取值于 $\mathcal{C}([0,T], \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d))$ 且由 $\tilde{\mu}: t \in [0,T] \mapsto \tilde{\mu}_t := \tilde{\rho}_t^1 \otimes \tilde{\rho}_t^2$ 定义的随机变量，则 μ^N 作为取值于 $\mathcal{C}([0,T], \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d))$ 的随机变量，几乎处处收敛到 $\tilde{\mu}$。我们的目标是证明 $\tilde{\mu}$ 实际上是确定性的，且等于 μ，其中 $\mu: t \in [0,T] \mapsto \rho_t^1 \otimes \rho_t^2$，而 ρ_t^1, ρ_t^2 满足福克-普朗克方程 (8)。

注意到，由上文可知，对所有的 $t \in [0,T]$ 和 $\phi \in \mathcal{C}_b(\mathbb{R}^d \times \mathbb{R}^d)$，

$$\lim_{N\to+\infty} |\langle \phi, \mu_t^N - \tilde{\mu}_t \rangle| + |m_{L_1}^\alpha[\rho_t^N] - m_{L_1}^\alpha[\tilde{\rho}_t]| + |m_{L_2}^\alpha[\rho_t^N] - m_{L_2}^\alpha[\tilde{\rho}_t]| = 0 \quad \text{a.s.} \tag{44}$$

事实上，根据假设 1，有 $\theta e^{-\alpha L_k(\theta)}\, e^{-\alpha L_k(\theta)} \in \mathcal{C}_b(\mathbb{R}^d)$ 对于 $k = 1, 2$. 另一方面，$\rho_t^N = \frac{N_1}{N}\rho^{1,}{}_t{}^N + \frac{N_2}{N}\rho^{2,}{}_t{}^N$ 几乎必然收敛至 $\tilde{\rho}_t = w_1\tilde{\rho}_t^1 + w_2\tilde{\rho}_t^2$（见下文注释 9）。然后，根据连续映射定理，我们有

$$\lim_{N\to+\infty} m_{L_k}^\alpha[\rho_t^N] = \lim_{N\to+\infty} \frac{\langle \theta e^{-\alpha L_k(\theta)}, \rho_t^N(d\theta)\rangle}{\langle e^{-\alpha L_k(\theta)}, \rho_t^N(d\theta)\rangle} = \frac{\langle \theta e^{-\alpha L_k(\theta)}, \tilde{\rho}_t(d\theta)\rangle}{\langle e^{-\alpha L_k(\theta)}, \tilde{\rho}_t(d\theta)\rangle} = m_{L_k}^\alpha[\tilde{\rho}_t] \quad \text{a.s.} \quad \text{for } k = 1, 2.$$

对于每个 $A > 0$，取 $\phi = |\cdot|^4 \wedge A \in \mathcal{C}_b(\mathbb{R}^d \times \mathbb{R}^d)$。由（44）可得

$$\begin{aligned}\mathbb{E}\Big[\int_{\mathbb{R}^d\times\mathbb{R}^d} (|(\theta_1,\theta_2)|^4 \wedge A)\tilde{\mu}_t(d\theta_1, d\theta_2)\Big] &= \mathbb{E}\Big[\lim_{N\to+\infty}\int_{\mathbb{R}^d\times\mathbb{R}^d} (|(\theta_1,\theta_2)|^4 \wedge A)\mu_t^N(d\theta_1, d\theta_2)\Big] \\ &\leq \liminf_{N\to+\infty} \frac{1}{N}\Big(\sum_{i=1}^{N_1} \mathbb{E}[|\theta_t^{1,(i,N)}|^4] + \sum_{j=1}^{N_2} \mathbb{E}[|\theta_t^{2,(j,N)}|^4]\Big) \leq K,\end{aligned}$$

其中用到了引理 5.1 以及以下事实：得益于注释 10，我们可以假设 μ^N，它是定义在 Skorohod 引理所给出的概率空间上的随机变量，仍可表示为定义在新概率空间上的布朗运动所驱动的（4）型扩散过程的轨迹的经验测度。

令上式中的 $A \to +\infty$，由单调收敛定理可得，。 t

$$\sup_{t\in[0,T]} \mathbb{E}\Big[\int_{\mathbb{R}^d\times\mathbb{R}^d} (|\theta_1|^4 + |\theta_2|^4)\tilde{\mu}_t(d\theta_1, d\theta_2)\Big] \leq K. \tag{45}$$

类似地，我们也有

$$\sup_{t\in[0,T]} \mathbb{E}\Big[\int_{\mathbb{R}^d} |\theta|^4 \tilde{\rho}_t(d\theta)\Big] \leq K. \tag{46}$$

引理B.3 进而推出

$$\mathbb{E}[|m_{L_k}^\alpha[\tilde{\rho}_t]|^4] < +\infty, \tag{47}$$

对于 $k = 1, 2$ 以及所有 $t \in [0,T]$。根据 $\langle \phi, \mu_t^N - \tilde{\mu}_t\rangle$ 的几乎必然收敛，以及引理 5.1 和（47）中的一致估计，我们可以推导出

$$\lim_{N\to+\infty} \mathbb{E}\big[|\langle \phi, \mu_t^N - \tilde{\mu}_t\rangle|^2 + |m_{L_1}^\alpha[\rho_t^N] - m_{L_1}^\alpha[\tilde{\rho}_t]|^2 + |m_{L_2}^\alpha[\rho_t^N] - m_{L_2}^\alpha[\tilde{\rho}_t]|^2\big] = 0. \tag{48}$$

注 9。对于 $k = 1, 2$，$\rho^{k,N}$ 作为取值于 $\mathcal{P}(\mathcal{C}([0,T], \mathbb{R}^d))$ 中的随机变量，在弱收敛拓扑下几乎处处收敛于 $\tilde{\rho}^k$。特别地，存在可测集合 $\Omega^k \subseteq \Omega$，使得 $\mathbb{P}(\Omega \backslash \Omega^k) = 0$ 且

$$\lim_{N\to+\infty} \int \phi(x) d\rho_t^{k,N}(\omega) = \int \phi(x) d\tilde{\rho}_t^k(\omega),$$

对于任意 $\phi \in \mathcal{C}_b(\mathbb{R}^d)$、任意 $t \in [0,T]$，以及任意 $\omega \in \Omega^k$。令 $\tilde{\Omega} := \Omega^1 \cap \Omega^2$。那么

$$\begin{aligned}\lim_{N\to+\infty} \int \phi(x) d\big(\frac{N_1}{N}\rho_t^{1,N} + \frac{N_2}{N}\rho_t^{2,N}\big)(\omega) &= \lim_{N\to+\infty} \frac{N_1}{N}\int \phi(x) d\rho_t^{1,N}(\omega) + \frac{N_2}{N}\int \phi(x) d\rho_t^{2,N}(\omega) \\ &= \int \phi(x) d\big(w_1\tilde{\rho}_t^1 + w_2\tilde{\rho}_t^2\big)(\omega),\end{aligned}$$

对于任意 $\phi\in \mathcal{C}_b(\mathbb{R}^d)$，$t\in[0,T]$ 以及 $\omega\in\Omega$。注意，$\mathbb{P}(\Omega\backslash\tilde{\Omega})=0$。由上述内容可推知

$$\rho_t^N=\frac{N_1}{N}\rho_t^{1,N}+\frac{N_2}{N}\rho_t^{2,N}\xrightarrow{a.s.}\tilde{\rho}_t=w_1\tilde{\rho}_t^1+w_2\tilde{\rho}_t^2,$$

其中，收敛必须在弱意义下理解。

我们现在可以证明定理2.2。

定理 2.2 的证明。我们可以继续在上述公共概率空间 $(\Omega\ \mathcal{F},\mathbb{P})$ 上进行讨论，其中 $(\rho^{1,N},\rho^{2,N})$ 收敛到 $(\tilde{\rho}^1,\tilde{\rho}^2)$ 是几乎处处意义下的。显然，几乎必然地，$\tilde{\mu}_t$ 在 (43) 的意义下关于时间连续。

Now,

$$\begin{aligned}
\mathbb{E}|F_\varphi(\mu^N,\ N_1/N)\ -\ F_\varphi(\tilde{\mu},w_1)|-|N_1/N-w_1|\le\mathbb{E}\Big|\Big(&\big\langle\varphi(\theta_1,\theta_2),\mu_t^N(d\theta_1,d\theta_2)\big\rangle-\big\langle\varphi(\theta_1,\theta_2),\mu_0^N(d\theta_1,d\theta_2)\big\rangle\\
&+\int_0^t\big\langle\big(\lambda_1(\theta_1-m_{L_1}^\alpha[\rho_s^N])+\lambda_2\nabla L_1(\theta_1)\big)\cdot\nabla_{\theta_1}\varphi\\
&\qquad+\big(\lambda_1(\theta_2-m_{L_2}^\alpha[\rho_s^N])+\lambda_2\nabla L_2(\theta_2)\big)\cdot\nabla_{\theta_2}\varphi,\mu_s^N(d\theta_1,d\theta_2)\big\rangle ds\\
&-\int_0^t\big\langle\big(\frac{\sigma_1^2}{2}|\theta_1-m_{L_1}^\alpha[\rho_s^N]|^2+\frac{\sigma_2^2}{2}|\nabla L_1(\theta_1)|^2\big)\Delta_{\theta_1}\varphi\\
&\qquad+\big(\frac{\sigma_1^2}{2}|\theta_2-m_{L_2}^\alpha[\rho_s^N]|^2+\frac{\sigma_2^2}{2}|\nabla L_2(\theta_2)|^2\big)\Delta_{\theta_2}\varphi,\mu_s^N(d\theta_1,d\theta_2)\big\rangle ds\Big)\\
-\Big(&\big\langle\varphi(\theta_1,\theta_2),\tilde{\mu}_t(d\theta_1,d\theta_2)\big\rangle-\big\langle\varphi(\theta_1,\theta_2),\tilde{\mu}_0(d\theta_1,d\theta_2)\big\rangle\\
&+\int_0^t\big\langle\big(\lambda_1(\theta_1-m_{L_1}^\alpha[\tilde{\rho}_s])+\lambda_2\nabla L_1(\theta_1)\big)\cdot\nabla_{\theta_1}\varphi\\
&\qquad+\big(\lambda_1(\theta_2-m_{L_2}^\alpha[\tilde{\rho}_s])+\lambda_2\nabla L_2(\theta_2)\big)\cdot\nabla_{\theta_2}\varphi,\tilde{\mu}_s(d\theta_1,d\theta_2)\big\rangle ds\\
&-\int_0^t\big\langle\big(\frac{\sigma_1^2}{2}|\theta_1-m_{L_1}^\alpha[\tilde{\rho}_s]|^2+\frac{\sigma_2^2}{2}|\nabla L_1(\theta_1)|^2\big)\Delta_{\theta_1}\varphi\\
&\qquad+\big(\frac{\sigma_1^2}{2}|\theta_2-m_{L_2}^\alpha[\tilde{\rho}_s]|^2+\frac{\sigma_2^2}{2}|\nabla L_2(\theta_2)|^2\big)\Delta_{\theta_2}\varphi,\tilde{\mu}_s(d\theta_1,d\theta_2)\big\rangle ds\Big)\Big|\\
\le&\mathbb{E}\Big[\Big|\big\langle\varphi(\theta_1,\theta_2),\mu_t^N(d\theta_1,d\theta_2)-\mu_0^N(d\theta_1,d\theta_2)\big\rangle-\big\langle\varphi(\theta_1,\theta_2),\tilde{\mu}_t(d\theta_1,d\theta_2)-\tilde{\mu}_0(d\theta_1,d\theta_2)\big\rangle\Big|\Big]\\
&+\mathbb{E}\Big[\Big|\int_0^t\lambda_1\big\langle(\theta_1-m_{L_1}^\alpha[\rho_s^N])\cdot\nabla_{\theta_1}\varphi+(\theta_2-m_{L_2}^\alpha[\rho_s^N])\cdot\nabla_{\theta_2}\varphi,\mu_s^N(d\theta_1,d\theta_2)\big\rangle ds\\
&\quad-\int_0^t\lambda_1\big\langle(\theta_1-m_{L_1}^\alpha[\tilde{\rho}_s])\cdot\nabla_{\theta_1}\varphi+(\theta_2-m_{L_2}^\alpha[\tilde{\rho}_s])\cdot\nabla_{\theta_2}\varphi,\tilde{\mu}_s(d\theta_1,d\theta_2)\big\rangle ds\Big|\Big]\\
&+\mathbb{E}\Big[\Big|\int_0^t\frac{\sigma_1^2}{2}\big\langle|\theta_1-m_{L_1}^\alpha[\rho_s^N]|^2\Delta_{\theta_1}\varphi+|\theta_2-m_{L_2}^\alpha[\rho_s^N]|^2\Delta_{\theta_2}\varphi,\mu_s^N(d\theta_1,d\theta_2)\big\rangle ds\\
&\quad-\int_0^t\frac{\sigma_1^2}{2}\big\langle|\theta_1-m_{L_1}^\alpha[\tilde{\rho}_s]|^2\Delta_{\theta_1}\varphi+|\theta_2-m_{L_2}^\alpha[\tilde{\rho}_s]|^2\Delta_{\theta_2}\varphi,\tilde{\mu}_s(d\theta_1,d\theta_2)\big\rangle ds\Big|\Big]\\
&+\mathbb{E}\Big[\Big|\int_0^t\big\langle\lambda_2(\nabla L_1(\theta_1)\cdot\nabla_{\theta_1}\varphi+\nabla_{L_2}(\theta_2)\cdot\nabla_{\theta_2}\varphi)\\
&\quad+\frac{\sigma_2^2}{2}(|\nabla L_1(\theta_1)|^2\Delta_{\theta_1}\varphi+|\nabla L_2(\theta_2)|^2\Delta_{\theta_2}\varphi),\mu_s^N(d\theta_1,d\theta_2)-\tilde{\mu}_s(d\theta_1,d\theta_2)\big\rangle ds\Big|\Big]\\
=:&\ I_1^N+I_2^N+I_3^N+I_4^N.
\end{aligned}$$

