廣東工業大學

QG 中期考核详细报告书

题　　目　《Differentially private average consensus: Obstructions, trade-offs, and optimal algorithm design》

学　　院　计算机学院

专　　业　计算机类

年级班别　24 7 班

学　　号　3124004280

学生姓名　何晨洋

2025 年 4 月 3 日

目录

1、文摘

本文献主要讲述了在对数据进行充分保护的情况下，多智能体系统的平均共识问题，即实现了保护隐私与智能体状态对初始状态平均值的共识。主要贡献有：

1. 理论结果：证明了在保护差分隐私的前提下智能体无法实现（0，0）精确收敛到初始平均值.
2. 算法设计：提出了基于拉普拉斯算法的新的差分隐私共识算法。该算法通过线性状态变化和消息生成函数中加入一个呈指数衰减的拉普拉斯噪声用于保护隐私。并证明了该算法几乎必然收敛到智能体初始状态平均值的无偏估计.
3. 性能优化：分析了算法中参数的最优选择，保证了收敛点的方差最小值，并将其与其他方法进行比较，证明最有设计应是对初始状态的一次性扰动.
4. 模拟验证：通过模拟实验验证了理论结果，展示了算法在不同隐私保护水平下的性能表现.

2、 文献简介

2.1 文献研究的核心与面临的问题

Differentially private average consensus: Obstructions, trade-offs, and optimal algorithm design

Erfan Nozari, Pavankumar Tallapragada, Jorge Cortés

文献基于多智能体系统的差分隐私保护下的平均共识问题开展研究，即如何在保护每个智能体初始状态的数据隐私的前提下，使得智能体们能达成对其初始状态平均值的共识。

在解决该问题过程中，需要解决隐私与共识的矛盾。基于差分隐私的要求，智能体的初

始状态信息不能被准确推断，这恰好与达成平均共识的目标，即精确收敛到初始状态平均值存在矛盾。而就此矛盾而建立的算法需要同时兼顾二者，即需要能够充分保护智能体初始状态的隐私，又要保证模型最终收敛到接近初始状态平均值，即初始状态平均值的无偏估计。

提出算法过后还需建立实验来验证算法的性能足够优秀.

3、 文献提出的理论及其核心算法介绍

3.1 基本理论与方法

差分隐私:差分隐私是一种隐私保护技术,通过在数据分析过程中人为设置可控的噪声，用于平衡数据的可用性与个体隐私之间的关系。它是一种数学框架，在确保数据集中个人的隐私同时允许对数据进行分析而不泄露任何的个人敏感信息。通常来说，满足（1）的算法可认为满足差分隐私，其中 ε 为隐私预算.

$$\boldsymbol{P}\left[\boldsymbol{x}_{\boldsymbol{\theta}_0^{(1)}} \in \boldsymbol{\mathcal{O}}\right] \leq \boldsymbol{e}^{\boldsymbol{\epsilon}} \boldsymbol{P}\left[\boldsymbol{x}_{\boldsymbol{\theta}_0^{(2)}} \in \boldsymbol{\mathcal{O}}\right] \qquad (1)$$

拉普拉斯噪声机制：一种常用的差分隐私保护方法，其概率密度函数为（2），其中 b 是噪声的尺度参数，和 ε 相关。通过向数据中添加拉普拉斯噪声可以有效保护隐私.

$$\mathbf{Lap}(\boldsymbol{b}) = \frac{\mathbf{1}}{\mathbf{2}\boldsymbol{b}} \mathbf{exp}\left(-\frac{|\boldsymbol{x}|}{\boldsymbol{b}}\right) \qquad (2)$$

3.2 核心算法与公式

状态更新方程：

$$\boldsymbol{\theta}(\boldsymbol{k}+\mathbf{1}) = \boldsymbol{\theta}(\boldsymbol{k}) - \boldsymbol{hLX}(\boldsymbol{k}) + \boldsymbol{S\eta}(\boldsymbol{k}) \qquad (3)$$

其中 θ（k）是智能体在时间 k 的状态，h 是学习率，L 是损失函数的梯度，表示边权（各个智能体之间的连系关系），X（k）是时间 k 智能体传输的消息，S 是噪声缩放因子，η（k）表示第 k 次迭代添加的噪声.

噪声生成算法：

$$\boldsymbol{\eta_i(k) \sim \mathrm{lap}\left(c_i q_i^k\right)} \qquad (4)$$

其中 k 是迭代次数，c_i 表示第 i 个样本的敏感度，q_i 表示第 i 个样本的噪声比例.

收敛点方差公式.

$$\boldsymbol{\mathrm{var}\left(\theta_\infty\right) = \frac{2}{n^2}\sum_{i=1}^{n}\frac{s_i^2 c_i^2}{1-q_i^2}} \qquad (5)$$

其中 θ 表示参数值，无穷表示 θ 迭代无穷多次后，var 表示方差，s_i 表示第 i 个样本的缩放因子.

