

一种扩展的分布式非精确多更新单组合算法

郝晓惠, 肖艳阳

河北工业大学理学院, 天津

收稿日期: 2024年12月27日; 录用日期: 2025年1月18日; 发布日期: 2025年1月29日

摘要

在这项工作中, 我们提出了一种基于扩展梯度的分布式非精确单更新多组合算法(简称ExtendMUSIC)。该方法利用前两次迭代的梯度和来构造搜索方向, 再利用多次局部更新然后定期组合的策略来求解光滑强凸的分布式优化问题。此外, 我们证明了所提出的算法能够实现线性收敛。数值实验结果表明, 与不使用扩展梯度的分布式多更新单组合算法相比, 扩展的分布式多更新单组合算法可以实现加速收敛。

关键词

分布式优化, 多更新单组合, 扩展梯度下降法, 线性收敛

An Extended Distributed Inexact Method with Multi-Update and Single-Combination Strategy

Xiaohui Hao, Yanyang Xiao

College of Science, Hebei University of Technology, Tianjin

Received: Dec. 27th, 2024; accepted: Jan. 18th, 2025; published: Jan. 29th, 2025

Abstract

In this work, we propose a distributed inexact multi-update and single-combination algorithm based on extended gradients (ExtendMUSIC). This method utilizes the gradient sum of the first two iterations to construct the search direction, and then uses the strategy of multiple local updates and periodic combinations to solve smooth and strongly convex distributed optimization problems. In addition, we prove that the proposed algorithm can achieve linear convergence. The numerical experimental results show that compared with the distributed multi-update and single-combination

algorithm without using extended gradients, the extended distributed multi-update and single-combination algorithm can achieve accelerated convergence.

Keywords

Distributed Optimization, Multi-Update and Single-Combination, Extended Gradient Descent Method, Linear Convergence

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

随着多智能体系统协调技术的快速发展, 布式优化在分布式机器学习[1] [2]、传感器网络[3]、多智能体系统[4] [5]、电力系统控制[6]等诸多领域显示出巨大的潜力。考虑一个由 n 个节点组成的无向连接网络, 每个节点都有私有的局部目标函数。分布式优化的全局目标函数是对所有局部目标函数的平均值求最小。分布式优化算法的挑战在于如何通过局部计算和通信使整个系统协同工作以找到全局最优解或近似最优解。局部计算是指每个节点仅根据局部目标函数更新自己的变量, 局部通信是指每个节点只与相邻节点交换信息。

许多分布式优化算法已经被开发出来。由于计算成本低、易于实现, 分布式梯度型方法已经成为最受欢迎的算法之一。早期的分布式梯度下降(DGD)方法[7] [8]可以使用衰减步长, 缓慢地收敛到最优解。当使用固定步长时, DGD 可以实现线性收敛, 但是只能收敛到最优解的邻域。因此, 当不需要高精度时, 固定步长的 DGD 方法是不错的选择。文献[9]-[11]等结合了更多的技术, 可以线性收敛到最优解, 这些技术包括两次连续迭代的差值[9]、梯度跟踪技术[10]和多步一致性[11]等, 但是这些技术会带来更多的计算和通信负担。当前, 有一类加速方案是利用前几次迭代的梯度来更新变量, 比如乐观梯度下降-上升(OGDA)方法[12]和 Hessian 采样的惯性梯度算法(IGAHD) [13]通过前两次迭代的梯度差来更新变量。此外, 文献[14]中的 Anderson 加速梯度下降法(AA-PGA)通过使用前几次迭代的梯度的凸组合来更新变量。在文献[15]中, Tang 等人采用了改进的分布式 Anderson 加速法(ELLADA)来减少迭代次数, 加速收敛。近期, Zhang 等[16]人引入了一种扩展的分布式梯度下降方法, 即沿着前两次迭代的梯度和的方向更新变量。对于分布式强凸优化, 当步长小于给定上界时, 扩展的分布式梯度下降方法可以实现线性收敛性。数值实验也表明, 使用扩展梯度的梯度下降方法比传统的梯度下降方法具有更好的搜索方向。在文献[17]中, Wu 等人提出了一种非精确的多更新单组合(MUSIC)策略的分布式优化方法并证明了其线性收敛性, 然后基于非精确的 MUSIC 框架, 结合局部梯度矫正技术, 设计了一种精确的 MUSIC 加速算法, 该方法可以实现快速的精确收敛以及高效通信。多更新的思想来源于集中式局部更新随机梯度下降方法(局部更新 SGD) [18], 它允许节点执行多个局部更新, 然后定期进行组合。众所周知的联邦平均(FedAvg)算法[19]就是局部更新 SGD 的衍生物, 专门为不平衡的参与设备设计。

受扩展梯度法具有更好的搜索方向的启发, 结合多更新单组合策略(MUSIC), 针对无向网络上的分布式优化问题, 我们提出了一种扩展的分布式多更新单组合算法(ExtendMUSIC)。具体来说, 该方法沿着前两次迭代的梯度和的方向下降, 允许每个节点执行多个局部更新, 然后定期进行组合。对于光滑强凸

函数, 我们证明了 ExtendMUSIC 能够线性收敛到最优解的邻域。最后, 数值实验验证了 ExtendMUSIC 的有效性。

本文剩余部分组织如下: 第二章主要介绍了分布式优化问题以及我们所提出的方法。第三章提供了 ExtendMUSIC 的收敛性分析。第四章进行数值实验验证算法的有效性。最后, 我们对全文进行总结并提出展望。

本文主要符号如下: 为了更好地理解这项工作, 在整个论文中, 所涉及的矩阵和向量分别用大写字母和小写字母表示, 标量用标准字体表示, 其中, x^T 表示向量 x 的转置, 算子 \otimes 表示克罗内克积。 $\|\bullet\|$ 表示向量的欧氏范数和矩阵的谱范数, $\lfloor x \rfloor$ 表示不超过 x 的最大整数, 其在欧几里得空间中的内积用 $\langle \bullet, \bullet \rangle$ 表示。

2. 问题描述及扩展的先适应再结合分布式梯度下降方法

本章首先介绍分布式优化问题的数学模型, 并给出了主要假设。然后, 简要回顾了扩展的分布式梯度下降方法和带有多更新单组合策略(MUSIC)的非精确分布式优化算法。最后, 提出了扩展的分布式多更新单组合算法。

