

基于外推自适应步长的压缩联邦学习优化算法

张耀威

河北工业大学理学院, 天津

收稿日期: 2025年8月20日; 录用日期: 2025年9月20日; 发布日期: 2025年10月9日

摘要

由于联邦学习中的通信成本高昂, 依赖通信压缩的方法正日益受到关注。本文基于外推自适应步长和压缩联邦学习框架提出了外推压缩联邦学习算法(ExpFedCom)。ExpFedCom算法支持通信压缩并在凸、强凸以及非凸条件下实现了快速收敛。数值实验表明ExpFedCom算法既实现了快速收敛又节省了通信成本, 这在实际生活中有着较好的应用价值。

关键词

联邦学习, 通信成本, 通信压缩, 外推自适应步长

Extrapolation Adaptive Step Size for Compressed Federated Learning Optimization Algorithm

Yaowei Zhang

School of Sciences, Hebei University of Technology, Tianjin

Received: August 20, 2025; accepted: September 20, 2025; published: October 9, 2025

Abstract

The high communication costs in federated learning have motivated increasing interest in communication compression methods. This paper proposes Extrapolated Federated Compression (ExpFedCom), a novel algorithm that integrates extrapolation adaptive step sizes with a communication compression federated learning framework. The ExpFedCom algorithm supports communication compression while achieving faster convergence rates under convex, strongly convex, and nonconvex settings. Numerical experiments demonstrate that ExpFedCom simultaneously attains faster convergence and significant communication savings, exhibiting strong practical potential for real-world applications.

Keywords

Federated Learning, Communication Costs, Communication Compression, Extrapolation Adaptive Step Sizes

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

传统的集中式学习通常将数据集中存储在中心服务器上，并在该服务器上完成模型训练。然而，这种训练方式往往伴随着严重的数据隐私风险。与之相比，联邦学习[1][2]在多个设备或节点上对分布式数据集进行模型训练，即实现跨地域分布异构设备的协同模型训练，且在全局模型训练过程中，无需将本地数据传输至中心服务器。因此，联邦学习能够在保护数据隐私[3]的前提下，利用分布式数据进行高效的模型训练，现已成为大规模机器学习领域的重要研究方向。联邦学习通常解决的是以下优化问题：

$$\min_{x \in \mathbb{R}^d} f(x) := \frac{1}{N} \sum_{i=1}^N f_i(x), \quad (1)$$

其中 $f_i(x) := E_{d_j \sim D_i} f_{i,j}(x; d_j)$ 是第 i 个客户端在本地数据集 D_i 上的损失函数。

在联邦学习系统中，由于本地数据集的规模越来越大且服务器和边缘设备之间的带宽有限，联邦学习面临着通信成本的问题，所以实现通信效率、隐私保护和模型性能三者之间的最优平衡是联邦学习的一个关键目标。

为了解决联邦通信成本问题，通常有两种解决方法：第一种是在本地进行多次更新，例如联邦平均算法(FedAvg) [4]，随机控制平均算法(SCAFFOLD) [5]；第二种是对本地通信信息进行压缩，例如分布式压缩梯度算法(DCGD) [6]，带压缩的联邦平均算法(FedCOM) [7]，基于方差缩减和梯度追踪的 FedCOM 算法[8]。

由于自适应的方法广泛应用于联邦学习当中，例如，FEDADAM 算法[9]和 FedExp 算法[10]。本文受外推自适应步长的启发将其推广到了压缩联邦学习当中，并提出了外推压缩联邦学习算法(ExpFedCom)。

$$\lambda_t = \frac{\sum_{i=1}^N \|f_i(x^t) - x^t\|^2}{N \left\| \frac{1}{N} \sum_{i=1}^N f_i(x^t) - x^t \right\|^2}. \quad (\text{外推法步长})$$

2. ExpFedCom 算法

首先给出无偏压缩算子的定义：

定义 1 如果一个随机映射 $C: \mathbb{R}^d \mapsto \mathbb{R}^d$ 是 ω 压缩算子，若对任意 $x \in \mathbb{R}^d$ ，满足：

$$E[C(x)] = x, \quad E[\|C(x) - x\|^2] \leq \omega \|x\|^2.$$

若模型未压缩，即 $C(x) \equiv x$ ，则 $\omega = 0$ 。

考虑基于压缩的外推自适应步长：

$$\eta_t = \frac{\sum_{i=1}^N \|C(x^t - x'_{i,K})\|^2}{N \left\| \frac{1}{N} \sum_{i=1}^N C(x^t - x'_{i,K}) \right\|^2},$$

其中 K 是本地更新次数。提出的这一新步长通过比较客户端本地更新与全局平均更新的差异，来度量模型更新的“有效性”。如果客户端的更新方向高度一致，平均更新的范数会接近于单个更新的均值，比值较小，说明更新是有效的；反之，如果本地更新存在明显分歧，则平均更新的合力较弱，比值增大，提示更新效果不足。此外这一新步长不依赖于特定参数(例如李普希茨常数或压缩系数)，还有效节省了调节最优步长的时间。我们将提出的外推自适应步长与压缩联邦平均框架相结合，提出外推压缩联邦平均算法(ExpFedCom)，见算法 1。

算法 1: 外推压缩联邦平均算法(ExpFedCom)

输入: 服务器初始化全局模型 x^0 和本地步长 α

- 1: **for** $t = 0, 1, 2, \dots$ **do**
 - 2: 客户端 $i \in [N]$ 并行如下步骤
 - 3: 初始化本地模型 $x'_{i,0} = x^t$
 - 4: **for** $k = 0, 1, 2, \dots, K-1$ **do**
 - 5: 计算梯度信息 $\nabla f_i(x'_{i,k})$
 - 6: 更新本地模型 $x'_{i,k+1} = x'_{i,k} - \alpha \nabla f_i(x'_{i,k})$
 - 7: **end for**
 - 8: 将压缩信息 $\Delta'_i = C(x^t - x'_{i,K})$ 传输给服务器
 - 9: 服务器聚合 $\Delta^t = \frac{1}{N} \sum_{i=1}^N \Delta'_i$
 - 10: 计算外推步长 $\eta_t = \max \left\{ 1, \frac{\sum_{i=1}^N \|\Delta'_i\|^2}{2N(\|\Delta^t\|^2 + \varepsilon)} \right\}$
 - 11: 服务器更新全局模型 $x^{t+1} = x^t - \eta_t \Delta^t$
 - 12: **end for**
-

在算法中，我们在提出的新步长中引入了 $\frac{1}{2}$ ，这可以抑制过大的步长的扩张，使得算法在面对高度异质的数据和压缩时更稳定，同时仍能保留自适应调节的特性。在步长分母中引入 ε 是为了防止分母为 0 的情况。将步长与 1 比较取最大，是为了避免步长过小，而导致收敛过慢或更新停滞，这也使得步长有下界。注意到，当 η_t 取常数步长时，ExpFedCom 算法可以退化为 FedCOM 算法[7]。

3. 收敛性分析

假设 x^* 是问题(1)的全局最优解且满足 $f(x^*) = f^*$ 以及对任意 $i \in [N]$ ， $\nabla f_i(x^*) = 0$ 。文章出现的 E 记为对所有随机变量取期望。本节通过一些引理，在凸以及强凸情况下对 ExpFedCom 算法做出了收敛性分析。

3.1. 凸情况分析

引理 1 假设对任意 $i \in [N]$, f_i 是 L -光滑的, 则对任意 $t \geq 0$ 和 $\alpha \leq \frac{1}{L}$ 有

$$\|x'_{i,K} - x^*\|^2 \leq \|x' - x^*\|^2 - \frac{1}{KL\alpha} \|x' - x'_{i,K}\|^2.$$

