arXiv:2304.01779v1 [math.OC] 4 Apr 2023

具有线性收敛性的压缩差分隐私分布式优化

Antai Xie * Xinlei Yi ** Xiaofan Wang * Ming Cao ***
Xiaoqiang Ren *

*上海大学机电工程与自动化学院，中国上海。电子邮箱：{xatai、xfwang、xqren}@shu.edu.cn。

**麻省理工学院信息与决策系统实验室，美国马萨诸塞州剑桥市 02139。电子邮箱：xinleiyi@mit.edu。

***格罗宁根大学理学与工程学院，荷兰格罗宁根。电子邮箱：m.cao@rug.nl。

摘要：本文研究了通信受限条件下的差分隐私分布式优化问题，其中每个智能体都希望在最小化所有智能体代价函数之和的同时，保护自身代价函数的隐私。为此，我们提出了一种新型的压缩差分隐私分布式梯度跟踪算法（CPGT）。我们证明，即使使用一类有偏但收缩的压缩器，CPGT 对光滑强凸成本函数仍能实现线性收敛，并达到与理想化通信算法相同的精度。此外，我们严格证明了 CPGT 能够确保差分隐私。最后，通过仿真验证了所提算法的有效性。

关键词：压缩通信、分布式优化、差分隐私、线性收敛。

1. 引言

近几十年来，分布式优化在多智能体系统中的问题因其在多个领域的广泛应用而受到广泛关注，例如传感器网络、大规模机器学习（Tsianos 等人，2012；Dougherty 和 Guay，2016）以及在线优化（Li 等人，2020）。在典型的分布式共识优化设置中，目标是在连通网络上最小化所有智能体局部代价函数之和，其中每个智能体只能访问自己的局部代价函数。该问题已得到广泛研究，并由此发展出多种方法，例如分布式（次）梯度下降法（Nedic 和 Ozdaglar，2009）、EXTRA（Shi 等人，2015）以及梯度跟踪法（Qu 和 Li，2017）。

无线网络上的信息交换可能容易受到恶意对手的攻击。近期文献表明，恶意对手能够从机器学习中的共享梯度中获取私人训练集（Zhu 等人，2019），进而可能导致医疗记录和金融交易等敏感数据泄露。因此，亟须解决分布式优化中的隐私问题。

为保护每个智能体的隐私，Huang 等（2015）；Zhu 等人（2018）；Ding 等（2021）；Chen 等（2021）通过引入差分隐私（Dwork，2008）的概念，提出了几种差分隐私分布式优化算法，该概念具有可由数学证明的安全性。更具体地说，Huang 等（2015）通过使用拉普拉斯噪声掩盖状态，提出了差分隐私梯度下降方法。Zhu 等人（2018）将上述结果扩展到了时变有向网络。 然而，由于递减步长的限制，这两种方法都只能保证次线性收敛。为此，Ding 等（2021）通过同时向状态和方向添加噪声，并采用常数步长，实现了线性收敛和差分隐私。Chen 等（2021）进一步考虑了有向图的情形。此外，实现隐私保护的另一种主要方法是加密。例如，Lu 和 Zhu（2018）提出了一种使用同态加密的隐私保护分布式优化方法。尽管这些基于加密的方法能够使解收敛到精确最优解，但它们需要大量的计算资源。

上述大多数方法研究的都是理想化通信网络下的隐私保护分布式优化算法。在实际应用中，由于通信带宽有限，必须考虑压缩信息。为此，研究者提出了多种研究成果。Alistarh 等人（2017）利用无偏压缩器提出了通信高效的随机梯度下降算法。Kajiyama 等人（2020）和 Liao 等人（2022）分别结合梯度跟踪算法与具有有界绝对压缩误差的压缩器，以及一类具有有界相对压缩误差的压缩器，实现了线性收敛

。Xiong 等人（2021）将 Kajiyama 等人（2020）提出的方法扩展到了有向图。

压缩信息有许多优点，例如降低通信成本，因此将通信压缩与隐私保护结合起来是很自然的。为此，Wang 和 Ba˛sar（2022）提出了一种压缩差分隐私梯度下降算法，将压缩与隐私保护结合起来。然而，他们没有提供线性收敛分析。

受 Ding 等（2021）的启发，本文提出了一种压缩差分隐私分布式梯度跟踪算法，该算法实现了线性收敛并保护差分隐私。本文的主要贡献总结如下:

（1）针对一类有偏但收缩的压缩器，我们提出了一种新颖的压缩差分隐私梯度跟踪算法（CPGT）。我们证明，CPGT 实现线性收敛（定理 1），且精度与理想化通信网络上的算法相同（Ding 等人，2021）。（2）我们证明，CPGT 能够保护每个智能体的局部代价函数的差分隐私（定理2）。值得注意的是，CPGT 适用于一类压缩器。与 Wang 和 Ba˛sar（2022）不同，CPGT 不依赖于任何特定的压缩器。

本文其余部分安排如下。第2节介绍必要的定义并表述所研究的问题。第3节提出 CPGT，并分析所提出 CPGT 的收敛性和隐私性。第4节通过一个数值示例说明所得结果。最后，第5节总结全文。

记号： $\mathbb{R}$ ($\mathbb{R}_+$) 表示（正）实数集合。 $\mathbb{Z}$表示整数集合， $\mathbb{N}$ 表示自然数集合。 $\mathbb{R}^n$ 和 $\mathbb{R}^{n\times d}$ 分别表示取实数值的 n 维向量集合和 $n\times d$维矩阵集合。矩阵 P的转置记为 P^T，我们用 $[P]_{ij}$表示第 i行、第 j列的元素。全 1 列向量和全 0 列向量分别记为 $\mathbf{1}$ 和 $\mathbf{0}$。单位矩阵记为 I。接下来引入堆叠矩阵：对于矩阵 $\mathbf{x}\in\mathbb{R}^{n\times d}$，$\bar{\mathbf{x}}$ $\frac{\mathbf{1}\mathbf{1}^T}{n}\mathbf{x}$。 $|\cdot|$、 $\|\cdot\|$ $\|\cdot\|_F$分别表示绝对值、l_2 范数和弗罗贝尼乌斯范数。对于矩阵W，我们用 λ_W 表示其谱半径。给定常数 $\theta>0$，Lap(θ) 表示拉普拉斯分布，其概率密度函数为 $f_L(x,\theta)=\frac{1}{2\theta}e^{\frac{|x|}{\theta}}$。对于任意向量$\eta=[\eta_1,\ldots,\eta_d]^T\in\mathbb{R}^d$，若其每个分量 $\eta_i\sim$均服从 Lap(θ)， $i=1,\ldots,d$，则称 $\eta\sim$服从 Lap^d(θ)。此外，我们分别用 $\mathbb{E}(x)$ 和 $\mathbb{P}(x)$ 表示随机变量 x 的期望和概率。另外，若 $x\sim$服从 Lap(θ)，则有 $\mathbb{E}(x^2)=\theta^2$ 和 $\mathbb{E}(|x|)=\theta$。

2. 预备知识与问题表述

2.1 分布式优化

我们考虑一个由 n 个智能体组成的网络，其中每个智能体都有一个私有的凸成本函数 $f_i:\mathbb{R}^d\mapsto\mathbb{R}$。所有智能体协同求解以下优化问题:

$$\min_{x\in\mathbb{R}^d}\sum_{i=1}^{n}f_i(x), \tag{1}$$

其中， x 是全局决策变量，每个智能体都不知道该变量，只能在本地进行估计。更具体地说，我们假设每个智能体 i 在时间步 k 维护其局部估计 $x_i(k)\in\mathbb{R}^d$，并用 $\nabla f_i(x_i(k))$表示 f_i 关于 $x_i(k)$的梯度。此外，我们对局部代价函数 f_i作出以下假设:

假设1. 每个局部代价函数 f_i 都是 μ-强凸且 L-光滑的，其中 $0<\mu\le L$。也就是说，对于任意 x、$y\in\mathbb{R}^d$，

$$f_i(y)\ge f_i(x)+\nabla f_i(x)^T(y-x)+\frac{\mu}{2}\|x-y\|^2,$$
$$\|\nabla f_i(x)-\nabla f_i(y)\|\le L\|x-y\|.$$

假设1 是分布式优化中实现线性收敛的标准假设，例如 Kajiyama 等人（2020年）； Qu 和 Li（2017）； Xiong 等人（2021年）； Xu 等人（2017）。此外，在假设1 下，问题(1) 具有唯一的最优解 x^*（Boyd 和 Vandenberghe，2004）。