2.1. 问题描述

我们关注一个由 n 个节点组成的无向网络, 这个网络可以用无向图 $G=(V, E)$ 来表示, 其中 V 表示节点的集合, E 表示边 (i, j) 的集合, 其中 $i, j \in V$ 。如果 $i, j \in E$, 则 i 与 j 之间的信息可以相互传递。定义权重矩阵 $W = [w_{ij}] \in \mathbb{R}^{n \times n}$, 其中 w_{ij} 表示智能体 i 传递给其邻居 j 的信息权重。分布式优化问题可以表述为:

$$\min_{x \in \mathbb{R}^N} f(x) = \frac{1}{n} \sum_{i=1}^n f_i(x) \quad (2.1)$$

其中 $f_i: \mathbb{R}^N \rightarrow \mathbb{R}$ 是一个连续可微函数。接下来, 我们给出以下几个基本假设。

假设 1: (光滑性)局部目标函数 f_i 是 L -光滑的, 也就是说, 对任意的 $\tilde{x}, \tilde{y} \in \mathbb{R}^N$, 有

$$\|\nabla f_i(\tilde{x}) - \nabla f_i(\tilde{y})\| \leq L \|\tilde{x} - \tilde{y}\|. \quad (2.2)$$

假设 2: (强凸性)局部目标函数 f_i 是 μ 强凸的, 也就是说, 对任意的 $\tilde{x}, \tilde{y} \in \mathbb{R}^N$, 有

$$f_i(\tilde{y}) - f_i(\tilde{x}) \geq \langle \nabla f_i(\tilde{x}), \tilde{y} - \tilde{x} \rangle + \frac{\mu}{2} \|\tilde{y} - \tilde{x}\|^2. \quad (2.3)$$

假设 1 和 2 中的常数 L 和 μ 满足 $0 < \mu \leq L$ 。假设 1 和 2 表明全局目标函数 f 也是 L -光滑和 μ -强凸的, 因此, f 有唯一最优解, 记为 x^* 。本文中设 $f^* = f(x^*)$ 。

假设 3: 权重矩阵 W 具有下列性质:

(1) 如果 $(i, j) \in E$, 那么 $w_{ij} > 0$ 。对任意的 $i \in V$, $w_{ii} > 0$ 。否则, $w_{ij} = 0$ 。

(2) 权重矩阵 W 是双随机, 也就是说对任意的 $i, j \in V$, $\sum_{i=1}^n w_{ij} = \sum_{j=1}^n w_{ij} = 1$ 。

记 W 的第二大奇异值为 $\sigma \in (0, 1)$, σ 可以用来表示图参数[10]。对任意的 $\omega \in \mathbb{R}^n$,

$$\|W\omega - \mathbf{1}\bar{\omega}\| \leq \sigma \|\omega - \mathbf{1}\bar{\omega}\|, \quad \text{其中 } \bar{\omega} = \frac{1}{n} \mathbf{1}^T \omega.$$

假设 4: (梯度有界) 假设 f_i 在 $\tilde{x} \in \mathbb{R}^N$ 处的梯度是有界的, 即对于所有 $i=1, \dots, n$, 有

$$0 \leq G_{\min} \leq \nabla f_i(x_i^k) \leq G_{\max}.$$

2.2. 扩展的分布式梯度下降方法

具体来说, 扩展的非精确 MUSIC 算法由两个循环迭代组成, 即节点内的计算循环和节点之间通信循环。下面, 我们将算法过程中的总迭代次数表示为 K , 每 K 个局部更新步骤进行一次组合, 由于通信只发生在组合步骤中, 因此每个节点的通信轮数等于 $\lfloor K/E \rfloor$, 其中一轮意味着节点 i 向其相邻节点发送当前估计值 x_k^i , 并接收 x_k^j , 其中 $j \in \mathcal{N}_i$ 。设 \mathcal{I}_E 是组合步骤的集合, 即 $\mathcal{I}_E = \{tE | t = 0, 1, 2, \dots, \lfloor K/E \rfloor\}$ 。因此, 存在 $k^0 = tE \leq k$, 对于任意 $k > 0$, 满足 $k - k^0 \leq E$ 。我们可以用下面的迭代来描述扩展的非精确 MUSIC 算法:

$$x_{k+1}^i = \begin{cases} x_k^i - \eta(\nabla f_i(x_k^i) + \nabla f_i(x_{k-1}^i)), & k+1 \notin \mathcal{I}_E. \\ \sum_{j \in \mathcal{N}_i} w_{ij} \left(x_k^j - \eta(\nabla f_j(x_k^j) + \nabla f_i(x_{k-1}^j)) \right), & k+1 \in \mathcal{I}_E. \end{cases} \quad (2.4)$$

如果我们引入中间变量 v_k^i , 那么(2.4)可以重写为:

$$\begin{aligned} v_{k+1}^i &= x_k^i - \eta(\nabla f_i(x_k^i) + \nabla f_i(x_{k-1}^i)), \\ x_{k+1}^i &= \begin{cases} v_{k+1}^i, & k+1 \notin \mathcal{I}_E. \\ \sum_{j \in \mathcal{N}_i} w_{ij} v_{k+1}^j, & k+1 \in \mathcal{I}_E. \end{cases} \end{aligned} \quad (2.5)$$

这对后续的收敛分析很有用。其中, v_{k+1}^i 是扩展的梯度下降在 x_k^i 处的结果。值得注意的是, 当 $E=1$ 时, 扩展的非精确 MUSIC 方法可以被简化为使用扩展梯度的先适应再结合(ATC)方法, 即

$$x_{k+1}^i = \sum_{j \in \mathcal{N}_i} w_{ij} \left(x_k^j - \eta(\nabla f_j(x_k^j) + \nabla f_i(x_{k-1}^j)) \right).$$

3. 收敛性分析

为方便记述, 我们引入以下在分析中使用的变量: $\bar{v}_k^i = \sum_{j=1}^N w_{ij} v_k^j$, $\bar{x}_k^i = \sum_{j=1}^N w_{ij} x_k^j$ 以及 $\bar{g}_k^i = \sum_{j=1}^N w_{ij} \nabla f_j(x_k^j)$ 。