证明: 对任意 $t \geq 0$, 有

$$\begin{aligned} \|x'_{i,k} - x^*\|^2 &= \|x'_{i,k-1} - \alpha \nabla f_i(x'_{i,k-1}) - x^*\|^2 \\ &= \|x'_{i,k-1} - x^*\|^2 - 2\alpha \langle \nabla f_i(x'_{i,k-1}), x'_{i,k-1} - x^* \rangle + \alpha^2 \|\nabla f_i(x'_{i,k-1})\|^2 \\ &\leq \|x'_{i,k-1} - x^*\|^2 - \alpha \left(\frac{2}{L} - \alpha \right) \|\nabla f_i(x'_{i,k-1})\|^2 \\ &\leq \|x'_{i,k-1} - x^*\|^2 - \frac{\alpha}{L} \|\nabla f_i(x'_{i,k-1})\|^2 \\ &\leq \|x'_{i,0} - x^*\|^2 - \frac{\alpha}{L} \sum_{j=0}^{k-1} \|\nabla f_i(x'_{i,j})\|^2, \end{aligned}$$

其中第一个不等式利用了 f_i 是 L -光滑的, 第二个不等式利用了 $\alpha \leq \frac{1}{L}$ 。令 $k = K$ 则有

$$\begin{aligned} \|x'_{i,K} - x^*\|^2 &\leq \|x'_{i,0} - x^*\|^2 - \frac{\alpha}{L} \sum_{k=0}^{K-1} \|\nabla f_i(x'_{i,k})\|^2 \\ &= \|x' - x^*\|^2 - \frac{1}{L\alpha} \sum_{k=0}^{K-1} \|x'_{i,k} - x'_{i,k+1}\|^2 \\ &\leq \|x' - x^*\|^2 - \frac{1}{KL\alpha} \left\| \sum_{k=0}^{K-1} (x'_{i,k} - x'_{i,k+1}) \right\|^2 \\ &= \|x' - x^*\|^2 - \frac{1}{KL\alpha} \|x' - x'_{i,K}\|^2, \end{aligned}$$

其中第三个不等式利用了均值不等式, 证毕。

引理 2 假设对任意 $i \in [N]$, f_i 是 L -光滑的, 则对任意 $t \geq 0$ 和 $\alpha \leq \min\left\{\frac{1}{L}, \frac{1}{KL\omega}\right\}$ 有

$$\frac{\sum_{i=1}^N \|\Delta'_i\|^2}{2N \|\Delta'\|^2} \leq \eta_t \leq \frac{\sum_{i=1}^N \|\Delta'_i\|^2}{N \|\Delta'\|^2}.$$

证明: 对任意 $t \geq 0$, 有

$$\begin{aligned} \mathbb{E}[\|x^{t+1} - x^*\|^2] &= \mathbb{E}[\|x^t - \eta_t \Delta' - x^*\|^2] \\ &= \mathbb{E}[\|x^t - x^*\|^2 - 2\eta_t \langle \Delta', x^t - x^* \rangle + \eta_t^2 \|\Delta'\|^2]. \end{aligned}$$

对 η_t 取微分, 有

$$\mathbb{E}[2\eta_t \|\Delta'\|^2 - 2\langle \Delta', x^t - x^* \rangle] = 0,$$

所以

$$\begin{aligned}\eta_t &= \frac{\mathbb{E} \left[2 \left\langle \frac{1}{N} \sum_{i=1}^N \Delta'_i, x' - x^* \right\rangle \right]}{2 \|\Delta^t\|^2} \\ &= \frac{\sum_{i=1}^N \mathbb{E} \left[\|\Delta'_i\|^2 + \|x' - x^*\|^2 - \|\Delta'_i - (x' - x^*)\|^2 \right]}{2N \|\Delta^t\|^2}.\end{aligned}$$

结合 Δ'_i 的定义, 有

$$\begin{aligned}\eta_t &= \frac{\sum_{i=1}^N \mathbb{E} \left[\|\Delta'_i\|^2 + \|x' - x^*\|^2 - \|x' - x^* - C(x' - x'_{i,K})\|^2 \right]}{2N \|\Delta^t\|^2} \\ &= \frac{\sum_{i=1}^N \mathbb{E} \left[\|\Delta'_i\|^2 + \|x' - x^*\|^2 - \|x' - x'_{i,K} - C(x' - x'_{i,K})\|^2 - \|x'_{i,K} - x^*\|^2 \right]}{2N \|\Delta^t\|^2} \\ &\geq \frac{\sum_{i=1}^N \left(\|\Delta'_i\|^2 + \|x' - x^*\|^2 - \omega \|x' - x'_{i,K}\|^2 - \|x'_{i,K} - x^*\|^2 \right)}{2N \|\Delta^t\|^2},\end{aligned}$$

其中第四个等式和最后一个不等式利用了定义 1。结合引理 1, 则有

$$\begin{aligned}\eta_t &\geq \frac{\sum_{i=1}^N \left(\|\Delta'_i\|^2 + \|x' - x^*\|^2 - \omega \|x' - x'_{i,K}\|^2 - \|x' - x^*\|^2 + \frac{1}{KL\alpha} \|x' - x'_{i,K}\|^2 \right)}{2N \|\Delta^t\|^2} \\ &= \frac{\sum_{i=1}^N \left(\|\Delta'_i\|^2 - \left(\omega - \frac{1}{KL\alpha} \right) \|x' - x'_{i,K}\|^2 \right)}{2N \|\Delta^t\|^2} \\ &\geq \frac{\sum_{i=1}^N \|\Delta'_i\|^2}{2N \|\Delta^t\|^2},\end{aligned}$$

其中最后一个不等式利用了 $\alpha \leq \min \left\{ \frac{1}{L}, \frac{1}{KL\omega} \right\}$ 。又因为 $\eta_t = \max \left\{ 1, \frac{\sum_{i=1}^N \|\Delta'_i\|^2}{2N (\|\Delta^t\|^2 + \varepsilon)} \right\}$, 且下述不等式成立

$$\frac{\sum_{i=1}^N \|\Delta'_i\|^2}{2N \|\Delta^t\|^2} \leq \frac{\sum_{i=1}^N \|\Delta'_i\|^2}{N \|\Delta^t\|^2}, \quad \frac{\sum_{i=1}^N \|\Delta'_i\|^2}{N \|\Delta^t\|^2} \geq 1,$$

所以有

$$\frac{\sum_{i=1}^N \|\Delta'_i\|^2}{2N \|\Delta^t\|^2} \leq \eta_t \leq \frac{\sum_{i=1}^N \|\Delta'_i\|^2}{N \|\Delta^t\|^2}.$$

证毕。

引理 3 假设对任意 $i \in [N]$, f_i 是 L -光滑的, 则对任意 $t \geq 0$ 有

$$\frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \|\nabla f_i(x'_{i,k})\|^2 \leq \frac{2L^2}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \|x^t - x'_{i,k}\|^2 + 4KL(f(x^t) - f^*).$$

证明：对任意 $t \geq 0$ ，根据均值不等式有

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \|\nabla f_i(x'_{i,k})\|^2 &\leq \frac{2}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \left(\|\nabla f_i(x'_{i,k}) - \nabla f_i(x^t)\|^2 + \|\nabla f_i(x^t)\|^2 \right) \\ &= \frac{2}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \left(\|\nabla f_i(x'_{i,k}) - \nabla f_i(x^t)\|^2 + \|\nabla f_i(x^t) - \nabla f_i(x^*)\|^2 \right). \end{aligned}$$