2.2 图论基础

智能体之间的信息交换由一个无向图 $\mathcal{G}(\mathcal{V},\mathcal{E})$ 表示，其中包含 n 个智能体，$\mathcal{V}=\{1,2,\ldots,n\}$ 是智能体索引的集合，$\mathcal{E}\subseteq\mathcal{V}\times\mathcal{V}$ 是边的集合。当且仅当智能体 i 和 j 可以相互通信时，边 $(i,j)\in\mathcal{E}$。令 $W=[w_{ij}]_{n\times n}\in\mathbb{R}^{n\times n}$ 为$\mathcal{G}$ 的邻接矩阵，即当 $(i,j)\in\mathcal{E}$ 或 $i=j$时， $w_{ij}>0$；否则为 $w_{ij}=0$。我们使用 $\mathcal{N}_i=\{j\in\mathcal{V}|\,(i,j)\in\mathcal{E}\}$ 表示智能体 i的邻居集合。

假设2。无向图 $\mathcal{G}(\mathcal{V},\mathcal{E})$ 是连通的，且 W是一个双随机矩阵，即 $\mathbf{1}^TW=\mathbf{1}^T$和$W\mathbf{1}=\mathbf{1}$。

2.3 压缩方法

由于通信信道容量有限，我们考虑智能体在发送信息前对其进行压缩的情形。更具体地，对于任意 $x\in\mathbb{R}^d$，我们考虑一类随机压缩器 $C(x,\varrho)$，并用$f_c(x,\varrho)$表示相应的概率密度函数，其中 ϱ 是一个随机扰动变量。当 ϱ 的分布已知时，$C(x,\varrho)$可以简写为 $C(x)$。接下来，我们给出以下假设。

假设3。对于某个 $\varphi\in[0,1)$，压缩器 $C(\cdot):\mathbb{R}^d\mapsto\mathbb{R}^d$ 满足

$$\mathbb{E}_C\left[\|C(x)-x\|^2\right]\le\varphi\|x\|^2,\forall x\in\mathbb{R}^d. \tag{2}$$

由（2）和詹森不等式可得

$$\mathbb{E}_C\left[\|C(x)-x\|\right]\le\sqrt{\varphi}\|x\|,\forall x\in\mathbb{R}^d. \tag{3}$$

备注1. 满足假设3的压缩器非常常见，例如 Koloskova 等（2019）、Reisizadeh 等（2019）和 Taheri 等（2020）。请注意，与 Wang 和 Ba˙sar（2022）不同，假设3不要求压缩器是无偏的。

2.4 差分隐私

为评估隐私性能，我们采用分布式优化中的(ϵ, δ)-差分隐私概念，该概念近期已由 Huang 等（2015）和 Ding 等（2021）研究。具体而言，我们引入以下定义。

定义 1.（(δ-邻接（Ding 等人，2021））给定$\delta > 0$，若存在某个$i_0 \in \{1, 2, \ldots, n\}$使得函数集 $\mathcal{S}^{(1)} = \{f_i^{(1)}\}_{i=1}^n$ 和 $\mathcal{S}^{(2)} = \{f_i^{(2)}\}_{i=1}^n$ 满足以下条件，则称它们是 δ-邻接的:

$$f_i^{(1)} = f_i^{(2)} \ \forall i \neq i_0, \text{ and } D(f_{i_0}^{(1)}, f_{i_0}^{(2)}) \leq \delta,$$

其中，$D(f_{(1)i_0}, f_{(2)i_0}) \triangleq \max_{x \in \mathbb{R}^d} \|\nabla f_{i_0}^{(1)}(x) - \nabla f_{(2)i_0}(x)\|$表示函数 $f_{i_0}^{(1)}$ 和$f_{i_0}^{(2)}$ 两个函数之间的距离。

定义 2.（差分隐私（Chen 等人，2021年））给定 δ、$\epsilon > 0$ 和随机机制 M，对于任意两个 δ-相邻函数集 $\mathcal{S}^{(1)}$ 和 $\mathcal{S}^{(2)}$，以及任意观测 $\mathcal{H} \subseteq \text{Range}(M)$，若随机机制 M 满足

$$\mathbb{P}\{M(\mathcal{S}^{(1)}) \in \mathcal{H}\} \leq e^{\epsilon} \mathbb{P}\{M(\mathcal{S}^{(2)}) \in \mathcal{H}\}. \tag{4}$$

定义 2 表明，若对于任意一对 δ-相邻函数集，其观测的概率密度函数相似，则随机机制 M 满足差分隐私。直观而言，如果相应机制 M 满足差分隐私，那么对手仅凭观测很难区分两个 δ-相邻函数集。

3. 主要结果

本节介绍压缩差分隐私梯度跟踪算法（CPGT），如算法 1 所示。

3.1 算法描述

所提出的 CPGT 算法受到 DiaDSP（Ding 等人，2021年）的启发。我们假设每个智能体 $i \in \mathcal{V}$ 都维护一个估计值 $x_i(k)$ 和一个辅助变量 $y_i(k)$，用于跟踪全局梯度。为实现差分隐私，每一步中，每个智能体 i 都会向其邻居 $\mathcal{N}_i$ 广播添加噪声后的信息 $x_i^a(k)$ 和 $y_i^a(k)$，其中

$$x_i^a(k) = x_i(k) + \eta_{x_i}(k), \tag{5}$$
$$y_i^a(k) = y_i(k) + \eta_{y_i}(k), \tag{6}$$

其中，$\eta_{x_i}(k)$ 和 $\eta_{y_i}(k)$ 为拉普拉斯噪声。类似于 DiaDSP 算法（Ding 等人，2021），我们设置 $\eta_{x_i} \sim \text{Lap}_d(d_{\eta_{x_i}} q_i^k)$ 和 $\eta_{y_i} \sim \text{Lap}_d(d_{\eta_{y_i}} q_i^k)$，$\forall i \in \mathcal{V}$，其中$d_{\eta_{x_i}} > 0$、$d_{\eta_{y_i}} > 0$ 和 $0 < q_i < 1$。完成信息交换后，智能体 i 执行以下更新:

算法1 CPGT 算法

1: 输入：停止时间 K、邻接矩阵 W，以及正参数 $\alpha, \gamma, d_{\eta_{x_i}}, d_{\eta_{y_i}}, q_i, \forall i \in \mathcal{V}$。
2: 初始化：每个 $i \in \mathcal{V}$ 任意选择 $x_i(0) \in \mathbb{R}^d$, $x_i^c(-1) = \mathbf{0}$, $y_i^c(-1) = \mathbf{0}$，并计算 $y_i(0) = \nabla f_i(x_i(0))$。
3: 对于 $k = 0, 1, \ldots, K-1$ 执行
4: 并行执行 for $i \in \mathcal{V}$ in
5: 生成拉普拉斯噪声 $\eta_{x_i}(k) \sim \text{Lap}_d(d_{\eta_{x_i}} q_i^k)$ 以及 $\eta_{y_i}(k) \sim \text{Lap}_d(d_{\eta_{y_i}} q_i^k)$.
6: 分别从 (5) 和 (6) 中获得 $x_i^a(k)$ 和 $y_i^a(k)$, 分别地。
7: 计算 $C(x_i^a(k) - x_i^c(k-1))$ 和 $C(y_i^a(k) - y_i^c(k-1))$，然后将它们广播给其邻居 $\mathcal{N}_i$.
8: 接收 $C(x_j^a(k) - x_j^c(k-1))$ 和 $C(y_j^a(k) - y_j^c(k-1))$ 来自 $j \in \mathcal{N}_i$。
9: 分别根据 (11) 和 (12) 更新 $x_j^c(k)$ 和 $y_j^c(k)$, 分别。
10: 根据 (9) 和 (10) 更新 $x_i(k+1)$ 和 $y_i(k+1)$, 分别。
11: 结束循环
12: 结束循环
13: 输出：$\{x_i(K)\}$。

$$x_i(k+1) = \sum_{j=1}^n w_{ij} x_j^a(k) - \alpha y_i(k), \tag{7}$$

$$y_i(k+1) = \sum_{j=1}^n w_{ij} y_j^a(k) + \nabla f_i(x_i(k+1)) - \nabla f_i(x_i(k)), \tag{8}$$