3.1. 关键引理

下面, 给出几个关键引理, 以建立与网络最优性间隙 $\|\bar{v}_k^i - x^*\|^2$ 相关的一般动力系统。

首先, 求出一步扩展的梯度下降(2.4)中第一个等式的有界结果, 这为后面的应用提供了一个重要的关系式。

引理 1: 在光滑性和强凸性的假设下, 对于一步扩展的梯度下降(2.4)中第一个等式, 如果步长 η 满足 $\eta \leq \frac{1}{2L}$, 有

$$\begin{aligned} \|\bar{v}_i^{t+1} - x^*\|^2 &\leq (1 - \mu\eta) \|\bar{x}_k^i - x^*\|^2 + \sum_{j=1}^N w_{ij} \|x_k^j - \bar{x}_k^i\|^2 + \sum_{j=1}^N w_{ij} \|x_{k-1}^j - \bar{x}_k^i\|^2 \\ &\quad + \gamma \sum_{j=1}^N w_{ij} \|x_k^j - \bar{x}_k^j\|^2 + \gamma \sum_{j=1}^N w_{ij} \|x_{k-1}^j - \bar{x}_k^j\|^2 + 4\eta\tau, \end{aligned} \quad (3.1)$$

其中, 对任意节点 i , $\gamma = \frac{\eta(1-3L\eta)}{\pi}$ 以及 $0 < \pi < \frac{1}{L}$, $f_i(x^*) - f_i^* \leq \tau$ 。

证明: 基于(2.4)中第一个等式以及 \bar{v}_k^i, \bar{x}_k^i 和 \bar{g}_k^i 的定义, 有

$$\begin{aligned} \|\bar{v}_{k+1}^i - x^*\|^2 &= \|\bar{x}_k^i - \eta(\bar{g}_k^i + \bar{g}_{k-1}^i) - x^*\|^2 \\ &= \|\bar{x}_k^i - x^*\|^2 \underbrace{- 2\eta \langle \bar{x}_k^i - x^*, \bar{g}_k^i + \bar{g}_{k-1}^i \rangle}_{A_1} + \eta^2 \underbrace{\|\bar{g}_k^i + \bar{g}_{k-1}^i\|^2}_{A_2}. \end{aligned} \quad (3.2)$$

首先求 A_2 的界。

$$\begin{aligned} A_2 &= \eta^2 \|\bar{g}_k^i + \bar{g}_{k-1}^i\|^2 \\ &\leq 2\eta^2 \|\bar{g}_k^i\|^2 + 2\eta^2 \|\bar{g}_{k-1}^i\|^2 \\ &= 2\eta^2 \left\| \sum_{j=1}^N w_{ij} \nabla f_j(x_k^j) \right\|^2 + 2\eta^2 \left\| \sum_{j=1}^N w_{ij} \nabla f_j(x_{k-1}^j) \right\|^2 \\ &\leq 2\eta^2 \sum_{j=1}^N w_{ij} \|\nabla f_j(x_k^j)\|^2 + 2\eta^2 \sum_{j=1}^N w_{ij} \|\nabla f_j(x_{k-1}^j)\|^2 \\ &\leq 4L\eta^2 \sum_{j=1}^N w_{ij} (f_j(x_k^j) - f_j^*) + 4L\eta^2 \sum_{j=1}^N w_{ij} (f_j(x_{k-1}^j) - f_j^*), \end{aligned} \quad (3.3)$$

其中第一个不等式是根据 $\forall a, b \in \mathbb{R}^N, (a+b)^2 \leq 2a^2 + 2b^2$ 这一事实得到, 第二个不等式是根据 f_j 的凸性得到的, 最后一个不等式基于 f_j 的 L -光滑性得到的。

接下来, 求出 A_1 的界。

$$\begin{aligned} A_1 &= -2\eta \langle \bar{x}_k^i - x^*, \bar{g}_k^i + \bar{g}_{k-1}^i \rangle \\ &= -2\eta \left\langle \bar{x}_k^i - x^*, \sum_{j=1}^N w_{ij} g_k^j \right\rangle - 2\eta \left\langle \bar{x}_k^i - x^*, \sum_{j=1}^N w_{ij} g_{k-1}^j \right\rangle \\ &= -2\eta \sum_{j=1}^N w_{ij} \langle \bar{x}_k^i - x^*, g_k^j \rangle - 2\eta \sum_{j=1}^N w_{ij} \langle \bar{x}_k^i - x^*, g_{k-1}^j \rangle \\ &= -2\eta \sum_{j=1}^N w_{ij} \langle \bar{x}_k^i - x_k^j + x_k^j - x^*, g_k^j \rangle - 2\eta \sum_{j=1}^N w_{ij} \langle \bar{x}_k^i - x_{k-1}^j + x_{k-1}^j - x^*, g_{k-1}^j \rangle \\ &= -2\eta \sum_{j=1}^N w_{ij} \langle \bar{x}_k^i - x_k^j, g_k^j \rangle - 2\eta \sum_{j=1}^N w_{ij} \langle x_k^j - x^*, g_k^j \rangle \\ &\quad - 2\eta \sum_{j=1}^N w_{ij} \langle \bar{x}_k^i - x_{k-1}^j, g_{k-1}^j \rangle - 2\eta \sum_{j=1}^N w_{ij} \langle x_{k-1}^j - x^*, g_{k-1}^j \rangle, \end{aligned} \quad (3.4)$$

其中在第二个等式中我们用到了 $g_k^j \triangleq \nabla f_j(x_k^j)$ 。根据 μ -强凸性, 有

$$-\langle x_k^j - x^*, g_k^j \rangle \leq -(f_j(x_k^j) - f_j(x^*)) - \frac{\mu}{2} \|x_k^j - x^*\|^2, \quad (3.5)$$

$$-\langle x_{k-1}^j - x^*, g_{k-1}^j \rangle \leq -\left(f_j(x_{k-1}^j) - f_j(x^*)\right) - \frac{\mu}{2} \|x_{k-1}^j - x^*\|^2. \quad (3.6)$$

根据 AM-GM 不等式, 有对任意向量 a 以及 b , $\pm 2\langle a, b \rangle \leq \eta \|a\|^2 + \eta^{-1} \|b\|^2$. 因此, 下式成立,

$$-2\langle \bar{x}_k^i - x_k^j, g_k^j \rangle \leq \eta^{-1} \|\bar{x}_k^i - x_k^j\|^2 + \eta \|g_k^j\|^2, \quad (3.7)$$

$$-2\langle \bar{x}_k^i - x_{k-1}^j, g_{k-1}^j \rangle \leq \eta^{-1} \|\bar{x}_k^i - x_{k-1}^j\|^2 + \eta \|g_{k-1}^j\|^2. \quad (3.8)$$