结合 f_i 是 L -光滑的，有

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \|\nabla f_i(x'_{i,k})\|^2 &\leq \frac{2L^2}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \|x^t - x'_{i,k}\|^2 + \frac{4KL}{N} \sum_{i=1}^N (f_i(x^t) - f_i(x^*)) \\ &= \frac{2L^2}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \|x^t - x'_{i,k}\|^2 + 4KL(f(x^t) - f^*). \end{aligned}$$

证毕。

引理 4 假设对任意 $i \in [N]$ ， f_i 是 L -光滑的，则对任意 $t \geq 0$ 和 $\alpha \leq \min\left\{\frac{1}{2KL}, \frac{1}{KL\omega}\right\}$ 有

$$\frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \|x^t - x'_{i,k}\|^2 \leq 8L(K-1)K^2\alpha^2(f(x^t) - f^*).$$

证明：对任意 $t \geq 0$ ，根据均值不等式有

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \|x^t - x'_{i,k}\|^2 &= \alpha^2 \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \left\| \sum_{c=0}^{k-1} \nabla f_i(x'_{i,c}) \right\|^2 \\ &\leq \alpha^2 \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} k \sum_{c=0}^{k-1} \|\nabla f_i(x'_{i,c})\|^2 \\ &\leq K(K-1)\alpha^2 \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \|\nabla f_i(x'_{i,k})\|^2, \end{aligned}$$

其中第一个不等式利用了均值不等式，第二个不等式利用了下述等式关系：

$$\sum_{k=0}^{K-1} k \sum_{c=0}^{k-1} \|\nabla f_i(x'_{i,c})\|^2 = \sum_{k=0}^{K-2} \left(\frac{K(K-1)}{2} - \frac{k(k+1)}{2} \right) \|\nabla f_i(x'_{i,k})\|^2.$$

结合引理 3 有

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \|x^t - x'_{i,k}\|^2 &\leq K(K-1)\alpha^2 \left(\frac{2L^2}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \|x^t - x'_{i,k}\|^2 + 4KL(f(x^t) - f^*) \right) \\ &= 2K(K-1)L^2\alpha^2 \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \|x^t - x'_{i,k}\|^2 + 4L(K-1)K^2\alpha^2(f(x^t) - f^*) \\ &\leq \frac{1}{2N} \sum_{i=1}^N \sum_{k=0}^{K-1} \|x^t - x'_{i,k}\|^2 + 4L(K-1)K^2\alpha^2(f(x^t) - f^*), \end{aligned}$$

其中最后一个不等式利用了 $\alpha \leq \min\left\{\frac{1}{2KL}, \frac{1}{KL\omega}\right\}$ 。因此

$$\frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \|x^t - x'_{i,k}\|^2 \leq 8L(K-1)K^2\alpha^2(f(x^t) - f^*).$$

证毕。

定理 1 假设对任意 $i \in [N]$, f_i 是光滑的且凸的, 则对任意 $t \geq 0$ 和 $\alpha \leq \frac{1}{7KL(1+\omega)}$ 有

$$\mathbb{E}[f(\bar{x}) - f^*] \leq \frac{7\|x^0 - x^*\|^2}{\alpha K \inf_t \eta_t T},$$

其中设 $p_t = \frac{\eta_t}{\sum_{i=0}^{T-1} \eta_i}$, $\bar{x}$ 是从 $\{x^0, x^1, \dots, x^{T-1}\}$ 中以 p_t 概率选取。

证明: 对任意 $t \geq 0$, 有

$$\begin{aligned} \mathbb{E}[\|x^{t+1} - x^*\|^2] &= \mathbb{E}\left[\|x^t - x^*\|^2 - 2\eta_t \left\langle x^t - x^*, \frac{1}{N} \sum_{i=1}^N C(x^t - x'_{i,K}) \right\rangle + \eta_t^2 \left\| \frac{1}{N} \sum_{i=1}^N C(x^t - x'_{i,K}) \right\|^2\right] \\ &= \mathbb{E}\left[\|x^t - x^*\|^2 - 2\eta_t \frac{1}{N} \sum_{i=1}^N \langle x^t - x^*, x^t - x'_{i,K} \rangle + \eta_t^2 \left\| \frac{1}{N} \sum_{i=1}^N C(x^t - x'_{i,K}) \right\|^2\right]. \end{aligned}$$

结合引理 2, 有

$$\begin{aligned} \mathbb{E}[\|x^{t+1} - x^*\|^2] &\leq \mathbb{E}\left[\|x^t - x^*\|^2 - 2\eta_t \frac{1}{N} \sum_{i=1}^N \langle x^t - x^*, x^t - x'_{i,K} \rangle + \eta_t \frac{1}{N} \sum_{i=1}^N \|C(x^t - x'_{i,K})\|^2\right] \\ &\leq \mathbb{E}\left[\|x^t - x^*\|^2 - 2\eta_t \frac{1}{N} \sum_{i=1}^N \langle x^t - x^*, x^t - x'_{i,K} \rangle + (1+\omega)\eta_t \frac{1}{N} \sum_{i=1}^N \|x^t - x'_{i,K}\|^2\right] \\ &= \mathbb{E}\left[\|x^t - x^*\|^2 - 2\eta_t \alpha \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \langle x^t - x^*, \nabla f_i(x'_{i,k}) \rangle + (1+\omega)\eta_t \alpha^2 \frac{1}{N} \sum_{i=1}^N \left\| \sum_{k=0}^{K-1} \nabla f_i(x'_{i,k}) \right\|^2\right] \\ &\leq \mathbb{E}\left[\|x^t - x^*\|^2 - 2\eta_t \alpha \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \langle x^t - x^*, \nabla f_i(x'_{i,k}) \rangle + (1+\omega)K\eta_t \alpha^2 \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \|\nabla f_i(x'_{i,k})\|^2\right], \end{aligned}$$

其中第二个不等式利用了定义 1, 最后一个不等式利用了均值不等式。因此

$$\begin{aligned} \mathbb{E}[\|x^{t+1} - x^*\|^2] &\leq \mathbb{E}\left[\|x^t - x^*\|^2 - 2\eta_t \alpha \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \left(\langle x^t - x'_{i,k}, \nabla f_i(x'_{i,k}) \rangle + \langle x'_{i,k} - x^*, \nabla f_i(x'_{i,k}) \rangle \right) \right. \\ &\quad \left. + (1+\omega)K\eta_t \alpha^2 \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \|\nabla f_i(x'_{i,k})\|^2 \right] \\ &\leq \mathbb{E}\left[\|x^t - x^*\|^2 + 2\eta_t \alpha \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \left(f_i(x'_{i,k}) - f_i(x^t) + \frac{L}{2} \|x^t - x'_{i,k}\|^2 + \langle x^* - x'_{i,k}, \nabla f_i(x'_{i,k}) \rangle \right) \right. \\ &\quad \left. + (1+\omega)K\eta_t \alpha^2 \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \|\nabla f_i(x'_{i,k})\|^2 \right] \\ &\leq \mathbb{E}\left[\|x^t - x^*\|^2 + 2\eta_t \alpha \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \left(f_i(x'_{i,k}) - f_i(x^t) + \frac{L}{2} \|x^t - x'_{i,k}\|^2 + f_i(x^*) - f_i(x'_{i,k}) \right) \right. \\ &\quad \left. + (1+\omega)K\eta_t \alpha^2 \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \|\nabla f_i(x'_{i,k})\|^2 \right] \\ &\leq \mathbb{E}\left[\|x^t - x^*\|^2 + 2\eta_t \alpha \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \left(f_i(x^*) - f_i(x^t) + \frac{L}{2} \|x^t - x'_{i,k}\|^2 \right) \right. \\ &\quad \left. + (1+\omega)K\eta_t \alpha^2 \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \|\nabla f_i(x'_{i,k})\|^2 \right], \end{aligned}$$