其中，步长 α 为常数，初始值$y_i(0) = \nabla f_i(x_i(0))$ $\forall i \in \mathcal{V}$。为提高通信效率，我们使用压缩信息 $x_i^c(k)$、$y_i^c(k)$ 分别替代 $x_i^a(k)$ 和 $y_i^a(k)$。然后，我们设计智能体 $i \in \mathcal{V}$ 的更新如下:

$$x_i(k+1) = x_i^a(k) + \gamma \sum_{j=1}^n w_{ij}(x_j^c(k) - x_i^c(k)) - \alpha y_i(k), \tag{9}$$

$$y_i(k+1) = y_i^a(k) + \gamma \sum_{j=1}^n w_{ij}(y_j^c(k) - y_i^c(k)) + \nabla f_i(x_i(k+1)) - \nabla f_i(x_i(k)), \tag{10}$$

其中

$$x_i^c(k) = x_i^c(k-1) + C(x_i^a(k) - x_i^c(k-1)), \tag{11}$$
$$y_i^c(k) = y_i^c(k-1) + C(y_i^a(k) - y_i^c(k-1)), \tag{12}$$

其中，γ 为正参数。我们假设$x_i^c(k) = \mathbf{0}$且 $y_i^c(k) = \mathbf{0}$，其中 $k < 0$, $\forall i \in \mathcal{V}$。令 W_γ，$(1-\gamma)I + \gamma W$，式（9）和（10）可改写为以下矩阵形式

$$\mathbf{x}(k+1) = W_\gamma(\mathbf{x}(k) + \eta_{\mathbf{x}}(k)) + \gamma(W - I)\sigma_{\mathbf{x}}(k) - \alpha \mathbf{y}(k), \tag{13}$$

$$\mathbf{y}(k+1) = W_\gamma(\mathbf{y}(k) + \eta_{\mathbf{y}}(k)) + \gamma(W - I)\sigma_{\mathbf{y}}(k) + \nabla \mathbf{f}(\mathbf{x}(k+1)) - \nabla \mathbf{f}(\mathbf{x}(k)), \tag{14}$$

其中 $\mathbf{x}(k) \ [x_1(k)\ x_2(k) \ldots, x_n(k)]^T \in \mathbb{R}^{n \times d}$, $\mathbf{y}(k) \triangleq [y_1(k)\ x_2(k) \ldots, y_n(k)]^T \in \mathbb{R}^{n \times d}$, $\sigma_{\mathbf{x}}(k) \ [x_1^c(k) - x_1^a(k), \ldots, x_n^c(k) - x_n^a(k)]^T \in \mathbb{R}^{n \times d}$, $\sigma_{\mathbf{y}}(k) \ [y_1^c(k) - y_1^a(k) \ldots,$

$y_n^c(k) - y_n^a(k)]^T \in \mathbb{R}^{n\times d}$，$\nabla \mathbf{f}(\mathbf{x}(k))$ $[\nabla f_1(x_1(k))\ldots,$ $\nabla f_n(x_n(k))]^T \in \mathbb{R}^{n\times d}$，$\eta_{\mathbf{x}}(k)$ $[\eta_{x_1}(k)\ldots,$ $\eta_{x_n}(k)]^T \in \mathbb{R}^{n\times d}$ 以及 $\eta_{\mathbf{y}}(k)$ $[\eta_{y_1}(k)\ldots,$ $\eta_{y_n}(k)]^T \in \mathbb{R}^{n\times d}$。

3.2 CPGT 的收敛性分析

本节首先证明，在满足假设3的压缩器下，CPGT 具有线性收敛性。其次，证明了 CPGT 算法下所有代价函数的差分隐私。需要指出的是，CPGT 的收敛精度与采用理想化通信的算法相同，即 DiaDSP（Ding 等人，2021）。

令 $\Theta(k) \triangleq [\Omega_{\mathbf{x}}(k),\ \Omega_{\bar{\mathbf{x}}}(k),\ \Omega_{\mathbf{y}}(k),\ \Omega_{\sigma_{\mathbf{x}}}(k),\ \Omega_{\sigma_{\mathbf{y}}}(k)]^T$，其中 $\Omega_{\mathbf{x}}(k) = \|\mathbf{x}(k) - \bar{\mathbf{x}}(k)\|_F$，$\Omega_{\bar{\mathbf{x}}}(k) = \|\bar{\mathbf{x}}(k) - \mathbf{x}^\infty\|_F$，$\Omega_{\mathbf{y}}(k) = \|\mathbf{y}(k) - \bar{\mathbf{y}}(k)\|_F$，$\Omega_{\sigma_{\mathbf{x}}}(k) = \|\sigma_{\mathbf{x}}(k)\|_F$，且 $\Omega_{\sigma_{\mathbf{y}}}(k) = \|\sigma_{\mathbf{y}}(k)\|_F$，其中 $\mathbf{x}^\infty \in \mathbb{R}^{n\times d}$ 由引理1给出。下面的引理构造了一个与 $\Theta(k)$ 相关的线性不等式系统。

引理1. 令 $\bar{q} = \max_i\{q_i\}$ 和 $d_\eta = \max_i\{d_{\eta_{x_i}}, d_{\eta_{y_i}}\}$。假设假设 1–3成立。在算法1下，若 $\alpha < \frac{1}{\mu+L}$，则有以下线性不等式:

$$\mathbb{E}[\Theta(k+1)] \preceq G\mathbb{E}[\Theta(k)] + \vartheta \bar{q}^k \bar{d}_\eta, \tag{15}$$

其中，符号 $\preceq$ 表示逐元素小于或等于，矩阵 $G \in \mathbb{R}^{5\times 5}$ 和常向量 $\vartheta \in \mathbb{R}^{5\times 1}$ 的元素见附录A，且 $\mathbf{x}^\infty = \mathbf{1}(x^\infty)^T$，其中 $x^\infty \in \mathbb{R}^d$ 满足

$$\sum_{i=1}^n \nabla f_i(x^\infty) = -\sum_{i=1}^n \sum_{k=0}^\infty \eta_{y_i}(k). \tag{16}$$

证明。见附录A。

根据引理1可知，若矩阵 G 的谱半径严格小于1，即 $\lambda_G < 1$，则CPGT可以线性收敛到 $\mathbf{x}^\infty$。因此，我们通过以下引理证明CPGT的线性收敛。

引理2. 假设假设 1–3成立。在算法1下，对于某个 $m \in (0,1)$，$\lambda_G \le 1 - \frac{m}{2}\alpha\mu$ 成立，若参数 γ 和步长 α 满足

$$\gamma \le \min\left\{\frac{(1-\sqrt{\varphi})\zeta_5}{2\sqrt{\varphi}\bar{\lambda}_{W-I}(2\zeta_1+\zeta_2+2\zeta_3+\zeta_4+\zeta_5)}, \frac{(1-\sqrt{\varphi})L}{\bar{\lambda}_{W-I}m\mu}, 1\right\}, \tag{17}$$

$$\alpha \le \min\left\{\frac{\gamma(1-\rho_w)\zeta_1}{L(2\zeta_3+m\zeta_1)}, \frac{\gamma(1-\rho_w)\zeta_3}{L(2\zeta_1+2\zeta_2+(2+m)\zeta_3)}, \frac{\gamma\bar{\lambda}_{W-I}}{2L}\right\}, \tag{18}$$

$$\zeta_1+\zeta_4+\zeta_5 \le \frac{(1-\rho_w)\zeta_3}{2\bar{\lambda}_{W-I}}, \zeta_5 \le \zeta_4, \zeta_4 \le \frac{(1-\rho_w)\zeta_1}{2\bar{\lambda}_{W-I}},$$

$$\zeta_1 \le \frac{m\mu\zeta_2}{2L}, \tag{19}$$

其中，ζ_1、$\ldots$、ρ_w =、ζ_5 是一些正常数，且 $\lambda_{W-\frac{\mathbf{1}\mathbf{1}^T}{n}}$。

证明。参见附录 B。

下面的定理表明，通过为 ζ_1、…、ζ_5 取某些具体值，CPGT 可以在均值意义下线性收敛到解 $\mathbf{x}^\infty$。

定理 1. 假设 1–3 成立。在算法 1 下，对于某个 $m \in (0, 1)$，如果参数 γ 和步长 α 满足

$$\gamma \le \min\left\{\frac{\kappa(1-\rho_w)^2 m\mu}{4\sqrt{\varphi}s_1}, \frac{\kappa L}{m\mu}, 1\right\}, \tag{20}$$

$$\alpha \le \min\left\{\frac{m\mu(1-\rho_w)^2}{s_2}, \frac{\bar{\lambda}_{W-I}(1-\bar{q})}{2}\right\}\frac{\gamma}{L}, \tag{21}$$

