

有向网络中一种快速的分布式随机算法

宫秀慧

河北工业大学理学院, 天津

收稿日期: 2024年1月28日; 录用日期: 2024年2月22日; 发布日期: 2024年2月29日

摘要

本文考虑有向网络上多智能体系统中的分布式优化问题。其全局目标函数可表示为网络中所有局部目标函数有限和的形式。通过利用 Nesterov 动量技巧方法和单循环的方差缩减技术 LSVRG, 本文提出了有向网络中一种快速的分布式随机算法 AB-LSVRG。对光滑和强凸的目标函数, 理论分析证明所提出的算法可以线性收敛到最优解。基于分布式逻辑回归问题, 数值实验表明本文所提出的算法与现有的一些分布式算法相比表现效果更好。

关键词

分布式优化, 多智能系统, 有向网络, 分布式方差缩减, 动量加速

A Fast Distributed Stochastic Algorithm over Directed Networks

Xiuhui Gong

School of Science, Hebei University of Technology, Tianjin

Received: Jan. 28th, 2024; accepted: Feb. 22nd, 2024; published: Feb. 29th, 2024

Abstract

This paper considers the distributed optimization in a multi-agent system over balanced directed networks. The global objective function describes a finite sum of all local objective functions on the networks. Combining the distributed loopless variance

reduction method with Nesterov momentum strategy, a fast distributed stochastic algorithm is developed, named ABN-LSVRG. For smooth and strongly convex objective functions, it is proved that ABN-LSVRG has a linear convergence rate. Based on the distributed logistic problem, simulation results show that ABN-LSVRG performs better in comparison with some distributed algorithms.

Keywords

Distributed Optimization, Multi-Agent System, Directed Networks, Distributed Variance Reduction, Momentum Acceleration

Copyright © 2024 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/>



1. 引言

分布式优化已被广泛应用于机器学习 [1,2], 网络系统控制 [3] 和资源分配 [4] 等领域. 这些实际问题通常可以建模为最小化网络中多个局部目标函数平均值的形式, 每个局部目标函数 f_i 仅被智能体 i 所知. 网络中所有智能体的目标是通过局部通信和计算找到全局最优解. 网络中智能体之间的通信可以抽象为一个网络拓扑图, 包括无向网络和有向网络. 在无向网络中, 智能体之间的通信是双向和平衡的. 但在实际应用中, 智能体之间的通信往往是不平衡且有向的. 因此基于有向网络的分布式优化问题受到了广泛关注. 本文主要考虑有向网络中的分布式随机优化算法.

许多分布式确定性的算法被提出, 例如分布式梯度下降算法 (DGD) [5], 梯度追踪算法 [6–8]. 基于这些算法, 一些带有动量加速的高效的分布式优化算法得到了发展. 由于上述算法考虑的是局部目标函数的全梯度, 当网络中的节点在梯度估计时受到噪声的干扰, 精确的梯度无法计算. 因此, 有学者考虑用随机梯度来近似全梯度.

DGD 算法的随机变体分布式随机梯度下降方法 DSGD [9,10] 用于解决这类问题. 进一步, 将梯度跟踪方法 [11,12] 和 DSGD 算法相结合提出了分布式随机梯度跟踪方法 (DSGT) [13], 该方法改善了 DSGD 算法的收敛. 然而, 由于随机梯度方差的存在, DSGD 算法和 DSGT 算法在固定步长下只能收敛到最优解的领域内. 受集中式随机优化算法的启发 [14–16], 许多带有方差缩减的分布式随机优化算法被提出. 文献 [17] 在梯度跟踪方法的基础下利用方差缩减技术提出了两个快速的分布式随机算法 GT-SAGA 和 GT-SVRG. 这两个算法在固定步长下可以线性收敛到最优解.

以上所提到的分布式随机算法仅适用于无向网络或者权重平衡的有向网络, 这些算法需要构造双随机权重矩阵. 然而, 基于有向网络的分布式优化问题, 构造双随机矩阵是很困难的. 为了解决有向网络中的这类问题, 将 push-sum 方法 [18] 与 DSGD 算法相结合提出了分布式随机 gradient-push (SGP) 算法 [19,20]. 该方法通过利用 push-sum 技巧进行特征向量的估计, 从而消除了有向网络引起的不平衡性. 进一步, 文献 [21] 将梯度跟踪方法与 SGP 相结合提出了 SADDOPT 算法, 该算法在

梯度跟踪方法的帮助下改善了 SGP 算法的收敛. 通过同时使用行随机和列随机权重矩阵, 文献 [22] 提出了一个分布式随机一阶方法 S-AB. 类似地, 以上有向网络中的分布式随机算法受到了随机梯度方差的影响, 在固定步长下只能收敛到最优解的领域内. 因此, 文献 [23] 利用 SAGA 方差缩减技术 [14] 提出了第一个线性收敛到最优解的分布式随机方法 Push-SAGA. 后来, 文献 [24] 在 S-AB 算法中引入 SAGA 方法提出了 AB-SAGA 算法, 该算法避免了估计特征向量, 减少了计算和通信. 基于 SAGA 方差缩减的分布式随机算法需要一个梯度表来储存梯度, 这会产生存储的负担. 当数据量很大时, 存储大量的梯度是难以实现的. 文献 [25] 利用双循环的方差缩减技术 SVRG [15] 提出了 Push-SVRG 和 AB-SVRG 算法, 这两种算法相较于 Push-SAGA 和 AB-SAGA 收敛速度略慢, 但不需要存储空间. 最近, 文献 [26] 采用单循环的方差缩减技术 [27] 提出了 Push-LSVRG 算法, 该算法具有加速线性收敛和较低的存储成本.

为了加速算法收敛, 带有动量加速的分布式优化算法得到了发展. 文献 [28] 将 GT-SAGA 算法与 heavy-ball 动量加速相结合提出了一种带有梯度跟踪和分布式重球加速的分布式随机优化算法. 对光滑和强凸的目标函数, 该算法可以实现更快的收敛. 通过将 S-AB 算法与 Nesterov 动量相结合, 有向网络中一种分布式随机 Nesterov 算法 [29] 被提出. 受动量加速方法的启发, 利用 Nesterov 动量加速技巧, 本文提出了有向网络中一种快速的方差缩减的分布式随机优化算法.

用小写粗体字母表示向量, 用大写斜体字母表示矩阵. I_n 表示 $n \times n$ 阶单位矩阵, $\mathbf{1}_n$ 表示 n 个元素全为 1 的列向量. 矩阵 X 和 Y 的克罗内克积定义为 $X \otimes Y$, 对一个任意向量 \mathbf{x} , 用 \mathbf{x}_i 表示它的第 i 个元素, $\bar{\mathbf{x}}$ 和 $\underline{\mathbf{x}}$ 分别表示它的最大元素和最小元素, $\text{diag}(\mathbf{x})$ 表示以向量 \mathbf{x} 的元素生成的对角矩阵. 对一个矩阵 X , $\rho(X)$ 表示它的谱半径并且 X_∞ 表示它的无限次幂 (如果它存在), 即 $X_\infty = \lim_{k \rightarrow \infty} X^k$. 对一个原始的行随机矩阵 A , 用 π_r 和 $\mathbf{1}_n$ 分别表示它的对应于特征值为 1 的左和右特征向量, 即 $\pi_r \mathbf{1}_n^T = 1$ 且 $A_\infty = \mathbf{1}_n \pi_r^T$. 对于一个原始的列随机矩阵 B , 有 $B_\infty = \pi_c \mathbf{1}_n^T$. 对任意向量 $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$ 引入两个权重内积, $\langle \mathbf{x}, \mathbf{y} \rangle_{\pi_r} = \mathbf{x}^T \text{diag}(\pi_r) \mathbf{y}$ 和 $\langle \mathbf{x}, \mathbf{y} \rangle_{\pi_c} = \mathbf{x}^T \text{diag}(\pi_c)^{-1} \mathbf{y}$. 用 $\|\mathbf{x}\|_{\pi_r}$ 和 $\|\mathbf{x}\|_{\pi_c}$ 表示权重欧式范数. 用 $\|\cdot\|_{\pi_r}$ 和 $\|\cdot\|_{\pi_c}$ 分别表示由 $\|\cdot\|_{\pi_r}$ 和 $\|\cdot\|_{\pi_c}$ 诱导的矩阵范数, 即 $\forall X \in \mathbb{R}^{n \times n}$, 有 $\|X\|_{\pi_r} = \|\text{diag}(\sqrt{\pi_r}) X \text{diag}(\sqrt{\pi_r})^{-1}\|_2$ 和 $\|X\|_{\pi_c} = \|\text{diag}(\sqrt{\pi_c})^{-1} X \text{diag}(\sqrt{\pi_c})\|_2$.

2. 问题描述

考虑由 n 个智能体构成的有向网络, 网络中的所有智能体合作求解以下分布式优化问题:

$$\min_{\mathbf{x} \in \mathbb{R}^p} F(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n f_i(\mathbf{x}), f_i(\mathbf{x}) = \frac{1}{m_i} \sum_{j=1}^n f_{i,j}(\mathbf{x}), \quad (1)$$

其中 $\mathbf{x} \in \mathbb{R}^p$ 是全局决策变量, 局部目标函数 $f_i: \mathbb{R}^p \rightarrow \mathbb{R}$ 仅被智能体 i 所知. 每个局部目标函数进一步可以表示为 m_i 个分量目标函数的平均值的形式. 网络中所有智能体的目标是通过局部通信和计算找到全局最优解. 智能体的信息交换通过有向网络 $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ 进行. 其中 $\mathcal{V} = \{1, \dots, n\}$ 表示网络中智能体的集合, \mathcal{E} 表示网络中智能体之间通信链路的集合. 用 \mathcal{N}_i^{in} 表示智能体 i 入邻居的集合, 即智能体 i 能够收到集合中每个智能体发送的信息. 同理, 用 \mathcal{N}_i^{out} 表示智能体 i 出邻居的集合, 即智能体 i 能够将信息发送到集合中的每个智能体. 接下来, 给出一些标准的假设.