收敛点表达公式：

$$\boldsymbol{\theta_\infty = \mathrm{Ave}(\theta_0) + \frac{1}{n}\sum_{i=1}^{n} s_i \sum_{j=0}^{\infty}\eta_i(j)} \qquad (6)$$

ave 表示均值，n 表示样本数量，`η_i(j)`表示第 i 个样本的噪声.

参数设置公式：

$$\boldsymbol{s_i = 1, \quad q_i = \alpha + (1-\alpha)|s_i - 1|} \qquad (7)$$

α 是一个属于[0,1]中的参数，用于调整噪声比例 q_i.

收敛时间公式：

$$\boldsymbol{\left|\left|\theta_k\text{-}\theta_\infty\right|\right| < \mathrm{tol}} \qquad (8)$$

小于号左边表示第 k 次迭代的参数值与最终收敛值的距离，tol 表示容忍度，即认为算法收敛的阈值.

局部目标公式：

$$\boldsymbol{\phi(\alpha, s) = \frac{s^2(\alpha+(1-\alpha)|s-1|)^2}{\alpha^2(1-|s-1|)^2\left[1-(\alpha+(1-\alpha)|s-1|)^2\right]}} \qquad (9)$$

s 是噪声-状态增益参数，α 是与噪声衰减率 q 相关的参数，且 α=（1−｜s−1｜）/（q−｜s−1｜），q 是噪声衰减率，且 q∈(｜s−1｜,1).

4、复现实验结果与流程

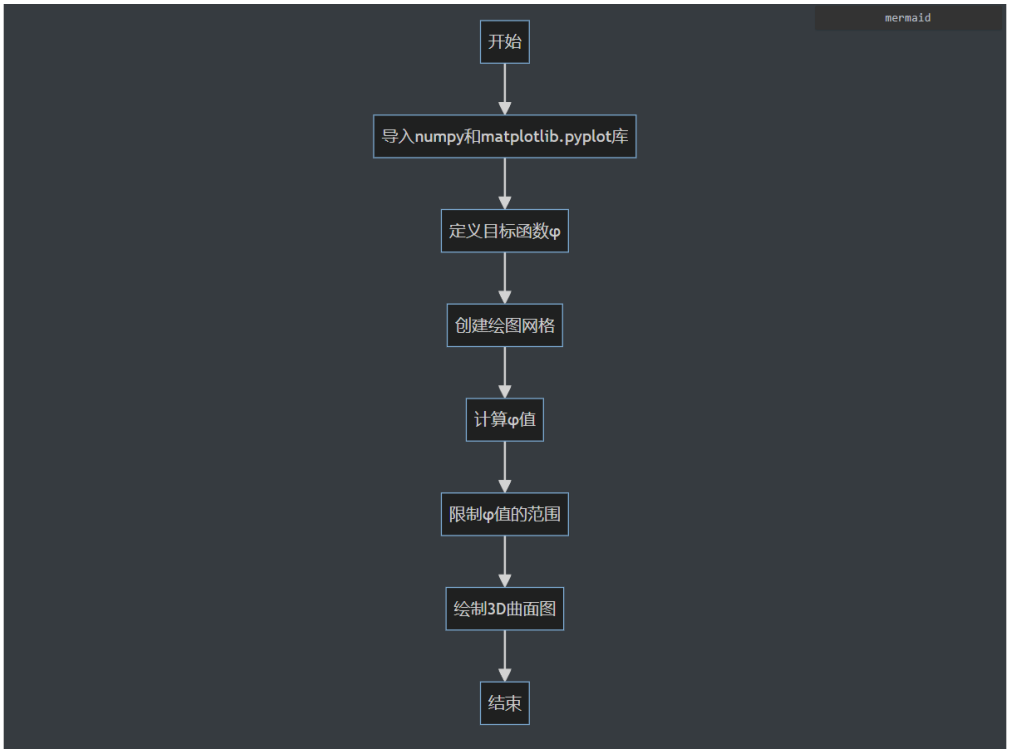
4.1 公式与参数设置

图 1 的复现运用了公式（9），关键公式如图

```
def phi(alpha,s):  1个用法
    numerator=s**2*(alpha+(1-alpha)*np.abs(s-1))**2
    denominator=alpha**2*(1-np.abs(s-1))**2*(1-(alpha+(1-alpha)*np.abs(s-1))**2)#公式33上
    return numerator/denominator
```