其中，$\kappa = \frac{1-\sqrt{\varphi}}{\lambda_{W-I}}$、$s_1 = 4\bar{\lambda}_{W-I}^2 m\mu + (1-\rho_w)^2 m\mu + (6m\mu + 2L)\bar{\lambda}_{W-I}(1-\rho_w$ $s_2 = 6m\mu + 2m^2\mu +)$，且$(4L)(1-\rho_w) + 2m\mu\lambda_{W-I}(2+m)$，那么，我们有 $\mathbb{E}[\|\mathbf{x}(k) - \mathbf{x}^\infty\|_F] = \mathcal{O}((1-\frac{m}{2}\alpha\mu)^k)$。

证明。令 $\zeta_1 = 1$、$\zeta_2 = \frac{2L}{m\mu}$、$\zeta_3 = \frac{2\bar{\lambda}_{W-I}-2\rho_w+2}{1-\rho_w}$、$\zeta_4 = \zeta_5 = \frac{1-\rho_w}{2\bar{\lambda}_{W-I}}$，容易验证(19)成立。然后，通过将 ζ_1、…、ζ_5 替换为上述取值，我们从 (20)–(21) 推导出 (17)–(18)。根据引理 1 和 2，令$h = 1 - \frac{m}{2}\alpha\mu$，我们有

$$\begin{aligned}\mathbb{E}[\|\Theta(k+1)\|] &\le h^k\mathbb{E}[\|\Theta(0)\|] + \vartheta\bar{d}_\eta\sum_{t=0}^{k-1}h^{k-1-t}\bar{q}^t \\ &\le h^k\mathbb{E}[\|\Theta(0)\|] + \vartheta\bar{d}_\eta h^{k-1}\sum_{t=0}^{k-1}\left(\frac{\bar{q}}{h}\right)^t \\ &\le (\mathbb{E}[\|\Theta(0)\|] + \frac{\vartheta\bar{d}_\eta}{h-\bar{q}})h^k.\end{aligned}$$

由于 $\mathbb{E}[\|\mathbf{x}(k) - \mathbf{x}^\infty\|_F] \le \mathbb{E}[\|\mathbf{x}(k) - \bar{\mathbf{x}}(k)\|_F] + \mathbb{E}[\|\bar{\mathbf{x}}(k) - \mathbf{x}^\infty\|_F]$，可得在线观看中文字幕

$$\mathbb{E}[\|\mathbf{x}(k) - \mathbf{x}^\infty\|_F] = \mathcal{O}(h^k), \tag{22}$$

至此，证明得证。

注 2. 注意到差分隐私与精确收敛不可能同时实现（参见 Ding 等人，2021，命题 1）。如式（16）所示，收敛点 x^∞ 与最优解 x^* 之间的距离受到添加到梯度中的噪声总和的影响。此外，CPGT 的收敛点 x^∞ 与 DiaDSP（Ding 等人，2021）相同，这意味着所提出的 CPGT 具有与采用理想化通信的算法相同的精度。

3.3 差分隐私

在本节中，我们证明在 CPGT 下可以保留所有代价函数的差分隐私性。

我们用 $\mathcal{H}_k$ 表示在时间步 k 智能体之间传输的信息，即 $\mathcal{H}_k = \{C(x_i^a(k) - x_i^c(k-1))\ C(y_i^a(k) - y_i^c(k-1)) \mid \forall i \in \mathcal{V}\}$。不失一般性，我们假设对手旨在推断智能体 i_0 的代价函数。考虑任意两个 δ-相邻函数集 $\mathcal{S}^{(1)}$ 和 $\mathcal{S}^{(2)}$，且两个集合之间只有代价函数f_{i_0}不同，即 $f_{i_0}^1 \ne f_{i_0}^{(2)})$和$f_i^{(1)} = f_i^{(2)}$，$\forall i \ne i_0$。

假设 4. 对于任意 x_1、$x_2 \in \mathbb{R}^d$，均有

$$\nabla f_{i_0}^{(1)}(x_1) - \nabla f_{i_0}^{(1)}(x_2) = \nabla f_{i_0}^{(2)}(x_1) - \nabla f_{i_0}^{(2)}(x_2).$$

注3. 假设4是实现无限时间差分隐私所必需的，同样的假设也可见于相关文献（Ding 等人，2021）。

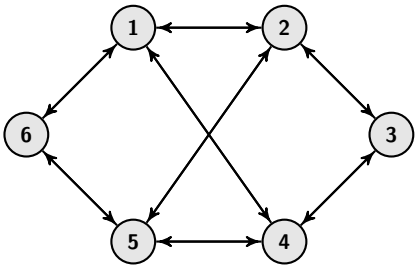


图1. 由6个智能体组成的连通无向图。

定理2. 假设假设1–4成立。对于任意智能体 i_0的代价函数，若$\alpha < \frac{1}{2L}$且 $q_{i_0} \in (\frac{\alpha L+\sqrt{\alpha^2L^2+4\alpha L}}{2}, 1)$，其中$\epsilon_{i_0}$ 由下式给出，则CPGT可保持 ϵ_{i_0}-差分隐私

$$\epsilon_{i_0} = \frac{\tau_{i_0} q_{i_0}^2 \delta}{q_{i_0}^2 - \alpha L - q_{i_0}\alpha L}, \quad \forall i_0 \in \mathcal{V}, \tag{23}$$

其中 $\tau_{i_0} = \frac{\alpha}{d_{\eta_{x_{i_0}}}} + \frac{1}{d_{\eta_{y_{i_0}}}}$。

证明。见附录C。

注4. 与 Wang 和 Ba¸sar (2022) 的方法不同，CPGT 适用于一类压缩器，并实现线性收敛。

4. 仿真

本节通过仿真验证 CPGT 的有效性。我们考虑一个包含 $n = 6$ 个智能体的分布式估计问题，这些智能体通过一个连通无向图进行通信，其拓扑结构如图1所示。具体而言，假设每个智能体 i 旨在估计未知参数 $x \in \mathbb{R}^{10}$，并拥有含噪测量值$b_i = A_i x + e_i$，其中 $A_i \in \mathbb{R}^{6\times 10}$ 是智能体 i 的非奇异测量矩阵，$e_i \in \mathbb{R}^6$ 是噪声向量。该问题可重写为以下形式

$$\min_x f(x) = \frac{1}{6}\sum_{i=1}^{6} \|A_i x - b_i\|^2.$$

在此示例中，矩阵 A_i 和向量 b_i 的元素由服从高斯分布的 $N(0, 1)$ 随机生成。此外，每个智能体的初始值 $x_i(0)$ 在 $[0, 1]^{10}$ 中随机选取。

我们考虑以下两种有偏但收缩的压缩器:

- 贪心（Top-k）量化器（Beznosikov 等，2020年）：

$$C_1(x) := \sum_{i_s=1}^{k} x_{(i_s)} e_{i_s},$$

其中，$x_{(i_s)}$ 是 x 的第 i_s 个坐标，i_1、……、i_k 是 x 中幅值最大的 k 个坐标的索引，而 e_1、……、e_d 是 $\mathbb{R}^d$ 中的标准单位基向量。

- 有偏 b 位量化器（Koloskova 等，2019年）：

$$C_2(x) := \frac{\|x\|}{\xi} \cdot \text{sign}(x) \cdot 2^{-(b-1)} \circ \left\lfloor \frac{2^{(b-1)|x|}}{\|x\|} + u \right\rfloor,$$

其中，$\xi = 1 + \min\{\frac{d}{2^{2(b-1)}}, \frac{\sqrt{d}}{2^{(b-1)}}\}$，$u$ 是从 $[0, 1]^d$、$\circ$ 中均匀采样的随机抖动向量

算法	压缩器	γ	α	d_η	q
CPGT-C1	C_1	0.05	0.1	100	0.99
CPGT-C2-1	C_2	0.2	0.1	100	0.99
CPGT-C2-2	C_2	0.05	0.15	100	0.99
DiaDSP	—	—	0.15	100	0.99