对于 $\varphi \in \mathcal{C}_c^2(\mathbb{R}^d \times \mathbb{R}^d)$，使用收敛（48），我们得到

$$\begin{aligned}\lim_{N\to+\infty} I_1^N &= \lim_{N\to+\infty} \mathbb{E}\Big[\Big|\big\langle \varphi(\theta_1,\theta_2), \mu_t^N(d\theta_1,d\theta_2) - \mu_0^N(d\theta_1,d\theta_2)\big\rangle - \big\langle \varphi(\theta_1,\theta_2), \tilde{\mu}_t(d\theta_1,d\theta_2) - \tilde{\mu}_0(d\theta_1,d\theta_2)\big\rangle\Big|\Big] \\ &= 0.\end{aligned}$$

对于项 I_2^N，我们有

$$\begin{aligned}&\Big| \int_0^t \langle (\theta_1 - m_{L_1}^\alpha[\rho_s^N]) \cdot \nabla_{\theta_1}\varphi + (\theta_2 - m_{L_2}^\alpha[\rho_s^N]) \cdot \nabla_{\theta_2}\varphi, \mu_s^N(d\theta_1,d\theta_2)\rangle ds \\ &\quad - \int_0^t \langle (\theta_1 - m_{L_1}^\alpha[\tilde{\rho}_s]) \cdot \nabla_{\theta_1}\varphi + (\theta_2 - m_{L_2}^\alpha[\tilde{\rho}_s]) \cdot \nabla_{\theta_2}\varphi, \tilde{\mu}_s(d\theta_1,d\theta_2)\rangle ds \Big| \\ \le &\int_0^t \Big| \langle (\theta_1 - m_{L_1}^\alpha[\rho_s^N]) \cdot \nabla_{\theta_1}\varphi + (\theta_2 - m_{L_2}^\alpha[\rho_s^N]) \cdot \nabla_{\theta_2}\varphi, \mu_s^N(d\theta_1,d\theta_2) - \tilde{\mu}_s(d\theta_1,d\theta_2)\rangle \Big| ds \\ &\quad + \int_0^t \Big| \langle (m_{L_1}^\alpha[\tilde{\rho}_s] - m_{L_1}^\alpha[\rho_s^N]) \cdot \nabla_{\theta_1}\varphi + (m_{L_2}^\alpha[\tilde{\rho}_s] - m_{L_2}^\alpha[\rho_s^N]) \cdot \nabla_{\theta_2}\varphi, \tilde{\mu}_s(d\theta_1,d\theta_2)\rangle \Big| ds \\ =: &\int_0^t |I_2^{N,1}(s)| ds + \int_0^t |I_2^{N,2}(s)| ds.\end{aligned}$$

可以计算得

$$\begin{aligned}\mathbb{E}[|I_2^{N,1}(s)|] &\le \mathbb{E}\big[\big|\langle \theta_1 \cdot \nabla_{\theta_1}\varphi + \theta_2 \cdot \nabla_{\theta_2}\varphi, \mu_s^N(d\theta_1,d\theta_2) - \tilde{\mu}_s(d\theta_1,d\theta_2)\rangle\big|\big] \\ &\quad + \mathbb{E}\big[\big|\langle m_{L_1}^\alpha[\rho_s^N] \cdot \nabla_{\theta_1}\varphi, \mu_s^N(d\theta_1,d\theta_2) - \tilde{\mu}_s(d\theta_1,d\theta_2)\rangle\big|\big] \\ &\quad + \mathbb{E}\big[\big|\langle m_{L_2}^\alpha[\rho_s^N] \cdot \nabla_{\theta_2}\varphi, \mu_s^N(d\theta_1,d\theta_2) - \tilde{\mu}_s(d\theta_1,d\theta_2)\rangle\big|\big] \\ &\le \mathbb{E}\big[\big|\langle \theta_1 \cdot \nabla_{\theta_1}\varphi + \theta_2 \cdot \nabla_{\theta_2}\varphi, \mu_s^N(d\theta_1,d\theta_2) - \tilde{\mu}_s(d\theta_1,d\theta_2)\rangle\big|\big] \\ &\quad + K^{\frac{1}{2}}\big(\mathbb{E}\big[\big|\langle \nabla_{\theta_1}\varphi, \mu_s^N(d\theta_1,d\theta_2) - \tilde{\mu}_s(d\theta_1,d\theta_2)\rangle\big|^2\big]\big)^{\frac{1}{2}} \\ &\quad + K^{\frac{1}{2}}\big(\mathbb{E}\big[\big|\langle \nabla_{\theta_2}\varphi, \mu_s^N(d\theta_1,d\theta_2) - \tilde{\mu}_s(d\theta_1,d\theta_2)\rangle\big|^2\big]\big)^{\frac{1}{2}},\end{aligned}$$

其中，第二个不等式中用到了引理5.1 。由于 φ 具有紧支集，应用（48）可得

$$\lim_{N\to+\infty} \mathbb{E}\big[|I_2^{N,1}(s)|\big] = 0.$$

此外， $\mathbb{E}\big[|I_1^N(s)|\big]$ 关于 s 的一致有界性可直接由（45）以及引理5.1中的估计得到。因此，由控制收敛定理

$$\lim_{N\to+\infty} \int_0^t \mathbb{E}\big[|I_2^{N,1}(s)|\big] ds = 0. \tag{49}$$

对于 $I^{N,22}$，我们知道

$$\begin{aligned}&\big|\langle (m_{L_1}^\alpha[\tilde{\rho}_s] - m_{L_1}^\alpha[\rho_s^N]) \cdot \nabla_{\theta_1}\varphi + (m_{L_1}^\alpha[\tilde{\rho}_s] - m_{L_1}^\alpha[\rho_s^N]) \cdot \nabla_{\theta_1}\varphi, \tilde{\mu}_s(d\theta_1,d\theta_2)\rangle\big| \\ &\le \|\nabla\varphi\|_\infty \big(\big|m_{L_1}^\alpha[\tilde{\rho}_s] - m_{L_1}^\alpha[\rho_s^N]\big| + \big|m_{L_2}^\alpha[\tilde{\rho}_s] - m_{L_2}^\alpha[\rho_s^N]\big|\big).\end{aligned}$$

因此，由（48）可得

$$\lim_{N\to+\infty} \mathbb{E}\big[|I_2^{N,2}(s)|\big] = 0.$$

再次根据控制收敛定理，我们有

$$\lim_{N\to+\infty} \int_0^t \mathbb{E}\big[|I_2^{N,2}(s)|ds\big] = 0.$$

这与（49）结合后得到

$$\begin{aligned}\lim_{N\to+\infty} I_2^N &= \lim_{N\to+\infty} \mathbb{E}\Big[\Big| \int_0^t \langle (\theta_1 - m_{L_1}^\alpha[\rho_s^N]) \cdot \nabla_{\theta_1}\varphi + (\theta_2 - m_{L_2}^\alpha[\rho_s^N]) \cdot \nabla_{\theta_2}\varphi, \mu_s^N(d\theta_1, d\theta_2)\rangle ds \\ &\qquad - \int_0^t \langle (\theta_1 - m_{L_1}^\alpha[\tilde{\rho}_s]) \cdot \nabla_{\theta_1}\varphi + (\theta_2 - m_{L_2}^\alpha[\tilde{\rho}_s]) \cdot \nabla_{\theta_2}\varphi, \tilde{\mu}_s(d\theta_1, d\theta_2)\rangle ds \Big|\Big] \\ &= 0.\end{aligned} \tag{50}$$

类似地，对于术语 I_3^N，我们将误差分为

$$\begin{aligned}&\Big| \int_0^t \langle |\theta_1 - m_{L_1}^\alpha[\rho_s^N]|^2 \Delta_{\theta_1}\varphi + |\theta_2 - m_{L_2}^\alpha[\rho_s^N]|^2 \Delta_{\theta_2}\varphi, \mu_s^N(d\theta_1, d\theta_2)\rangle ds \\ &\quad - \int_0^t \langle |\theta_1 - m_{L_1}^\alpha[\tilde{\rho}_s]|^2 \Delta_{\theta_1}\varphi + |\theta_2 - m_{L_2}^\alpha[\tilde{\rho}_s]|^2 \Delta_{\theta_2}\varphi, \tilde{\mu}_s(d\theta_1, d\theta_2)\rangle ds \Big| \\ =& \int_0^t \Big| \langle |\theta_1 - m_{L_1}^\alpha[\rho_s^N]|^2 \Delta_{\theta_1}\varphi, \mu_s^N(d\theta_1, d\theta_2) - \tilde{\mu}_s(d\theta_1, d\theta_2)\rangle \Big| ds \\ &+ \int_0^t \Big| \langle |\theta_2 - m_{L_2}^\alpha[\rho_s^N]|^2 \Delta_{\theta_2}\varphi, \mu_s^N(d\theta_1, d\theta_2) - \tilde{\mu}_s(d\theta_1, d\theta_2)\rangle \Big| ds \\ &+ \int_0^t \Big| \langle \big(|\theta_1 - m_{L_1}^\alpha[\tilde{\rho}_s]|^2 - |\theta_1 - m_{L_1}^\alpha[\rho_s^N]|^2\big) \Delta_{\theta_1}\varphi, \tilde{\mu}_s(d\theta_1, d\theta_2)\rangle \Big| ds \\ &+ \int_0^t \Big| \langle \big(|\theta_2 - m_{L_2}^\alpha[\tilde{\rho}_s]|^2 - |\theta_2 - m_{L_2}^\alpha[\rho_s^N]|^2\big) \Delta_{\theta_2}\varphi, \tilde{\mu}_s(d\theta_1, d\theta_2)\rangle \Big| ds \\ =:& \int_0^t |I_3^{N,1}(s)| ds + \int_0^t |I_3^{N,2}(s)| ds + \int_0^t |I_3^{N,3}(s)| ds + \int_0^t |I_3^{N,4}(s)| ds.\end{aligned}$$

可以计算

$$\begin{aligned}\mathbb{E}\big[|I_3^{N,1}(s)|\big] &\le 2\mathbb{E}\big[\big|\langle (|\theta_1|^2 + |m_{L_1}^\alpha[\rho_s^N]|^2) \Delta_{\theta_1}\varphi, \mu_s^N(d\theta_1, d\theta_2) - \tilde{\mu}_s(d\theta_1, d\theta_2)\rangle\big|\big] \\ &\le 2\mathbb{E}\big[\big|\langle |\theta_1|^2 \Delta_{\theta_1}\varphi, \mu_s^N(d\theta_1, d\theta_2) - \tilde{\mu}_s(d\theta_1, d\theta_2)\rangle\big|\big] \\ &\quad + 2\mathbb{E}\big[|m_{L_1}^\alpha[\rho_s^N]|^2 |\langle \Delta_{\theta_1}\varphi, \mu_s^N(d\theta_1, d\theta_2) - \tilde{\mu}_s(d\theta_1, d\theta_2)\rangle|\big] \\ &\le 2\mathbb{E}\big[\big|\langle |\theta_1|^2 \Delta_{\theta_1}\varphi, \mu_s^N(d\theta_1, d\theta_2) - \tilde{\mu}_s(d\theta_1, d\theta_2)\rangle\big|\big] \\ &\quad + 2K\big(\mathbb{E}\big[\big|\langle \Delta_{\theta_1}\varphi, \mu_s^N(d\theta_1, d\theta_2) - \tilde{\mu}_s(d\theta_1, d\theta_2)\rangle\big|^2\big]\big)^{\frac{1}{2}},\end{aligned}$$