其中第二个不等式利用了 f_i 是 L -光滑的, 第三个不等式利用了 f_i 是凸的。结合引理 3 有

$$\begin{aligned}
\mathbb{E}\left[\|x^{t+1}-x^*\|^2\right] &\leq \mathbb{E}\left[\|x^t-x^*\|^2-2K\eta_t\alpha\left(f(x^t)-f^*\right)+L\eta_t\alpha\frac{1}{N}\sum_{i=1}^N\sum_{k=0}^{K-1}\|x^t-x_{i,k}^t\|^2\right. \\
&\quad \left.+(1+\omega)K\eta_t\alpha^2\left(\frac{2L^2}{N}\sum_{i=1}^N\sum_{k=0}^{K-1}\|x^t-x_{i,k}^t\|^2+4KL\left(f(x^t)-f^*\right)\right)\right] \\
&\leq \mathbb{E}\left[\|x^t-x^*\|^2-2K\eta_t\alpha(1-2KL(1+\omega)\alpha)\left(f(x^t)-f^*\right)\right. \\
&\quad \left.+L\eta_t\alpha(1+2KL(1+\omega)\alpha)\frac{1}{N}\sum_{i=1}^N\sum_{k=0}^{K-1}\|x^t-x_{i,k}^t\|^2\right].
\end{aligned}$$

结合引理 4 有

$$\begin{aligned}
\mathbb{E}\left[\|x^{t+1}-x^*\|^2\right] &\leq \mathbb{E}\left[\|x^t-x^*\|^2-2K\eta_t\alpha(1-2KL(1+\omega)\alpha)\left(f(x^t)-f^*\right)\right. \\
&\quad \left.+8\eta_t(K-1)L^2K^2\alpha^3(1+2KL(1+\omega)\alpha)\left(f(x^t)-f^*\right)\right].
\end{aligned}$$

取 $\alpha \leq \min\left\{\frac{1}{3KL}, \frac{1}{KL\omega}\right\}$, 则有

$$\begin{aligned}
\mathbb{E}\left[\|x^{t+1}-x^*\|^2\right] &\leq \mathbb{E}\left[\|x^t-x^*\|^2-K\eta_t\alpha(2-4KL(1+\omega)\alpha)\left(f(x^t)-f^*\right)\right. \\
&\quad \left.+K\eta_t\alpha(1+2KL(1+\omega)\alpha)\left(f(x^t)-f^*\right)\right] \\
&= \mathbb{E}\left[\|x^t-x^*\|^2-K\eta_t\alpha(1-6KL(1+\omega)\alpha)\left(f(x^t)-f^*\right)\right] \\
&\leq \mathbb{E}\left[\|x^t-x^*\|^2-\frac{K\eta_t\alpha}{7}\left(f(x^t)-f^*\right)\right],
\end{aligned}$$

其中最后一个不等式利用了 $\alpha \leq \frac{1}{7KL(1+\omega)}$ 。因此有

$$\mathbb{E}\left[\sum_{t=0}^{T-1}\eta_t\left(f(x^t)-f^*\right)\right] \leq \frac{7\|x^0-x^*\|^2}{\alpha K}.$$

对任意 $t=0,1,\dots,T-1$, 设 $p_t = \frac{\eta_t}{\sum_{t=0}^{T-1}\eta_t}$ 且, $\bar{x}$ 是从 $\{x^0, x^1, \dots, x^{T-1}\}$ 中以 p_t 概率选取, 则有

$$\mathbb{E}\left[f(\bar{x})-f^*\right] \leq \frac{7\|x^0-x^*\|^2}{\alpha K \sum_{t=0}^{T-1}\eta_t} \leq \frac{7\|x^0-x^*\|^2}{\alpha K \inf_t \eta_t T}.$$

证毕。

3.2. 强凸情况分析

定理 2 假设对任意 $i \in [N]$, f_i 是 L -光滑的且 μ -强凸的, 则对任意 $t \geq 0$ 和

$$\alpha \leq \min\left\{\frac{1}{7KL(1+\omega)}, \frac{14}{\mu K \inf_t \eta_t}\right\} \text{ 有}$$

$$\mathbb{E}\left[\|x^{t+1}-x^*\|^2\right] \leq \left(1-\frac{\mu K \inf_t \eta_t \alpha}{14}\right)^{t+1} \|x^0-x^*\|^2.$$

证明: 根据定理 1 证明, 当 $\alpha \leq \frac{1}{7KL(1+\omega)}$, 有

$$\begin{aligned}
\mathbb{E}[\|x^{t+1} - x^*\|^2] &\leq \mathbb{E}\left[\|x^t - x^*\|^2 - \frac{K \inf_t \eta_t \alpha}{7} (f(x^t) - f^*)\right] \\
&\leq \mathbb{E}\left[\left(1 - \frac{\mu K \inf_t \eta_t \alpha}{14}\right) \|x^t - x^*\|^2\right] \\
&\leq \left(1 - \frac{\mu K \inf_t \eta_t \alpha}{14}\right)^{t+1} \|x^0 - x^*\|^2,
\end{aligned}$$

其中第二个不等式利用了 f_t 的强凸性，最后一个不等式利用了 $\alpha \leq \min\left\{\frac{1}{7KL(1+\omega)}, \frac{14}{\mu K \inf_t \eta_t}\right\}$ ，证毕。

3.3. 非凸情况分析

假设 1 假设梯度方差有界，即对任意 $i \in [N]$ ，存在 $\sigma \geq 0$ 使得下式成立

$$\frac{1}{N} \sum_{i=1}^N \|\nabla f_i(x) - \nabla f(x)\|^2 \leq \sigma^2.$$

引理 5 假设对任意 $i \in [N]$ ， f_i 是 L -光滑的，假设梯度方差有界，则对任意 $t \geq 0$ 和 $\alpha \leq \frac{1}{2KL}$ 有

$$\frac{1}{N} \sum_{i=1}^N \left\| \nabla f_i(x^t) - \frac{1}{K} \sum_{k=0}^{K-1} \nabla f_i(x_{i,k}^t) \right\|^2 \leq \frac{3K(K-1)L^2\alpha^2\sigma^2}{1-3K(K-1)L^2\alpha^2} + \frac{3K(K-1)L^2\alpha^2 \|\nabla f(x^t)\|^2}{1-3K(K-1)L^2\alpha^2}.$$

证明：利用均值不等式，有

$$\begin{aligned}
\frac{1}{N} \sum_{i=1}^N \left\| \nabla f_i(x^t) - \frac{1}{K} \sum_{k=0}^{K-1} \nabla f_i(x_{i,k}^t) \right\|^2 &\leq \frac{1}{N} \sum_{i=1}^N \frac{1}{K} \sum_{k=0}^{K-1} \left\| \nabla f_i(x^t) - \nabla f_i(x_{i,k}^t) \right\|^2 \\
&\leq \frac{L^2}{K} \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \|x^t - x_{i,k}^t\|^2 \\
&= \frac{L^2\alpha^2}{K} \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \left\| \sum_{c=0}^{k-1} \nabla f_i(x_{i,c}^t) \right\|^2 \\
&\leq \frac{L^2\alpha^2}{K} \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} k \sum_{c=0}^{k-1} \left\| \nabla f_i(x_{i,c}^t) \right\|^2
\end{aligned}$$