假设1. 每个全局目标 F 是 μ 强凸的, 即 $\forall i \in \mathcal{V}$ 和 $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^p$, 存在常数 $\mu > 0$, 有

$$F(\mathbf{x}_2) \geq F(\mathbf{x}_1) + \nabla F(\mathbf{x}_1)^T (\mathbf{x}_2 - \mathbf{x}_1) + \frac{\mu}{2} \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2.$$

假设2. 每个局部目标 f_i 是 L -光滑的, 即 $\forall i \in \mathcal{V}$ 和 $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^p$, 存在常数 $L > 0$, 有

$$\|\nabla f_i(\mathbf{x}_1) - \nabla f_i(\mathbf{x}_2)\|_2 \leq L\|\mathbf{x}_1 - \mathbf{x}_2\|_2.$$

假设3. 图 \mathcal{G} 是强连通的.

假设 1 表明全局目标函数存在唯一的最优解. 假设 2 表明全局目标函数 F 是 L 光滑的, 用 $\kappa = \frac{L}{\mu}$ 表示全局目标函数 F 的条件数.

3. 算法

本节首先介绍单循环的方差缩减方法 (LSVRG), 然后将其引入到 S-AB 算法中, 并结合 Nesterov 动量加速方法提出了 ABN-LSVRG 算法.

令 τ_k^i 是计算局部全梯度 ∇f 的参考点. 在每次迭代中, 每个智能体 i 从 $\{1, \dots, n\}$ 中随机选取 l_k^i , 参考点 τ_k^i 的更新规则为: 以 $\frac{1}{m}$ 的概率取 \mathbf{x}_{k-1}^i , 否则 $\tau_k^i = \tau_{k-1}^i$. 在第 k 次迭代中, 分布式 LSVRG 算法的迭代格式为:

$$\mathbf{g}_k^i = \nabla f_{i,l_k^i}(\mathbf{x}_k^i) - \nabla f_{i,l_k^i}(\tau_k^i) + \nabla f_i(\tau_k^i). \tag{2}$$

针对分布式优化问题, 文献 [22] 同时使用行随机和列随机权重矩阵提出了 S-AB 算法, 该算法与 SADDOPT 算法相比避免特征向量的计算, 减少计算和通信成本. 在分布式优化中, 动量加速技巧受到了广泛的关注, 例如: 分布式 heavy-ball 算法和分布式 Nesterov 算法. 本文将分布式 LSVRG 算法与 S-AB 算法相结合, 并引入 Nesterov 动量项, 提出了一种快速的分布式随机方差缩减方法, 简称 ABN-LSVRG. 在第 k 次迭代中, 每个智能体更新四个变量, 即 $\mathbf{y}_k^i, \mathbf{s}_k^i, \mathbf{g}_k^i$ 和 \mathbf{x}_k^i . 任意的初始迭代 $\mathbf{s}_0^i = \mathbf{x}_0^i$ 和 $\mathbf{g}_{-1}^i = \mathbf{y}_{-1}^i = \nabla f(\mathbf{x}_0)$, $\forall i \in \mathcal{V}$. 所提出的算法 ABN-LSVRG 迭代格式为:

$$\begin{aligned} \mathbf{y}_k^i &= \sum_{j=1}^n b_{ij} \mathbf{y}_{k-1}^j + \mathbf{g}_k - \mathbf{g}_{k-1}, \\ \mathbf{s}_{k+1}^i &= \sum_{j=1}^n a_{ij} \mathbf{x}_k^j - \alpha \mathbf{y}_k^i, \\ \mathbf{x}_{k+1}^i &= \mathbf{s}_{k+1}^i + \beta(\mathbf{s}_{k+1}^i - \mathbf{s}_k^i). \end{aligned} \tag{3}$$

其实 a_{ij} 和 b_{ij} 分别满足以下性质:

$$\begin{aligned} a_{ij} &= \begin{cases} > 0, j \in \mathcal{N}_i^{\text{in}}, \\ 0, \text{ 否则}, \end{cases} & \sum_{j=1}^n a_{ij} = 1, \forall i, \\ b_{ij} &= \begin{cases} > 0, i \in \mathcal{N}_j^{\text{out}}, \\ 0, \text{ 否则}, \end{cases} & \sum_{i=1}^n b_{ij} = 1, \forall j. \end{aligned} \tag{4}$$

权重矩阵 $A = [a_{ij}] \in \mathbb{R}^{n \times n}$ 是行随机矩阵, $B = [b_{ij}] \in \mathbb{R}^{n \times n}$ 是列随机矩阵.

定义 $\mathbf{s}_k, \mathbf{x}_k, \mathbf{g}_k, \mathbf{y}_k \in \mathbb{R}^p$, i.e. $\mathbf{s}_k = [\mathbf{s}_k^1, \dots, \mathbf{s}_k^n]^T, \mathbf{x}_k = [\mathbf{x}_k^1, \dots, \mathbf{x}_k^n]^T, \mathbf{g}_k = [\mathbf{g}_k^1, \dots, \mathbf{g}_k^n]^T, \mathbf{y}_k =$

$[\mathbf{y}_k^1, \dots, \mathbf{y}_k^n]^T$, $\mathbf{A} = A \otimes I_p$ 和 $\mathbf{B} = B \otimes I_p$. 利用如上定义将所提出算法 ABN-LSVRG 等价的写为

$$\begin{aligned} \mathbf{y}_k &= \mathbf{B}\mathbf{y}_{k-1} + \mathbf{g}_k - \mathbf{g}_{k-1}, \\ \mathbf{s}_{k+1} &= \mathbf{A}\mathbf{s}_k - \alpha\mathbf{y}_k, \\ \mathbf{x}_{k+1} &= \mathbf{s}_{k+1} + \beta(\mathbf{s}_{k+1} - \mathbf{s}_k). \end{aligned} \tag{5}$$

4. 收敛性分析

在进行收敛性分析之前, 定义一些符号

$$\bar{\mathbf{y}}_k = \frac{1}{n} \mathbf{1}_n^T \mathbf{y}_k, \nabla \bar{\mathbf{f}}(\mathbf{x}_k) = \frac{1}{n} \mathbf{1}_n^T \nabla f(\mathbf{x}_k), \hat{\mathbf{s}}_k = \boldsymbol{\pi}_r^T \mathbf{s}_k, \bar{\mathbf{g}}_k = \frac{1}{n} \mathbf{1}_n^T \mathbf{g}_k,$$

其中 $\nabla \mathbf{f}(\mathbf{x}_k) = [\nabla f_1(\mathbf{x}_k^1), \dots, \nabla f_n(\mathbf{x}_k^n)]$.

以下引理说明了行随机和列随机权重矩阵的收缩性质, 其证明可以参考 [22].

引理1. 若假设 3 成立, 对 $\forall k \geq 0$, 考虑权重矩阵 A, B 和任意向量 $\mathbf{x} \in \mathbb{R}^p$, 有

$$\begin{aligned} \|\mathbf{A}\mathbf{x} - A_\infty \mathbf{x}\|_{\boldsymbol{\pi}_r} &\leq \sigma_A \|\mathbf{x} - A_\infty \mathbf{x}\|_{\boldsymbol{\pi}_r}, \\ \|\mathbf{B}\mathbf{x} - B_\infty \mathbf{x}\|_{\boldsymbol{\pi}_c} &\leq \sigma_B \|\mathbf{x} - B_\infty \mathbf{x}\|_{\boldsymbol{\pi}_c}. \end{aligned} \tag{6}$$

其中 $\sigma_A = \|A - A_\infty\|_{\boldsymbol{\pi}_r} < 1$ 且 $\sigma_B = \|B - B_\infty\|_{\boldsymbol{\pi}_c} < 1$.

由范数的定义得

$$\begin{aligned} \|\cdot\|_{\boldsymbol{\pi}_r} &\leq \bar{\boldsymbol{\pi}}_r^{0.5} \|\cdot\|_2, \|\cdot\|_{\boldsymbol{\pi}_c} \leq \underline{\boldsymbol{\pi}}_c^{-0.5} \|\cdot\|_2, \\ \|\cdot\|_2 &\leq \underline{\boldsymbol{\pi}}_r^{-0.5} \|\cdot\|_{\boldsymbol{\pi}_r}, \|\cdot\|_2 \leq \bar{\boldsymbol{\pi}}_c^{0.5} \|\cdot\|_{\boldsymbol{\pi}_c}. \end{aligned} \tag{7}$$

下面的引理表明了跟踪梯度的平均 $\bar{\mathbf{y}}_k$ 和局部梯度估计的平均 $\bar{\mathbf{g}}_k$ 的一致.

引理2. [24] 若假设 4 成立, 对 $\forall k \geq 0$, 有

$$\begin{aligned} \bar{\mathbf{y}}_k &= \bar{\mathbf{g}}_k, \\ \mathbb{E}[\bar{\mathbf{y}}_k | \mathcal{F}_k] &= \mathbb{E}[\bar{\mathbf{g}}_k | \mathcal{F}_k] = \nabla \bar{\mathbf{f}}(\mathbf{x}_k). \end{aligned}$$

以下两个引理对收敛性证明至关重要, 可参考文献 [24, 29].

引理3. [29] 若假设 2 成立, 考虑由所提出的算法产生的迭代 $\{\mathbf{x}_k\}_{k \geq 0}$. 对 $\forall k \geq 0$, 有

$$\|\nabla \bar{\mathbf{f}}(\mathbf{x}_k) - \nabla F(\hat{\mathbf{s}}_k)\|_2 \leq \frac{L}{\sqrt{n}} \|\mathbf{s}_k - \mathbf{1}_n^T \hat{\mathbf{s}}_k\|_2^2 + \frac{L}{\sqrt{n}} \beta \|\hat{\mathbf{s}}_k - \mathbf{x}^*\|_2. \tag{8}$$

引理4. [24] 若假设 1 和 2 成立, 全局目标函数 F 是 μ 强凸且 L 光滑的. 如果 $0 < \alpha < \frac{1}{L}$, 对 $\forall \mathbf{s} \in \mathbb{R}^p$, 有

$$\|\mathbf{s} - \alpha \nabla F(\mathbf{s}) - \mathbf{x}^*\|_2 \leq (1 - \mu\alpha) \|\mathbf{s} - \mathbf{x}^*\|_2. \tag{9}$$

引理5. [17] 令 $G \in \mathbb{R}^{np \times np}$ 是一个非负矩阵且 $\boldsymbol{\gamma}$ 是一个正向量. 若 $G\boldsymbol{\gamma} < \omega\boldsymbol{\gamma} (\omega > 0)$, 有 $\rho(G) < \omega$.