其中我们使用了 引理 5.1 在最后一个不等式中。由于 φ 具有紧支集，应用（48）得到 t o

$$\lim_{N\to+\infty} \mathbb{E}\big[|I_3^{N,1}(s)|\big] = 0.$$

此外， $\mathbb{E}\big[|I_3^{N,1}(s)|\big]$ 的一致有界性可直接由（45）及引理5.1中的估计得到，由此根据控制收敛定理可知

$$\lim_{N\to+\infty} \int_0^t \mathbb{E}\big[|I_3^{N,1}|\big] ds = 0.$$

通过类似的计算，我们还可以得到

$$\lim_{N\to+\infty} \int_0^t \mathbb{E}\big[|I_3^{N,2}|\big] ds = 0.$$

至于 $I^{N,3}_3$ 和 $I^{N,4}_3$，

$$\begin{aligned}
\mathbb{E}\big[|I_3^{N,3}|\big] &= \mathbb{E}\big[\big|\big\langle\big(|\theta_1 - m^\alpha_{L_1}[\tilde\rho_s]|^2 - |\theta_1 - m^\alpha_{L_1}[\rho_s^N]|^2\big)\Delta_{\theta_1}\varphi, \tilde\mu_s(d\theta_1, d\theta_2)\big\rangle\big|\big] \\
&= \mathbb{E}\Big[\Big|\big[\langle|m^\alpha_{L_1}[\tilde\rho_s]|^2 - m^\alpha_{L_1}[\rho_s^N]\big|^2 + 2\theta_1\cdot(m^\alpha_{L_1}[\rho_s^N] - m^\alpha_{L_1}[\tilde\rho_s])\big]\Delta_{\theta_1}\varphi, \tilde\mu_s(d\theta_1, d\theta_2)\rangle\Big|\Big] \\
&\le \mathbb{E}\Big[\big(|m^\alpha_{L_1}[\tilde\rho_s]| + |m^\alpha_{L_1}[\rho_s^N]|\big)\big|m^\alpha_{L_1}[\rho_s^N] - m^\alpha_{L_1}[\tilde\rho_s]\big| + 2\|\theta_1\Delta_{\theta_1}\varphi\|_\infty\big|m^\alpha_{L_1}[\rho_s^N] - m^\alpha_{L_1}[\tilde\rho_s]\big|\Big] \\
&\le 4K^{\frac12}\big(\mathbb{E}\big[\big|m^\alpha_{L_1}[\rho_s^N] - m^\alpha_{L_1}[\tilde\rho_s]\big|^2\big]\big)^{\frac12} + 2\|\theta_1\Delta_{\theta_1}\varphi\|_\infty\mathbb{E}\big[\big|m^\alpha_{L_1}[\rho_s^N] - m^\alpha_{L_1}[\tilde\rho_s]\big|\big],
\end{aligned}$$

其中我们用到了　　引理5.1 中的最后一个不等式。由于 φ 具有紧支集，应用（48）可得　　o

$$\lim_{N\to+\infty}\mathbb{E}\big[|I_3^{N,3}(s)|\big] = 0.$$

此外，$\mathbb{E}[|I_3^{N,3}$（s）$|]$ 的一致有界性可直接由（47）以及引理5.1中的估计得出，再由控制收敛定理可知

$$\lim_{N\to+\infty}\int_0^t\mathbb{E}\big[|I_3^{N,3}(s)|\big]ds = 0.$$

与上述计算类似，我们还有

$$\lim_{N\to+\infty}\int_0^t\mathbb{E}\big[|I_3^{N,4}(s)|\big]ds = 0.$$

因此，我们得出

$$\begin{aligned}
\lim_{N\to+\infty} I_3^N = \lim_{N\to+\infty}\mathbb{E}\Big[\Big|&\int_0^t\langle|\theta_1 - m^\alpha_{L_1}[\rho_s^N]|^2\Delta_{\theta_1}\varphi + |\theta_2 - m^\alpha_{L_2}[\rho_s^N]|^2\Delta_{\theta_2}\varphi, \mu_s^N(d\theta_1, d\theta_2)\rangle ds \\
&- \int_0^t\langle|\theta_1 - m^\alpha_{L_1}[\tilde\rho_s]|^2\Delta_{\theta_1}\varphi + |\theta_2 - m^\alpha_{L_2}[\tilde\rho_s]|^2\Delta_{\theta_2}\varphi, \tilde\mu_s(d\theta_1, d\theta_2)\rangle ds\Big|\Big] \\
= 0.&
\end{aligned} \tag{51}$$

对于项 I_4^N，由于 L_1，L_2 具有有界梯度，且 $\varphi\in\mathcal{C}_c^2$（$\mathbb{R}^d\times\mathbb{R}^d$），由（48）可知

$$\begin{aligned}
\lim_{N\to+\infty}\mathbb{E}\Big[\Big|\big\langle&\lambda_2(\nabla L_1(\theta_1)\cdot\nabla_{\theta_1}\varphi + \nabla_{L_2}(\theta_2)\cdot\nabla_{\theta_2}\varphi) \\
&+ \frac{\sigma_2^2}{2}(|\nabla L_1(\theta_1)|^2\Delta_{\theta_1}\varphi + |\nabla L_2(\theta_2)|^2\Delta_{\theta_2}\varphi), \mu_s^N(d\theta_1, d\theta_2) - \tilde\mu_s(d\theta_1, d\theta_2)\big\rangle\Big|\Big] = 0.
\end{aligned}$$

另请注意，上述对象是一致有界的。因此，根据控制收敛定理，我们得到

$$\lim_{N\to+\infty} I_4^N = 0.$$

综合 I_1^N, I_2^N, I_3^N和 I_4^N的估计，以及假设 N_1/N 收敛到w_1 这一事实，我们推得

$$\lim_{N\to+\infty}\mathbb{E}\big[|F_\varphi(\mu^N, N_1/N) - F_\varphi(\tilde\mu, w_1)|\big] = 0.$$

于是我们有

$$\begin{aligned}
\mathbb{E}\big[|F_\varphi(\tilde\mu, w_1)|\big] &\le \mathbb{E}\big[|F_\varphi(\mu^N, N_1/N) - F_\varphi(\tilde\mu, w_1)|\big] + \mathbb{E}\big[|F_\varphi(\mu^N, N_1/N)|\big] \\
&\le \mathbb{E}\big[|F_\varphi(\mu^N, N_1/N) - F_\varphi(\tilde\mu, w_1)|\big] + \sqrt{\frac{C}{N_1N_2}} + 2|\frac{N_1}{N} - w_1|^2 \to 0
\end{aligned}$$

当 N_1, $N_2 \to +\infty$ 且 $N \to +\infty$，其中我们在最后一个不等式中使用了引理 5.4 。这意味着F_φ（$\tilde{\mu}$, w_1）$= 0$ 几乎必然成立。换言之，有

$$\begin{aligned}
&\langle \varphi(\theta_1,\theta_2), \tilde{\mu}_t(d\theta_1,d\theta_2)\rangle - \langle \varphi(\theta_1,\theta_2), \tilde{\mu}_0(d\theta_1,d\theta_2)\rangle \\
&\quad + \int_0^t \langle \big(\lambda_1(\theta_1 - m_{L_1}^\alpha[\tilde{\rho}_s]) + \lambda_2 \nabla L_1(\theta_1)\big)\cdot \nabla_{\theta_1}\varphi \\
&\qquad\qquad + \big(\lambda_1(\theta_2 - m_{L_2}^\alpha[\tilde{\rho}_s]) + \lambda_2 \nabla L_2(\theta_2)\big)\cdot \nabla_{\theta_2}\varphi, \tilde{\mu}_s(d\theta_1,d\theta_2)\rangle ds \\
&\quad - \int_0^t \langle \big(\frac{\sigma_1^2}{2}|\theta_1 - m_{L_1}^\alpha[\tilde{\rho}_s]|^2 + \frac{\sigma_2^2}{2}|\nabla L_1(\theta_1)|^2\big)\Delta_{\theta_1}\varphi \\
&\qquad\qquad + \big(\frac{\sigma_1^2}{2}|\theta_2 - m_{L_2}^\alpha[\tilde{\rho}_s]|^2 + \frac{\sigma_2^2}{2}|\nabla L_2(\theta_2)|^2\big)\Delta_{\theta_2}\varphi, \tilde{\mu}_s(d\theta_1,d\theta_2)\rangle ds = 0 \qquad \text{a.s.},
\end{aligned}$$

对于所有 $\varphi \in C_c^2(\mathbb{R}^d \times \mathbb{R}^d)$ 及所有 $t \in [0,T]$。

由上述结果，我们可以通过稠密性论证推知， $\tilde{\mu}$ 几乎必然是方程（42）的一个解。利用方程（42）的弱解的唯一性（参见附录中的引理 D.2 ）以及备注 8 ，我们可以得出$\tilde{\mu} = \mu$ 几乎必然成立。特别地，Law(μ^N) 弱收敛于 δ_μ，即位于 μ处的狄拉克 δ 分布。因此，后一结论在任何概率空间中均成立（而不仅仅是在斯科罗霍德定理所提供的概率空间中），只要该概率空间支持具有 μ^N所对应分布的随机变量。并且，由于依分布收敛的极限是一个狄拉克 δ 分布，我们可以利用斯卢茨基定理推知， μ^N 在原始概率空间中也依概率收敛于常数随机变量μ（而不仅仅是在斯科罗霍德定理所提供的空间中）。这里，依概率收敛是指：对于每一个 $\zeta > 0$ ，我们都有

$$\lim_{N\to\infty} \mathbb{P}\left(\sup_{t\in[0,T]} d_{LP}(\mu_t^N, \rho_t^1 \otimes \rho_t^2) \geq \zeta \right) = 0.$$

注意到，空间 $\mathcal{C}([0,T], \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d))$ 中的度量确实由下式给出：

$$d(\mu', \tilde{\mu}') = \sup_{t\in[0,T]} d_{LP}(\mu_t', \tilde{\mu}_t'),$$

其中， $d_{LP}(\cdot,\cdot)$ 是 $\mathbb{R}^d \times \mathbb{R}^d$ 上概率测度之间的列维-普罗霍罗夫度量。证明至此完成。

□

6. 结论

本文是将基于共识的优化领域的研究以及其他基于偏微分方程的优化方法与联邦学习问题相结合的初步探索。具体而言，我们提出了一种新型的 CBO 类相互作用粒子系统，可用于解决实际聚类联邦学习场景中出现的非凸优化问题。我们证明，当相互作用粒子的数量趋于无穷时，该粒子系统会收敛到适当的平均场极限。进而，我们分析平均场模型的时间演化，并探讨其如何促使每个簇中的粒子围绕该簇目标函数的全局极小值点达成共识。这种平均场视角或许与实际情况相去不远，尤其是在处理跨设备联邦学习问题时，因为此类问题中的用户数量确实相当庞大。受我们提出的新型 CBO 类粒子动力学启发，我们提出了 FedCBO 算法，并通过实验评估其性能。在实验中，我们表明，该算法优于当前联邦学习领域最先进的方法。

我们这项工作引出了一些值得进一步研究的重要问题。理论方面，平均场系统的长期稳定性仍是一个悬而未决的问题。尤其是，当方差（定义见定理 2.3）达到规定的容差水平 ε 后，模型会如何表现，目前尚不清楚。此外，直接研究有限粒子系统，而不经过平均场极限，也很有意义。另外，分析一种更现实的情形也很有意义：同一集群中的智能体可能具有不同但彼此相关的损失函数。实验方面，我们希望研究 FedCBO 算法对对抗攻击的鲁棒性。事实上，鉴于模型聚合步骤（19）中的加权平均机制，认为 FedCBO 算法能够对对抗攻击提供一定防护并非不合理。