其中第二个不等式利用了 f_i 是 L -光滑的，最后一个不等式利用了均值不等式。再根据下述等式关系：

$$\sum_{k=0}^{K-1} k \sum_{c=0}^{k-1} \left\| \nabla f_i(x_{i,c}^t) \right\|^2 = \sum_{k=0}^{K-2} \left(\frac{K(K-1)}{2} - \frac{k(k+1)}{2} \right) \left\| \nabla f_i(x_{i,k}^t) \right\|^2,$$

则有

$$\begin{aligned}
\frac{1}{N} \sum_{i=1}^N \left\| \nabla f_i(x^t) - \frac{1}{K} \sum_{k=0}^{K-1} \nabla f_i(x_{i,k}^t) \right\|^2 &\leq (K-1)L^2\alpha^2 \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \left\| \nabla f_i(x_{i,k}^t) \right\|^2 \\
&\leq 3(K-1)L^2\alpha^2 \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \left(\left\| \nabla f_i(x_{i,k}^t) - \nabla f(x_{i,k}^t) \right\|^2 \right. \\
&\quad \left. + \left\| \nabla f(x_{i,k}^t) - \nabla f(x^t) \right\|^2 + \left\| \nabla f(x^t) \right\|^2 \right) \\
&\leq 3(K-1)L^2\alpha^2 \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \left(\left\| \nabla f_i(x_{i,k}^t) - \nabla f(x_{i,k}^t) \right\|^2 \right. \\
&\quad \left. + \frac{1}{N} \sum_{i=1}^N \left\| \nabla f_i(x_{i,k}^t) - \nabla f_i(x^t) \right\|^2 + \left\| \nabla f(x^t) \right\|^2 \right),
\end{aligned}$$

其中最后两个不等式利用了均值不等式。结合假设 1 梯度方差有界, 则有

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \left\| \nabla f_i(x') - \frac{1}{K} \sum_{k=0}^{K-1} \nabla f_i(x'_{i,k}) \right\|^2 &\leq 3K(K-1)L^2\alpha^2\sigma^2 + 3K(K-1)L^2\alpha^2 \left\| \nabla f(x') \right\|^2 \\ &\quad + 3(K-1)L^2\alpha^2 \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \left\| \nabla f_i(x'_{i,k}) - \nabla f_i(x') \right\|^2 \\ &\leq 3K(K-1)L^2\alpha^2\sigma^2 + 3K(K-1)L^2\alpha^2 \left\| \nabla f(x') \right\|^2 \\ &\quad + 3(K-1)L^4\alpha^2 \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \left\| x'_{i,k} - x' \right\|^2, \end{aligned}$$

其中第二个不等式利用了 f_i 是 L -光滑的。注意到

$$\begin{aligned} \frac{L^2}{K} \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \left\| x' - x'_{i,k} \right\|^2 &\leq 3K(K-1)L^2\alpha^2\sigma^2 + 3K(K-1)L^2\alpha^2 \left\| \nabla f(x') \right\|^2 \\ &\quad + 3K(K-1)L^2\alpha^2 \left(\frac{L^2}{K} \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \left\| x' - x'_{i,k} \right\|^2 \right). \end{aligned}$$

取 $\alpha \leq \frac{1}{2KL}$, 则有

$$\frac{L^2}{K} \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} \left\| x' - x'_{i,k} \right\|^2 \leq \frac{3K(K-1)L^2\alpha^2\sigma^2}{1-3K(K-1)L^2\alpha^2} + \frac{3K(K-1)L^2\alpha^2 \left\| \nabla f(x') \right\|^2}{1-3K(K-1)L^2\alpha^2}.$$

因此有

$$\frac{1}{N} \sum_{i=1}^N \left\| \nabla f_i(x') - \frac{1}{K} \sum_{k=0}^{K-1} \nabla f_i(x'_{i,k}) \right\|^2 \leq \frac{3K(K-1)L^2\alpha^2\sigma^2}{1-3K(K-1)L^2\alpha^2} + \frac{3K(K-1)L^2\alpha^2 \left\| \nabla f(x') \right\|^2}{1-3K(K-1)L^2\alpha^2}.$$

证毕。

定理 3 假设对任意 $i \in [N]$, f_i 是 L -光滑的, 假设梯度方差有界, 则对任意 $t \geq 0$ 和 $\alpha \leq \frac{1}{6KL(1+\omega)}$,

有

$$\mathbb{E} \left[\left\| \nabla f(\bar{x}) \right\|^2 \right] \leq \frac{8(f(x^0) - f^*)}{\alpha K \inf_t \eta_t T} + 3\sigma^2,$$

其中设 $p_t = \frac{\eta_t}{\sum_{i=0}^{T-1} \eta_i}$, $\bar{x}$ 是从 $\{x^0, x^1, \dots, x^{T-1}\}$ 中以 p_t 概率选取。

证明: 根据 f_i 是 L -光滑的, 有

$$\begin{aligned} \mathbb{E} \left[f(x^{t+1}) - f^* \right] &\leq \mathbb{E} \left[f(x^t) - f^* + \left\langle \nabla f(x^t), x^{t+1} - x^t \right\rangle + \frac{L}{2} \left\| x^{t+1} - x^t \right\|^2 \right] \\ &= \mathbb{E} \left[f(x^t) - f^* - \eta_t \left\langle \nabla f(x^t), \frac{1}{N} \sum_{i=1}^N C(x^t - x'_{i,K}) \right\rangle + \frac{L\eta_t^2}{2} \left\| \frac{1}{N} \sum_{i=1}^N C(x^t - x'_{i,K}) \right\|^2 \right] \\ &\leq \mathbb{E} \left[f(x^t) - f^* - \eta_t \left\langle \nabla f(x^t), \frac{1}{N} \sum_{i=1}^N C(x^t - x'_{i,K}) \right\rangle + \frac{L\eta_t}{2} \frac{1}{N} \sum_{i=1}^N \left\| C(x^t - x'_{i,K}) \right\|^2 \right], \end{aligned}$$

其中最后一个不等式利用了引理 2。结合定义 1, 有

$$\begin{aligned}
\mathbb{E}[f(x^{t+1}) - f^*] &\leq \mathbb{E}\left[f(x^t) - f^* - \eta_t \left\langle \nabla f(x^t), \frac{1}{N} \sum_{i=1}^N (x^t - x_{i,K}^t) \right\rangle + \frac{(1+\omega)L\eta_t}{2} \frac{1}{N} \sum_{i=1}^N \|x^t - x_{i,K}^t\|^2\right] \\
&= \mathbb{E}\left[f(x^t) - f^* - \alpha K \eta_t \left\langle \nabla f(x^t), \frac{1}{N} \sum_{i=1}^N \frac{1}{K} \sum_{k=0}^{K-1} \nabla f_i(x_{i,k}^t) \right\rangle\right. \\
&\quad \left. + \frac{(1+\omega)LK^2\alpha^2\eta_t}{2} \frac{1}{N} \sum_{i=1}^N \left\| \frac{1}{K} \sum_{k=0}^{K-1} \nabla f_i(x_{i,k}^t) \right\|^2\right] \\
&= \mathbb{E}\left[f(x^t) - f^* - \frac{\alpha K \eta_t}{2} \left(\|\nabla f(x^t)\|^2 + \left\| \frac{1}{N} \sum_{i=1}^N \frac{1}{K} \sum_{k=0}^{K-1} \nabla f_i(x_{i,k}^t) \right\|^2 \right.\right. \\
&\quad \left. \left. - \left\| \nabla f(x^t) - \frac{1}{N} \sum_{i=1}^N \frac{1}{K} \sum_{k=0}^{K-1} \nabla f_i(x_{i,k}^t) \right\|^2 \right) + \frac{(1+\omega)LK^2\alpha^2\eta_t}{2} \frac{1}{N} \sum_{i=1}^N \left\| \frac{1}{K} \sum_{k=0}^{K-1} \nabla f_i(x_{i,k}^t) \right\|^2\right] \\
&\leq \mathbb{E}\left[f(x^t) - f^* - \frac{\alpha K \eta_t}{2} \|\nabla f(x^t)\|^2 + \frac{\alpha K \eta_t}{2} \left\| \nabla f(x^t) - \frac{1}{N} \sum_{i=1}^N \frac{1}{K} \sum_{k=0}^{K-1} \nabla f_i(x_{i,k}^t) \right\|^2\right. \\
&\quad \left. + \frac{(1+\omega)LK^2\alpha^2\eta_t}{2} \frac{1}{N} \sum_{i=1}^N \left\| \frac{1}{K} \sum_{k=0}^{K-1} \nabla f_i(x_{i,k}^t) \right\|^2\right].
\end{aligned}$$