将(3.5)~(3.8)代入到(3.4), 有

$$\begin{aligned} A_1 &\leq -2\eta \sum_{j=1}^N w_{ij} \langle \bar{x}_k^i - x_k^j, g_k^j \rangle - 2\eta \sum_{j=1}^N w_{ij} \langle x_k^j - x^*, g_k^j \rangle \\ &\quad - 2\eta \sum_{j=1}^N w_{ij} \langle \bar{x}_k^i - x_{k-1}^j, g_{k-1}^j \rangle - 2\eta \sum_{j=1}^N w_{ij} \langle x_{k-1}^j - x^*, g_{k-1}^j \rangle \\ &\leq \sum_{j=1}^N w_{ij} \left(\|\bar{x}_k^i - x_k^j\|^2 + \eta^2 \|g_k^j\|^2 \right) - 2\eta \sum_{j=1}^N w_{ij} \left(f_j(x_k^j) - f_j(x^*) + \frac{\mu}{2} \|x_k^j - x^*\|^2 \right) \\ &\quad + \sum_{j=1}^N w_{ij} \left(\|\bar{x}_k^i - x_{k-1}^j\|^2 + \eta^2 \|g_{k-1}^j\|^2 \right) - 2\eta \sum_{j=1}^N w_{ij} \left(f_j(x_{k-1}^j) - f_j(x^*) + \frac{\mu}{2} \|x_{k-1}^j - x^*\|^2 \right). \end{aligned} \quad (3.9)$$

结合 $\|\bar{x}_k^i - x^*\|^2 = \left\| \sum_{j=1}^N w_{ij} (x_k^j - x^*) \right\|^2 \leq \sum_{j=1}^N w_{ij} \|x_k^j - x^*\|^2$ 以及 A_2 的界(3.3), 可以求出 $A_1 + A_2$ 的界.

$$\begin{aligned} A_1 + A_2 &\leq -\mu\eta \|\bar{x}_k^i - x^*\|^2 + \sum_{j=1}^N w_{ij} \|\bar{x}_k^i - x_k^j\|^2 \\ &\quad + \underbrace{2\eta \sum_{j=1}^N w_{ij} \left[3L\eta \left(f_j(x_k^j) - f_j^* \right) - \left(f_j(x_k^j) - f_j(x^*) \right) \right]}_{B_1} \\ &\leq -\mu\eta \|\bar{x}_{k-1}^i - x^*\|^2 + \sum_{j=1}^N w_{ij} \|\bar{x}_k^i - x_{k-1}^j\|^2 \\ &\quad + \underbrace{2\eta \sum_{j=1}^N w_{ij} \left[3L\eta \left(f_j(x_{k-1}^j) - f_j^* \right) - \left(f_j(x_{k-1}^j) - f_j(x^*) \right) \right]}_{B_2}. \end{aligned} \quad (3.10)$$

根据 τ 的定义, 将 B_1 重写为

$$\begin{aligned} B_1 &= 2\eta \sum_{j=1}^N w_{ij} \left[(3L\eta - 1) \left(f_j(x_k^j) - f_j^* \right) + \left(f_j(x^*) - f_j^* \right) \right] \\ &\leq 2\eta (3L\eta - 1) \underbrace{\sum_{j=1}^N w_{ij} \left(f_j(x_k^j) - f_j^* \right)}_{C_1} + 2\eta\tau. \end{aligned} \quad (3.11)$$

下面, 求出 C_1 的界, 有

$$\begin{aligned}
 C &= \sum_{j=1}^N w_{ij} (f_j(x_k^j) - f_j^*) \\
 &= \sum_{j=1}^N w_{ij} \left[(f_j(x_k^j) - f_j(\bar{x}_k^j)) + (f_j(\bar{x}_k^j) - f_j^*) \right] \\
 &\geq \sum_{j=1}^N w_{ij} \left[\langle \nabla f_j(\bar{x}_k^j), x_k^j - \bar{x}_k^j \rangle + (f_j(\bar{x}_k^j) - f_j^*) \right] \\
 &\geq -\frac{1}{2} \sum_{j=1}^N w_{ij} \left[\pi \|\nabla f_j(\bar{x}_k^j)\|^2 + \frac{1}{\pi} \|x_k^j - \bar{x}_k^j\|^2 \right] + \sum_{j=1}^N w_{ij} (f_j(\bar{x}_k^j) - f_j^*) \\
 &\geq -\sum_{j=1}^N w_{ij} \left[L\pi (f_j(\bar{x}_k^j) - f_j^*) + \frac{1}{2\pi} \|x_k^j - \bar{x}_k^j\|^2 \right] + \sum_{j=1}^N w_{ij} (f_j(\bar{x}_k^j) - f_j^*) \\
 &\geq -\sum_{j=1}^N w_{ij} \left[(L\pi - 1)(f_j(\bar{x}_k^j) - f_j^*) + \frac{1}{2\pi} \|x_k^j - \bar{x}_k^j\|^2 \right],
 \end{aligned} \tag{3.12}$$

其中, 第一个不等式是基于 f_j 的凸性, 第二个不等式是基于下面这一事实得到的, 对任意向量 $a, b \in \mathbb{R}^N$ 以及 $\pi > 0$, 有 $2\langle a, b \rangle \geq -\pi \|a\|^2 + \pi^{-1} \|b\|^2$. 在第三个不等式中, 使用了 f_j 的 L -光滑性. 如果 $L\pi - 1 < 0$ (i.e., $\pi < \frac{1}{L}$) 成立, 那么根据 $f_j(\bar{x}_k^j) - f_j^* \geq 0$ 这一事实, C 可以被进一步限定为:

$$C_1 \geq -\frac{1}{2\pi} \sum_{j=1}^N w_{ij} \|x_k^j - \bar{x}_k^j\|^2. \tag{3.13}$$

基于 $\eta \leq \frac{1}{3L}$, 将(3.13)代入到(3.11), 有

$$B_1 \leq \gamma \sum_{j=1}^N w_{ij} \|x_k^j - \bar{x}_k^j\|^2 + 2\eta\tau, \tag{3.14}$$

同理, B_2 可以化简如下

$$\begin{aligned}
 B_2 &= 2\eta \sum_{j=1}^N w_{ij} \left[(3L\eta - 1)(f_j(x_{k-1}^j) - f_j^*) + (f_j(x^*) - f_j^*) \right] \\
 &\leq 2\eta (3L\eta - 1) \underbrace{\sum_{j=1}^N w_{ij} (f_j(x_{k-1}^j) - f_j^*)}_{C_2} + 2\eta\tau
 \end{aligned} \tag{3.15}$$

其中,

$$C_2 \geq -\frac{1}{2\pi} \sum_{j=1}^N w_{ij} \|x_{k-1}^j - \bar{x}_{k-1}^j\|^2. \tag{3.16}$$

将(3.10)以及(3.14)~(3.16)代入(3.2)可以得到(3.1)结果. 证毕.