参数设置为 α 是在(0,2)之间均匀取 100 个点，s 在(0,2)均匀取 100 个点，对应计算后取得的值进行统一绘图

图 3 的实验流程图如下



最终实验图如下

Fig 1

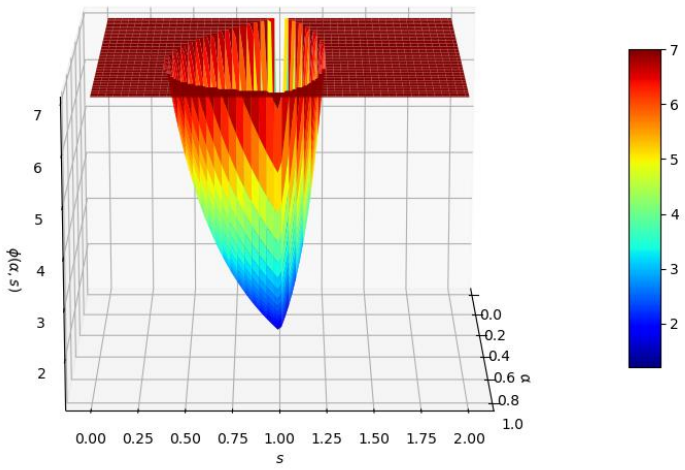
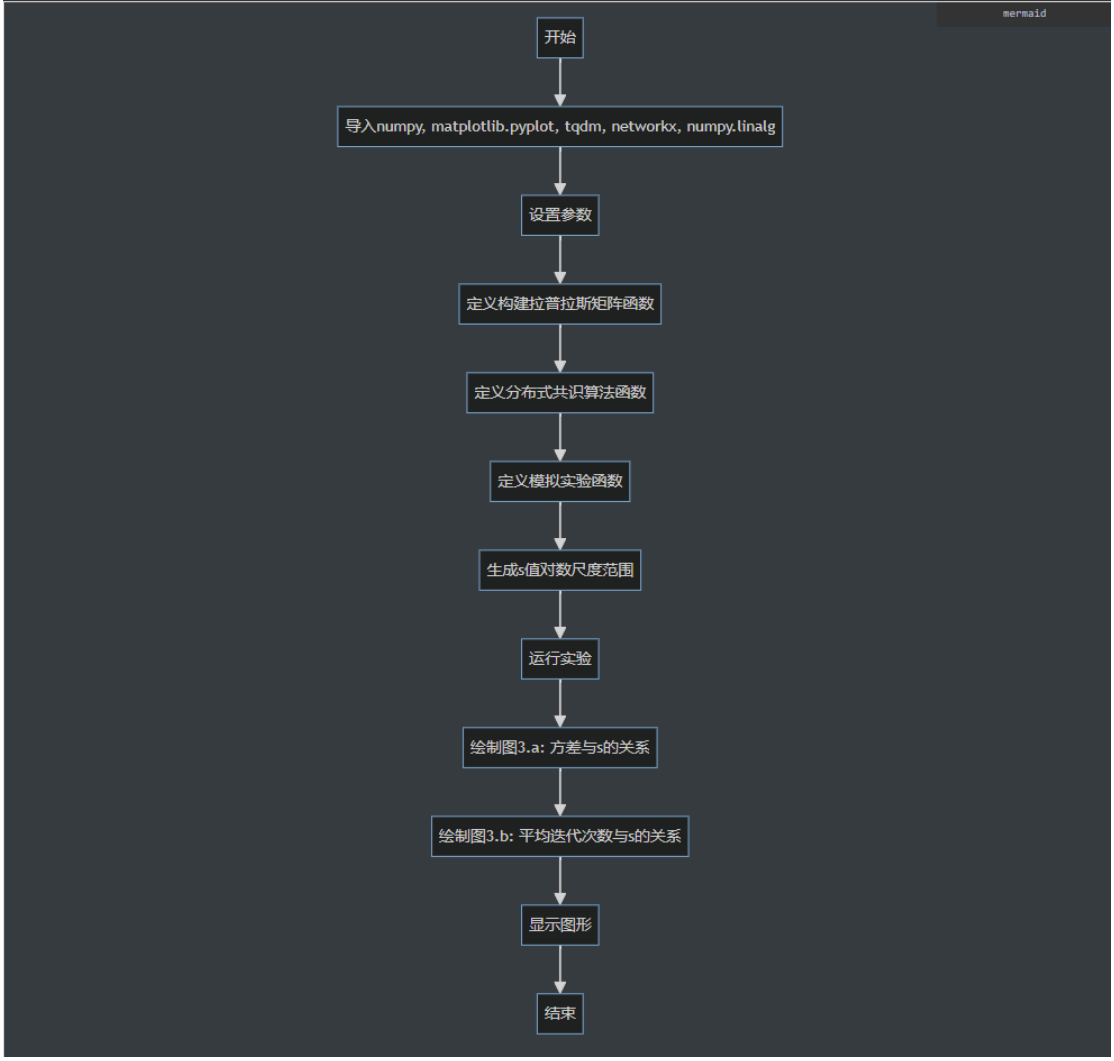