最后，进一步探索降低我们算法通信成本的策略，也是一个值得关注的研究课题。

References

[1] L. Arnold。随机微分方程。 *New York*，1974年。[2] P. Billingsley。 *Convergence of probability measures*。John Wiley & Sons，2013年。[3] G. Borghi、M. Herty 和 L. Pareschi。用于多目标优化并实现均匀帕累托前沿近似的自适应共识方法。 *arXiv preprint arXiv:2208.01362*，2022年。[4] G. Borghi、M. Herty 和 L. Pareschi。用于多目标优化的基于共识的算法及其平均场描述。见 *2022 IEEE 61st Conference on Decision and Control (CDC)*，第 4131–4136 页。IEEE，2022年。[5] L. Bungert、P. Wacker 和 T. Roith。用于优化和采样的极化的基于共识动力学。 *arXiv preprint arXiv:2211.05238* 2022 ，。[6] J. A. Carrillo、Y.-P. Choi、C. Totzeck 和 O. Tse。基于共识的全局优化方法的分析框架。 *Mathematical Models and Methods in Applied Sciences*，28(06):1037–1066，2018年。[7] J. A. Carrillo、F. Hoffmann、A. M. Stuart 和 U. Vaes。基于共识的采样。 *Studies in Applied Mathematics*，148 3 1069 1140，2022年（）：–。[8] J. A. Carrillo、S. Jin、L. Li 和 Y. Zhu。面向高维机器学习问题的基于共识的全局优化方法。 *ESAIM: Control, Optimisation and Calculus of Variations*，27:S5，2021年。[9] J. A. Carrillo、C. Totzeck 和 U. Vaes。用于带约束全局优化问题的基于共识的优化与集合卡尔曼反演。见 *Modeling and Simulation for Collective Dynamics Lecture Notes Series Institute for* ，，，，第 *Mathematical Sciences NUS* 40, 2023 卷。[10] A. Dembo。 *Large deviations techniques and applications*。Springer，2009年。[11] L. C. Evans。 *Partial differential equations*，第 19 卷。American Mathematical Soc.，2010年。[12] M. Fornasier、H. Huang、L. Pareschi 和 P. S¨ unnen。球面上的基于共识优化：收敛至全局极小值点及机器学习。 *The Journal of Machine Learning Research*，22(1):10722–10776，2021年。[13] M. Fornasier、T. Klock 和 K. Riedl。基于共识的优化算法在平均场定律意义下全局收敛。 *arXiv preprint arXiv:2103.15130* 2021 ，。[14] J. Geiping、H. Bauermeister、H. Dr¨oge 和 M. Moeller。反演梯度——在联邦学习中破解隐私有多容易? *Advances in Neural Information Processing Systems*，33:16937–16947，2020年。[15] A. Ghosh、J. Chung、D. Yin 和 K. Ramchandran。用于聚类联邦学习的高效框架。 *arXiv preprint arXiv:2006.04088* 2020 ，。[16] H. Huang 和 J. Qiu。关于基于共识优化的平均场极限。 *arXiv preprint arXiv:2105.12919*，2021年。[17] H. Huang、J. Qiu 和 K. Riedl。用于鞍点问题的基于共识的优化。 *arXiv preprint arXiv:2212.12334*，2022年。[18] Y. Huang、S. Gupta、Z. Song、K. Li 和 S. Arora。评估联邦学习中的梯度反演攻击与防御措施。*Advances in Neural Information Processing Systems*，34:7232–7241，2021年。[19] P. Kairouz、H. B. McMahan、B. Avent、A. Bellet、M. Bennis、A. N. Bhagoji、K. Bonawitz、Z. Charles、G. Cormode、R. Cummings 等。联邦学习的进展与开放问题。 *Foundations and Trends® in Machine Learning*，14 1 2 1(–): –210，2021年。[20] O. Kallenberg。关于经典随机微分方程普适函数解的存在性。 *Annals of Probability*，24:196–205，1996年。[21] S. P. Karimireddy、S. Kale、M. Mohri、S. Reddi、S. Stich 和 A. T. Suresh。Scaffold：用于联邦学习的随机控制平均。见 *International Conference on Machine Learning*，第 5132–5143 页。PMLR，2020年。[22] Y. Lecun、L. Bottou、Y. Bengio 和 P. Haffner。应用于文档识别的基于梯度的学习。 *Proceedings of the IEEE* 86 11 2278 2324, 1998 ，( ): –。[23] T. Li、A. K. Sahu、M. Zaheer、M. Sanjabi、A. Talwalkar 和 V. Smith。异构网络中的联邦优化。 *Proceedings of Machine learning and systems*，2:429–450，2020年。[24] Z. Li、J. Zhang、L. Liu 和 J. Liu。通过生成式梯度泄露审计联邦学习中的隐私防御措施。见 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*，第 10132–10142 页，2022年。[25] G. Long、M. Xie、T. Shen、T. Zhou、X. Wang 和 J. Jiang。多中心联邦学习：通过客户端聚类实现更好的个性化。 *World Wide Web*，26(1):481–500，2023年。[26] J. Ma、G. Long、T. Zhou、J. Jiang 和 C. Zhang。关于聚类联邦学习的收敛性。 *arXiv preprint arXiv:2202.06187* 2022 ，。[27] Y. Mansour、M. Mohri、J. Ro 和 A. T. Suresh。个性化的三种方法及其在联邦学习中的应用。 *arXiv preprint arXiv:2002.10619*，2020年。[28] B. McMahan、E. Moore、D. Ramage、S. Hampson 和 B. A. y Arcas。基于去中心化数据的深度网络通信高效学习。见 *Artificial intelligence and statistics*，第 1273–1282 页。PMLR，2017年。[29] P. D. Miller。 *Applied asymptotic analysis*，第 75 卷。American Mathematical Soc.，2006年。[30] M. Mohri、G. Sivek 和 A. T. Suresh。无先验联邦学习。见 *International Conference on Machine Learning*，第 4615–4625 页。PMLR，2019年。[31] R. Pinnau、C. Totzeck、O. Tse 和 S. Martin。用于全局优化的基于共识的模型及其平均场极限。*Mathematical Models and Methods in Applied Sciences*，27(01):183–204，2017年。[32] K. Riedl。在基于共识的优化中利用记忆效应和梯度信息：平均场定律意义下的全局收敛。 *arXiv preprint arXiv:2211.12184*

[33] Y. Ruan 和 C. Joe-Wong。Fedsoft：采用近端本地更新的软聚类联邦学习。载于 *Proceedings of the AAAI Conference on Artificial Intelligence* 36, 8124 8131, 2022，卷页码 – 。

[34] F. Sattler、K.-R. M¨uller 和 W. Samek。聚类联邦学习：隐私约束下与模型无关的分布式多任务优化。*IEEE transactions on neural networks and learning systems*, 32(8):3710–3722, 2020。[35] A.-S. Sznitman。混沌传播专题。载于 *Ecole d'été de probabilités de Saint-Flour XIX—1989*, 第 165–251 页。Springer, 1991。[36] C. Totzeck。基于共识的优化趋势。载于 *Active Particles Volume 3: Advances in Theory Models and Applications* 201 226 2021 第 – 页。Springer, 。[37] H. Yin、A. Mallya、A. Vahdat、J. M. Alvarez、J. Kautz 和 P. Molchanov。透视梯度：通过 gradinversion 恢复图像批次。载于 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 第 16337–16346 页，2021。[38] M. Zhang、K. Sapra、S. Fidler、S. Yeung 和 J. M. Alvarez。采用一阶模型优化的个性化联邦学习。*arXiv preprint arXiv:2012.08565*, 2020。[39] B. Zhao、K. R. Mopuri 和 H. Bilen。idlg：改进的梯度深度泄露。*arXiv preprint arXiv:2001.02610*, 2020。[40] L. Zhu 和 S. Han。梯度深度泄露。载于 *Federated learning*, 第 17–31 页。Springer, 2020。

AppendixA. Moment estimates for the Stochastic Empirical Measures对于粒子系统（$\boldsymbol{\theta}^{1,N} \in \mathcal{C}([0,T], \mathbb{R}^d)^{N_1}$, $\boldsymbol{\theta}^{2,N} \in \mathcal{C}([0,T], \mathbb{R}^d)^{N_2}$的解（4），我们记

$$\rho_t^{1,N_1} = \frac{1}{N_1}\sum_{i=1}^{N_1} \delta_{\theta_t^{1,i}} \qquad \rho_t^{2,N_2} = \frac{1}{N_2}\sum_{i=1}^{N_2} \delta_{\theta_t^{2,i}} \qquad \rho_t^N = \frac{N_1}{N}\rho_t^{1,N_1} + \frac{N_2}{N}\rho_t^{2,N_2}$$

对应于 $\boldsymbol{\theta}^{1,N}$、$\boldsymbol{\theta}^{2,N}$的经验测度，对每个 $t \in [0,T]$ 均成立。

引理 A.1。设 L_1、L_2 满足假设 *1*，并满足有界性 (1) 或无穷远处的二次增长 (2)，且 $\rho_0 \in \mathcal{P}_{2p}(\mathbb{R}^d)$ $p \geq 1$。此外，设 $\boldsymbol{\theta}^{1,N}$、$\boldsymbol{\theta}^{2,N}$ 是粒子系统 (4) 的解，其初始数据 $\boldsymbol{\theta}^{1,N}_0$、$\boldsymbol{\theta}^{2,N}_0$ 和 ρ^{1,N_1}、ρ^{2,N_2} 分别服从 $\rho_0^{\otimes N}$ 分布，且 ρ^N 为相应的经验测度。于是存在一个与 N 无关的常数 $K > 0$，使得

$$\sup_{t\in[0,T]} \mathbb{E}\int |\theta|^{2p} d\rho_t^N, \qquad \sup_{t\in[0,T]} \mathbb{E}|m^{\alpha}_{L_1}[\rho_t^N]|^{2p}, \qquad \sup_{t\in[0,T]} \mathbb{E}|m^{\alpha}_{L_2}[\rho_t^N]|^{2p} \leq K, \tag{52}$$

因此，以下估计也成立

$$\sup_{t\in[0,T]} \mathbb{E}\int |\theta|^{2} d\eta_{1,t}^{\alpha,N}, \qquad \sup_{t\in[0,T]} \mathbb{E}|\theta_t^{1,i}|^{2p} \leq K,$$

$$\sup_{t\in[0,T]} \mathbb{E}\int |\theta|^{2} d\eta_{2,t}^{\alpha,N}, \qquad \sup_{t\in[0,T]} \mathbb{E}|\theta_t^{2,j}|^{2p} \leq K,$$

对于 $i = 1、2、\ldots、N_1$ 和 $j = 1、2、\ldots、N_2$。

证明。设 $\boldsymbol{\theta}^{1,N}$, $\boldsymbol{\theta}^{2,N}$ 为粒子系统（4）的解。利用不等式 $(a+b)^q \le 2^{q-1}(a^q+b^q)$ $q \ge 1$ 以及伊藤等距公式和詹森不等式可得

$$
\begin{aligned}
\mathbb{E}|\theta_t^{1,i}|^{2p} &= \mathbb{E}\Big[\theta_0^{1,i} + \int_0^t \big(-\lambda_1(\theta_s^{1,i} - m_{L_1}^{\alpha}[\rho_s^N]) - \lambda_2 \nabla L_1(\theta_s^{1,i})\big)ds \\
&\quad + \int_0^t \sigma_1 |\theta_s^{1,i} - m_{L_1}^{\alpha}[\rho_s^N]| dB_s^{1,i} + \int_0^t \sigma_2 |\nabla L_1(\theta_s^{1,i})| d\widetilde{B}_s^{1,i}\Big]^{2p} \\
&\le 2^{2p-1}\mathbb{E}|\theta_0^{1,i}|^{2p} + 2^{3(2p-1)}\mathbb{E}\Big[\int_0^t \big(-\lambda_1(\theta_s^{1,i} - m_{L_1}^{\alpha}[\rho_s^N]) - \lambda_2 \nabla L_1(\theta_s^{1,i})\big)ds\Big]^{2p} \\
&\quad + 2^{3(2p-1)}\mathbb{E}\Big[\int_0^t \sigma_1 |\theta_s^{1,i} - m_{L_1}^{\alpha}[\rho_s^N]| dB_s^{1,i}\Big]^{2p} + 2^{3(2p-1)}\mathbb{E}\Big[\int_0^t \sigma_2 |\nabla L_1(\theta_s^{1,i})| d\widetilde{B}_s^{1,i}\Big]^{2p} \\
&\le 2^{2p-1}\mathbb{E}|\theta_0^{1,i}|^{2p} + 2^{4(2p-1)}\lambda_1^{2p}T^{2p-1}\int_0^t \big(\mathbb{E}|\theta_s^{1,i}|^{2p}ds + \mathbb{E}|m_{L_1}^{\alpha}[\rho_s^N]|^{2p}\big)ds \\
&\quad + 2^{3(2p-1)}\lambda_2^{2p}T^{2p-1}\int_0^t \mathbb{E}|\nabla L_1(\theta_s^{1,i})|^{2p}ds \\
&\quad + 2^{4(2p-1)}T^{p-1}\sigma_1^{2p}p(2p-1)^p\int_0^t \big(\mathbb{E}|\theta_s^{1,i}|^{2p}ds + \mathbb{E}|m_{L_1}^{\alpha}[\rho_s^N]|^{2p}\big)ds \\
&\quad + 2^{3(2p-1)}T^{p-1}\sigma_2^{2p}p(2p-1)^p\int \mathbb{E}|\nabla L_1(\theta_s^{1,i})|^{2p}ds \\
&= 2^{2p-1}\mathbb{E}|\theta_0^{1,i}|^{2p} + 2^{4(2p-1)}\big(\lambda_1^{2p}T^p + p(2p-1)^p\sigma_1^{2p}\big)T^{p-1}\int_0^t \big(\mathbb{E}|\theta_s^{1,i}|^{2p}ds + \mathbb{E}|m_{L_1}^{\alpha}[\rho_s^N]|^{2p}\big)ds \\
&\quad + 2^{3(2p-1)}\big(\lambda_2^{2p}T^p + \sigma_2^{2p}p(2p-1)^p\big)T^pC_{\nabla L_1}^{2p},
\end{aligned}
\tag{54}
$$