利用均值不等式, 有

$$\begin{aligned}
\mathbb{E}[f(x^{t+1}) - f^*] &\leq \mathbb{E}\left[f(x^t) - f^* - \frac{\alpha K \eta_t}{2} \|\nabla f(x^t)\|^2 + \frac{\alpha K \eta_t}{2} \frac{1}{N} \sum_{i=1}^N \left\| \nabla f_i(x^t) - \frac{1}{K} \sum_{k=0}^{K-1} \nabla f_i(x_{i,k}^t) \right\|^2\right. \\
&\quad \left. + \frac{(1+\omega)LK^2\alpha^2\eta_t}{2} \frac{1}{N} \sum_{i=1}^N \left\| \frac{1}{K} \sum_{k=0}^{K-1} \nabla f_i(x_{i,k}^t) \right\|^2\right] \\
&\leq \mathbb{E}\left[f(x^t) - f^* - \frac{\alpha K \eta_t}{2} \|\nabla f(x^t)\|^2 + \frac{\alpha K \eta_t}{2} \frac{1}{N} \sum_{i=1}^N \left\| \nabla f_i(x^t) - \frac{1}{K} \sum_{k=0}^{K-1} \nabla f_i(x_{i,k}^t) \right\|^2\right. \\
&\quad \left. + \frac{3(1+\omega)LK^2\alpha^2\eta_t}{2} \frac{1}{N} \sum_{i=1}^N \left(\left\| \frac{1}{K} \sum_{k=0}^{K-1} \nabla f_i(x_{i,k}^t) - \nabla f_i(x^t) \right\|^2 + \|\nabla f_i(x^t) - \nabla f(x^t)\|^2 + \|\nabla f(x^t)\|^2 \right)\right] \\
&\leq \mathbb{E}\left[f(x^t) - f^* - \frac{\alpha K \eta_t}{2} \|\nabla f(x^t)\|^2 + \frac{\alpha K \eta_t}{2} \frac{1}{N} \sum_{i=1}^N \left\| \nabla f_i(x^t) - \frac{1}{K} \sum_{k=0}^{K-1} \nabla f_i(x_{i,k}^t) \right\|^2\right. \\
&\quad \left. + \frac{3(1+\omega)LK^2\alpha^2\eta_t}{2} \frac{1}{N} \sum_{i=1}^N \left\| \nabla f_i(x^t) - \frac{1}{K} \sum_{k=0}^{K-1} \nabla f_i(x_{i,k}^t) \right\|^2\right. \\
&\quad \left. + \frac{3(1+\omega)LK^2\alpha^2\sigma^2\eta_t}{2} + \frac{3(1+\omega)LK^2\alpha^2\eta_t}{2} \|\nabla f(x^t)\|^2\right] \\
&= \mathbb{E}\left[f(x^t) - f^* - \frac{\alpha K \eta_t}{2} (1 - 3(1+\omega)LK\alpha) \|\nabla f(x^t)\|^2\right. \\
&\quad \left. + \frac{\alpha K \eta_t}{2} (1 + 3(1+\omega)LK\alpha) \frac{1}{N} \sum_{i=1}^N \left\| \nabla f_i(x^t) - \frac{1}{K} \sum_{k=0}^{K-1} \nabla f_i(x_{i,k}^t) \right\|^2 + \frac{3(1+\omega)LK^2\alpha^2\sigma^2\eta_t}{2}\right],
\end{aligned}$$

其中最后一个不等式利用了假设 1 梯度方差有界。结合引理 5, 则有

$$\begin{aligned} \mathbb{E}[f(x^{t+1}) - f^*] &\leq \mathbb{E} \left[f(x^t) - f^* - \frac{\alpha K \eta_t}{2} (1 - 3(1+\omega) L K \alpha) \|\nabla f(x^t)\|^2 + \frac{3(1+\omega) L K^2 \alpha^2 \sigma^2 \eta_t}{2} \right. \\ &\quad \left. + \frac{\alpha K \eta_t}{2} (1 + 3(1+\omega) L K \alpha) \left(\frac{3K(K-1)L^2 \alpha^2 \sigma^2}{1 - 3K(K-1)L^2 \alpha^2} + \frac{3K(K-1)L^2 \alpha^2 \|\nabla f(x^t)\|^2}{1 - 3K(K-1)L^2 \alpha^2} \right) \right]. \end{aligned}$$

利用 $\alpha \leq \frac{1}{6KL(1+\omega)}$, 则有

$$\begin{aligned} \mathbb{E}[f(x^{t+1}) - f^*] &\leq \mathbb{E} \left[f(x^t) - f^* - \frac{\alpha K \eta_t}{4} \|\nabla f(x^t)\|^2 + \frac{\alpha K \eta_t \sigma^2}{4} \right. \\ &\quad \left. + \frac{3\alpha K \eta_t}{4} \left(\frac{3K(K-1)L^2 \alpha^2 \sigma^2}{1 - 3K(K-1)L^2 \alpha^2} + \frac{3K(K-1)L^2 \alpha^2 \|\nabla f(x^t)\|^2}{1 - 3K(K-1)L^2 \alpha^2} \right) \right] \\ &= \mathbb{E} \left[f(x^t) - f^* - \frac{\alpha K \eta_t}{4} \left(1 - \frac{9K(K-1)L^2 \alpha^2}{1 - 3K(K-1)L^2 \alpha^2} \right) \|\nabla f(x^t)\|^2 \right. \\ &\quad \left. + \frac{\alpha K \eta_t \sigma^2}{4} \left(1 + \frac{9K(K-1)L^2 \alpha^2}{1 - 3K(K-1)L^2 \alpha^2} \right) \right]. \end{aligned}$$

取 $\alpha \leq \min \left\{ \frac{1}{5KL}, \frac{1}{6KL(1+\omega)} \right\} = \frac{1}{6KL(1+\omega)}$, 则有

$$\mathbb{E}[f(x^{t+1}) - f^*] \leq \mathbb{E} \left[f(x^t) - f^* - \frac{\alpha K \eta_t}{8} \|\nabla f(x^t)\|^2 + \frac{3\alpha K \eta_t \sigma^2}{8} \right].$$

因此有

$$\mathbb{E} \left[\sum_{t=0}^{T-1} \eta_t \|\nabla f(x^t)\|^2 \right] \leq \frac{8(f(x^0) - f^*)}{\alpha K} + 3\sigma^2 \sum_{t=0}^{T-1} \eta_t.$$