本文通过考虑以下 5 个量对 ABN-LSVRG 算法进行收敛性分析.

- 1) 一致误差: $\mathbb{E}[\|\mathbf{s}_{k+1} - \mathbf{1}_n \hat{\mathbf{s}}_{k+1}\|_{\pi_r}^2]$;
- 2) 最优间隙: $\mathbb{E}[\|\hat{\mathbf{s}}_{k+1} - \mathbf{x}^*\|_2^2]$;
- 3) 状态最优性误差: $\mathbb{E}[\mathbf{v}_{k+1}]$, 其中 $\mathbf{v}_{k+1} = \frac{1}{m_i} \sum_{j=1}^{m_i} \|f_{i,j}(\tau_{k+1}^{i,j}) - f_{i,j}(\mathbf{x}^*)\|_2^2$;
- 4) 状态误差: $\mathbb{E}[\|\mathbf{s}_{k+1} - \mathbf{s}_k\|_2^2]$;
- 5) 梯度跟踪误差: $\mathbb{E}[\|\mathbf{y}_{k+1} - \mathcal{B}_\infty \mathbf{y}_{k+1}\|_{\pi_c}^2]$.

为了证明 ABN-LSVRG 的收敛性, 首先给出梯度跟踪的上界.

引理6. 设假设 2 成立, 对 $\forall k \geq 0$, 以下不等式成立

$$\begin{aligned} \mathbb{E}[\|\mathbf{y}_k\|_2^2] &\leq \frac{45n\|\pi_c\|_2^2 L^2}{\pi_r} \mathbb{E}[\|\mathbf{s}_k - \mathbf{1}_n \hat{\mathbf{s}}_k\|_{\pi_r}^2] + 25n^2 \|\pi_c\|_2^2 L^2 \mathbb{E}[\|\hat{\mathbf{s}}_k - \mathbf{x}^*\|_2^2] \\ &\quad + 45n\|\pi_c\|_2^2 L^2 \beta^2 \mathbb{E}[\|\mathbf{s}_k - \mathbf{s}_{k-1}\|_2^2] + 10\|\pi_c\|_2^2 \mathbb{E}[\mathbf{v}_k] \\ &\quad + 5\pi_c \mathbb{E}[\|\mathbf{y}_k - \mathcal{B}_\infty \mathbf{y}_k\|_{\pi_c}^2]. \end{aligned} \tag{10}$$

证明: 注意到 $\mathcal{B}_\infty = \pi_c \mathbf{1}_n^T$, 有

$$\begin{aligned} \|\mathbf{y}_k\|_2 &= \|\mathbf{y}_k - \pi_c \mathbf{1}_n^T \mathbf{y}_k + \pi_c \mathbf{1}_n^T \mathbf{y}_k\|_2 \\ &\leq \bar{\pi}_c^{0.5} \|\mathbf{y}_k - \pi_c \mathbf{1}_n^T \mathbf{y}_k\|_{\pi_c} + \|\pi_c \mathbf{1}_n^T \mathbf{y}_k\|_2 \\ &\leq \bar{\pi}_c^{0.5} \|\mathbf{y}_k - \pi_c \mathbf{1}_n^T \mathbf{y}_k\|_{\pi_c} + n\|\pi_c\|_2 \|\bar{\mathbf{g}}_k - \nabla \bar{\mathbf{f}}(\mathbf{x}_k)\|_2 \\ &\quad + n\|\pi_c\|_2 \|\nabla \bar{\mathbf{f}}(\mathbf{x}_k) - \nabla F(\hat{\mathbf{s}}_k)\|_2 + n\|\pi_c\|_2 \|\nabla F(\hat{\mathbf{s}}_k) - \nabla F(\mathbf{x}^*)\|_2 \\ &\leq \bar{\pi}_c^{0.5} \|\mathbf{y}_k - \mathcal{B}_\infty \mathbf{y}_k\|_{\pi_c} + n\|\pi_c\|_2 \|\bar{\mathbf{g}}_k - \nabla \bar{\mathbf{f}}(\mathbf{x}_k)\|_2 + \frac{\sqrt{n}\|\pi_c\|_2 L}{\sqrt{\pi_r}} \|\mathbf{s}_k - \mathbf{1}_n \hat{\mathbf{s}}_k\|_{\pi_r} \\ &\quad + \sqrt{n}\beta \|\pi_r\|_2 L \|\mathbf{s}_k - \mathbf{s}_{k-1}\|_2 + n\|\pi_c\|_2 L \|\hat{\mathbf{s}}_k - \mathbf{x}^*\|_2. \end{aligned} \tag{11}$$

其中最后一个不等式利用了引理 3. 在上述不等式的两边同时平方并取条件期望可得

$$\begin{aligned} \mathbb{E}[\|\mathbf{y}_k\|_2^2 | \mathcal{F}_k] &\leq \frac{5n\|\pi_c\|_2^2 L^2}{\pi_r} \|\mathbf{s}_k - \mathbf{1}_n \hat{\mathbf{s}}_k\|_{\pi_r}^2 + 5n^2 \|\pi_c\|_2^2 L^2 \|\hat{\mathbf{s}}_k - \mathbf{x}^*\|_2^2 \\ &\quad + 5n\beta^2 \|\pi_c\|_2^2 L^2 \|\mathbf{s}_k - \mathbf{s}_{k-1}\|_2^2 + 5\pi_c \|\mathbf{y}_k - \mathcal{B}_\infty \mathbf{y}_k\|_{\pi_c}^2 \\ &\quad + 5n^2 \|\pi_c\|_2^2 \mathbb{E}[\|\bar{\mathbf{g}}_k - \nabla \bar{\mathbf{f}}(\mathbf{x}_k)\|_2^2 | \mathcal{F}_k]. \end{aligned} \tag{12}$$

其中 $\mathbb{E}[\|\bar{\mathbf{y}}_k - \nabla \bar{\mathbf{f}}(\mathbf{x}_k)\|_2^2 | \mathcal{F}_k] = \frac{1}{n^2} \mathbb{E}[\|\mathbf{g}_k - \nabla \mathbf{f}(\mathbf{x}_k)\|_2^2 | \mathcal{F}_k]$. 参考文献 [24], $\mathbb{E}[\|\mathbf{g}_k - \nabla \mathbf{f}(\mathbf{x}_k)\|_2^2 | \mathcal{F}_k]$ 的表达式如下所示

$$\begin{aligned} \mathbb{E}[\|\mathbf{g}_k - \nabla \mathbf{f}(\mathbf{x}_k)\|_2^2 | \mathcal{F}_k] &= \frac{8L^2}{\pi_r} \mathbb{E}[\|\mathbf{s}_k - \mathbf{1}_n \hat{\mathbf{s}}_k\|_{\pi_r}^2] + 4nL^2 \mathbb{E}[\|\hat{\mathbf{s}}_k - \mathbf{x}^*\|_2^2] \\ &\quad + 8\beta^2 L^2 \mathbb{E}[\|\mathbf{s}_k - \mathbf{s}_{k-1}\|_2^2] + 2\mathbb{E}[\mathbf{v}_k]. \end{aligned} \tag{13}$$

将 (13) 带入 (12) 并在式子两边同时取全期望完成证明.

引理7. 若假设 2 和 3 成立. 如果 $0 < \alpha \leq \frac{1}{\|\pi_c\|_2 L} \sqrt{\frac{h}{10n}}$ 并且 $0 \leq \beta < 1$, 对 $\forall k \geq 0$, 网络中的一

致误差如下

$$\begin{aligned} \mathbb{E}[\|\mathbf{s}_{k+1} - \mathbf{1}_n \hat{\mathbf{s}}_{k+1}\|_{\pi_r}^2] &\leq \left(\frac{1 + \sigma_A^2}{2} + \beta + \frac{90nh\|\pi_c\|_2^2 L^2}{1 - \sigma_A^2} \alpha^2\right) \mathbb{E}[\|\mathbf{s}_k - \mathbf{1}_n \hat{\mathbf{s}}_k\|_{\pi_r}^2] \\ &\quad + \frac{50n^2\|\pi_c\|_2^2 L^2 \bar{\pi}_r}{1 - \sigma_A^2} \alpha^2 \mathbb{E}[\|\hat{\mathbf{s}}_k - \mathbf{x}^*\|_2^2] \\ &\quad + ((1 + \sigma_A^2)\bar{\pi}_r + \frac{90n\|\pi_c\|_2^2 L^2 \bar{\pi}_r}{1 - \sigma_A^2} \alpha^2) \beta \mathbb{E}[\|\mathbf{s}_k - \mathbf{s}_{k-1}\|_2^2] \\ &\quad + \frac{20\|\pi_c\|_2^2 \bar{\pi}_r}{1 - \sigma_A^2} \alpha^2 \mathbb{E}[\mathbf{v}_k] + \frac{10\bar{\pi}_r \bar{\pi}_c}{1 - \sigma_A^2} \alpha^2 \mathbb{E}[\|\mathbf{y}_k - \mathcal{B}_\infty \mathbf{y}_k\|_{\pi_c}^2], \end{aligned} \tag{14}$$

其中 $h = \bar{\pi}_r / \bar{\pi}_r$.