接下来, 约束不等式(3.1)右边的第二项和第三项.

引理 2: 在有界梯度的假设下, 对于扩展的非精确 MUSIC 方法的 $\|x_k^j - \bar{x}_k^j\|$, 有

$$\|x_k^j - \bar{x}_k^j\| \leq 4\eta(k - k^0)G_{\max} \tag{3.17}$$

其中, $k^0 \in \mathcal{I}_E$ 表示一次组合时间, 对任意的 k , 满足 $0 \leq k - k^0 \leq E - 1$.

证明: 首先, 在 $k = k^0$ 的情况下, 由于 $x_j^{k^0} = \bar{x}_j^{k^0}$. 基于组合策略((2.5)中的第二个等式), 不等式(3.17)恒成立. 其次, 可以将 $\|x_k^j - \bar{x}_k^j\|$ 写为

$$\begin{aligned}\|x_k^j - \bar{x}_k^j\| &= \|x_k^j - x_{k^0}^j + x_{k^0}^j - \bar{x}_k^j\| \\ &\leq \|x_k^j - x_{k^0}^j\| + \|\bar{x}_k^j - x_{k^0}^j\|.\end{aligned}\quad (3.18)$$

对于从 k^0 到 $k \leq k^0 + E - 1$ 的内循环迭代, 有

$$\begin{aligned}x_{k^0+1}^j &= x_{k^0}^j - \eta \nabla f_j(x_{k^0}^j) - \eta \nabla f_j(x_{k^0-1}^j), \\ x_{k^0+2}^j &= x_{k^0+1}^j - \eta \nabla f_j(x_{k^0+1}^j) - \eta \nabla f_j(x_{k^0}^j), \\ &\vdots \\ x_k^j &= x_{k-1}^j - \eta \nabla f_j(x_{k-1}^j) - \eta \nabla f_j(x_{k-2}^j).\end{aligned}\quad (3.19)$$

对(3.19)求和得出

$$x_k^j - x_{k^0}^j = -\eta \sum_{s=k^0}^k \nabla f_j(x_s^j) - \eta \sum_{s=k^0-1}^{k-1} \nabla f_j(x_s^j).\quad (3.20)$$

根据梯度有界假设, 我们可以得出

$$\begin{aligned}\|x_k^j - x_{k^0}^j\| &= \left\| \eta \sum_{s=k^0}^k \nabla f_j(x_s^j) + \eta \sum_{s=k^0-1}^{k-1} \nabla f_j(x_s^j) \right\| \\ &\leq \left\| \eta \sum_{s=k^0}^k \nabla f_j(x_s^j) \right\| + \left\| \eta \sum_{s=k^0-1}^{k-1} \nabla f_j(x_s^j) \right\| \\ &\leq 2\eta(k - k^0)G_{\max}.\end{aligned}\quad (3.21)$$

通过对(3.19)进行加权求和, 可以得到

$$\begin{aligned}\bar{x}_{k^0+1}^j &= \bar{x}_{k^0}^j - \eta \sum_{l=1}^N w_{jl} \nabla f_l(x_{k^0}^j) - \eta \sum_{l=1}^N w_{jl} \nabla f_l(x_{k^0-1}^j), \\ \bar{x}_{k^0+2}^j &= \bar{x}_{k^0+1}^j - \eta \sum_{l=1}^N w_{jl} \nabla f_l(x_{k^0+1}^j) - \eta \sum_{l=1}^N w_{jl} \nabla f_l(x_{k^0}^j), \\ &\vdots \\ \bar{x}_k^j &= \bar{x}_{k-1}^j - \eta \sum_{l=1}^N w_{jl} \nabla f_l(x_{k-1}^j) - \eta \sum_{l=1}^N w_{jl} \nabla f_l(x_{k-2}^j).\end{aligned}\quad (3.22)$$

用同样的求和以及二次正则的方法, 我们同样可以得到

$$\begin{aligned}\|\bar{x}_k^j - \bar{x}_{k^0}^j\| &= \left\| \eta \sum_{s=k^0}^k \sum_{l=1}^N w_{jl} \nabla f_l(x_s^j) + \eta \sum_{s=k^0-1}^{k-1} \sum_{l=1}^N w_{jl} \nabla f_l(x_s^j) \right\| \\ &\leq \eta \sum_{s=k^0}^k \sum_{l=1}^N w_{jl} \|\nabla f_l(x_s^j)\| + \eta \sum_{s=k^0-1}^{k-1} \sum_{l=1}^N w_{jl} \|\nabla f_l(x_s^j)\| \\ &\leq 2\eta(k - k^0)G_{\max}.\end{aligned}\quad (3.23)$$

由于 $\bar{x}_{k^0}^j = x_{k^0}^j$, 将(3.23)重写为

$$\|\bar{x}_k^j - x_{k^0}^j\| \leq 2\eta(k - k^0)G_{\max}\quad (3.24)$$

把(3.21)和(3.24)代入(3.18), 证明完成。

在确定 $\|x_j^k - \bar{x}_i^k\|$ 的界之前, 我们还要引入一个有界的假设。

假设 5: 对于扩展的非精确 MUSIC (2.5)中的任意迭代 $k^0 \in \mathcal{I}_E$, 任意两个代理 i 和 j 之间的偏差都是有界的, 即 $\|x_{k^0}^j - x_{k^0}^i\| \leq \varepsilon$, 其中 ε 是一个很小的非负常数。

先前的许多研究都清楚地表明, 由组合(共识)步骤((2.5)中的第二个等式)生成的所有迭代点的估计值之间的差值几乎肯定为零, 也就是说: 当网络连通性, 双随机权重 w_{ij} 和有界梯度假设成立时, $\lim_{k \rightarrow \infty} \|x_k^j - x_k^i\| = 0$ 的概率为 1。因此, 有限的 $\|x_j^0 - x_i^0\|$ 是一个合理的假设。从而, 我们可以得到下面的引理。