表 1。

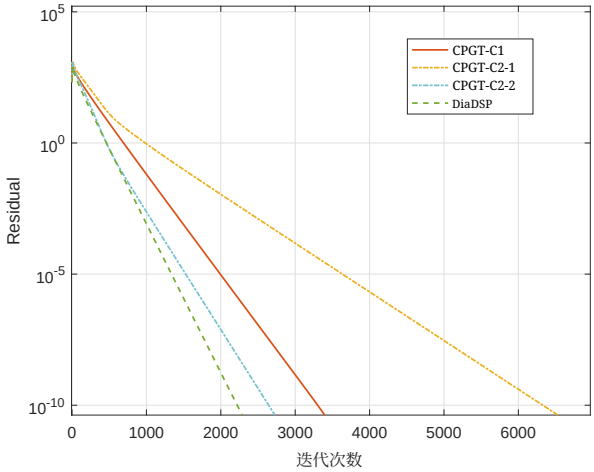


图 2. CPGT 与未压缩方法 DiaDSP 下残差的变化。

表示 Hadamard 乘积，sign(·)、$|\cdot|$、$\lfloor\cdot\rceil$ 分别表示逐元素符号函数、绝对值函数和向下取整函数。

正如 Koloskova 等（2019）所指出的，上述两种压缩器都满足假设3。具体而言，在以下仿真中，我们选择 $k = 2$ 和 $b = 2$。

我们首先验证 CPGT 在不同压缩器下的收敛速度。我们设定 $d_{\eta_{x_i}} = d_{\eta_{x_i}} = d_\eta$ 和 $q_i = q$，不同算法的参数见表1。残差的计算定义为 residual $\triangleq \frac{1}{T}\sum_{t=1}^{T} \|\mathbf{x}_t(k) - \mathbf{x}^\infty\|_F$，其中 $T = 1000$ 和 $\mathbf{x}_t(k)$ 分别表示时间步 k时的状态 $\mathbf{x}$。图2显示，在采用不同常数步长和压缩器的 CPGT 下，$\mathbf{x}(k)$ 线性收敛至点 $\mathbf{x}^\infty$。此外，选择合适的参数和压缩器后，CPGT 的收敛速度可接近 DiaDSP（Ding 等人，2021）。

我们进一步仿真噪声衰减率对收敛精度的影响。设 $\mathbf{x}^* = \mathbf{1} x^* \quad x^* \in (\mathbb{R}^d)^\infty$ 为最优解。我们使用 $\|\mathbf{x} - \mathbf{x}^*\|_F$ 衡量不同算法的收敛精度。我们设定 $d_\eta = 5$，其他参数 γ 和 α 与表1相同。精度与衰减率 q 之间的关系如图3所示，其中 q =0.18、0.26、0.34、0.42、0.5、0.58、0.66、0.74、0.82、0.9。可以看出，CPGT 的精度与 DiaDSP 几乎相同。此外，精度仅取决于噪声，与步长、γ和压缩器无关。

5. 结论

在本文中，我们研究了通信受限条件下的差分隐私分布式优化问题。具体而言，针对一类有偏但收缩的压缩器，我们提-

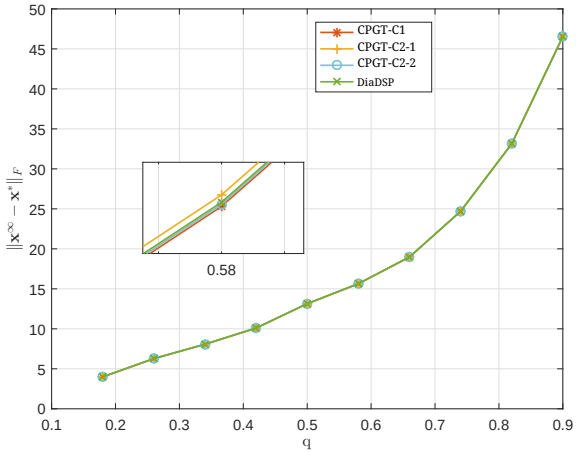


图3. 噪声衰减率对收敛精度的影响。

出了新颖的压缩差分隐私梯度跟踪算法（CPGT）。当所有局部代价函数均光滑且强凸时，我们证明了 CPGT 的线性收敛性。所提出的 CPGT 具有与采用理想通信的算法相同的准确性。不同于以往的研究，CPGT 使用一类有偏但收缩的压缩器，为每个智能体的局部代价函数保留了差分隐私。未来的工作包括将其扩展到有向图，并考虑更一般的压缩器。

附录A. 引理1的证明

A.1 辅助引理

引理3. 假设假设1成立，若 $\alpha \le \frac{1}{\mu+L}$，则可得

$$\|\bar{\mathbf{x}}(k)-\mathbf{x}^{\infty}-\alpha(\nabla\bar{\mathbf{f}}(\bar{\mathbf{x}}(k))-\nabla\bar{\mathbf{f}}(\mathbf{x}^{\infty})\|_F \le (1-\alpha\mu)\|\bar{\mathbf{x}}(k)-\mathbf{x}^{\infty}\|_F.$$

证明. 根据性质 $\frac{\mathbf{1}\mathbf{1}^T}{n}\bar{\mathbf{x}}(k) = \bar{\mathbf{x}}(k)$ $\frac{\mathbf{1}\mathbf{1}^T}{n}\mathbf{x}^{\infty} = \mathbf{x}^{\infty}$，有

$$\|\bar{\mathbf{x}}(k)-\mathbf{x}^{\infty}-\alpha(\nabla\bar{\mathbf{f}}(\bar{\mathbf{x}}(k))-\nabla\bar{\mathbf{f}}(\mathbf{x}^{\infty})\|_F \le \|\bar{\mathbf{x}}(k)-\mathbf{x}^{\infty}-\alpha(\nabla\mathbf{f}(\bar{\mathbf{x}}(k))-\nabla\mathbf{f}(\mathbf{x}^{\infty})\|_F.$$

然后，按照 Qu 和 Li（2017）中引理10的相同推导即可完成证明。

引理4.（Liao等人，2022）假设假设2成立。对于 $\gamma \in (0,1]$ 以及任意 $\omega \in \mathbb{R}^{n\times d}$，有 $\|W_\gamma\omega-\bar{\omega}\|_F \le \hat{\lambda}\|\omega-\bar{\omega}\|_F$，其中 $\lambda = 1-\gamma(1-\rho_w)$，且 $\rho_w = \lambda_{W-\frac{\mathbf{1}\mathbf{1}^T}{n}}$。

A.2 引理1的证明

我们通过分别构造$\mathbb{E}[\Omega_{\mathbf{x}}(k+1)]$, $\mathbb{E}[\Omega_{\bar{\mathbf{x}}}(k+1)]$, $\mathbb{E}[\Omega_{\mathbf{y}}(k+1)]$, $\mathbb{E}[\Omega_{\sigma_{\mathbf{x}}}(k+1)]$,和 $\mathbb{E}[\Omega(\sigma_{\mathbf{y}})k+1]$, 的上界来证明引理1。

(a) 根据 (13) 和引理4，我们得到

$$\mathbb{E}[\Omega_{\mathbf{x}}(k+1)] \le \hat{\lambda}\mathbb{E}[\Omega_{\mathbf{x}}(k)]+\alpha\mathbb{E}[\Omega_{\mathbf{y}}(k)] + \gamma\bar{\lambda}_{W-I}\mathbb{E}[\Omega_{\sigma_{\mathbf{x}}}(k)]+\nu_1(k), \tag{A.1}$$

其中 $\nu_1(k) = \hat{\lambda}\mathbb{E}[\|\eta_{\mathbf{x}}(k)-\bar{\eta}_{\mathbf{x}}(k)\|_F]$。

（b）由（14）可知，我们首先介绍 CPGT 的一个关键性质，即对于 $k \ge 0$,

$$\mathbf{1}^T\mathbf{y}(k) = \mathbf{1}^T\nabla\mathbf{f}(\mathbf{x}(k))+\mathbf{1}^T\sum_{t=0}^{k-1}\eta_{\mathbf{y}}(t). \tag{A.2}$$

由 (13)、(16) 和 (A.2)，我们有

$$\bar{\mathbf{x}}(k+1)-\mathbf{x}^{\infty} = \bar{\mathbf{x}}(k)-\mathbf{x}^{\infty}+\bar{\eta}_{\mathbf{x}}(k)+\alpha\sum_{t=k}^{\infty}\bar{\eta}_{\mathbf{y}}(t) - \alpha(\nabla\bar{\mathbf{f}}(\mathbf{x}(k))-\nabla\bar{\mathbf{f}}(\mathbf{x}^{\infty})).$$

由引理3，我们有

$$\mathbb{E}[\Omega_{\bar{\mathbf{x}}}(k+1)] \le (1-\alpha\mu)\mathbb{E}[\Omega_{\bar{\mathbf{x}}}(k)]+\alpha L\mathbb{E}[\Omega_{\mathbf{x}}(k)]+\nu_2(k), \tag{A.3}$$

其中 $\nu_2(k) = \mathbb{E}[\|\bar{\eta}_{\mathbf{x}}(k)\|_F]+\alpha\sum_{t=k}^{\infty}\mathbb{E}[\|\bar{\eta}_{\mathbf{y}}(t)\|_F]$。