对任意 $t = 0, 1, \dots, T-1$, 设 $p_t = \frac{\eta_t}{\sum_{t=0}^{T-1} \eta_t}$ 且, $\bar{x}$ 是从 $\{x^0, x^1, \dots, x^{T-1}\}$ 中以 p_t 概率选取, 则有

$$\mathbb{E}[\|\nabla f(\bar{x})\|^2] \leq \frac{8(f(x^0) - f^*)}{\alpha K \sum_{t=0}^{T-1} \eta_t} + 3\sigma^2 \leq \frac{8(f(x^0) - f^*)}{\alpha K \inf_t \eta_t T} + 3\sigma^2.$$

证毕。

综合上述三小节收敛性分析, 我们可以得到以下三个结论:

(1) 在凸条件下, 如果 $\alpha = \frac{1}{7KL(1+\omega)}$, 那么 ExpFedCom 可以在 $T = O\left(\frac{L(1+\omega)}{\varepsilon}\right)$ 通信轮数内找到一个点 x , 使得 $\mathbb{E}[f(x)] - f^* \leq \varepsilon$ 。

(2) 在强凸条件下, 如果 $\alpha \leq \min \left\{ \frac{1}{7KL(1+\omega)}, \frac{14}{\mu K \inf_t \eta_t} \right\}$, 那么 ExpFedCom 可以在

$T = O\left(\frac{1}{\mu K \alpha} \log\left(\frac{1}{\varepsilon}\right)\right)$ 通信轮数内找到一个点 x ，使得 $E\left[\|x - x^*\|^2\right] \leq \varepsilon$ 。

(3) 在非凸条件下，如果 $\alpha = \frac{1}{6KL(1+\omega)}$ 以及 $\sigma \leq \frac{\varepsilon}{6}$ ，那么 ExpFedCom 可以在 $T = O\left(\frac{1+\omega}{\varepsilon^2}\right)$ 通信轮数内找到一个点 x ，使得 $E\left[\|\nabla f(x)\|\right] \leq \varepsilon$ 。

综合上述三小节收敛性分析，我们可以得到以下三个结论(表 1):

Table 1. Communication complexity comparison of algorithms
表 1. 算法通信复杂度比较

算法	凸	强凸	非凸
FedAvg	$T = O\left(\frac{L}{\varepsilon}\right)$	$T = O\left(\frac{L}{\mu} \log\left(\frac{1}{\varepsilon}\right)\right)$	$T = O\left(\frac{1}{\varepsilon^2}\right)$
FedCOM	$T = O\left(\frac{N+\omega}{N\varepsilon} \log\left(\frac{1}{\varepsilon}\right)\right)$	$T = O\left(\frac{L}{\mu} \left(1 + \frac{\omega}{N}\right) \log\left(\frac{1}{\varepsilon}\right)\right)$	$T = O\left(\frac{1}{\varepsilon^2}\right)$
Ours (ExpFedCom)	$T = O\left(\frac{L(1+\omega)}{\varepsilon}\right)$	$T = O\left(\frac{1}{\mu K \alpha} \log\left(\frac{1}{\varepsilon}\right)\right)$	$T = O\left(\frac{1+\omega}{\varepsilon^2}\right)$

4. 数值实验

我们分别在 MNIST 数据集上对 MLP 模型[7]和在 EMNIST 数据集上对 CNN 模型进行了训练。对于每个实验，我们都分别进行 100 回合的通信。对于 MNIST 数据集，每回合我们从 200 个客户端中随机均匀抽取 100 个客户端进行通信。对于 EMNIST 数据集，每回合我们从 200 个客户端中随机均匀抽取 20 个客户端进行通信。压缩方式与文献[7]方法一致，采取无偏量化的方式。本地损失函数采用交叉熵损失函数。每个算法选择理论上得到的最佳步长，最终实验结果如下(图 1~4):