引理 3: 在有界假设条件下, 对于扩展的非精确 MUSIC 算法(2.5), 对任意 $k^0 \in \mathcal{I}_E$ 以及 $0 \leq k - k^0 \leq E - 1$, $\|x_k^j - \bar{x}_k^i\|$ 是有界的, 可以表述为下式:

$$\|x_k^j - \bar{x}_k^i\| \leq 8\eta(k - k^0)G_{\max} + \varepsilon, \quad (3.25)$$

$$\|x_{k-1}^j - \bar{x}_k^i\| \leq 8\eta(k - k^0)G_{\max} + \varepsilon. \quad (3.26)$$

证明: 注意到, 对于网络上的任意两个代理 i 和 j ,

$$\begin{aligned} \|x_k^j - \bar{x}_k^i\| &= \|x_k^j - \bar{x}_k^j + \bar{x}_k^j - \bar{x}_k^i\| \\ &\leq \|x_k^j - \bar{x}_k^j\| + \|\bar{x}_k^j - \bar{x}_k^i\|. \end{aligned} \quad (3.27)$$

对于(3.27)右边的第二项, 有

$$\begin{aligned} \|\bar{x}_k^j - \bar{x}_k^i\| &= \|(\bar{x}_k^j - x_{k^0}^j) + (x_{k^0}^i - \bar{x}_k^i) + (x_{k^0}^j - x_{k^0}^i)\| \\ &\leq \|\bar{x}_k^j - x_{k^0}^j\| + \|x_{k^0}^i - \bar{x}_k^i\| + \|x_{k^0}^j - x_{k^0}^i\| \\ &\leq 4\eta(k - k^0)G_{\max} + \varepsilon, \end{aligned} \quad (3.28)$$

其中我们使用了不等式(3.24)和有界假设。将(3.17)和(3.28)代入(3.27)即可得到,

$$\|x_k^j - \bar{x}_k^i\| \leq 8\eta(k - k^0)G_{\max} + \varepsilon.$$

下面, 求出 $\|\bar{x}_{k-1}^j - \bar{x}_k^i\|$ 的界。

$$\begin{aligned} \|\bar{x}_{k-1}^j - \bar{x}_k^i\| &\leq \|x_{k-1}^j - \bar{x}_{k-1}^j\| + \|\bar{x}_{k-1}^j - \bar{x}_k^i\| \\ &\leq 4\eta(k - k^0)G_{\max} + \|\bar{x}_{k-1}^j - x_{k^0}^j\| + \|x_{k^0}^i - \bar{x}_k^i\| + \|x_{k^0}^j - x_{k^0}^i\| \\ &\leq 8\eta(k - k^0)G_{\max} + \varepsilon, \end{aligned} \quad (3.29)$$

证毕。

3.2. 收敛性结果

最后, 可以得到扩展的不精确 MUSIC 的收敛结果如下。

定理 1: 当假设 1~5 均成立且步长满足 $\eta \leq \frac{1}{3L}$ 时, 扩展的非精确 MUSIC 方法(2.5)在均方意义上线性收敛到最优解的邻域, 即对任意的 $t = 1, 2, \dots, \lfloor K/E \rfloor$, 有

$$\|\bar{x}_{tE}^i - x^*\|^2 \leq (1 - \mu\eta)^{tE} \|\bar{x}_0^i - x^*\|^2 + D_1, \quad (3.30)$$

其中

$$D_1 = \frac{1 - (1 - \mu\eta)^{kE}}{1 - (1 - \mu\eta)^E} \sum_{s=0}^{E-1} \xi^{E-1-s} (1 - \mu\eta)^s \xrightarrow{k \rightarrow \infty} \mathcal{O}\left(\frac{(E-1)^2 (128 + 32\gamma)\eta^2 G_{\max}^2 + 4\eta\tau}{\mu\eta}\right), \quad (3.31)$$

以及 $\xi^s = 4\gamma\eta^2 s^2 G_{\max}^2 + [4\eta s G_{\max} + \varepsilon]^2 + 2\eta\tau$ 。

证明：众所周知，不管是 $k \in \mathcal{I}_E$ 还是 $k \notin \mathcal{I}_E$ ， $\bar{x}_k^i = \bar{v}_k^i$ 总是成立的。因此，结合引理 1~3，我们有

$$\|\bar{x}_{k+1}^i - x^*\|^2 \leq (1 - \mu\eta) \|\bar{x}_k^i - x^*\|^2 + \xi^{k-k^0}. \quad (3.32)$$

为方便起见，定义 $\Delta^k = \|\bar{x}_k^i - x^*\|^2$ ，还可以将(3.32)写成

$$\Delta^{k+1} \leq (1 - \mu\eta) \Delta^k + \xi^{k-k^0}. \quad (3.33)$$

从 $k^0 + 1$ 到 $k^0 + E$ 通过递归应用(3.33)，有

$$\Delta^{k^0+E} \leq (1 - \mu\eta)^E \Delta^{k^0} + \sum_{s=0}^{E-1} \xi^{E-1-s} (1 - \mu\eta)^s, \quad (3.34)$$

或者

$$\|\bar{x}_{tE}^i - x^*\|^2 \leq (1 - \mu\eta)^E \|\bar{x}_{(t-1)E}^i - x^*\|^2 + D_2, \quad (3.35)$$

其中 $D_2 = \sum_{s=0}^{E-1} \xi^{E-1-s} (1 - \mu\eta)^s$ ， $t = 1, 2, \dots, \lfloor K/E \rfloor$ 。通过递归使用(3.35) t 次，我们得到

$$\|\bar{x}_{tE}^i - x^*\|^2 \leq (1 - \mu\eta)^{tE} \|\bar{x}_0^i - x^*\|^2 + D_1, \quad (3.36)$$

其中

$$D_1 = \frac{D_2 (1 - (1 - \mu\eta)^{tE})}{1 - (1 - \mu\eta)^E}. \quad (3.37)$$

当 $t \rightarrow \infty$ 时，我们可以得出结论：在标准共识策略下，局部估计值之间渐近地达成了共识(即 $\lim_{t \rightarrow \infty} \varepsilon = 0$)。此外，为了简化计算，我们还在(3.33)中使用了 $\xi^{k-k^0} = \xi^{E-1}$ ，从而得到

$$\limsup_{t \rightarrow \infty} \xi^{E-1} = (E-1)^2 (128 + 32\gamma)\eta^2 G_{\max}^2 + 4\eta\tau, \quad (3.38)$$

$$\limsup_{k \rightarrow \infty} D_2 = \frac{1 - (1 - \mu\eta)^E}{\mu\eta} \limsup_{k \rightarrow \infty} \xi^{E-1}. \quad (3.39)$$