(c) 根据 (14)、引理4 和假设1，可得

$$\mathbb{E}[\Omega_{\mathbf{y}}(k+1)] \le \hat{\lambda}\mathbb{E}[\Omega_{\mathbf{y}}(k)]+L\mathbb{E}[\|\mathbf{x}(k+1)-\mathbf{x}(k)\|_F] + \hat{\lambda}\mathbb{E}[\|\eta_{\mathbf{y}}(k)-\bar{\eta}_{\mathbf{y}}(k)\|_F] + \gamma\bar{\lambda}_{W-I}\mathbb{E}[\Omega_{\sigma_{\mathbf{y}}}(k)]. \tag{A.4}$$

由式(9)，可得

$$\mathbb{E}[\|\mathbf{x}(k+1)-\mathbf{x}(k)\|_F] \le \gamma\bar{\lambda}_{W-I}\mathbb{E}[\Omega_{\mathbf{x}}(k)] + \gamma\bar{\lambda}_{W-I}\mathbb{E}[\Omega_{\sigma_{\mathbf{x}}}(k)]+\alpha\mathbb{E}[\|\mathbf{y}(k)\|_F]+\hat{\lambda}\mathbb{E}[\|\eta_{\mathbf{x}}(k)\|_F]. \tag{A.5}$$

此外，由式(16)和式(A.2)可得

$$\mathbb{E}[\|\mathbf{y}(k)\|_F] \le \mathbb{E}[\Omega_{\mathbf{y}}(k)]+\mathbb{E}[\|\bar{\mathbf{y}}(k)\|_F] \le \mathbb{E}[\Omega_{\mathbf{y}}(k)]+\sum_{t=k}^{\infty}\mathbb{E}[\|\bar{\eta}_{\mathbf{y}}(t)\|_F]+L\mathbb{E}[\Omega_{\mathbf{x}}(k)] + L\mathbb{E}[\Omega_{\bar{\mathbf{x}}}(k)]. \tag{A.6}$$

结合式(A.4)–(A.6)，可得

$$\mathbb{E}[\Omega_{\mathbf{y}}(k+1)] \le L(\gamma\bar{\lambda}_{W-I}+\alpha L)\mathbb{E}[\Omega_{\mathbf{x}}(k)] + \alpha L^2\mathbb{E}[\Omega_{\bar{\mathbf{x}}}(k)]+(\alpha L+\hat{\lambda})\mathbb{E}[\Omega_{\mathbf{y}}(k)] + L\gamma\bar{\lambda}_{W-I}\mathbb{E}[\Omega_{\sigma_{\mathbf{x}}}(k)]+\gamma\bar{\lambda}_{W-I}\mathbb{E}[\Omega_{\sigma_{\mathbf{y}}}(k)]+\nu_3(k), \tag{A.7}$$

其中 $\nu_3(k) = \hat{\lambda}\mathbb{E}[\|\eta_{\mathbf{y}}(k)-\bar{\eta}_{\mathbf{y}}(k)\|_F]+\alpha L\sum_{t=k}^{\infty}\mathbb{E}[\|\bar{\eta}_{\mathbf{y}}(t)\|_F]+L\lambda\mathbb{E}[\|\eta_{\mathbf{x}}(k)\|_F]$。

(d) 由式(11)和假设3可得

$$\mathbb{E}[\Omega_{\sigma_{\mathbf{x}}}(k+1)] \le \sqrt{\varphi}\mathbb{E}[\|\mathbf{x}(k+1)+\eta_{\mathbf{x}}(k+1)-\mathbf{x}^c(k)\|_F] \le \sqrt{\varphi}\mathbb{E}[\Omega_{\sigma_{\mathbf{x}}}(k)]+\sqrt{\varphi}\mathbb{E}[\|\mathbf{x}(k+1)-\mathbf{x}(k)\|_F] + \sqrt{\varphi}\mathbb{E}[\|\eta_{\mathbf{x}}(k+1)-\eta_{\mathbf{x}}(k)\|_F]. \tag{A.8}$$

将 $\mathbb{E}[\|\mathbf{x}(k+1)-\mathbf{x}(k)\|_F]$ 代入 (A.5)–(A.6)，可得

$$\mathbb{E}[\Omega_{\sigma_{\mathbf{x}}}(k+1)] \le \sqrt{\varphi}(\gamma\bar{\lambda}_{W-I}+\alpha L)\mathbb{E}[\Omega_{\mathbf{x}}(k)] + \sqrt{\varphi}\alpha L\mathbb{E}[\Omega_{\bar{\mathbf{x}}}(k)]+\sqrt{\varphi}\alpha\mathbb{E}[\Omega_{\mathbf{y}}(k)] + \sqrt{\varphi}(\gamma\bar{\lambda}_{W-I}+1)\mathbb{E}[\Omega_{\sigma_{\mathbf{x}}}(k)]+\nu_4(k), \tag{A.9}$$

其中 $\nu_4(k) = \sqrt{\varphi}\lambda\mathbb{E}[\|\eta_{\mathbf{x}}(k)\|_F]+\sqrt{\varphi}\mathbb{E}[\|\eta_{\mathbf{x}}(k+1)-\eta_{\mathbf{x}}(k)\|_F]+\sqrt{\varphi}\alpha\sum_{t=k}^{\infty}\mathbb{E}[\|\bar{\eta}_{\mathbf{y}}(t)\|_F]$。 (e) 与 (A.8) 类似，我们有

$$\mathbb{E}[\Omega_{\sigma_{\mathbf{y}}}(k+1)] \le \sqrt{\varphi}\mathbb{E}[\Omega_{\sigma_{\mathbf{y}}}(k)] + \sqrt{\varphi}\mathbb{E}[\|\mathbf{y}(k+1)-\mathbf{y}(k)\|_F] + \sqrt{\varphi}\mathbb{E}[\|\eta_{\mathbf{y}}(k+1)-\eta_{\mathbf{y}}(k)\|_F]. \tag{A.10}$$

由 (10) 可得

$$\mathbb{E}[\|\mathbf{y}(k+1)-\mathbf{y}(k)\|_F] \le \gamma\bar{\lambda}_{W-I}\mathbb{E}[\Omega_{\mathbf{y}}(k)] + \gamma\bar{\lambda}_{W-I}\mathbb{E}[\Omega_{\sigma_{\mathbf{y}}}(k)] + \hat{\lambda}\mathbb{E}[\|\eta_{\mathbf{y}}(k)\|_F] + L\mathbb{E}[\|\mathbf{x}(k+1)-\mathbf{x}(k)\|_F]. \tag{A.11}$$

结合 (A.5)、(A.6)、(A.10) 和 (A.11)，可得

$$\begin{aligned}\mathbb{E}[\Omega_{\sigma_{\mathbf{y}}}(k+1)] &\le \sqrt{\varphi}L(\gamma\bar{\lambda}_{W-I}+\alpha L)\mathbb{E}[\Omega_{\mathbf{x}}(k)] \\ &+ \sqrt{\varphi}\alpha L^2\mathbb{E}[\Omega_{\bar{\mathbf{x}}}(k)] + \sqrt{\varphi}(\alpha L+\gamma\bar{\lambda}_{W-I})\mathbb{E}[\Omega_{\mathbf{y}}(k)] \\ &+ \sqrt{\varphi}L\gamma\bar{\lambda}_{W-I}\mathbb{E}[\Omega_{\sigma_{\mathbf{x}}}(k)] + \sqrt{\varphi}(\gamma\bar{\lambda}_{W-I}+1)\mathbb{E}[\Omega_{\sigma_{\mathbf{y}}}(k)] \\ &+ \nu_5(k),\end{aligned} \tag{A.12}$$