对于 $i \in [N_1]$。类似的计算可得，对于 $j \in [N_2]$

$$
\begin{aligned}
\mathbb{E}|\theta_t^{2,j}|^{2p} &\le 2^{2p-1}\mathbb{E}|\theta_0^{2,j}|^{2p} + 2^{4(2p-1)}\big(\lambda_1^{2p}T^p + p(2p-1)^p\sigma_1^{2p}\big)T^{p-1}\int_0^t \big(\mathbb{E}|\theta_s^{2,j}|^{2p}ds + \mathbb{E}|m_{L_2}^{\alpha}[\rho_s^N]|^{2p}\big)ds \\
&\quad + 2^{3(2p-1)}\big(\lambda_2^{2p}T^p + \sigma_2^{2p}p(2p-1)^p\big)T^pC_{\nabla L_2}^{2p}.
\end{aligned}
$$

将上述两个不等式分别对 $i \in [N_1]$ 和 $j \in [N_2]$ 求和，再除以 N，可得

$$
\begin{aligned}
\mathbb{E}\int |\theta|^{2p}d\rho_t^N &\le 2^{2p-1}\mathbb{E}\int |\theta|^{2p}d\rho_0^N \\
&\quad + 2^{4(2p-1)}\big(\lambda_1^{2p}T^p + p(2p-1)^p\sigma_1^{2p}\big)T^{p-1}\int_0^t \Big[\mathbb{E}\int |\theta|^{2p}d\rho_s^N \\
&\qquad + \frac{N_1}{N}\mathbb{E}|m_{L_1}^{\alpha}[\rho_s^N]|^{2p} + \frac{N_2}{N}\mathbb{E}|m_{L_2}^{\alpha}[\rho_s^N]|^{2p}\Big]ds \\
&\quad + 2^{3(2p-1)}\big(\lambda_2^{2p}T^p + \sigma_2^{2p}p(2p-1)^p\big)T^p\big(\frac{N_1}{N}C_{\nabla L_1}^{2p} + \frac{N_2}{N}C_{\nabla L_2}^{2p}\big).
\end{aligned}
\tag{55}
$$

如第 5 节和附录 B.2所示，对于满足假设 L_1, L_2 **1**，且有上界或在无穷远处呈二次增长的损失函数，我们有

$$
|m_{L_k}^{\alpha}[\rho_s^N]|^2 \le \int |\theta|^2 d\eta_{k,s}^{\alpha,N} \le c_1 + c_2\int |\theta|^2 d\rho_s^N,
\tag{56}
$$

对于 $k=1,2$ 以及适当的常数 c_1, c_2（与 N 无关），根据构造有 $m_{L_k}^{\alpha}[\rho_s^N] = \int \theta d\eta^{\alpha,k,N}_t$，其中 $\eta^{\alpha,k,N}_t = w_{L_k}^{\alpha}\rho_t^N / \|w_{L_k}^{\alpha}\|_{\mathbb{L}^1(\rho_t^N)}$。因此，我们进一步得到

$$
|m_{L_i}^{\alpha}[\rho_s^N]|^{2p} \le \big(c_1 + c_2\int |\theta|^2 d\rho_s^N\big)^p \le 2^{p-1}\big(c_1^p + c_2^p\int |\theta|^{2p}d\rho_s^N\big).
$$

将上述不等式代入（55），并应用 Gr¨onwall 不等式可得到一个常数 $K_p > 0$，与 N无关，使得 $\sup_{t\in[0,T]} \mathbb{E}\int |\theta|^{2p} d\rho_t^N \le K_p$成立，从而也有

$$\sup_{t\in[0,T]} \mathbb{E}|m_{L_i}^{\alpha}[\rho_t^N]|^{2p} \le 2^{p-1}(c_1^p + c_2^p K_p),$$

由此得出52式中的估计结果，方法是选择足够大的 K。另外两个估计结果则分别可由56式以及在¨式（54）上应用 Gronwall 不等式直接得到。□

Appendix B. Auxiliary Propositions and Lemma for Well-posedness of System

B.1. **损失函数有上界**。

引理 B.1. 设 L_1，L_2 满足假设 *1* 且 $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ 其中 $\int |\theta|^2 d\mu \le K$。则

$$\frac{e^{-\alpha \underline{L_1}}}{\|w_{L_1}^{\alpha}\|_{\mathbb{L}^1(\mu)}} \le \exp(\alpha C_{L_1}(1+K)) =: C_{K_1}, \qquad \frac{e^{-\alpha \underline{L_2}}}{\|w_{L_2}^{\alpha}\|_{\mathbb{L}^1(\mu)}} \le \exp(\alpha C_{L_2}(1+K)) =: C_{K_2}$$

证明。 证明与 [6, 引理 3.1] 中相同。 □

引理 B.2。 设 L_1，L_2 满足假设 *1* 且 μ，$\hat{}\mu \in \mathcal{P}_2(\mathbb{R}^d)$ 其中 $\int |\theta|^4 d\mu$，$\int |\hat{}\theta|^4 d\hat{}\mu \le K$。则以下稳定性估计成立：

$$|m_{L_1}^{\alpha}[\mu] - m_{L_1}^{\alpha}[\hat{\mu}]| \le c_{0,1} W_2(\mu,\hat{\mu}), \qquad |m_{L_2}^{\alpha}[\mu] - m_{L_2}^{\alpha}[\hat{\mu}]| \le c_{0,2} W_2(\mu,\hat{\mu}),$$

其中常数为 $c_{0,\ k} > 0$，且仅取决于 α、M_{L_k} 和 K。

证明。 证明与 [6, 引理 3.2] 中的证明相同。 □

B.2. **无穷远处具有二次增长的损失函数**。 本节将证明定理2.1在目标函数在无穷远处具有二次增长的情况下成立，即存在常数 $M > 0$和 $c_{q_k} > 0$，使得 $L_k(\theta) - L_k \ge c_{q_k}|\theta|^2$对所有 $|\theta| \ge M$均成立。

引理 B.3。 设 L_1，L_2 满足假设 *1* 且在无穷远处具有二次增长，并且 $\mu \in \mathcal{P}_2(\mathbb{R}^d)$。那么对于 $k = 1, 2$

$$\int |\theta|^2 d\eta_k^{\alpha} \le b_{k,1} + b_{k,2}\int |\theta|^2 d\mu, \qquad \eta_k^{\alpha} = \frac{w_{L_k}^{\alpha}\mu}{\|w_{L_k}^{\alpha}\|_{\mathbb{L}^1(\mu)}},$$

其中常数

$$b_{k,1} := M^2 + b_{k,2}, \qquad b_{k,2} = 2\frac{C_{L_k}}{C_{q_k}}\Big(1 + \frac{1}{\alpha C_{q_k} M^2}\Big).$$

证明。 证明与 [6, 引理 3.3] 中的证明相同。 □

定理 *2.1* 的证明。此处我们给出无穷远处二次增长情形的证明。由于步骤 1、2 和 4 保持不变，我们只展示步骤 3。

步骤 3：设 $(u^1,u^2) \in \mathcal{C}([0,T],\mathbb{R}^d) \times \mathcal{C}([0,T],\mathbb{R}^d)$ 满足 $(u^1,u^2) = \tau\mathcal{T}(u^1,u^2)$，其中 $\tau \in [0,1]$。特别地，存在 ν_1，$\nu_2 \in \mathcal{C}([0,T],\mathcal{P}_2(\mathbb{R}^d))$，分别满足 (26a) 和 (26b)，使得 $(u^1,u^2) = \tau(m_{L_1}^{\alpha}[\nu], m_{L_2}^{\alpha}[\nu])$，其中 $\nu = w_1\nu_1 + w_2\nu_2$。由引理 B.3 和詹森不等式，我们有

$$|u_t^1|^2 = \tau^2 |m_{L_1}^{\alpha}[\nu_t]|^2 \le \tau^2\big(b_{1,1} + b_{1,2}\int |\theta|^2 d\nu_t\big) \le \tau^2\big[b_{1,1} + b_{1,2}(w_1\int |\theta|^2 d\nu_t^1 + w_2\int |\theta|^2) d\nu_t^2\big]. \tag{57}$$

Th因此，与有界情形类似地计算二阶矩估计可得 s

$$\begin{aligned}\frac{d}{dt}\int |\theta|^2 d\nu_t^1 &\le (d\sigma_1^2 - 2\lambda_1 + |\gamma| + \lambda_2)\int |\theta|^2 d\nu_t^1 + (d\sigma_1^2 + |\gamma|)|u_t^1|^2 + (\lambda_2 + d\sigma_2^2)C_{\nabla L_1}^2 \\ &\le (d\sigma_1^2 + |\gamma| + \lambda_2)(1 + b_{1,2})\int |\theta|^2 d\nu_t^1 + (d\sigma_1^2 + |\gamma| + \lambda_2)b_{1,2}\int |\theta|^2 d\nu_t^2 \\ &\quad + (d\sigma_1^2 + |\gamma| + \lambda_2)b_{1,1} + (\lambda_2 + d\sigma_2^2)C_{\nabla L_1}^2.\end{aligned}$$

同样地，我们有

$$\begin{aligned}\frac{d}{dt}\int |\theta|^2 d\nu_t^2 &\le (d\sigma_1^2 + |\gamma| + \lambda_2)b_{2,2}\int |\theta|^2 d\nu_t^1 + (d\sigma_1^2 + |\gamma| + \lambda_2)(1 + b_{2,2})\int |\theta|^2 d\nu_t^2 \\ &\quad + (d\sigma_1^2 + |\gamma| + \lambda_2)b_{2,1} + (\lambda_2 + d\sigma_2^2)C_{\nabla L_2}^2.\end{aligned}$$

将上述两个不等式相加，得到

$$\frac{d}{dt}\Big(\int|\theta|^2d\nu_t^1+\int|\theta|^2d\nu_t^2\Big)\le C_1\Big(\int|\theta|^2d\nu_t^1+\int|\theta|^2d\nu_t^2\Big)+C_2.$$

于是由 Gr¨onwall 不等式可得

$$\int|\theta|^2d\nu_t^1+\int|\theta|^2d\nu_t^2\le\exp(C_1t)\Big(\int|\theta|^2d\nu_t^1+\int|\theta|^2d\nu_t^2\Big)+\frac{C_2}{C_1}\big(\exp(C_1t)-1\big).$$

因此，我们知道 $\|u^1\|_\infty$ 可通过（57）得到有界性。对于 $\|u^2\|_\infty$也有类似的有界性。然后，按照有界情形下 **步骤 3** 中的论证，我们完成证明。□

Appendix C. Auxiliary Lemma for Large-time behavior of the Mean-field Particle System

引理 *C.1*（方差演化）。对于 $k=1,2$ 令 $L_k:\ \mathbb{R}^d\to\mathbb{R}$ 并固定 α、λ_1、λ_2、σ_1、$\sigma_2>0$。此外，令 $T>0$ 和 $\rho^k\in\mathcal{C}$（$[0,T]$、$\mathcal{P}_4$（$\mathbb{R}$））分别为福克-普朗克方程（*8a*）和（*8b*）的弱解。则泛函 $\mathcal{V}$（ρ_t^k）满足

$$\begin{aligned}\frac{d}{dt}\mathcal{V}(\rho_t^k)\le&-\big(2\lambda_1-2\lambda_2M-d\sigma_1^2-d\sigma_2^2M^2\big)\mathcal{V}(\rho_t^k)\\&+\sqrt{2}(\lambda_1+d\sigma_1^2)|m_{L_k}^\alpha[\rho_t]-\theta_k^*|\sqrt{\mathcal{V}(\rho_t^k)}+\frac{d\sigma_1^2}{2}|m_{L_k}^\alpha[\rho_t^k]-\theta_k^*|^2,\end{aligned}$$

其中 $M:=$最大值$\{M_{\nabla L_1},M_{\nabla L_2}\}$。

证明。 由于 ρ^1 是福克-普朗克方程（8a）的弱解，因此 $\mathcal{V}(\rho_t^1)$ 的演化为

$$\begin{aligned}\frac{d}{dt}\mathcal{V}(\rho_t^1)&=\frac{1}{2}\frac{d}{dt}\int|\theta-\theta_1^*|^2d\rho_t^1(\theta)\\&=-\lambda_1\int(\theta-m_{L_1}^\alpha[\rho_t])\cdot(\theta-\theta_1^*)d\rho_t^1-\lambda_2\int\nabla L_1(\theta)\cdot(\theta-\theta_1^*)d\rho_t^1\\&\quad+\frac{d\sigma_1^2}{2}\int|\theta-m_{L_1}^\alpha[\rho_t]|^2d\rho_t^1+\frac{d\sigma_2^2}{2}\int|\nabla L_1(\theta)|^2d\rho_t^1\\&=:T_1+T_2+T_3+T_4.\end{aligned}$$