以及

$$\limsup_{k \rightarrow \infty} D_1 = \frac{(E-1)^2 (128 + 32\gamma)\eta G_{\max}^2 + 4\tau}{\mu}, \quad (3.40)$$

证毕。

定理 1 表明，扩展的非精确 MUSIC 方法以 $\mathcal{O}((1 - \mu\eta)^{kE})$ 的速率线性收敛，收敛速度分别随 E 和 η 的增大单调递增和单调递减，直到达到误差邻域，且误差邻域大小为

$$\mathcal{O}\left(\underbrace{\frac{(E-1)^2 (128 + 32\gamma)\eta G_{\max}^2}{\mu}}_{\text{局部漂移}} + \underbrace{\frac{4\tau}{\mu}}_{\text{不精确偏差}}\right). \quad (3.41)$$

当 $E=1$ 时, 扩展的非精确 MUSIC 会退化为使用扩展梯度的 ATC 版本, 收敛速率为 $\mathcal{O}\left((1-\mu\eta)^t\right)$, 渐近误差大小为 $\mathcal{O}\left(\frac{4\tau}{\mu}\right)$ 。

注: (η 和 E 的选择) 一方面, 根据定理 1, 局部更新的频率 E 没有限制, 这意味着 E 可以取一个大的值来加速收敛。另一方面, 过大的 E 值会显著扩大误差邻域的大小。因此, 参数 E 在平衡收敛速度和精度之间的权衡方面起着类似于步长 η 的作用。因此, 通过选择一个略大于 1 的 E (例如 2, 3, 4) 和较小的步长, 就可以实现收敛速度和精度的双赢。这有效地解决了传统优化技术 (例如 DGD 方法) 长期存在的挑战。

4. 数值实验

在本节中, 我们通过解决二进制的逻辑回归学习分类问题来测试扩展的非精确 MUSIC 的性能。每个节点有一个如下局部目标函数

$$f_i(x) = \frac{1}{m} \sum_{j=1}^m \ln(1 + \exp(-\gamma_{i,j} h_{i,j}^T x)) + \frac{\mu}{2} \|x\|^2. \quad (3.42)$$

其中 $\{h_{i,j} \in \mathbb{R}^N\}$ 为特征向量, $\gamma_{i,j} \in \{-1, 1\}$ 为对应的标签。利用 “letter” 数据集和平均度为 4 的 Erdos-Renyi 模型来生成连接网络, 使用第二个和第四个标签将 “letter” 数据子集划分为 $n=50$ 个节点, 其中每个节点接收 $m=30$ 个维度为 $N=16$ 的训练样本。在这个问题中, 由于最优解 x^* 是未知的, 我们用一个非常小的步长运行集中式梯度下降迭代 2×10^5 次近似求出 x^* 。

从图 1 中可以看出, 当迭代过程中步长固定 ($\eta=0.4$) 时, 参数 E 的作用与步长 η 类似, 即 E 或 η 越大 (越小) 收敛速度越快 (越慢), 精度越低 (越高)。这样验证了 ExtendMUSIC 框架可以在速度和精度之间实现平衡, 例如, 既可以实现快速的速度, 又可以实现良好的精度。

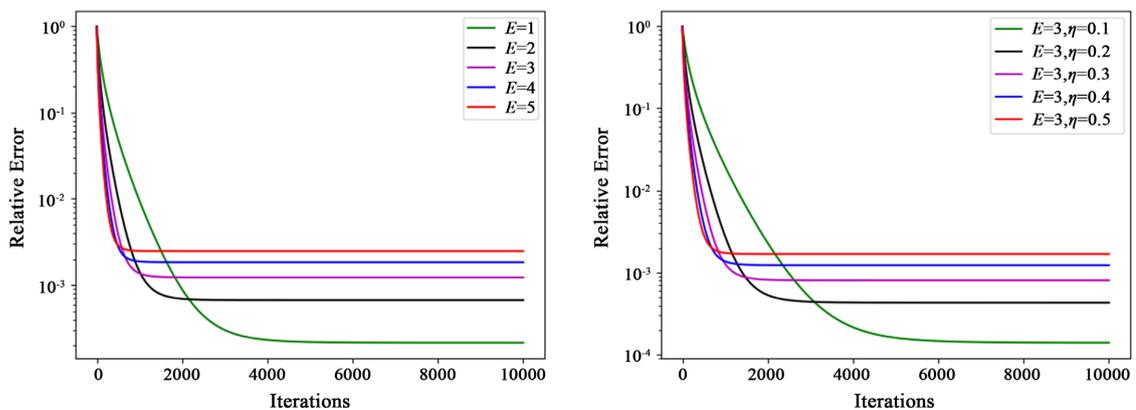


Figure 1. Effect of step size and E on the performance of the algorithm ExtendMUSIC. (a) The effect of different E on algorithm performance with a constant step size; (b) The effect of different step sizes on algorithm performance with a fixed E

图 1. 步长和 E 对算法 ExtendMUSIC 的性能影响。(a) 固定步长下, 不同的 E 对算法性能的影响; (b) 固定 E 下, 不同的步长对算法性能的影响

从图 2 中可以看出, 在相同步长下 ($\eta=0.2$), 使用扩展梯度的 ExtendMUSIC 方法比不使用扩展梯度的 MUSIC 方法可以更快找到非精确解, 同理, ATC 方法也是。此外, 在相同步长下 ($\eta=0.2$), 基于多更新单组合策略的方法比不使用多更新单组合策略的方法可以实现更快收敛。