其中 $\nu_5(k) = \sqrt{\varphi}L\lambda\mathbb{E}[\|\eta_{\mathbf{x}}(k)\|_F] + \sqrt{\varphi}\mathbb{E}[\|\eta_{\mathbf{y}}(k+1) - \eta_{\mathbf{y}}(k)\|_F] + \sqrt{\varphi}\alpha L\sum_{t=k}^{\infty}\mathbb{E}[\|\bar{\eta}_{\mathbf{y}}(t)\|_F] + \sqrt{\varphi}\lambda\mathbb{E}[\|\eta_{\mathbf{y}}(k)\|_F]$。那么，$G$ 的各元素分别对应 (A.1)、(A.3)、(A.7)、(A.9) 和 (A.12) 中的系数。令 $\nu(k) = [\nu_1\ k\ \ \nu_2\ k\ \ \nu_3\ k\ \ \nu_4\ k\ \ \nu_5\ k\ ]^T$ $\eta_{x_i} \sim {}_d\, d_{\eta_{x_i}} q_i^k$ ( ) ( ) ( ) ( ) ( )。由于 Lap ( ) 且 $\eta_{y_i} \sim \text{Lap}_d(d_{\eta_{y_i}} q_i^k)$，因此有 $\nu(k) \preceq \vartheta\bar{q}^k\bar{d}_\eta$，其中 ϑ 由下式给出

$$\vartheta = \left\{\hat{\lambda},\ \frac{1-\bar{q}+\alpha}{1-\bar{q}},\ (L+1)\hat{\lambda}+\frac{\alpha L}{1-\bar{q}},\ \sqrt{\varphi}(\hat{\lambda}+1)+\frac{\sqrt{\varphi}\alpha}{1-\bar{q}},\ \sqrt{\varphi}L\hat{\lambda}+\sqrt{\varphi}+\frac{\sqrt{\varphi}\alpha L}{1-\bar{q}}+\sqrt{\varphi}\hat{\lambda}\right\}n^2. \tag{A.13}$$

附录 B. 引理 2 的证明

B.1 辅助引理

我们首先介绍以下有用的引理。

引理 5.（Horn and Johnson（2012）中的推论 8.1.29）设 $A \in \mathbb{R}^{n\times n}$ 是一个非负矩阵，$v \in \mathbb{R}^n$ 是一个逐元素正向量。对于 $\beta \ge 0$、$\lambda_G \le \beta$，若 $Gv \le \beta v$。

B.2 引理2的证明

根据引理 5，显然只要以下线性不等式成立，就可以证明引理 2。

$$G\zeta \preceq (1-\frac{m}{2}\alpha\mu)\zeta, \tag{B.1}$$

其中 $\zeta = [\zeta_1, \zeta_2, L\zeta_3, \zeta_4, L\zeta_5]^T$ 与 ζ_1、……、$\zeta_5 > 0$。

(a) 由 (A.1) 可得，(B.1) 中的第一个不等式为

$$\hat{\lambda}\zeta_1 + \alpha L\zeta_3 + \gamma\bar{\lambda}_{W-I}\zeta_4 \le (1-\frac{m}{2}\alpha\mu)\zeta_1, \tag{B.2}$$

可将其改写为

$$\alpha(L\zeta_3 + \frac{m\mu}{2}\zeta_1) + \gamma\bar{\lambda}_{W-I}\zeta_4 \le \gamma(1-\rho_w)\zeta_1.$$

因此，当 $\alpha \le \frac{\gamma(1-\rho_w)\zeta_1}{L(2\zeta_3+m\zeta_1)}$ 且 $\zeta_4 \le \frac{(1-\rho_w)\zeta_1}{2\bar{\lambda}_{W-I}}$ 时，不等式 (B.2) 成立。

(b) 根据 (A.3)，(B.1) 中的第二个不等式为

$$\alpha L\zeta_1 + (1-\alpha\mu)\zeta_2 \le (1-\frac{m}{2}\alpha\mu)\zeta_2. \tag{B.3}$$

当 $\zeta_1 \le \frac{m\mu\zeta_2}{2L}$ 时，不等式 (B.3) 成立。

(c) 由 (A.7) 可得，(B.1) 中的第三个不等式为

$$L(\gamma\bar{\lambda}_{W-I}+\alpha L)\zeta_1 + \alpha L^2\zeta_2 + (\alpha L+\hat{\lambda})L\zeta_3 + L\gamma\bar{\lambda}_{W-I}\zeta_4 + L\gamma\bar{\lambda}_{W-I}\zeta_5 \le (1-\frac{m}{2}\alpha\mu)L\zeta_3. \tag{B.4}$$

将 (B.4) 两边除以 γ 和 L，可得

$$\frac{\alpha}{\gamma}(L\zeta_1 + L\zeta_2 + L\zeta_3 + \frac{m}{2}\mu\zeta_3) + \bar{\lambda}_{W-I}(\zeta_1+\zeta_4+\zeta_5) \le (1-\rho_w)\zeta_3. \tag{B.5}$$

由 (B.5) 可知，当 $\zeta_1+\zeta_4+\zeta_5 \le \frac{(1-\rho_w)\zeta_3}{2\bar{\lambda}_{W-I}}$ 时，不等式 (B.4) 成立 and $\alpha \le \frac{\gamma(1-\rho_w)\zeta_3}{L(2\zeta_1+2\zeta_2+(2+m)\zeta_3)}$.

(d) 由 (A.9) 可知，(B.1) 中的第四个不等式为

$$\sqrt{\varphi}(\gamma\bar{\lambda}_{W-I}+\alpha L)\zeta_1 + \sqrt{\varphi}\alpha L\zeta_2 + \sqrt{\varphi}\alpha L\zeta_3 + \sqrt{\varphi}(\gamma\bar{\lambda}_{W-I}+1)\zeta_4 \le (1-\frac{m}{2}\alpha\mu)\zeta_4. \tag{B.6}$$

若 $\alpha \le \frac{\gamma\bar{\lambda}_{W-I}}{L}$，则可得

$$\sqrt{\varphi}\gamma\bar{\lambda}_{W-I}(2\zeta_1+\zeta_2+\zeta_3+\zeta_4) + \frac{\gamma\bar{\lambda}_{W-I}m\mu\zeta_4}{2L} \le (1-\sqrt{\varphi})\zeta_4.$$

因此，当 $\alpha \le \frac{\gamma\lambda_{W-I}}{L}$，$\gamma \le \min\{\frac{(1-\sqrt{\varphi}\ \zeta_4)}{2\sqrt{\varphi}\bar{\lambda}_{W-I}}$ $\frac{}{(2\zeta_1+\zeta_2+\zeta_3+\zeta_4)}, \frac{(1-\sqrt{\varphi}\ L)}{\bar{\lambda}_{W-I}m\mu}\}$ 时，不等式 (B.6) 成立。

(e) 由 (A.12) 可知，(B.1) 中的第五个不等式为

$$\sqrt{\varphi}L(\gamma\bar{\lambda}_{W-I}+\alpha L)\zeta_1 + \sqrt{\varphi}\alpha L^2\zeta_2 + \sqrt{\varphi}(\alpha L+\gamma\bar{\lambda}_{W-I})L\zeta_3 + \sqrt{\varphi}L\gamma\bar{\lambda}_{W-I}\zeta_4 + \sqrt{\varphi}(\gamma\bar{\lambda}_{W-I}+1)L\zeta_5 \le (1-\frac{m}{2}\alpha\mu)L\zeta_5. \tag{B.7}$$

与 (B.6) 类似，当 $\alpha \le \frac{\gamma\bar{\lambda}_{W-I}}{L}$，$\gamma \le \min\{\frac{(1-\sqrt{\varphi}\ \zeta_5)(}{2\sqrt{\varphi}\bar{\lambda}_{W-I}}$ $\frac{}{2\zeta_1+\zeta_2+2\zeta_3+\zeta_4+\zeta_5)}, \frac{(1-\sqrt{\varphi}\ L)}{\bar{\lambda}_{W-I}m\mu}\}$ 时，不等式 (B.7) 成立。总之，当 $\zeta_1, \ldots, \zeta_5$，γ，且步长 α 满足 (17)– (19) 时，(B.1) 成立。

附录C. 定理2的证明

由 GTPT 可知，观测序列 $\mathcal{H} = \{\mathcal{H}_k\}_{k=0}^{\infty}$ 由噪声序列 $\eta_{\mathbf{x}} = \{\eta_{\mathbf{x}}(k)\}_{k=0}^{\infty}$、$\eta_{\mathbf{y}} = \{\eta_{\mathbf{y}}(k)\}_{k=0}^{\infty}$以及随机序列 $\varrho = \{\varrho(k)\}_{k=0}^{\infty}$唯一确定，其中 $\varrho(k) \in \mathbb{R}^{n\times d}$ 是一个矩阵，其元素 $[\varrho(k)]_{ij}$ 是 $[\sigma_{\mathbf{x}}(k)]_{ij}$的压缩扰动。我们用函数 $Z_{\mathcal{F}}$ 表示这一关系，即 $\mathcal{H} = Z_{\mathcal{F}}(\eta_{\mathbf{x}}, \eta_{\mathbf{y}}, \varrho)$，其中 $\mathcal{F} = \{\mathbf{x}(0)\ W, \mathcal{S}\}$。根据定义2，为证明代价函数 f_{i_0}的差分隐私性，我们需要证明，对于任意观测 $\mathcal{H} \subseteq$ Range(CPGT)以及任意一对相邻的代价函数集 $\mathcal{S}^{(1)}$ 和 $\mathcal{S}^{(2)}$，下列不等式成立:

$$\mathbb{P}\{(\eta_x, \eta_y, \varrho) \in \Psi | Z_{\mathcal{F}^{(1)}}(\eta_x, \eta_y, \varrho) \in \mathcal{H}\} \le e^{\epsilon}\mathbb{P}\{(\eta_x, \eta_y, \varrho) \in \Psi | Z_{\mathcal{F}^{(2)}}(\eta_x, \eta_y, \varrho) \in \mathcal{H}\},$$