向右扩展　　通过在 T_1 的积分中对内积的左侧减去并加上 θ_1^* 得到。　　s

$$\begin{aligned}T_1&=-\lambda\int\|\theta-\theta_1^*\|_2^2d\rho_t^1(\theta)-\lambda\int(\theta_1^*-m_{L_1}^\alpha[\rho_t])\cdot(\theta-\theta_1^*)d\rho_t^1(\theta)\\&\le-2\lambda\mathcal{V}(\rho_t^1)+\lambda\|\mathbb{E}[\rho_t^1]-\theta_1^*\|_2\|m_{L_1}^\alpha[\rho_t]-\theta_1^*\|_2,\end{aligned}\tag{58}$$

其中最后一步是由于柯西–施瓦茨不等式。此外请注意，

$$\|\mathbb{E}[\rho_t^1]-\theta_1^*\|_2\le\int\|\theta-\theta_1^*\|_2d\rho_t^1(\theta)\le\sqrt{\int\|\theta-\theta_1^*\|_2^2d\rho_t^1(\theta)}=\sqrt{2\mathcal{V}(\rho_t^1)}.\tag{59}$$

因此，我们可以得出
固有推论。

$$T_1\le-2\lambda_1\mathcal{V}(\rho_t^1)+\lambda_1\sqrt{2\mathcal{V}(\rho_t)}|m_{L_1}^\alpha[\rho_t]-\theta_1^*|$$

对于术语 T_2，可以计算得到

$$T_2=-\lambda_2\int\nabla L_1(\theta)\cdot(\theta-\theta_1^*)d\rho_t^1\le\lambda_2M_{\nabla L_1}\int|\theta-\theta_1^*|^2d\rho_t^1=2\lambda_2M_{\nabla L_1}\mathcal{V}(\rho_t^1).$$

对于项 T_3，再次通过减去并加上 θ_1^*，我们得到

$$\begin{aligned}T_2&=\frac{d\sigma^2}{2}\int\|\theta-m_{L_1}^\alpha[\rho_t]\|_2^2d\rho_t^1(\theta)\\&=\frac{d\sigma^2}{2}\Big(\int\|\theta-\theta_1^*\|_2^2d\rho_t^1(\theta)-2\int(\theta-\theta_1^*)\cdot(m_{L_1}^\alpha[\rho_t]-\theta_1^*)d\rho_t^1(\theta)+\|m_{L_1}^\alpha[\rho_t]-\theta_1^*\|_2^2\Big)\\&\le d\sigma^2\Big(\mathcal{V}(\rho_t^1)+\|m_{L_1}^\alpha[\rho_t]-\theta_1^*\|_2\int\|\theta-\theta_1^*\|_2d\rho_t^1(v)+\frac{1}{2}\|m_{L_1}^\alpha[\rho_t]-\theta_1^*\|_2^2\Big),\end{aligned}\tag{60}$$

with 最后一步用到了柯西–施瓦茨不等式。对于项 T_4，可以计算

$$T_4 = \frac{d\sigma_2^2}{2}\int |\nabla L_1(\theta)|^2 d\rho_t^1(\theta) \le \frac{d\sigma_2^2}{2}\int M_{\nabla L_1}^2 |\theta-\theta_1^*|^2 d\rho_t^1(\theta) = d\sigma_2^2 M_{\nabla L_1}^2 \mathcal{V}(\rho_t^1).$$

因此，结合 T_1、T_2、T_3 和 T_4 的估计，我们得到

$$\begin{aligned}\frac{d}{dt}\mathcal{V}(\rho_t^1) \le &-\big(2\lambda_1 - 2\lambda_2 M_{\nabla L_1} - d\sigma_1^2 - d\sigma_2^2 M_{\nabla L_1}^2\big)\mathcal{V}(\rho_t^1)\\ &+\sqrt{2}(\lambda_1 + d\sigma_1^2)|m_{L_1}^\alpha[\rho_t]-\theta_1^*|\sqrt{\mathcal{V}(\rho_t^1)} + \frac{d\sigma_1^2}{2}|m_{L_1}^\alpha[\rho_t^1]-\theta_1^*|^2.\end{aligned}$$

计算 $\mathcal{V}(\rho_t^2)$ 的过程类似，因此我们得到如陈述中所述的上界 . □

引理 *C.2*（定量拉普拉斯原理）。对于 $k=1,2$，记 $L_k := \inf_{\theta\in\mathbb{R}^d} L_k(\theta)$。令 $\rho\in\mathcal{P}(\mathbb{R}^d)$ 并固定 $\alpha>0$。对于任意 $r>0$，定义 $L_r^k := \sup_{\theta\in B_r(\theta_k^*)} L_k(\theta)$。那么，在假设 *2* 下，对于任意 $r\in(0, \min\{R_0^1, R_0^2\}]$ 以及满足 $q_k + L_r^k - L_k \le L_\infty^k$ 的 $q_k>0$，我们有

$$\|m_{L_k}^\alpha[\rho]-\theta_k^*\| \le \frac{(q_k + L_r^k - \underline{L_k})^{\nu_k}}{\eta_k} + \frac{\exp(-\alpha q_k)}{\rho(B_r(\theta_k^*))}\int \|\theta-\theta_k^*\|_2 d\rho(\theta).$$

证明。与 [13] 中命题 21 的证明相同。 □

定义 C.1（磨光函数）。对于 $k=1,2$，$r>0$，我们定义磨光函数 $\phi_r^k:\ \mathbb{R}^d\to\mathbb{R}$ *by*

$$\phi_r^k(\theta) := \begin{cases}\exp\big(-\frac{r^2}{r^2-\|\theta-\theta_k^*\|_2^2}\big), & \text{if } \|\theta-\theta_k^*\|_2<r,\\ 0, & \text{else}\end{cases} \tag{61}$$

我们有 $\phi_t^k(\theta_k^*)=1$、$Im(\phi_t^k)=[0,1]$、$supp(\phi_r^k)=B_r(\theta_k^*)$ $\phi_r^k\in\mathcal{C}_c^\infty(\mathbb{R}^d)$ 以及

$$\begin{aligned}\nabla\phi_r^k(\theta) &= -2r^2\frac{\theta-\theta_k^*}{(r^2-|\theta-\theta_k^*|^2)^2}\phi_r^k(\theta),\\ \Delta\phi_r^k(\theta) &= 2r^2\left(\frac{2(2|\theta-\theta_k^*|^2-r^2)|\theta-\theta_k^*|^2 - d(r^2-|\theta-\theta_k^*|^2)^2}{(r^2-|\theta-\theta_k^*|^2)^4}\right)\phi_r^k(\theta).\end{aligned}$$

引理 *C.3*。对于 $k=1,2$，令 $T>0$、$r>0$，并固定参数 α、λ_1、λ_2、σ_1、$\sigma_2>0$。假设 ρ^1、$\rho^2\in\mathcal{C}([0,T],\mathcal{P}(\mathbb{R}^d))$ 分别是方程 *(8a)* 和 *(8b)* 的福克-普朗克方程的弱解，且具有初始条件 ρ_0^1、$\rho_0^2\in\mathcal{P}(\mathbb{R}^d)$。此外，记 $B_k := \sup_{t\in[0,T]}|m_{L_k}^\alpha[\rho_t]-\theta_k^*|$。则对所有 $t\in[0,T]$，有

$$\rho_t^k(B_r(\theta_k^*)) \ge \Big(\int \phi_r^k(\theta)d\rho_0^k(\theta)\Big)\exp\big(-(q_k^l+q_k^g)t\big),$$

with

$$q_k^l := \max\left\{\frac{2\lambda_1(\sqrt{c}r+B_k)\sqrt{c}}{(1-c)^2 r} + \frac{2\sigma_1^2(cr^2+B_k^2)(2c+d)}{(1-c)^4 r^2}, \frac{4\lambda_1^2}{(2c-1)\sigma_1^2}\right\}, \tag{62}$$

$$q_k^g := \max\left\{\frac{2\lambda_2 c M_{\nabla L_k}}{(1-c)^2} + \frac{\sigma_2^2 M_{\nabla L_k}^2 c(2c+d)}{(1-c)^4}, \frac{4\lambda_2^2}{(2c-1)\sigma_2^2}\right\}, \tag{63}$$

其中，$c\in(\frac{1}{2},1)$ 可以是满足该不等式的任意常数

$$(2c-1)c \ge d(1-c)^2. \tag{64}$$

证明。这里我们将证明 $\rho_t^k(B_r(\theta_1^*))$ 的情形，另一个情形的计算类似。根据定义 C.1 中磨光函数的性质，我们有 $0\le\phi_r^1(\theta)\le 1$ 且 $\text{supp}(\phi_r^1) = B_r(\theta_1^*)$。这意味着 $\rho_t^1(B_r(\theta_1^*)) = \rho_t^1\big(\{\theta\in\mathbb{R}^d:\ |\theta-\theta_1^*|\le r\}\big) \ge \int\phi_r^1(\theta)\, d\rho_t^1(\theta)$。类似于 [13, 32], 中的证明，我们将为该不等式的右侧推导一个下界。由于 ρ^1 是方程 (8a) 的弱解，且 $\phi_r^1\in\mathcal{C}_c^\infty(\mathbb{R}^d)$，我们有

$$\frac{d}{dt}\int\phi_r^1(\theta)d\rho_t^1(\theta) = \int\big(T_1(\theta)+T_2(\theta)+T_3(\theta)+T_4(\theta)\big)d\rho_t^1(\theta),$$

with

$$T_1(\theta) := -\lambda_1(\theta - m^{\alpha}_{L_1}[\rho_t]) \cdot \nabla\phi^1_r(\theta), \qquad T_2(\theta) := -\lambda_2 \nabla L_1(\theta) \cdot \nabla\phi^1_r(\theta),$$
$$T_3(\theta) := \frac{\sigma_1^2}{2}|\theta - m^{\alpha}_{L_1}[\rho_t]|^2 \Delta\phi^1_r(\theta), \qquad T_4(\theta) := \frac{\sigma_2^2}{2}|\nabla L_1(\theta)|^2 \Delta\phi^1_r(\theta).$$

根据 [13,32], 中的证明，我们知道

$$T_1(\theta) + T_3(\theta) \geq -q_1^l \phi^1_r(\theta) \qquad \text{for all } \theta \in \mathbb{R}^d, \tag{65}$$

其中

$$q_1^l := \max\left\{ \frac{2\lambda_1(\sqrt{c}r + B_1)\sqrt{c}}{(1-c)^2 r} + \frac{2\sigma_1^2(cr^2 + B_1^2)(2c+d)}{(1-c)^4 r^2}, \frac{4\lambda_1^2}{(2c-1)\sigma_1^2} \right\}.$$

现在我们要证明，对于所有 $\theta \in \mathbb{R}^d$ 和某些常数 $q_1^g > 0$， $T_2(\theta) + T_4(\theta) \geq -q_1^g \phi^1_r(\theta)$ 成立。由于磨光函数 ϕ^1_r 及其一阶和二阶导数在 $\Omega_r := \{\theta \in \mathbb{R}^d$ 外均为零， $|\theta - \theta_1^*| < r\}$，因此我们可以将注意力集中在开球 Ω_r 上。为了在 Ω_r 上获得下界，我们引入子集

$$K_1 := \big\{\theta \in \mathbb{R}^d : |\theta - \theta_1^*| > \sqrt{c}r\big\},$$
$$K_2 := \big\{\theta \in \mathbb{R}^d : -\lambda_2 \nabla L_1(\theta) \cdot (\theta - \theta_1^*)(r^2 - |\theta - \theta_1^*|^2)^2 > (2c-1)r^2 \frac{\sigma_2^2}{2} |\nabla L_1(\theta)|^2 |\theta - \theta_1^*|^2\big\},$$

其中 c 是满足（64）的常数。现在，我们根据Ω_r 对其进行分解

$$\Omega_r = (K_1^c \cap \Omega_r) \cup (K_1 \cap K_2^c \cap \Omega_r) \cup (K_1 \cap K_2 \cap \Omega_r).$$

下面我们分别处理这三个子集中的每一个。**子集** $K_1^c \cap \bar{\Omega}_r$**:** 在这个子集上，我们有 $|\theta - \theta_1^*| \leq \sqrt{c}r$，然后可以计算得到

$$\begin{aligned}
T_2(\theta) &= 2\lambda_2 r^2 \frac{\nabla L_1(\theta) \cdot (\theta - \theta_1^*)}{(r^2 - |\theta - \theta_1^*|^2)^2} \phi^1_r(\theta) \\
&\geq -2\lambda_2 r^2 \frac{M_{\nabla L_1}|\theta - \theta_1^*|^2}{(r^2 - |\theta - \theta_1^*|^2)^2} \phi^1_r(\theta) \\
&\geq -2\lambda_2 r^2 \frac{M_{\nabla L_1} c r^2}{(1-c)^2 r^4} \phi^1_r(\theta) \\
&= \frac{-2c\lambda_2 M_{\nabla L_1}}{(1-c)^2} \phi^1_r(\theta) =: -q^{g,1}\phi^1_r(\theta).
\end{aligned}$$