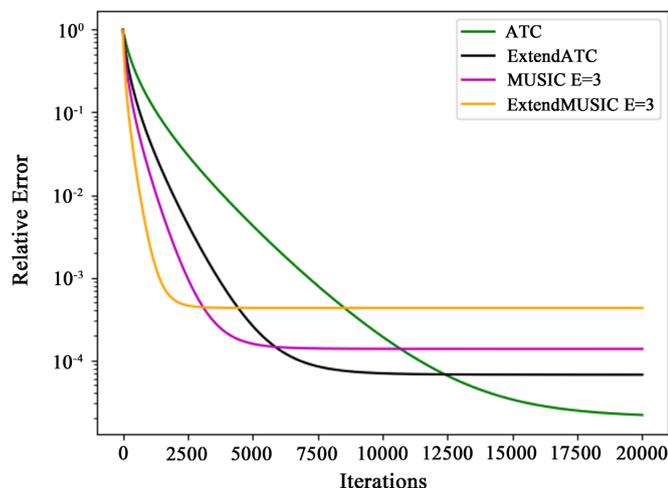


Figure 2. The performance comparison of different methods with the same step size
图 2. 相同步长下, 不同方法的性能比较

5. 总结与展望

本文首先介绍了一种基于多更新单组合策略的分布式扩展梯度方法 ExtendMUSIC。然后, 建立了该方法的线性收敛性。最后, 通过数值算例验证了该方法的收敛性以及不同的组合次数 E 对算法的影响, 并与现有的经典梯度方法进行了比较。在未来, 这项工作中使用的加速技术可以应用于更多的优化问题。

参考文献

- [1] Nedic, A. (2020) Distributed Gradient Methods for Convex Machine Learning Problems in Networks: Distributed Optimization. *IEEE Signal Processing Magazine*, **37**, 92-101. <https://doi.org/10.1109/msp.2020.2975210>
- [2] Duan, Y. and Wu, J. (2023) Accelerating Distributed Machine Learning with Model Compression and Graph Partition. *Journal of Parallel and Distributed Computing*, **179**, Article 104705. <https://doi.org/10.1016/j.jpdc.2023.04.006>
- [3] Yan, J., Jiao, H., Pu, W., Shi, C., Dai, J. and Liu, H. (2022) Radar Sensor Network Resource Allocation for Fused Target Tracking: A Brief Review. *Information Fusion*, **86**, 104-115. <https://doi.org/10.1016/j.inffus.2022.06.009>
- [4] Guo, S., Zhao, X., Wang, H. and Xu, N. (2023) Distributed Consensus of Heterogeneous Switched Nonlinear Multiagent Systems with Input Quantization and Dos Attacks. *Applied Mathematics and Computation*, **456**, Article 128127. <https://doi.org/10.1016/j.amc.2023.128127>
- [5] Li, Y., He, X. and Xia, D. (2022) Distributed Fixed-Time Optimization for Multi-Agent Systems with Time-Varying Objective Function. *International Journal of Robust and Nonlinear Control*, **32**, 6523-6538. <https://doi.org/10.1002/rnc.6157>
- [6] Mbungu, N.T., Ismail, A.A., AlShabi, M., Bansal, R.C., Elnady, A. and Hamid, A.K. (2023) Control and Estimation Techniques Applied to Smart Microgrids: A Review. *Renewable and Sustainable Energy Reviews*, **179**, Article 113251. <https://doi.org/10.1016/j.rser.2023.113251>
- [7] Nedic, A. and Ozdaglar, A. (2009) Distributed Subgradient Methods for Multi-Agent Optimization. *IEEE Transactions on Automatic Control*, **54**, 48-61. <https://doi.org/10.1109/tac.2008.2009515>
- [8] Yuan, K., Ling, Q. and Yin, W. (2016) On the Convergence of Decentralized Gradient Descent. *SIAM Journal on Optimization*, **26**, 1835-1854. <https://doi.org/10.1137/130943170>
- [9] Shi, W., Ling, Q., Wu, G. and Yin, W. (2015) EXTRA: An Exact First-Order Algorithm for Decentralized Consensus Optimization. *SIAM Journal on Optimization*, **25**, 944-966. <https://doi.org/10.1137/14096668x>
- [10] Qu, G. and Li, N. (2018) Harnessing Smoothness to Accelerate Distributed Optimization. *IEEE Transactions on Control of Network Systems*, **5**, 1245-1260. <https://doi.org/10.1109/tcns.2017.2698261>
- [11] Gao, J., Liu, X., Dai, Y., Huang, Y. and Yang, P. (2021) Achieving Geometric Convergence for Distributed Optimization

- with Barzilai-Borwein Step Sizes. *Science China Information Sciences*, **65**, Article No. 149204. <https://doi.org/10.1007/s11432-020-3256-x>
- [12] Popov, L.D. (1980) A Modification of the Arrow-Hurwicz Method for Search of Saddle Points. *Mathematical Notes of the Academy of Sciences of the USSR*, **28**, 845-848. <https://doi.org/10.1007/bf01141092>
- [13] Attouch, H., Chbani, Z., Fadili, J. and Riahi, H. (2020) First-Order Optimization Algorithms via Inertial Systems with Hessian Driven Damping. *Mathematical Programming*, **193**, 113-155. <https://doi.org/10.1007/s10107-020-01591-1>
- [14] Mai, V. and Johansson, M. (2020) Anderson Acceleration of Proximal Gradient Methods. In *International Conference on Machine Learning*, Online, 13-18 July 2020, 6620-6629.
- [15] Tang, W. and Daoutidis, P. (2021) Fast and Stable Nonconvex Constrained Distributed Optimization: The ELLADA Algorithm. *Optimization and Engineering*, **23**, 259-301. <https://doi.org/10.1007/s11081-020-09585-w>
- [16] Zhang, X., Liu, S. and Zhao, N. (2023) An Extended Gradient Method for Smooth and Strongly Convex Functions. *Mathematics*, **11**, Article 4771. <https://doi.org/10.3390/math11234771>
- [17] Wu, M., Liao, H., Ding, Z. and Xiao, Y. (2024) MUSIC: Accelerated Convergence for Distributed Optimization with Inexact and Exact Methods. *IEEE Transactions on Neural Networks and Learning Systems*, 1-15. <https://doi.org/10.1109/tnnls.2024.3376421>
- [18] Mangasarian, L.O. (1995) Parallel Gradient Distribution in Unconstrained Optimization. *SIAM Journal on Control and Optimization*, **33**, 1916-1925. <https://doi.org/10.1137/s0363012993250220>
- [19] Kairouz, P., McMahan, H.B., Avent, B., Bellet, A., Bennis, M., Nitin Bhagoji, A., *et al.* (2021) Advances and Open Problems in Federated Learning. Now Publishers. <https://doi.org/10.1561/9781680837896>