其中 $\mathcal{F}^{(l)} = \{\mathbf{x}(0)\ W、\mathcal{S}^{(l)}\}$、$l = 1$、2 和 Ψ 表示样本空间。因此，必须保证$Z_{\mathcal{F}^{(1)}}(\eta_x$、$\eta_y$、$\varrho) = Z_{\mathcal{F}^{(2)}}(\eta_x$、$\eta_y$、$\varrho)$，即，

$$C(x_i^{a-c,(1)}(k),\varrho(k)) = C(x_i^{a-c,(2)}(k),\varrho(k)),$$
$$C(y_i^{a-c,(1)}(k),\varrho(k)) = C(y_i^{a-c,(2)}(k),\varrho(k)),$$

对于 $\forall i \in \mathcal{V}$ 以及任意 $k \geq 0$，其中

$$x_i^{a-c,(1)}(k) = x_i^{a,(1)}(k) - x_i^{c,(1)}(k-1),$$
$$y_i^{a-c,(1)}(k) = y_i^{a,(1)}(k) - y_i^{c,(1)}(k-1).$$

由于 $x^{c}{}_{,i}{}^{(l)}(k)) = y^{c}{}_{,i}{}^{(l)}(k) = 0, k < 0, l = 1, 2$，根据 (11)– (12)，我们有

$$x_i^{c,(1)}(k)) = x_i^{c,(2)}(k)), y_i^{c,(1)}(k)) = y_i^{c,(2)}(k)).$$

于是可得

$$\begin{aligned} f_c(x_i^{a-c,(1)}(k),\varrho(k)) &= f_c(x_i^{a-c,(2)}(k),\varrho(k)), \\ f_c(y_i^{a-c,(1)}(k),\varrho(k)) &= f_c(y_i^{a-c,(2)}(k),\varrho(k)), \end{aligned} \tag{C.1}$$

若 $x^{a}{}_{,i}{}^{(1)}(k) = x^{a}{}_{,i}{}^{(2)}(k)$ 且 $y^{a}{}_{,i}{}^{(1)}(k) = y^{a}{}_{,i}{}^{(2)}(k)$，对于 $\forall i \in \mathcal{V}$。那么根据条件概率的性质，我们有

$$\begin{aligned} &\frac{\mathbb{P}\{(\eta_x,\eta_y,\varrho) \in \Psi | Z_{\mathcal{F}^{(1)}}(\eta_x,\eta_y,\varrho) \in \mathcal{H}\}}{\mathbb{P}\{(\eta_x,\eta_y,\varrho) \in \Psi | Z_{\mathcal{F}^{(2)}}(\eta_x,\eta_y,\varrho) \in \mathcal{H}\}} \\ &\leq \frac{\mathbb{P}\{(\eta_x,\eta_y,\varrho) \in \Psi | Z_{\mathcal{F}^{(1)}}(\eta_x,\eta_y,\varrho) \in \mathcal{H}\}}{\mathbb{P}\{(\eta_x,\eta_y,\varrho) \in \Psi | Z_{\mathcal{F}^{(2)}}(\eta_x,\eta_y,\varrho) \in \mathcal{H}, E_1\}}, \end{aligned} \tag{C.2}$$

其中 $E_1 = \cup_{k=0}^{\infty}\{x_i^{a,(2)}(k) = x^{a}{}_{,i}{}^{(1)}(k)\ y^{a}{}_{,i}{}^{(2)}(k) = y^{a}{}_{,i}{}^{(1)}(k)\ \forall i \in \mathcal{V}\}$ 是一个事件。因此，根据 (C.1)–(C.2)，可以按照 Ding 等（2021）中定理5的证明方式完成证明。

参考文献

Alistarh, D., Grubic, D., Li, J., Tomioka, R., and Vojnovic, M.（2017）。QSGD：通过梯度量化与编码实现高效通信的随机梯度下降。载于《神经信息处理系统进展》，1707–1718。Beznosikov, A., Horv´ath, S., Richt´arik, P., and Safaryan, M.（2020）。关于分布式学习中的有偏压缩。arXiv 预印本 arXiv:2002.12410。Boyd, S. and Vandenberghe, L.（2004）。《凸优化》。剑桥大学出版社。Chen, X., Huang, L., He, L., Dey, S., and Shi, L.（2021）。一种通过状态分解在有向网络中实现分布式优化的差分隐私方法。arXiv 预印本 arXiv:2107.04370。Ding, T., Zhu, S., He, J., Chen, C., and Guan, X.（2021）。多智能体系统中通过状态与方向扰动实现差分隐私分布式优化。IEEE 自动控制汇刊，67(2)，722–737。Dougherty, S. and Guay, M.（2016）。一种用于传感器网络分布式优化的极值搜索控制器。IEEE 自动控制汇刊，62(2)，928–933。Dwork, C.（2008）。差分隐私：研究成果综述。载于《计算模型理论与应用国际会议》，1–19。Horn, R.A. and Johnson, C.R.（2012）。《矩阵分析》。剑桥大学出版社。Huang, Z., Mitra, S., and Vaidya, N.（2015）。差分隐私分布式优化。载于《分布式计算与网络国际会议论文集》，1–10。Kajiyama, Y., Hayashi, N., and Takai, S.（2020）。基于共识的量化优化的线性收敛适用于光滑强凸成本函数。IEEE 自动控制汇刊，66(3)，1254–1261。Koloskova, A., Lin, T., Stich, S. U., and Jaggi, M.（2019）。采用任意通信压缩的去中心化深度学习。载于《国际学习表征会议》。Li, X., Yi, X., and Xie, L.（2020）。具有耦合不等式约束的多智能体网络分布式在线优化。IEEE 自动控制汇刊，66(8)，3575–3591。Liao, Y., Li, Z., Huang, K., and Pu, S.（2022）。一种具有线性收敛性的去中心化优化压缩梯度跟踪方法。IEEE 自动控制汇刊，67(10)，1254–1261。Lu, Y. and Zhu, M.（2018）。使用同态加密保护隐私的分布式优化。Automatica，96，314–325。Nedic, A. and Ozdaglar, A.（2009）。多智能体优化的分布式次梯度方法。IEEE 自动控制汇刊，54(1)，48–61。Qu, G. and Li, N.（2017）。利用光滑性加速分布式优化。IEEE 网络控制系统汇刊，5(3)，1245–1260。Reisizadeh, A., Mokhtari, A., Hassani, H., and Pedarsani, R.（2019）。一种精确的量化去中心化梯度下降算法。IEEE 信号处理汇刊，67(19)，4934–4947。Shi, W., Ling, Q., Wu, G., and Yin, W.（2015）。EXTRA：一种用于去中心化共识优化的精确一阶算法。SIAM 优化期刊，25(2)，944–966。Taheri, H., Mokhtari, A., Hassani, H., and Pedarsani, R.（2020）。有向图上的量化去中心化随机学习。载于《国际机器学习会议》，9324–9333。Tsianos, K.I., Lawlor, S., and Rabbat, M.G.（2012）。基于共识的分布式优化：大规模机器学习中的实际问题与应用。载于《年度阿勒顿通信、控制与计算会议》，1543–1550。Wang, Y. and Ba¸sar, T.（2022）。量化赋能的去中心化随机优化隐私保护。IEEE 自动控制汇刊。Xiong, Y., Wu, L., You, K., and Xie, L.（2021）。有向网络中具有线性收敛性的量化分布式梯度跟踪算法。arXiv 预印本 arXiv:2104.03649。Xu, J., Zhu, S., Soh, Y.C., and Xie, L.（2017）。随机网络上异步分布式梯度方法的收敛性。IEEE 自动控制汇刊，63(2)，434–448。Zhu, J., Xu, C., Guan, J., and Wu, D.O.（2018）。时变有向网络上的差分隐私分布式在线算法。IEEE 信号与信息处理网络汇刊，4(1)，4–17。Zhu, L., Liu, Z., and Han, S.（2019）。梯度信息的深度泄露。载于《神经信息处理系统进展》，14774–14784。