对于项 T_4，我们推导得出

$$\begin{aligned}
T_4(\theta) &= \frac{\sigma_2^2}{2}|\nabla L_1|^2 2r^2 \left(\frac{2(2|\theta - \theta_1^*|^2 - r^2)|\theta - \theta_1^*|^2 - d(r^2 - |\theta - \theta_1^*|^2)^2}{(r^2 - |\theta - \theta_1^*|^2)^4} \right) \phi^1_r(\theta) \\
&\leq -\sigma_2^2 M^2_{\nabla L_1} |\theta - \theta_1^*|^2 r^2 \left(\frac{2c(2c-1)r^4 - d(1-c)^2 r^4}{(1-c)^4 r^8} \right) \phi^1_r(\theta) \\
&\leq -\frac{\sigma_2^2 M^2_{\nabla L_1} c(2c+d)}{(1-c)^4} \phi^1_r(\theta) =: -q^{g,2}\phi^1_r(\theta).
\end{aligned}$$

子集 $K_1 \cap K_2^c \cap \bar{\Omega}_r$**:** 根据 K_1 和 K_2 的定义，我们有$|\theta - \theta_1^*| > \sqrt{c}r$ 和

$$-\lambda_2 \nabla L_1(\theta) \cdot (\theta - \theta_1^*)(r^2 - |\theta - \theta_1^*|^2)^2 \leq (2c-1)r^2 \frac{\sigma_2^2}{2} |\nabla L_1|^2 |\theta - \theta_1^*|^2,$$

分别。O 我们现在的目标是证明，对于该子集中的所有 θ，都有 $T_2(\theta) + T_4(\theta) \geq 0$。我们首先计算e

$$\begin{aligned}
\frac{T_2(\theta) + T_4(\theta)}{2r^2 \phi^1_r(\theta)} &= \frac{\lambda_2 \nabla L_1(\theta) \cdot (\theta - \theta_1^*)(r^2 - |\theta - \theta_1^*|^2)^2}{(r^2 - |\theta - \theta_1^*|^2)^4} \\
&\quad + \frac{\sigma_2^2}{2}|\nabla L_1|^2 \frac{2(2|\theta - \theta_1^*|^2 - r^2)|\theta - \theta_1^*|^2 - d(r^2 - |\theta - \theta_1^*|^2)^2}{(r^2 - |\theta - \theta_1^*|^2)^4}.
\end{aligned}$$

因此，只要我们能够证明 $T_2(\theta) + T_4(\theta) \geq 0$

$$\big(-\lambda_2 \nabla L_1(\theta) \cdot (\theta - \theta_1^*) + \frac{d\sigma_2^2}{2}|\nabla L_1|^2\big)(r^2 - |\theta - \theta_1^*|^2)^2 \leq \sigma_2^2 |\nabla L_1|^2 \big(2|\theta - \theta_1^*|^2 - r^2\big)|\theta - \theta_1^*|^2.$$

左侧第一项的上界为

$$\begin{aligned}-\lambda_2\nabla L_1(\theta)\cdot(\theta-\theta_1^*)(r^2-|\theta-\theta_1^*|^2)^2 &\le (2c-1)r^2\frac{\sigma_2^2}{2}|\nabla L_1(\theta)|^2|\theta-\theta_1^*|^2\\ &\le \frac{\sigma_2^2}{2}|\nabla L_1|^2(2|\theta-\theta_1^*|^2-r^2)|\theta-\theta_1^*|^2.\end{aligned}$$

对于左侧的第二项，我们可以使用 $d(1-c)^2 \le (2c-1)c$，得到

$$\begin{aligned}\frac{d\sigma_2^2}{2}|\nabla L_1|^2(r^2-|\theta-\theta_1^*|^2)^2 &\le \frac{d\sigma_2^2}{2}|\nabla L_1|^2(1-c)^2r^4\\ &\le \frac{\sigma_2^2}{2}|\nabla L_1|^2(2c-1)r^2cr^2\\ &\le \frac{\sigma_2^2}{2}|\nabla L_1|^2(2|\theta-\theta_1^*|^2-r^2)|\theta-\theta_1^*|^2.\end{aligned}$$

因此，在该子集上，我们有 $T_2(\theta)+T_4(\theta)\ge 0$ 一致成立。

子集 $K_1\cap K_2\cap\Omega_r$： 在该子集上，我们有 $|\theta-\theta_1^*|>\sqrt{c}r$，并且

$$-\lambda_2\nabla L_1(\theta)\cdot(\theta-\theta_1^*)(r^2-|\theta-\theta_1^*|^2)^2 > (2c-1)r^2\frac{\sigma_2^2}{2}|\nabla L_1(\theta)|^2|\theta-\theta_1^*|^2.$$

首先注意到，当 $\lambda_2^2|\nabla L_1(\theta)|^2=0$ 且满足 $\lambda_2>0$ 时，$T_2(\theta)=0$。另一方面，如果$\lambda_2^2|\nabla L_1(\theta)|^2>0$, 则可计算得

$$\begin{aligned}T_2(\theta) &= 2\lambda_2r^2\frac{\nabla L_1(\theta)\cdot(\theta-\theta_1^*)}{(r^2-|\theta-\theta_1^*|^2)^2}\phi_r^1(\theta)\\ &\ge 2\lambda_2r^2\frac{-\lambda_2(\nabla L_1(\theta)\cdot(\theta-\theta_1^*))^2}{(2c-1)r^2\frac{\sigma_2^2}{2}|\nabla L_1(\theta)|^2|\theta-\theta_1^*|^2}\phi_r^1(\theta)\\ &\ge -2\lambda_2^2r^2\frac{|\nabla L_1(\theta)|^2|\theta-\theta_1^*|^2}{(2c-1)r^2\frac{\sigma_2^2}{2}|\nabla L_1(\theta)|^2|\theta-\theta_1^*|^2}\phi_r^1(\theta)\\ &= -\frac{4\lambda_2^2}{(2c-1)\sigma_2^2}\phi_r^1(\theta) =: -q^{g,3}\phi_r^1(\theta).\end{aligned}$$

对于项 T_4，我们可以推得

$$\begin{aligned}T_4(\theta) &= \frac{\sigma_2^2}{2}|\nabla L_1|^2r^2\left(\frac{2(2|\theta-\theta_1^*|^2-r^2)|\theta-\theta_1^*|^2-d(r^2-|\theta-\theta_1^*|^2)^2}{(r^2-|\theta-\theta_1^*|^2)^4}\right)\phi_r^1(\theta)\\ &\ge \frac{\sigma_2^2}{2}|\nabla L_1|^2r^2\left(\frac{2(2c-1)r^2cr^2-d(1-c)^2r^4}{(r^2-|\theta-\theta_1^*|^2)^4}\right)\phi_r^1(\theta)\\ &= \frac{\sigma_2^2}{2}|\nabla L_1|^2r^2\left(\frac{[2(2c-1)c-d(1-c)^2]r^4}{(r^2-|\theta-\theta_1^*|^2)^4}\right)\phi_r^1(\theta)\ge 0,\end{aligned}$$

只要 c 满足 $2(2c-1)c\ge d(1-c)^2$。

证明的总结： 根据上述计算，可以得到

$$\begin{aligned}\int\big(T_2(\theta)+T_4(\theta)\big)d\rho_t^1(\theta) &= \int_{K_1^c\cap\Omega_r}\underbrace{\big(T_2(\theta)+T_4(\theta)\big)}_{\ge-(q^{g,1}+q^{g,2})\phi_r^1(\theta)}d\rho_t^1(\theta)+\int_{K_1\cap K_2^c\cap\Omega_r}\underbrace{\big(T_2(\theta)+T_4(\theta)\big)}_{\ge 0}d\rho_t^1(\theta)\\ &\quad+\int_{K_1\cap K_2\cap\Omega_r}\underbrace{\big(T_2(\theta)+T_4(\theta)\big)}_{-q^{g,3}\phi_r^1(\theta)}d\rho_t^1(\theta)\\ &\ge\int -q_1^g\phi_r^1(\theta)d\rho_t^1(\theta),\end{aligned}$$

其中

$$q_1^g := \max\left\{q^{g,1}+q^{g,2},q^{g,3}\right\} = \max\left\{\frac{2\lambda_2cM_{\nabla L_1}}{(1-c)^2}+\frac{\sigma_2^2M_{\nabla L_1}^2c(2c+d)}{(1-c)^4},\frac{4\lambda_2^2}{(2c-1)\sigma_2^2}\right\}.$$

将上述估计与（65）结合，可得

$$\frac{d}{dt}\int \phi_r^1(\theta)d\rho_t^1(\theta) \geq -(q_1^l + q_1^g)\int \phi_r^1(\theta)d\rho_t^1(\theta).$$

应用 Gr¨onwall 不等式并将两边乘以 (-1) 得

$$\int \phi_r^1(\theta)d\rho_t^1(\theta) \geq \left(\int \phi_r^1(\theta)d\rho_0^1(\theta)\right)\exp\big(-(q_1^l + q_1^g)t\big)$$

因此，我们得到

$$\rho_t^1(B_r(\theta_1^*)) \geq \left(\int \phi_r^1(\theta)d\rho_0^1(\theta)\right)\exp\big(-(q_1^l + q_1^g)t\big).$$

□

Appendix D. Auxiliary Lemmas for Mean-field Limit

引理 D.1。设 $\{X^n\}_{n\in\mathbb{N}}$ 是定义在概率空间 $(\Omega\ \mathcal{F}、\ \mathbb{P})$ 上且取值于 $\mathcal{C}([0,T]、\ \mathbb{R}^d)$ 中的随机变量序列。若以下两个条件成立，则 $\{X^n\}_{n\in\mathbb{N}}$ 的概率分布序列 $\{\mu_{X^n}\}_{n\in\mathbb{N}}$ 在 $\mathcal{C}([0,T]、\ \mathbb{R}^d)$ 上是紧的。

- *(*条件（*Con1*）*)* 对所有 $t\in[0,T]$，X_t^n 的分布集合（记为 $\{\mu_{X_t^n}\}_{n\in\mathbb{N}}$）作为定义在 $\mathbb{R}^d$ 上的概率测度序列是紧的。
- *(*条件（*Con2*）*)* 对所有 $\varepsilon>0$、$\eta>0$，存在 $\delta_0>0$ 和 $n_0\in\mathbb{N}$，使得对于所有 $n\geq n_0$ 以及满足 $0\leq\beta+\delta_0\leq T$ 的离散值 $\sigma(X_s^n;\ s\in[0,T])$-停时 β，均有

$$\sup_{\delta\in[0,\delta_0]}\mathbb{P}\big(|X_{\beta+\delta_0}^n - X_\beta^n|\geq\eta\big)\leq\varepsilon.$$

下面我们将详细说明如何在第 5.2 节中应用 Skorokhod 引理。

注 *10.* 令 $B := \{B^{1,i}, B^{2,j}\}_{i,j=1}^{\infty}$ 表示分别出现在描述类别*1*和类别*2*粒子演化的随机微分方程中的所有布朗运动的集合。此外，令 $(\theta^{1,i}, B^{1,i})$ 和 $(\theta^{2,j}, B^{2,j})$ 分别为 $i=1,\ldots,N_1$ 和 $j=1,\ldots,N_2$ 对应的随机微分方程 *(4a)* 和 *(4b)* 的解。那么，根据经典随机微分方程存在普适函数解这一结论 *[20]*，我们知道存在（确定性的）泛函 $F_N^k:\ B\to F_N^k(B)\in\mathcal{P}(\mathcal{C}([0,T],\mathbb{R}^d))$，使得 $\rho^{k,N} := F_N^k(B)$ 对 $k=1,2$ 成立。这仅仅说明，经验测度 $\rho^{k,N} := \frac{1}{N_k}\sum_{i=1}^{N_k}\delta_{X^{k,i}}$ 是 B 的确定性泛函。在第 *5.2* 节应用 *Skorohod* 引理时，我们应将其应用于随机变量集合 $\{((\rho^{1,N}, B), (\rho^{2,N}, B))\}_N$，该集合可写为 $\{((F_N^1(B)\ B), (F_N^2(B)\ B))\}_N$。由于 F_N^1 和 F_N^2 是确定性的，我们推断，由 *Skorohod* 引理给出的新概率空间中 $\rho^{k,N}$ 的版本，可以表示为式 *(5)* 的形式，其中布朗运动取自该新概率空间。