证明: 注意到 $\mathcal{A}_\infty = \mathbf{1}_n \pi_r^T$ 和 $\hat{\mathbf{x}}_k = \pi_r^T \mathbf{x}_k$, 由 \mathbf{s}_k 的定义可知

$$\begin{aligned} \|\mathbf{s}_{k+1} - \mathbf{1}_n \hat{\mathbf{s}}_{k+1}\|_{\pi_r}^2 &= \|\mathcal{A} \mathbf{x}_k - \mathcal{A}_\infty \mathbf{x}_k - \alpha(\mathbf{I}_n - \mathcal{A}_\infty) \mathbf{y}_k\|_{\pi_r}^2 \\ &\leq \|\mathcal{A} \mathbf{x}_k - \mathcal{A}_\infty \mathbf{x}_k\|_{\pi_r}^2 + \alpha^2 \|\mathbf{y}_k\|_{\pi_r}^2 + 2\|\mathcal{A} \mathbf{x}_k - \mathcal{A}_\infty \mathbf{x}_k\|_{\pi_r} \|\alpha \mathbf{y}_k\|_{\pi_r} \\ &\leq \|\mathcal{A} \mathbf{x}_k - \mathcal{A}_\infty \mathbf{x}_k\|_{\pi_r}^2 + \alpha^2 \|\mathbf{y}_k\|_{\pi_r}^2 \\ &\quad + (\lambda \|\mathcal{A} \mathbf{x}_k - \mathcal{A}_\infty \mathbf{x}_k\|_{\pi_r}^2 + \frac{1}{\lambda} \alpha^2 \|\mathbf{y}_k\|_{\pi_r}^2), \end{aligned} \tag{15}$$

在上述不等式中分别令 $\lambda = \frac{1 - \sigma_A^2}{2\sigma_A^2}$ 和 1 有

$$\|\mathbf{s}_{k+1} - \mathbf{1}_n \hat{\mathbf{s}}_{k+1}\|_{\pi_r}^2 \leq \frac{1 + \sigma_A^2}{2\sigma_A^2} \|\mathcal{A} \mathbf{x}_k - \mathcal{A}_\infty \mathbf{x}_k\|_{\pi_r}^2 + \frac{2\alpha^2 \bar{\pi}_r}{1 - \sigma_A^2} \|\mathbf{y}_k\|_2^2, \tag{16}$$

$$\|\mathbf{s}_{k+1} - \mathbf{1}_n \hat{\mathbf{s}}_{k+1}\|_{\pi_r}^2 \leq 2\|\mathcal{A} \mathbf{x}_k - \mathcal{A}_\infty \mathbf{x}_k\|_{\pi_r}^2 + 2\bar{\pi}_r \alpha^2 \|\mathbf{y}_k\|_2^2. \tag{17}$$

由 \mathbf{s}_k 的定义可知

$$\begin{aligned} \|\mathcal{A} \mathbf{x}_k - \mathcal{A}_\infty \mathbf{x}_k\|_{\pi_r}^2 &\leq \sigma_A^2 \|\mathbf{s}_k - \mathbf{1}_n \hat{\mathbf{s}}_k\|_{\pi_r}^2 + \beta^2 \sigma_A^2 \bar{\pi}_r \|\mathbf{s}_k - \mathbf{s}_{k-1}\|_2^2 \\ &\quad + 2\beta \sigma_A^2 \|\mathbf{s}_k - \mathbf{1}_n \hat{\mathbf{s}}_k\|_{\pi_r} \|\mathbf{s}_k - \mathbf{s}_{k-1}\|_{\pi_r} \\ &\leq \sigma_A^2 (1 + \beta) \|\mathbf{s}_k - \mathbf{1}_n \hat{\mathbf{s}}_k\|_{\pi_r}^2 + 2\sigma_A^2 \bar{\pi}_r \beta \|\mathbf{s}_k - \mathbf{s}_{k-1}\|_2^2. \end{aligned} \tag{18}$$

通过将 (10) 和 (18) 带入 (16) 中并在不等式两边同时取全期望可得

$$\begin{aligned} \mathbb{E}[\|\mathbf{s}_{k+1} - \mathbf{1}_n \hat{\mathbf{s}}_{k+1}\|_{\pi_r}^2] &\leq \left(\frac{1 + \sigma_A^2}{2} + \beta + \frac{90nh\|\pi_c\|_2^2 L^2}{1 - \sigma_A^2} \alpha^2\right) \mathbb{E}[\|\mathbf{s}_k - \mathbf{1}_n \hat{\mathbf{s}}_k\|_{\pi_r}^2] \\ &\quad + \frac{50n^2\|\pi_c\|_2^2 L^2 \bar{\pi}_r}{1 - \sigma_A^2} \alpha^2 \mathbb{E}[\|\hat{\mathbf{s}}_k - \mathbf{x}^*\|_2^2] \\ &\quad + ((1 + \sigma_A^2)\bar{\pi}_r + \frac{90n\|\pi_c\|_2^2 L^2 \bar{\pi}_r}{1 - \sigma_A^2} \alpha^2) \beta \mathbb{E}[\|\mathbf{s}_k - \mathbf{s}_{k-1}\|_2^2] \\ &\quad + \frac{20\|\pi_c\|_2^2 \bar{\pi}_r}{1 - \sigma_A^2} \alpha^2 \mathbb{E}[\mathbf{v}_k] + \frac{10\bar{\pi}_r \bar{\pi}_c}{1 - \sigma_A^2} \alpha^2 \mathbb{E}[\|\mathbf{y}_k - \mathcal{B}_\infty \mathbf{y}_k\|_{\pi_c}^2]. \end{aligned} \tag{19}$$

此外, 若将 (10) 和 (17) 带入 (16) 中并在不等式两边同时取全期望可得

$$\begin{aligned} \mathbb{E}[\|\mathbf{s}_{k+1} - \mathbf{1}_n \hat{\mathbf{s}}_{k+1}\|_{\pi_r}^2] &\leq (2\sigma_A^2(1 + \beta) + 90nh\|\boldsymbol{\pi}_c\|_2^2 L^2 \alpha^2) \mathbb{E}[\|\mathbf{s}_k - \mathbf{1}_n \hat{\mathbf{s}}_k\|_{\pi_r}^2] \\ &\quad + 50n^2 \|\boldsymbol{\pi}_c\|_2^2 L^2 \bar{\boldsymbol{\pi}}_r \alpha^2 \mathbb{E}[\|\hat{\mathbf{s}}_k - \mathbf{x}^*\|_2^2] \\ &\quad + (4\sigma_A^2 \bar{\boldsymbol{\pi}}_r + 90n\|\boldsymbol{\pi}_c\|_2^2 L^2 \bar{\boldsymbol{\pi}}_r \alpha^2) \beta \mathbb{E}[\|\mathbf{s}_k - \mathbf{s}_{k-1}\|_2^2] \\ &\quad + 20\|\boldsymbol{\pi}_c\|_2^2 \bar{\boldsymbol{\pi}}_r \alpha^2 \mathbb{E}[\|\mathbf{v}_k\|_2^2] + 10\bar{\boldsymbol{\pi}}_r \bar{\boldsymbol{\pi}}_c \alpha^2 \mathbb{E}[\|\mathbf{y}_k - \mathcal{B}_\infty \mathbf{y}_k\|_{\pi_c}^2]. \end{aligned} \tag{20}$$

如果 $0 < \alpha \leq \frac{1}{\|\boldsymbol{\pi}_c\|_2 L} \sqrt{\frac{h}{10n}}$, 有

$$\begin{aligned} \mathbb{E}[\|\mathbf{s}_{k+1} - \mathbf{1}_n \hat{\mathbf{s}}_{k+1}\|_{\pi_r}^2] &\leq 11\mathbb{E}[\|\mathbf{s}_k - \mathbf{1}_n \hat{\mathbf{s}}_k\|_{\pi_r}^2] + 5n\bar{\boldsymbol{\pi}}_r \mathbb{E}[\|\hat{\mathbf{s}}_k - \mathbf{x}^*\|_2^2] \\ &\quad + 13\bar{\boldsymbol{\pi}}_r \beta \mathbb{E}[\|\mathbf{s}_k - \mathbf{s}_{k-1}\|_2^2] + 20\|\boldsymbol{\pi}_c\|_2^2 \bar{\boldsymbol{\pi}}_r \alpha^2 \mathbb{E}[\|\mathbf{v}_k\|_2^2] \\ &\quad + 10\bar{\boldsymbol{\pi}}_r \bar{\boldsymbol{\pi}}_c \alpha^2 \mathbb{E}[\|\mathbf{y}_k - \mathcal{B}_\infty \mathbf{y}_k\|_{\pi_c}^2]. \end{aligned} \tag{21}$$

此公式可用于后续的证明. 证明完成.

引理8. 若假设 1-3 成立. 如果 $0 < \alpha \leq \min\{\frac{1}{2n\boldsymbol{\pi}_r^T \boldsymbol{\pi}_c L}, \frac{1}{5\mu n \boldsymbol{\pi}_r^T \boldsymbol{\pi}_c}, \frac{1}{4\boldsymbol{\pi}_r^T \boldsymbol{\pi}_c L^2}\}$ 并且 $0 \leq \beta < 1$, 对 $\forall k \geq 0$, 网络中最优间隙如下

$$\begin{aligned} \mathbb{E}[\|\hat{\mathbf{s}}_{k+1} - \mathbf{x}^*\|_2^2] &\leq \frac{5\boldsymbol{\pi}_r^T \boldsymbol{\pi}_c L^2}{\mu \bar{\boldsymbol{\pi}}_r} \alpha \mathbb{E}[\|\mathbf{s}_k - \mathbf{1}_n \hat{\mathbf{s}}_k\|_{\pi_r}^2] + (1 - \frac{\mu n \boldsymbol{\pi}_r^T \boldsymbol{\pi}_c \alpha}{4}) \mathbb{E}[\|\hat{\mathbf{s}}_k - \mathbf{x}^*\|_2^2] \\ &\quad + \frac{5\boldsymbol{\pi}_r^T \boldsymbol{\pi}_c L^2}{\mu} \alpha \beta^2 \mathbb{E}[\|\mathbf{s}_k - \mathbf{s}_{k-1}\|_2^2] + 2.5(\boldsymbol{\pi}_r^T \boldsymbol{\pi}_c)^2 \alpha^2 \mathbb{E}[\|\mathbf{v}_k\|_2^2] \\ &\quad + \frac{5\|\boldsymbol{\pi}_r\|_2^2 \bar{\boldsymbol{\pi}}_c}{\mu n \boldsymbol{\pi}_r^T \boldsymbol{\pi}_c} \alpha \mathbb{E}[\|\mathbf{y}_k - \mathcal{B}_\infty \mathbf{y}_k\|_{\pi_c}^2]. \end{aligned} \tag{22}$$