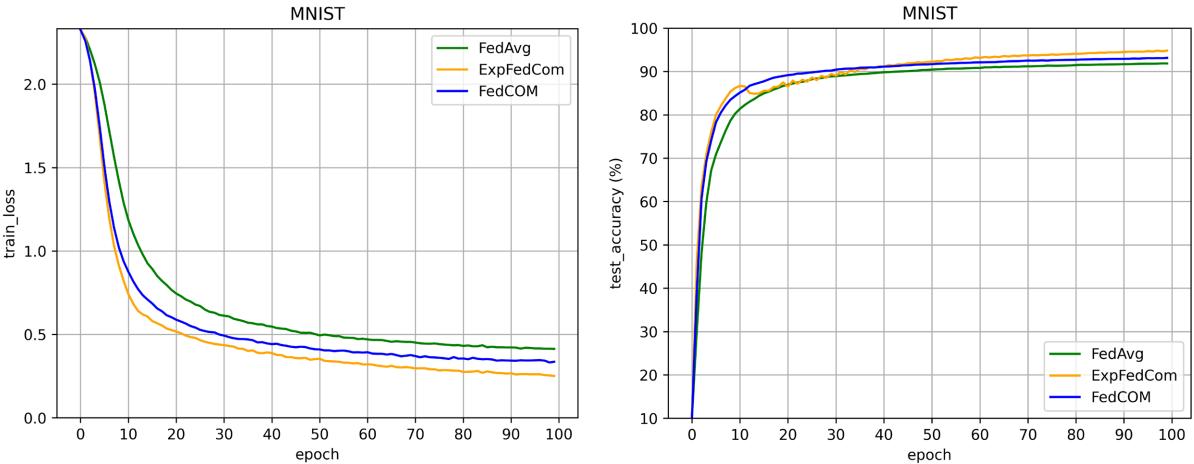


Figure 1. The communication rounds versus loss and accuracy on the MNIST dataset

图 1. MNIST 数据集下的通信轮数与损失和精度实验图

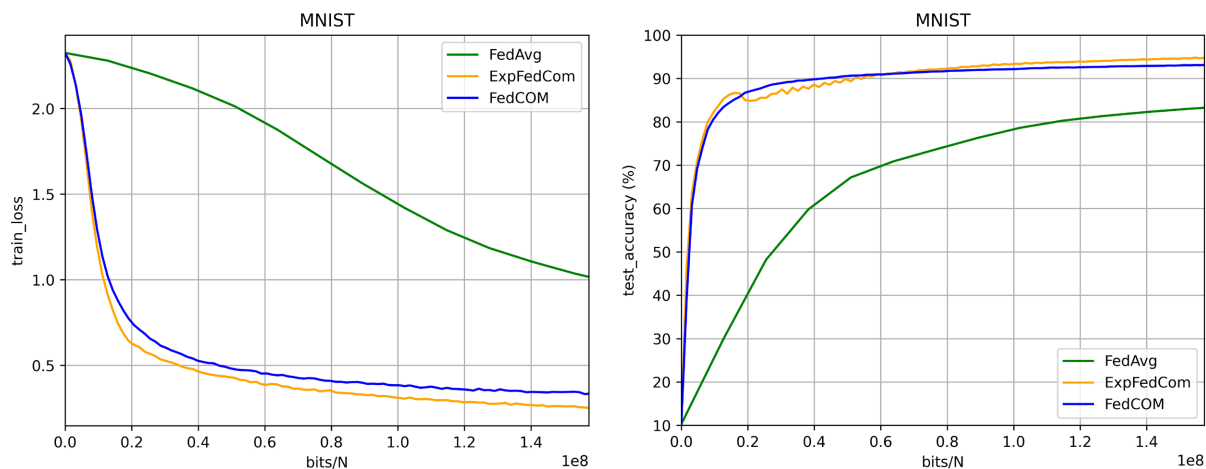


Figure 2. The communication bits versus loss and accuracy on the MNIST dataset

图 2. MNIST 数据集下的通信比特数与损失和精度实验图

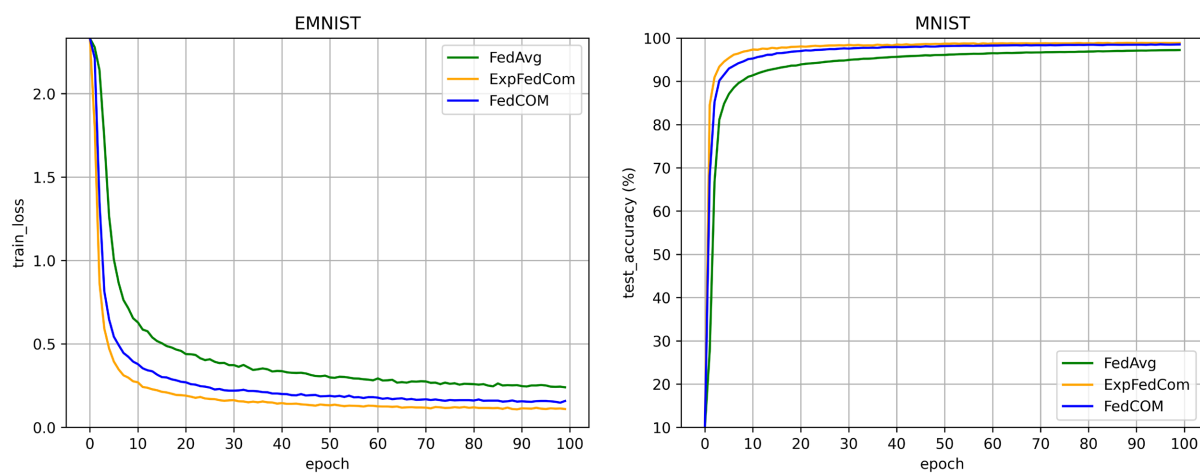


Figure 3. The communication rounds versus loss and accuracy on the EMNIST dataset

图 3. EMNIST 数据集下的通信轮数与损失和精度实验图

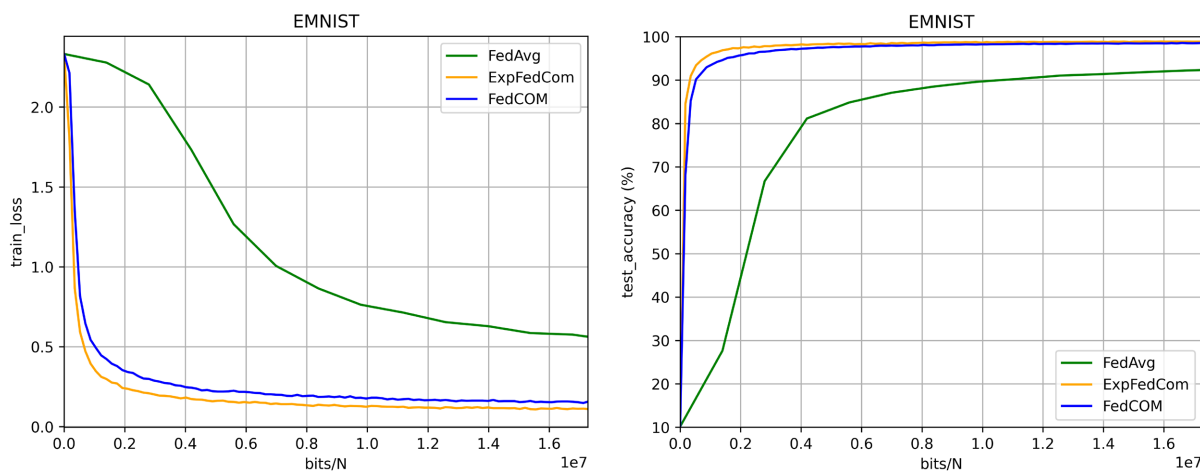


Figure 4. The communication bits versus loss and accuracy on the EMNIST dataset

图 4. EMNIST 数据集下的通信比特数与损失和精度实验图

5. 讨论

在本工作中，我们在联邦学习优化的标准假设下对所提出的算法进行了理论分析。值得注意的是，真实数据往往呈现高度非独立同分布(Non-IID)的特性，客户端之间存在显著的异质性。此外，深度神经网络本质上是而非凸的，这使得我们的收敛性结论在应用到实际任务时存在一定局限。尽管存在这些差异，我们的实验结果表明，所提出的算法在现实条件下依旧取得了较好的性能，同时显著节省了通信成本，并优于一些主流基线方法。

尽管如此，我们承认，在实践中，数据异质性、客户端部分参与以及模型的非凸性等因素可能导致收敛速度变慢或出现残余误差邻域。我们认为，尽管我们的理论发现源于理想化的设定，它们仍然为算法设计提供了宝贵的见解和指导。未来的工作值得去开发一个更通用的分析框架，以更好地兼顾理论严谨性与实际适用性，从而为联邦学习的探索和发展提供更强有力的保证。

6. 总结

基于外推的自适应步长应用于压缩联邦学习当中表现出了良好的性质，并且通过数值实验可以看到ExpFedCom 较不带压缩或者不带自适应步长的方法展现出了较好的优势。ExpFedCom 既支持通信压缩又实现了在凸、强凸以及非凸条件下的快速收敛，这大大提高了通信效率，节省了通信成本，具有广泛的应用价值。

参考文献

- [1] Kairouz, P., McMahan, H.B., Aven, B., *et al.* (2021) Advances and Open Problems in Federated Learning. *Foundations and Trends® in Machine Learning*, **14**, 1-210. <https://doi.org/10.1561/22000000083>
- [2] Klietk, T., Zvarikova, K. and Lazaroiu, G. (2022) Data-Driven Machine Learning and Neural Network Algorithms in the Retailing Environment: Consumer Engagement, Experience, and Purchase Behaviors. *Economics, Management and Financial Markets*, **17**, 57-69. <https://doi.org/10.22381/emfm17120224>
- [3] Liu, J., Huang, J., Zhou, Y., *et al.* (2022) From Distributed Machine Learning to Federated Learning: A Survey. *Knowledge and Information Systems*, **64**, 885-917. <https://doi.org/10.1007/s10115-022-01664-x>
- [4] McMahan, B., Moore, E., Ramage, D., *et al.* (2017) Communication-Efficient Learning of Deep Networks from Decentralized Data. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, **54**, 1273-1282.
- [5] Karimireddy, S.P., Kale, S., Mohri, M., *et al.* (2020) Scaffold: Stochastic Controlled Averaging for Federated Learning. *Proceedings of the 37th International Conference on Machine Learning*, **119**, 5132-5143.
- [6] Khirirat, S., Feyzmahdavian, H.R. and Johansson, M. (2018) Distributed Learning with Compressed Gradients. arXiv:1806.06573.
- [7] Haddadpour, F., Kamani, M.M., Mokhtari, A. and Mahdavi, M. (2021) Federated Learning with Compression: Unified Analysis and Sharp Guarantees. *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*, **130**, 2350-2358.
- [8] 贾泽慧, 李登辉, 刘治宇, 等. 基于数据压缩和梯度追踪的方差缩减的联邦优化算法[J]. 南京理工大学学报, 2025, 49(2): 155-166.
- [9] Reddi, S.J., Charles, Z., Zaheer, M., *et al.* (2020) Adaptive Federated Optimization. arXiv:2003.00295.
- [10] Jhunjhunwala, D., Wang, S. and Joshi, G. (2023) FedExP: Speeding up Federated Averaging via Extrapolation. arXiv:2301.09604.