引理 D.2. 假设 $\mu^1,\ \mu^2\in\mathcal{C}([0,T],\ \mathbb{R}^d\times\mathbb{R}^d)$ 是偏微分方程 (42) 的两个弱解，且具有相同的初始数据 μ_0。则有

$$\sup_{t\in[0,T]} W_2(\mu_t^1,\mu_t^2) = 0,$$

其中，W_2 是2-*Wasserstein* 距离。

证明。我们构造两个满足以下条件的线性过程 $(\widehat{\theta}^{1,\cdot,k}_t,\ \widehat{\theta}^{2,\cdot,k}_t)_{t\in[0,T]},\ k=1,\ 2$

$$\begin{aligned}
d\begin{pmatrix}\widehat{\theta}_t^{1,k}\\ \widehat{\theta}_t^{2,k}\end{pmatrix} = &-\begin{pmatrix}\lambda_1(\widehat{\theta}_t^{1,k} - m_{L_1}^\alpha[\rho_t^k]) + \lambda_2\nabla L_1(\widehat{\theta}_t^{1,k})\\ \lambda_1(\widehat{\theta}_t^{1,k} - m_{L_2}^\alpha[\rho_t^k]) + \lambda_2\nabla L_2(\widehat{\theta}_t^{2,k})\end{pmatrix}dt\\
&+\sigma_1\begin{pmatrix}\mathrm{diag}(|\widehat{\theta}_t^{1,k} - m_{L_1}^\alpha[\rho_t^k]|) & 0\\ 0 & \mathrm{diag}(|\widehat{\theta}_t^{2,k} - m_{L_2}^\alpha[\rho_t^k]|)\end{pmatrix}\begin{pmatrix}dB_t^{1,k}\\ dB_t^{2,k}\end{pmatrix}\\
&+\sigma_2\begin{pmatrix}\mathrm{diag}(|\nabla L_1(\widehat{\theta}_t^{1,k})|) & 0\\ 0 & \mathrm{diag}(|\nabla L_2(\widehat{\theta}_t^{2,k})|)\end{pmatrix}\begin{pmatrix}d\widetilde{B}_t^{1,k}\\ d\widetilde{B}_t^{2,k}\end{pmatrix},
\end{aligned}$$

它们具有共同的初始数据 $(\widehat{\theta}_0^1,\ \widehat{\theta}_0^2)$，其分布为 μ_0。上述过程是线性的，因为$\mu_t^k=(\rho^{1,\cdot}_{t}{}^k,\ \rho^{2,\cdot}_{t}{}^k)$ 和 $\rho_t^k=w_1\rho^{1,\cdot}_{t}{}^k+w_2\rho^{2,\cdot}_{t}{}^k$是预先给定的。记 $\widehat{\mu}_t^k:=$分布$(\widehat{\theta}^{1,\cdot}_{t}{}^k,\ \widehat{\theta}^{2,\cdot}_{t}{}^k)$ $k=1,\ 2$，它们是以下线性偏微分方程的弱解

$$\begin{aligned}\partial_t\widehat{\mu}_t^k(\theta)=&\nabla\cdot\left(\big(\lambda_1(\theta_1-m_{L_1}^{\alpha}[\rho_t^k])+\lambda_2\nabla L_1(\theta_1),\lambda_1(\theta_2-m_{L_2}^{\alpha}[\rho_t^k])+\lambda_2\nabla L_2(\theta_2)\big)\widehat{\mu}_t^k\right)\\&+\Delta_{\theta_1}\left(\big(\frac{\sigma_1^2}{2}|\theta_1-m_{L_1}^{\alpha}[\rho_t^k]|^2+\frac{\sigma_2^2}{2}|\nabla L_1(\theta_1)|^2\big)\widehat{\mu}_t^k\right)+\Delta_{\theta_2}\left(\big(\frac{\sigma_1^2}{2}|\theta_2-m_{L_2}^{\alpha}[\rho_t^k]|^2+\frac{\sigma_2^2}{2}|\nabla L_2(\theta_2)|^2\big)\widehat{\mu}_t^k\right),\end{aligned}$$

其中 $\theta=(\theta_1,\ \theta_2)\in\mathbb{R}^d\times\mathbb{R}^d$。根据上述线性偏微分方程弱解的唯一性（见定理D.3），并且由于 μ^k 也是上述偏微分方程的弱解，因此可得 $\widehat{\mu}_t^k=\mu_t^k$ 对于 $k=1,\ 2$。因此，过程 $\big(\widehat{\theta}^{1,\cdot}_{t}{}^k,\ \widehat{\theta}^{2,\cdot}_{t}{}^k\big)_{t\in[0,T]}=\big(\theta^{1,\cdot}_{t}{}^k,\ \theta^{2,\cdot}_{t}{}^k\big)_{t\in[0,T]}$是非线性随机微分方程（41）的解，其唯一性已由定理 2.1得到。特别地，有

$$\sup_{t\in[0,T]}\mathbb{E}\big[|\overline{\theta}_t^{1,1}-\overline{\theta}_t^{1,2}|^2+|\overline{\theta}_t^{2,1}-\overline{\theta}_t^{2,2}|^2\big]=0,$$

根据 Wasserstein 距离的定义，这意味着

$$\begin{aligned}\sup_{t\in[0,T]}W_2(\mu_t^1,\mu_t^2)=&\sup_{t\in[0,T]}W_2(\widehat{\mu}_t^1,\widehat{\mu}_t^2)\\\le&\sup_{t\in[0,T]}\mathbb{E}\big[|\widehat{\theta}_t^{1,1}-\widehat{\theta}_t^{1,2}|^2+|\widehat{\theta}_t^{2,1}-\widehat{\theta}_t^{2,2}|^2\big]\\=&\sup_{t\in[0,T]}\mathbb{E}\big[|\overline{\theta}_t^{1,1}-\overline{\theta}_t^{1,2}|^2+|\overline{\theta}_t^{2,1}-\overline{\theta}_t^{2,2}|^2\big]=0.\end{aligned}$$

因此，唯一性得以证明。 □

定理 D.3（线性偏微分方程的存在唯一性）。对于任意 $T>0$，令 $b^1,\ b^2\in\mathcal{C}([0,T],\ \mathbb{R}^d)$ 和$\mu_0\in\mathcal{P}_2(\mathbb{R}^d)\times\mathcal{P}_2(\mathbb{R}^d)$。则下列线性偏微分方程

$$\begin{aligned}\partial_t\mu_t=&-\nabla\cdot\left(\big(\lambda_1(x_1-b_t^1)+\lambda_2\nabla L_1(x_1),\lambda_1(x_2-b_t^2)+\lambda_2\nabla L_2(x_2)\big)\mu_t\right)\\&+\Delta_{x_1}\left(\big(\frac{\sigma_1^2}{2}|x_1-b_t^1|^2+\frac{\sigma_2^2}{2}|\nabla L_1(x_1)|^2\big)\mu_t\right)+\Delta_{x_2}\left(\big(\frac{\sigma_1^2}{2}|x_2-b_t^2|^2+\frac{\sigma_2^2}{2}|\nabla L_2(x_2)|^2\big)\mu_t\right),\end{aligned}\tag{66}$$

具有唯一的弱解 $\mu\in\mathcal{C}\left([0,T],\mathcal{P}_2\left(\mathbb{R}^d\right)\times\mathcal{P}_2\left(\mathbb{R}^d\right)\right.$。

证明概要。存在性显然，它可以通过将相关线性随机微分方程的解的分布作为其解来获得。为证明唯一性，我们可以采用对偶论证。对于每个 $t_0\in(0,\ T]$ 以及函数 $\varphi\in\mathcal{C}_c^{\infty}(\mathbb{R}^d\times\mathbb{R}^d)$，我们考虑如下向后偏微分方程

$$\begin{aligned}\partial_t h_t=&-\big(\lambda_1(x_1-b_t^1)+\lambda_2\nabla L_1(x_1),\lambda_1(x_2-b_t^2)+\lambda_2\nabla L_2(x_2)\big)\cdot\nabla h_t\\&-\big[(\frac{\sigma_1^2}{2}|x_1-b_t^1|^2+\frac{\sigma_2^2}{2}|\nabla L_1(x_1)|^2)\Delta_{x_1}h_t+(\frac{\sigma_1^2}{2}|x_2-b_t^2|^2+\frac{\sigma_2^2}{2}|\nabla L_2(x_2)|^2)\Delta_{x_2}h_t\big],\end{aligned}\tag{67}$$

其中 $(t,\ x_1,\ x_2)\in[0,t_0]\times\mathbb{R}^d\times\mathbb{R}^d$；$h_{t_0}=\varphi$。它存在一个经典解 $h\in\mathcal{C}^1([0,t_0],\ \mathcal{C}^2(\mathbb{R}^d\times\mathbb{R}^d))$。事实上，根据 Kolmogorov 向后方程，我们可以显式构造一个解

$$h_t(x_1,x_2)=\mathbb{E}\big[\varphi(X_{t_0}^{1,t,x_1},X_{t_0}^{2,t,x_2})\big]\qquad t\in[0,t_0],$$

其中 $(X^{1,\,t,\,s\quad x_1},\ X^{2,\,t,\,s\quad x_2})_{0\le t\le s\le t_0}$ 是以下随机微分方程的强解

$$
\begin{aligned}
d\begin{pmatrix} X_s^{1,t,x_1} \\ X_s^{2,t,x_2} \end{pmatrix} &= \begin{pmatrix} \lambda_1(X_s^{1,t,x_1}-b_s^1)+\lambda_2\nabla L_1(X_s^{1,t,x_1}) \\ \lambda_1(X_s^{2,t,x_2}-b_s^2)+\lambda_2\nabla L_2(X_s^{2,t,x_2}) \end{pmatrix} ds \\
&\quad + \sigma_1 \begin{pmatrix} \mathrm{diag}(|X_s^{1,t,x_1}-b_s^1|) & 0 \\ 0 & \mathrm{diag}(|X_s^{2,t,x_2}-b_s^2|) \end{pmatrix} \begin{pmatrix} dB_s^1 \\ dB_s^2 \end{pmatrix} \\
&\quad + \sigma_2 \begin{pmatrix} \mathrm{diag}(|\nabla L_1(X_s^{1,t,x_1})|) & 0 \\ 0 & \mathrm{diag}(|\nabla L_2(X_s^{2,t,x_2})|) \end{pmatrix} \begin{pmatrix} d\widetilde{B}_s^1 \\ d\widetilde{B}_s^2 \end{pmatrix}, \\
\begin{pmatrix} X_t^{1,t,x_1} \\ X_t^{2,t,x_2} \end{pmatrix} &= \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},
\end{aligned}
$$

其中 B^1，B^2，$\widetilde{B}^1$ 和 $\widetilde{B}^2$ 是相互独立的 d 维布朗运动。

假设 μ^1 和 μ^2 是方程（66）的两个弱解，且具有相同的初始条件 $\mu_0^1=\mu_0^2$。记$\delta_\mu:=\mu^1-\mu^2$。将上述定义的向后偏微分方程（67）的解 h 作为测试函数，可得

$$
\begin{aligned}
&\big\langle h_{t_0}(x_1,x_2),\delta_{\mu_{t_0}}(dx_1,dx_2)\big\rangle \\
&=\int_0^{t_0}\big\langle \partial_s h_s(x_1,x_2),\delta_{\mu_s}(dx_1,dx_2)\big\rangle ds \\
&\quad+\int_0^{t_0}\big\langle \big(\lambda_1(x_1-b_s^1)+\lambda_2\nabla L_1(x_1),\lambda_1(x_2-b_s^2)+\lambda_2\nabla L_2(x_2)\big)\cdot\nabla h_s(x_1,x_2),\delta_{\mu_s}(dx_1,dx_2)\big\rangle ds \\
&\quad+\int_0^{t_0}\big\langle (\frac{\sigma_1^2}{2}|x_1-b_s^1|^2+\frac{\sigma_2^2}{2}|\nabla L_1(x_1)|^2)\Delta_{x_1}h_s(x_1,x_2) \\
&\qquad\quad+(\frac{\sigma_1^2}{2}|x_2-b_s^2|^2+\frac{\sigma_2^2}{2}|\nabla L_2(x_2)|^2)\Delta_{x_2}h_s(x_1,x_2),\delta_{\mu_s}(dx_1,dx_2)\big\rangle ds \\
&=\int_0^{t_0}\big\langle \partial_s h_s(x_1,x_2),\delta_{\mu_s}(dx_1,dx_2)\big\rangle ds+\int_0^{t_0}\big\langle -\partial_s h_s(x_1,x_2),\delta_{\mu_s}(dx_1,dx_2)\big\rangle ds \\
&=0,
\end{aligned}
$$

由此得到 $\int_{\mathbb{R}^d\times\mathbb{R}^d}\varphi(x_1,\ x_2)\delta_{\mu_{t_0}}(dx_1,\ dx_2)=0$，对任意 $\varphi\in\mathcal{C}_c^\infty(\mathbb{R}^d\times\mathbb{R}^d)$ 均成立。这意味着 $\delta_{\mu_{t_0}}=0$，由 t_0 的任意性可得唯一性。□

Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK.
Email address: carrillo@maths.ox.ac.uk

Department of Statistics, University of Wisconsin-Madison, 1300 University Avenue, Madison, Wisconsin 53706, USA.*Email address*: garciatrillo@wisc
.edu

Department of Statistics, University of Wisconsin-Madison, 1300 University Avenue, Madison, Wisconsin 53706, USA.*Email address*: sli739@wisc.edu

Department of Mathematics, Halıcıoğlu Data Science Institute, University of California, San Diego, La Jolla, California 92093, USA*Email address*: yuz244@ucsd.edu