证明: 由 $\hat{\mathbf{s}}_k$ 的定义得

$$\begin{aligned} \|\hat{\mathbf{s}}_{k+1} - \mathbf{x}^*\|_2^2 &= \|\hat{\mathbf{s}}_k - \mathbf{x}^* - n\alpha \boldsymbol{\pi}_r^T \boldsymbol{\pi}_c \bar{\mathbf{y}}_k - \alpha \boldsymbol{\pi}_r^T (\mathbf{y}_k - \mathcal{B}_\infty \mathbf{g}_k)\|_2^2 \\ &\leq \|\hat{\mathbf{x}}_k - \mathbf{x}^* - \tilde{\alpha} \bar{\mathbf{g}}_k\|_2^2 - 2 \langle \hat{\mathbf{x}}_k - \mathbf{x}^* - \tilde{\alpha} \bar{\mathbf{y}}_k, \alpha \boldsymbol{\pi}_r^T (\mathbf{y}_k - \mathcal{B}_\infty \mathbf{y}_k) \rangle \\ &\quad + \|\boldsymbol{\pi}_r\|_2^2 \bar{\boldsymbol{\pi}}_c \alpha^2 \|\mathbf{y}_k - \mathcal{B}_\infty \mathbf{y}_k\|_{\pi_c}^2 \\ &= \mathbf{I}_1 + \mathbf{I}_2 + \|\boldsymbol{\pi}_r\|_2^2 \bar{\boldsymbol{\pi}}_c \alpha^2 \|\mathbf{y}_k - \mathcal{B}_\infty \mathbf{y}_k\|_{\pi_c}^2. \end{aligned} \tag{23}$$

其中 $\tilde{\alpha} = n\alpha \boldsymbol{\pi}_r^T \boldsymbol{\pi}_c$. 在 (23) 的两边同时取条件期望有

$$\mathbb{E}[\|\hat{\mathbf{s}}_{k+1} - \mathbf{x}^*\|_2^2] \leq \mathbb{E}[\mathbf{I}_1 | \mathcal{F}_k] + \mathbb{E}[\mathbf{I}_2 | \mathcal{F}_k] + \|\boldsymbol{\pi}_r\|_2^2 \bar{\boldsymbol{\pi}}_c \alpha^2 \mathbb{E}[\|\mathbf{y}_k - \mathcal{B}_\infty \mathbf{y}_k\|_{\pi_c}^2 | \mathcal{F}_k]. \tag{24}$$

首先, 考虑 (25) 中的第一项, 则

$$\begin{aligned}
 \mathbb{E}[\mathbf{I}_1|\mathcal{F}_k] &= \mathbb{E}[\|\hat{\mathbf{s}}_k - \mathbf{x}^* - \tilde{\alpha}\bar{\mathbf{g}}_k\|_2^2|\mathcal{F}_k] \\
 &= \mathbb{E}[\|\hat{\mathbf{s}}_k - \tilde{\alpha}\nabla F(\hat{\mathbf{s}}_k) - \mathbf{x}^* + \tilde{\alpha}(\nabla F(\hat{\mathbf{s}}_k) - \bar{\mathbf{g}}_k)\|_2^2|\mathcal{F}_k] \\
 &= \mathbb{E}[\|\hat{\mathbf{s}}_k - \tilde{\alpha}\nabla F(\hat{\mathbf{s}}_k) - \mathbf{x}^*\|_2^2|\mathcal{F}_k] + \tilde{\alpha}^2\mathbb{E}[\|\nabla F(\hat{\mathbf{s}}_k) - \bar{\mathbf{g}}_k\|_2^2|\mathcal{F}_k] \\
 &\quad + 2\mathbb{E}[\langle \hat{\mathbf{s}}_k - \tilde{\alpha}\nabla F(\hat{\mathbf{s}}_k) - \mathbf{x}^*, \tilde{\alpha}(\nabla F(\hat{\mathbf{s}}_k) - \bar{\mathbf{g}}_k) \rangle|\mathcal{F}_k] \\
 &\leq (1 - \mu\tilde{\alpha})\mathbb{E}[\|\hat{\mathbf{s}}_k - \mathbf{x}^*\|_2^2] + \tilde{\alpha}^2\mathbb{E}[\|\nabla F(\hat{\mathbf{s}}_k) - \bar{\mathbf{g}}_k\|_2^2|\mathcal{F}_k] \\
 &\quad + \lambda_1\tilde{\alpha}(1 - \mu\tilde{\alpha})\mathbb{E}[\|\hat{\mathbf{s}}_k - \mathbf{x}^*\|_2^2] + \frac{\tilde{\alpha}}{\lambda_1}\mathbb{E}[\|\nabla F(\hat{\mathbf{s}}_k) - \nabla\bar{\mathbf{f}}(\mathbf{x}_k)\|_2^2] \\
 &\leq (1 - \mu\tilde{\alpha})\mathbb{E}[\|\hat{\mathbf{s}}_k - \mathbf{x}^*\|_2^2] + \tilde{\alpha}^2\mathbb{E}[\|\nabla F(\hat{\mathbf{s}}_k) - \bar{\mathbf{g}}_k\|_2^2] \\
 &\quad + \frac{2\tilde{\alpha}L^2}{\mu n\pi_r}\mathbb{E}[\|\mathbf{s}_k - \mathbf{1}_n\hat{\mathbf{s}}_k\|_{\pi_r}^2] + \frac{2\tilde{\alpha}\beta^2L^2}{\mu n}\mathbb{E}[\|\mathbf{s}_k - \mathbf{s}_{k-1}\|_2^2].
 \end{aligned} \tag{25}$$

其中在最后一个不等式中令 $\lambda_1 = \mu$ 并利用引理 3. 接下来, 推导 (25) 中第二项的上界

$$\begin{aligned}
 \mathbb{E}[\|\nabla F(\hat{\mathbf{s}}_k) - \bar{\mathbf{g}}_k\|_2^2|\mathcal{F}_k] &= \mathbb{E}[\|\nabla F(\hat{\mathbf{s}}_k) - \nabla\bar{\mathbf{f}}(\mathbf{x}_k) + \nabla\bar{\mathbf{f}}(\mathbf{x}_k) - \bar{\mathbf{g}}_k\|_2^2|\mathcal{F}_k] \\
 &= \|\nabla F(\hat{\mathbf{s}}_k) - \nabla\bar{\mathbf{f}}(\mathbf{x}_k)\|_2^2 + \frac{1}{n^2}\mathbb{E}[\|\nabla\mathbf{f}(\mathbf{x}_k) - \mathbf{g}_k\|_2^2|\mathcal{F}_k].
 \end{aligned} \tag{26}$$

利用公式 (13) 和引理 3 可得

$$\begin{aligned}
 \mathbb{E}[\|\nabla F(\hat{\mathbf{s}}_k) - \bar{\mathbf{g}}_k\|_2^2|\mathcal{F}_k] &\leq \frac{10L^2}{n\pi_r}\mathbb{E}[\|\mathbf{s}_k - \mathbf{1}_n\hat{\mathbf{s}}_k\|_{\pi_r}^2] + \frac{4L^2}{n}\mathbb{E}[\|\hat{\mathbf{s}}_k - \mathbf{x}^*\|_2^2] \\
 &\quad + \frac{10L^2}{n}\beta^2\mathbb{E}[\|\mathbf{s}_k - \mathbf{s}_{k-1}\|_2^2] + \frac{2}{n^2}\mathbb{E}[\|\mathbf{v}_k\|_2^2].
 \end{aligned} \tag{27}$$

若 $0 < \tilde{\alpha} \leq \frac{1}{5\mu}$, 将 (27) 和 (25) 相结合有

$$\begin{aligned}
 \mathbb{E}[\mathbf{I}_1|\mathcal{F}_k] &\leq \frac{4L^2}{\mu n\pi_r}\tilde{\alpha}\mathbb{E}[\|\mathbf{s}_k - \mathbf{1}_n\hat{\mathbf{s}}_k\|_{\pi_r}^2] + (1 - \mu\tilde{\alpha} + \frac{4\tilde{\alpha}^2L^2}{n})\mathbb{E}[\|\hat{\mathbf{s}}_k - \mathbf{x}^*\|_2^2] \\
 &\quad + \frac{4L^2}{\mu n}\tilde{\alpha}\beta^2\mathbb{E}[\|\mathbf{s}_k - \mathbf{s}_{k-1}\|_2^2] + \frac{2}{n^2}\tilde{\alpha}^2\mathbb{E}[\|\mathbf{v}_k\|_2^2].
 \end{aligned} \tag{28}$$

现在考虑 \mathbf{I}_2

$$\begin{aligned}
 \mathbf{I}_2 &= -2 \langle \hat{\mathbf{s}}_k - \mathbf{x}^* - \tilde{\alpha}\bar{\mathbf{y}}_k, \alpha\pi_r^T(\mathbf{y}_k - \mathcal{B}_\infty\mathbf{y}_k) \rangle \\
 &\leq 2\alpha\|\hat{\mathbf{s}}_k - \mathbf{x}^* - \tilde{\alpha}\bar{\mathbf{y}}_k\|_2\|\pi_r^T(\mathbf{y}_k - \mathcal{B}_\infty\mathbf{y}_k)\|_2 \\
 &\leq \alpha(\lambda_2\mathbf{I}_1 + \frac{\|\pi_r\|_2^2\bar{\pi}_c}{\lambda_2}\|\mathbf{y}_k - \mathcal{B}_\infty\mathbf{y}_k\|_{\pi_c}^2)
 \end{aligned} \tag{29}$$

在 (29) 中令 $\lambda_2 = \frac{\mu n\pi_r^T\pi_c}{4}$ 并在 (29) 两边同时取条件期望可得

$$\mathbb{E}[\mathbf{I}_2|\mathcal{F}_k] \leq \frac{\mu\tilde{\alpha}}{4}\mathbb{E}[\mathbf{I}_1|\mathcal{F}_k] + \frac{4\|\pi_r\|_2^2\bar{\pi}_c}{\mu n\pi_r^T\pi_c}\alpha\mathbb{E}[\|\mathbf{y}_k - \mathcal{B}_\infty\mathbf{y}_k\|_{\pi_c}^2|\mathcal{F}_k]. \tag{30}$$