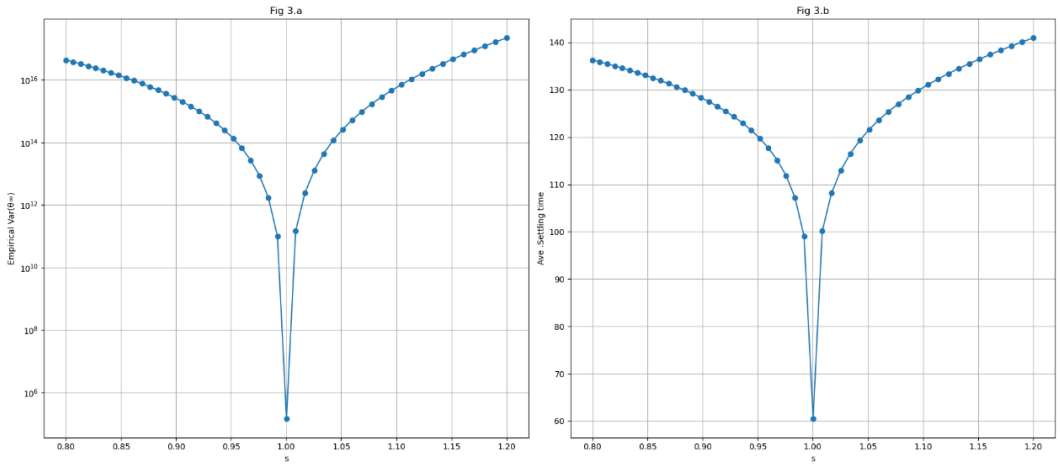
图 e3 的复现用了公式（5），关键公式如图

```
var=(2*(delta**2)/(n**2))*np.sum((s**2)*(q_i**2)/((epsilon**2)*((q_i-abs(s-1))**2)*(1-q_i**2)))#公式30
```

参数设置是 δ=1， ε=0.1，tol=1e-2，模拟次数为 10000，s 为(0.8,1.2)按对数取 50 个数，流程图如下



最终实现图如下



启示与思考

对本文献阅读过后，我不仅在相关知识方面有了更深刻的理解，同时在研究与论文撰写方面有了新的认知，例如隐私保护的重要性：在解决多智能体问题的时候，必须将数据的隐私纳入考虑范围，防止攻击者通过最终输出推断出初始状态从而对系统进行攻击，但同时不能把

数据的可利用率降低。而差分隐私为其提供了一个比较优秀的选项。而在算法设计的过程中，需要对关键矛盾进行平衡，例如本文中就隐私保护程度与算法性能进行了权衡，并就最优结果提出了相对应的公式与参数设置，当然，这个参数设置需要根据实际应用自行调整。在寻找最优参数的过程中，需要灵活运用绘图、理论进行相互对应，最终找到对整体最合适的参数设置，并通过可视化等方法证明。

在阅读的过程中，有很多暂时无法解决的问题，例如对于差分隐私的证明（4.1 节），其中出现许多变量与公式都是我未曾学习的，但在我大致阅读前后文后发现，只需了解差分隐私的定义，那么就算读不懂这个证明过程对全文献理解并没有太大障碍，遂选择了跳跃该部分继续阅读。这是面对对全文贡献不是很大的部分的方法，但一些关键部分，例如（5.10 节）最佳参数的设置部分，初次阅读只觉晦涩难懂，但发现其为关键部分后，反复运用翻译软件与搜索引擎，不断理解，甚至手推其中公式辅助理解。

遇到最大的问题，不是如何复现结果，不是如何编写代码，而是如何开始阅读、如何开始编写文档，首次面对全英文的文档，令人眼花的公式与证明过程，还有完全没有想法的复现，最开始我完全是退缩的，完全不信任自己能够完成如此一个看似无法解决的任务，但沉没成本已经太高了，我此刻需要肩负起之前的我的信任，不能就这么放下，硬着头皮开始读，读不懂就翻译，翻译还不懂就使用大数据，连续三天的高强度阅读让我真正理解了文献的大致内容与公式，开始着手写代码，这个过程其实很轻松，偶尔遇到无法解决的问题也有 ai 借助，在有自身理解的基础下使用 ai 就像本就锋利的快刀有沾上了火焰，满课的两天就解决了代码的复现，虽然在复现图 4 的过程中遇到了无法解决的问题，但至少图一与图三是比较顺利完成了。

在完成之后，看着代码冗长的运行时间，我会在想如何加快运行速度，如何充分发挥算法性能，如何使用这个算法到实际场景中，当然，没有什么结果。