如果 $0 < \tilde{\alpha} \leq \frac{n}{4L^2}$, 将 (30), (29), (28) 带入 (24). 注意到 $1 + \frac{\mu\tilde{\alpha}}{4} \leq \frac{5}{4}$ 和 $(1 + \frac{\mu\tilde{\alpha}}{4})(1 - \frac{\mu\tilde{\alpha}}{2}) \leq 1$, 用 $n\alpha\boldsymbol{\pi}_r^T\boldsymbol{\pi}_c$ 替换 $\tilde{\alpha}$ 可以得到

$$\begin{aligned} \mathbb{E}[\|\mathbf{s}_{k+1} - \mathbf{x}^*\|_2^2] &\leq \frac{5\boldsymbol{\pi}_r^T\boldsymbol{\pi}_cL^2}{\mu\boldsymbol{\pi}_r}\alpha\mathbb{E}[\|\mathbf{s}_k - \mathbf{1}_n\hat{\mathbf{s}}_k\|_{\boldsymbol{\pi}_r}^2] + \mathbb{E}[\|\hat{\mathbf{s}}_k - \mathbf{x}^*\|_2^2] \\ &\quad + \frac{5\boldsymbol{\pi}_r^T\boldsymbol{\pi}_cL^2}{\mu}\alpha\beta^2\mathbb{E}[\|\mathbf{s}_k - \mathbf{s}_{k-1}\|_2^2] + 2.5(\boldsymbol{\pi}_r^T\boldsymbol{\pi}_c)^2\alpha^2\mathbb{E}[\mathbf{v}_k] \\ &\quad + \frac{5\|\boldsymbol{\pi}_r\|_2^2\bar{\boldsymbol{\pi}}_c}{\mu n\boldsymbol{\pi}_r^T\boldsymbol{\pi}_c}\alpha\mathbb{E}[\|\mathbf{y}_k - \mathcal{B}_\infty\mathbf{y}_k\|_{\boldsymbol{\pi}_c}^2]. \end{aligned} \tag{31}$$

如果在 (25) 中令 $\lambda_1 = \frac{1}{\tilde{\alpha}}$, 则 $\mathbb{E}[\mathbf{I}_1|\mathcal{F}_k]$ 表达如下

$$\begin{aligned} \mathbb{E}[\mathbf{I}_1|\mathcal{F}_k] &= \mathbb{E}[\|\hat{\mathbf{s}}_k - \mathbf{x}^* - \tilde{\alpha}\bar{\mathbf{y}}_k\|_2^2|\mathcal{F}_k] \\ &\leq 2\mathbb{E}[\|\hat{\mathbf{s}}_k - \mathbf{x}^*\|_2^2] + 2\tilde{\alpha}^2\mathbb{E}[\|\nabla F(\hat{\mathbf{s}}_k) - \bar{\mathbf{g}}_k\|_2^2|\mathcal{F}_k] \\ &\leq \frac{20L^2}{n\boldsymbol{\pi}_r}\tilde{\alpha}^2\mathbb{E}[\|\mathbf{s}_k - \mathbf{1}_n\hat{\mathbf{s}}_k\|_{\boldsymbol{\pi}_r}^2] + (2 + \frac{8L^2}{n}\tilde{\alpha}^2)\mathbb{E}[\|\hat{\mathbf{s}}_k - \mathbf{x}^*\|_2^2] \\ &\quad + \frac{20L^2}{n}\tilde{\alpha}^2\beta^2\mathbb{E}[\|\mathbf{s}_k - \mathbf{s}_{k-1}\|_2^2] + \frac{2}{n}\tilde{\alpha}^2\mathbb{E}[\mathbf{v}_k] \end{aligned} \tag{32}$$

如果 $0 < \tilde{\alpha} \leq \frac{\sqrt{n}}{2L}$, 可以得到

$$\begin{aligned} \mathbb{E}[\mathbf{I}_1|\mathcal{F}_k] &\leq \frac{20L^2}{n\boldsymbol{\pi}_r}\tilde{\alpha}^2\mathbb{E}[\|\mathbf{s}_k - \mathbf{1}_n\hat{\mathbf{s}}_k\|_{\boldsymbol{\pi}_r}^2] + 4\mathbb{E}[\|\hat{\mathbf{s}}_k - \mathbf{x}^*\|_2^2] \\ &\quad + \frac{20L^2}{n}\tilde{\alpha}^2\beta^2\mathbb{E}[\|\mathbf{s}_k - \mathbf{s}_{k-1}\|_2^2] + \frac{2}{n}\tilde{\alpha}^2\mathbb{E}[\mathbf{v}_k] \end{aligned} \tag{33}$$

在 (29) 中令 $\lambda_2 = \frac{1}{\alpha}$ 有

$$\mathbb{E}[\mathbf{I}_2] \leq \mathbb{E}[\mathbf{I}_1] + \|\boldsymbol{\pi}_r\|_2^2\bar{\boldsymbol{\pi}}_c\alpha^2\mathbb{E}[\|\mathbf{y}_k - \mathcal{B}_\infty\mathbf{y}_k\|_{\boldsymbol{\pi}_c}^2] \tag{34}$$

将 (33) 和 (34) 带入 (24) 中得

$$\begin{aligned} \mathbb{E}[\|\mathbf{s}_{k+1} - \mathbf{x}^*\|_2^2] &\leq \frac{40L^2}{n\boldsymbol{\pi}_r}\tilde{\alpha}^2\mathbb{E}[\|\mathbf{s}_k - \mathbf{1}_n\hat{\mathbf{s}}_k\|_{\boldsymbol{\pi}_r}^2] + 8\mathbb{E}[\|\hat{\mathbf{s}}_k - \mathbf{x}^*\|_2^2] \\ &\quad + \frac{40L^2}{n}\tilde{\alpha}^2\beta^2\mathbb{E}[\|\mathbf{s}_k - \mathbf{s}_{k-1}\|_2^2] + \frac{4}{n^2}\tilde{\alpha}^2\mathbb{E}[\mathbf{v}_k] \\ &\quad + 2\|\boldsymbol{\pi}_r\|_2^2\bar{\boldsymbol{\pi}}_c\alpha^2\mathbb{E}[\|\mathbf{y}_k - \mathcal{B}_\infty\mathbf{y}_k\|_{\boldsymbol{\pi}_c}^2]. \end{aligned} \tag{35}$$

将 $\tilde{\alpha}$ 替换为 $n\alpha\boldsymbol{\pi}_r^T\boldsymbol{\pi}_c$, 证明完成.

引理9. 若假设 2 成立. 如果 $0 \leq \beta < 1$, 对 $\forall k \geq 0$, 以下不等式成立

$$\begin{aligned} &\mathbb{E}[\|\mathbf{s}_{k+1} - \mathbf{s}_k\|_2^2] \\ &\leq (\frac{16}{\boldsymbol{\pi}_r} + \frac{90n\|\boldsymbol{\pi}_c\|_2^2L^2}{\boldsymbol{\pi}_r}\alpha^2)\mathbb{E}[\|\mathbf{s}_k - \mathbf{1}_n\hat{\mathbf{s}}_k\|_{\boldsymbol{\pi}_r}^2] + 50n^2\|\boldsymbol{\pi}_c\|_2^2L^2\alpha^2\mathbb{E}[\|\hat{\mathbf{s}}_k - \mathbf{x}^*\|_2^2] \\ &\quad + 2(2h + 45n\|\boldsymbol{\pi}_c\|_2^2L^2\alpha^2)\beta^2\mathbb{E}[\|\mathbf{s}_k - \mathbf{s}_{k-1}\|_2^2] + 20\|\boldsymbol{\pi}_c\|_2^2\alpha^2\mathbb{E}[\mathbf{v}_k] \\ &\quad + 10\bar{\boldsymbol{\pi}}_c\alpha^2\mathbb{E}[\|\mathbf{y}_k - \mathcal{B}_\infty\mathbf{y}_k\|_{\boldsymbol{\pi}_c}^2]. \end{aligned} \tag{36}$$

证明: 由 \mathbf{s}_k 的定义可知

$$\begin{aligned}
 & \|\mathbf{s}_{k+1} - \mathbf{s}_k\|_2^2 \\
 &= \|\mathcal{A}\mathbf{x}_k - \alpha\mathbf{y}_k - \mathbf{s}_k\|_2^2 \\
 &\leq \|(\mathcal{A} - I_n)(\mathbf{s}_k - \mathcal{A}_\infty\mathbf{x}_k) - \alpha\mathbf{y}_k + \beta\mathcal{A}(\mathbf{s}_k - \mathbf{s}_{k-1})\|_2^2 \\
 &\leq \frac{4}{\underline{\pi}_r} \|\mathcal{A} - I_n\|_{\underline{\pi}_r}^2 \|\mathbf{s}_k - \mathbf{1}_n\hat{\mathbf{s}}_k\|_{\underline{\pi}_r}^2 + 4\beta^2h \|\mathcal{A}\|_{\underline{\pi}_r}^2 \|\mathbf{s}_k - \mathbf{s}_{k-1}\|_2^2 + 2\alpha^2\|\mathbf{y}_k\|_2^2 \\
 &\leq \left(\frac{16}{\underline{\pi}_r} + \frac{90n\|\underline{\pi}_c\|_2^2L^2}{\underline{\pi}_r}\alpha^2\right)\|\mathbf{s}_k - \mathbf{1}_n\hat{\mathbf{s}}_k\|_{\underline{\pi}_r}^2 + 50n^2\|\underline{\pi}_c\|_2^2L^2\alpha^2\|\hat{\mathbf{s}}_k - \mathbf{x}^*\|_2^2 \\
 &\quad + 2(2h + 45n\|\underline{\pi}_c\|_2^2L^2\alpha^2)\beta^2\|\mathbf{s}_k - \mathbf{s}_{k-1}\|_2^2 + 20\|\underline{\pi}_c\|_2^2\alpha^2\mathbf{v}_k \\
 &\quad + 10\underline{\pi}_c\alpha^2\|\mathbf{y}_k - \mathcal{B}_\infty\mathbf{y}_k\|_{\underline{\pi}_c}^2.
 \end{aligned} \tag{37}$$

其中在第二个不等式中利用了 $\|\mathcal{A} - I_n\|_{\underline{\pi}_r} \leq 2$, 在上述不等式的两边同时取全期望完成证明.

引理10. 设假设 2 成立, 对 $\forall k \geq 0$, 以下不等式成立

$$\begin{aligned}
 \mathbb{E}[\mathbf{v}_{k+1}] &\leq \frac{4L^2}{m\underline{\pi}_r} \mathbb{E}[\|\mathbf{s}_k - \mathbf{1}_n\hat{\mathbf{s}}_k\|_{\underline{\pi}_r}^2] + \frac{2nL^2}{m} \mathbb{E}[\|\hat{\mathbf{s}}_k - \mathbf{x}^*\|_2^2] \\
 &\quad + \frac{4L^2}{m} \beta^2 \mathbb{E}[\|\mathbf{s}_k - \mathbf{s}_{k-1}\|_2^2] + \left(1 - \frac{1}{m}\right) \mathbb{E}[\mathbf{v}_k].
 \end{aligned} \tag{38}$$

证明: 由 \mathbf{v}_k 的定义可知

$$\begin{aligned}
 \mathbb{E}[\mathbf{v}_{k+1}^i | \mathcal{F}_k] &= \frac{1}{m_i} \sum_{j=1}^{m_i} \mathbb{E}[\|f_{i,j}(\tau_{k+1}^{i,j}) - f_{i,j}(\mathbf{x}^*)\|_2^2 | \mathcal{F}_k] \\
 &= \frac{1}{m_i} \sum_{j=1}^{m_i} \left(\frac{1}{m} \|f_{i,j}(\mathbf{x}_k^i) - f_{i,j}(\mathbf{x}^*)\|_2^2 + \left(1 - \frac{1}{m}\right) \|f_{i,j}(\tau_k^{i,j}) - f_{i,j}(\mathbf{x}^*)\|_2^2\right) \\
 &\leq \frac{4L^2}{m\underline{\pi}_r} \|\mathbf{s}_k^i - \hat{\mathbf{s}}_k\|_{\underline{\pi}_r}^2 + \frac{2L^2}{m} \|\hat{\mathbf{s}}_k - \mathbf{x}^*\|_2^2 + \frac{4L^2}{m} \beta^2 \|\mathbf{s}_k^i - \mathbf{s}_{k-1}^i\|_2^2 \\
 &\quad + \left(1 - \frac{1}{m}\right) \mathbf{v}_k^i.
 \end{aligned} \tag{39}$$

在上述不等式的两边同时取全期望得

$$\begin{aligned}
 \mathbb{E}[\mathbf{v}_{k+1}] &\leq \frac{4L^2}{m\underline{\pi}_r} \mathbb{E}[\|\mathbf{s}_k - \mathbf{1}_n\hat{\mathbf{s}}_k\|_{\underline{\pi}_r}^2] + \frac{2nL^2}{m} \mathbb{E}[\|\hat{\mathbf{s}}_k - \mathbf{x}^*\|_2^2] \\
 &\quad + \frac{4L^2}{m} \beta^2 \mathbb{E}[\|\mathbf{s}_k - \mathbf{s}_{k-1}\|_2^2] + \left(1 - \frac{1}{m}\right) \mathbb{E}[\mathbf{v}_k].
 \end{aligned} \tag{40}$$

证明完成.

引理11. 若假设 1-3 成立. 如果 $0 < \alpha \leq \min\left\{\frac{1}{4L\sqrt{10n\|\pi_c\|_2^2(h+\beta^2)+2n^2(\pi_r^T \pi_c)^2}}, \frac{(1-\sigma_B^2)^2}{16L\sqrt{(15(h+\beta^2)+n\|\pi_r\|_2^2)g}}\right\}$ 并且 $0 \leq \beta < 1$, 对 $\forall k \geq 0$, 以下不等式成立

$$\begin{aligned} & \mathbb{E}[\|\mathbf{y}_{k+1} - \mathcal{B}_\infty \mathbf{y}_{k+1}\|_{\pi_c}^2] \\ & \leq \frac{1946L^2}{\underline{\pi}_r \underline{\pi}_c (1 - \sigma_B^2)} \mathbb{E}[\|\mathbf{s}_k - \mathbf{1}_n \hat{\mathbf{s}}_k\|_{\pi_r}^2] + \frac{338nL^2}{\underline{\pi}_c (1 - \sigma_B^2)} \mathbb{E}[\|\hat{\mathbf{s}}_k - \mathbf{x}^*\|_2^2] \\ & \quad + \frac{802hL^2}{\underline{\pi}_c (1 - \sigma_B^2)} \beta \mathbb{E}[\|\mathbf{s}_k - \mathbf{s}_{k-1}\|_2^2] + \frac{44}{\underline{\pi}_c (1 - \sigma_B^2)} \mathbb{E}[\mathbf{v}_k] \\ & \quad + \frac{3 + \sigma^2}{4} \mathbb{E}[\|\mathbf{y}_k - \mathcal{B}_\infty \mathbf{y}_k\|_{\pi_c}^2]. \end{aligned} \tag{41}$$

其中 $g = \bar{\pi}_c / \underline{\pi}_c$.

证明: 由 \mathbf{y}_k 的定义可知

$$\begin{aligned} \mathbb{E}[\|\mathbf{y}_{k+1} - \mathcal{B}_\infty \mathbf{y}_{k+1}\|_{\pi_c}^2 | \mathcal{F}_k] &= \mathbb{E}[\|\mathcal{B} \mathbf{y}_k - \mathcal{B}_\infty + (I_n - \mathcal{B}_\infty) \mathbf{g}_{k+1} - \mathbf{g}_k\|_{\pi_c}^2 | \mathcal{F}_k] \\ &\leq \frac{1 + \sigma_B^2}{2} \mathbb{E}[\|\mathbf{y}_k - \mathcal{B}_\infty \mathbf{y}_k\|_{\pi_c}^2 | \mathcal{F}_k] \\ &\quad + \frac{2}{1 - \sigma_B^2} \mathbb{E}[\|\mathbf{g}_{k+1} - \mathbf{g}_k\|_{\pi_c}^2 | \mathcal{F}_k]. \end{aligned} \tag{42}$$

其中在不等式中利用 Young 不等式. 接下来, 首先考虑 (42) 中的第二项

$$\begin{aligned} & \mathbb{E}[\|\mathbf{g}_{k+1} - \mathbf{g}_k\|_{\pi_c}^2 | \mathcal{F}_k] \\ &= \mathbb{E}[\|\mathbf{g}_{k+1} - \mathbf{g}_k - \nabla \mathbf{f}(\mathbf{x}_{k+1}) + \nabla \mathbf{f}(\mathbf{x}_k) + \nabla \mathbf{f}(\mathbf{x}_{k+1}) - \nabla \mathbf{f}(\mathbf{x}_k)\|_{\pi_c}^2 | \mathcal{F}_k] \\ &\leq 2\mathbb{E}[\|\nabla \mathbf{f}(\mathbf{x}_{k+1}) - \nabla \mathbf{f}(\mathbf{x}_k)\|_{\pi_c}^2 | \mathcal{F}_k] \\ &\quad + 2\mathbb{E}[\|\mathbf{g}_{k+1} - \mathbf{g}_k - \nabla \mathbf{f}(\mathbf{x}_{k+1}) + \nabla \mathbf{f}(\mathbf{x}_k)\|_{\pi_c}^2 | \mathcal{F}_k] \\ &\leq \frac{2L^2}{\underline{\pi}_c} \mathbb{E}[\|\mathbf{x}_{k+1} - \mathbf{x}_k\|_2^2 | \mathcal{F}_k] + \frac{4}{\underline{\pi}_c} \mathbb{E}[\|\mathbf{g}_k - \nabla \mathbf{f}(\mathbf{x}_k)\|_2^2 | \mathcal{F}_k] \\ &\quad + \frac{4}{\underline{\pi}_c} \mathbb{E}[\|\mathbf{g}_{k+1} - \nabla \mathbf{f}(\mathbf{x}_{k+1})\|_2^2 | \mathcal{F}_k] \\ &\leq \frac{4L^2}{\underline{\pi}_c} \beta^2 \mathbb{E}[\|\mathbf{s}_k - \mathbf{s}_{k-1}\|_2^2 | \mathcal{F}_k] + \frac{16L^2}{\underline{\pi}_c} \mathbb{E}[\|\mathbf{s}_{k+1} - \mathbf{s}_k\|_2^2 | \mathcal{F}_k] \\ &\quad + \frac{4}{\underline{\pi}_c} \mathbb{E}[\|\mathbf{g}_k - \nabla \mathbf{f}(\mathbf{x}_k)\|_2^2 | \mathcal{F}_k] + \frac{4}{\underline{\pi}_c} \mathbb{E}[\|\mathbf{g}_{k+1} - \nabla \mathbf{f}(\mathbf{x}_{k+1})\|_2^2 | \mathcal{F}_k] \end{aligned} \tag{43}$$

其中在最后一个不等式中利用以下不等式

$$\mathbb{E}[\|\mathbf{x}_{k+1} - \mathbf{x}_k\|_2^2] \leq 2(1 + \beta)^2 \mathbb{E}[\|\mathbf{s}_{k+1} - \mathbf{s}_k\|_2^2] + 2\beta^2 \mathbb{E}[\|\mathbf{s}_k - \mathbf{s}_{k-1}\|_2^2]. \quad (44)$$

下面进一步考虑 (43) 中最后一项, 根据 $\mathbb{E}[\|\mathbf{g}_k - \nabla f(\mathbf{x}_k)\|_2^2 | \mathcal{F}_k]$ 的表达式可得

$$\begin{aligned} & \mathbb{E}[\|\mathbf{g}_{k+1} - \nabla f(\mathbf{x}_{k+1})\|_2^2 | \mathcal{F}_k] \\ & \leq \frac{8L^2}{\underline{\pi}_r} \mathbb{E}[\|\mathbf{s}_{k+1} - \mathbf{1}_n \hat{\mathbf{s}}_{k+1}\|_{\underline{\pi}_r}^2] + 4nL^2 \|\hat{\mathbf{s}}_{k+1} - \mathbf{x}^*\|_2^2 \\ & \quad + 8\beta^2 L^2 \mathbb{E}[\|\mathbf{s}_{k+1} - \mathbf{s}_k\|_2^2] + 2\mathbb{E}[\|\mathbf{v}_{k+1}\|_2^2] \\ & \leq \left(\frac{168L^2}{\underline{\pi}_r} + \frac{720n\|\underline{\pi}_c\|_2^2 L^4}{\underline{\pi}_r} (h + \beta^2)\alpha^2 + \frac{160n^2 L^4 (\underline{\pi}_r^T \underline{\pi}_c)^2}{\underline{\pi}_r} \alpha^2 \right) \mathbb{E}[\|\mathbf{s}_k - \mathbf{1}_n \hat{\mathbf{s}}_k\|_{\underline{\pi}_r}^2] \\ & \quad + (36nL^2 + 400n^2 \|\underline{\pi}_c\|_2^2 L^4 \alpha^2 (h + \beta^2)) \mathbb{E}[\|\hat{\mathbf{s}}_k - \mathbf{x}^*\|_2^2] \\ & \quad + (72hL^2 + 720n\|\underline{\pi}_c\|_2^2 L^4 (h + \beta^2)\alpha^2 + 160n^2 L^4 (\underline{\pi}_r^T \underline{\pi}_c)^2 \alpha^2) \beta \mathbb{E}[\|\mathbf{s}_k - \mathbf{s}_{k-1}\|_{\underline{\pi}_r}^2] \\ & \quad + (160\|\underline{\pi}_c\|_2^2 L^2 (h + \beta^2)\alpha^2 + 16n(\underline{\pi}_r^T \underline{\pi}_c)^2 L^2 \alpha^2 + 2 - \frac{2}{m}) \mathbb{E}[\|\mathbf{v}_k\|_2^2] \\ & \quad + (80(h + \beta^2) + 8n\|\underline{\pi}_r\|_2^2) \bar{\pi}_c L^2 \alpha^2 \mathbb{E}[\|\mathbf{y}_k - \mathcal{B}_\infty \mathbf{y}_k\|_{\underline{\pi}_c}^2 | \mathcal{F}_k]. \end{aligned} \quad (45)$$

若 $0 < \alpha \leq \min\left\{\frac{1}{4L\sqrt{45n\|\underline{\pi}_c\|_2^2(h+\beta^2)+10n^2(\underline{\pi}_r^T \underline{\pi}_c)^2}}, \frac{1}{20\|\underline{\pi}_c\|_2 L^2 \sqrt{n(h+\beta^2)}}\right\}$, 则

$$\begin{aligned} & \mathbb{E}[\|\mathbf{g}_{k+1} - \nabla f(\mathbf{x}_{k+1})\|_2^2 | \mathcal{F}_k] \\ & \leq \frac{169L^2}{\underline{\pi}_r} \mathbb{E}[\|\mathbf{s}_k - \mathbf{1}_n \hat{\mathbf{s}}_k\|_{\underline{\pi}_r}^2] + 37nL^2 \|\hat{\mathbf{s}}_k - \mathbf{x}^*\|_2^2 + 73hL^2 \beta \mathbb{E}[\|\mathbf{s}_k - \mathbf{s}_{k-1}\|_2^2] \\ & \quad + 3\mathbb{E}[\|\mathbf{v}_k\|_2^2] + (80(h + \beta^2) + 8n\|\underline{\pi}_r\|_2^2) \bar{\pi}_c L^2 \alpha^2 \mathbb{E}[\|\mathbf{y}_k - \mathcal{B}_\infty \mathbf{y}_k\|_{\underline{\pi}_c}^2 | \mathcal{F}_k]. \end{aligned} \quad (46)$$

将 (36) 和 (46) 带入 (43) 有

$$\begin{aligned} \mathbb{E}[\|\mathbf{g}_{k+1} - \mathbf{g}_k\|_{\underline{\pi}_c}^2 | \mathcal{F}_k] & \leq \left(\frac{964L^2}{\underline{\pi}_r \underline{\pi}_c} + \frac{1440n\|\underline{\pi}_c\|_2^2 L^4}{\underline{\pi}_r \underline{\pi}_c} \alpha^2 \right) \mathbb{E}[\|\mathbf{s}_k - \mathbf{1}_n \hat{\mathbf{s}}_k\|_{\underline{\pi}_r}^2] \\ & \quad + \left(\frac{164nL^2}{\underline{\pi}_c} + \frac{800n^2 \|\underline{\pi}_c\|_2^2 L^4}{\underline{\pi}_c} \alpha^2 \right) \mathbb{E}[\|\hat{\mathbf{s}}_k - \mathbf{x}^*\|_2^2] \\ & \quad + \left(\frac{392hL^2}{\underline{\pi}_c} + \frac{1440n\|\underline{\pi}_c\|_2^2 L^4}{\underline{\pi}_c} \alpha^2 \right) \beta \mathbb{E}[\|\mathbf{s}_k - \mathbf{s}_{k-1}\|_2^2] \\ & \quad + \left(\frac{20}{\underline{\pi}_c} + \frac{320\|\underline{\pi}_c\|_2^2 L^4}{\underline{\pi}_c} \alpha^2 \right) \mathbb{E}[\|\mathbf{v}_k\|_2^2] \\ & \quad + 32(15(h + \beta^2) + n\|\underline{\pi}_r\|_2^2) g L^2 \alpha^2 \mathbb{E}[\|\mathbf{y}_k - \mathcal{B}_\infty \mathbf{y}_k\|_{\underline{\pi}_c}^2 | \mathcal{F}_k]. \end{aligned} \quad (47)$$

结合 (47)和 (42)可以得到

$$\begin{aligned}
 \mathbb{E}[\|\mathbf{y}_{k+1} - \mathcal{B}_\infty \mathbf{y}_{k+1}\|_{\boldsymbol{\pi}_c}^2 | \mathcal{F}_k] &\leq \left(\frac{1928L^2}{\boldsymbol{\pi}_r \boldsymbol{\pi}_c (1 - \sigma_B^2)} + \frac{2880n \|\boldsymbol{\pi}_c\|_2^2 L^4}{\boldsymbol{\pi}_r \boldsymbol{\pi}_c (1 - \sigma_B^2)} \alpha^2 \right) \mathbb{E}[\|\mathbf{s}_k - \mathbf{1}_n \hat{\mathbf{s}}_k\|_{\boldsymbol{\pi}_r}^2] \\
 &+ \left(\frac{328nL^2}{\boldsymbol{\pi}_c (1 - \sigma_B^2)} + \frac{1600n^2 \|\boldsymbol{\pi}_c\|_2^2 L^4}{\boldsymbol{\pi}_c (1 - \sigma_B^2)} \alpha^2 \right) \mathbb{E}[\|\hat{\mathbf{s}}_k - \mathbf{x}^*\|_2^2] \\
 &+ \left(\frac{784hL^2}{\boldsymbol{\pi}_c (1 - \sigma_B^2)} + \frac{2880n \|\boldsymbol{\pi}_c\|_2^2 L^4}{\boldsymbol{\pi}_c (1 - \sigma_B^2)} \alpha^2 \right) \beta \mathbb{E}[\|\mathbf{s}_k - \mathbf{s}_{k-1}\|_2^2] \\
 &+ \left(\frac{40}{\boldsymbol{\pi}_c (1 - \sigma_B^2)} + \frac{640 \|\boldsymbol{\pi}_c\|_2^2 L^4}{\boldsymbol{\pi}_c (1 - \sigma_B^2)} \alpha^2 \right) \mathbb{E}[\mathbf{v}_k] \\
 &+ \left(\frac{1 + \sigma_B^2}{2} + \frac{64(15(h + \beta^2) + n \|\boldsymbol{\pi}_r\|_2^2)}{1 - \sigma_B^2} g L^2 \alpha^2 \right) \\
 &\times \mathbb{E}[\|\mathbf{y}_k - \mathcal{B}_\infty \mathbf{y}_k\|_{\boldsymbol{\pi}_c}^2 | \mathcal{F}_k].
 \end{aligned} \tag{48}$$

如果 $\alpha \leq \left\{ \frac{1}{4L \|\boldsymbol{\pi}_c\|_2 \sqrt{10n}}, \frac{(1 - \sigma_B^2)^2}{16L \sqrt{(15(h + \beta^2) + n \|\boldsymbol{\pi}_r\|_2^2)g}} \right\}$, 以下式子成立

$$\begin{aligned}
 \mathbb{E}[\|\mathbf{y}_{k+1} - \mathcal{B}_\infty \mathbf{y}_{k+1}\|_{\boldsymbol{\pi}_c}^2 | \mathcal{F}_k] &\leq \frac{1946L^2}{\boldsymbol{\pi}_r \boldsymbol{\pi}_c (1 - \sigma_B^2)} \mathbb{E}[\|\mathbf{s}_k - \mathbf{1}_n \hat{\mathbf{s}}_k\|_{\boldsymbol{\pi}_r}^2] + \frac{338nL^2}{\boldsymbol{\pi}_c (1 - \sigma_B^2)} \mathbb{E}[\|\hat{\mathbf{s}}_k - \mathbf{x}^*\|_2^2] \\
 &+ \frac{802hL^2}{\boldsymbol{\pi}_c (1 - \sigma_B^2)} \beta \mathbb{E}[\|\mathbf{s}_k - \mathbf{s}_{k-1}\|_2^2] + \frac{44}{\boldsymbol{\pi}_c (1 - \sigma_B^2)} \mathbb{E}[\mathbf{v}_k] \\
 &+ \frac{3 + \sigma_B^2}{4} \mathbb{E}[\|\mathbf{y}_k - \mathcal{B}_\infty \mathbf{y}_k\|_{\boldsymbol{\pi}_c}^2 | \mathcal{F}_k].
 \end{aligned} \tag{49}$$

在上述不等式的两边同时取全期望完成证明.

基于引理 7-11 的结果, 可以得到如下线性不等式.

性质1. 设假设 1-3 成立. 如果 $0 < \alpha \leq \min\left\{ \frac{1}{4L \sqrt{10n} \|\boldsymbol{\pi}_c\|_2^2 (h + \beta^2) + 2n^2 (\boldsymbol{\pi}_r^T \boldsymbol{\pi}_c)^2}, \frac{1}{5n\kappa L \boldsymbol{\pi}_r^T \boldsymbol{\pi}_c}, \frac{(1 - \sigma_B^2)^2}{16L \sqrt{(15(h + \beta^2) + n \|\boldsymbol{\pi}_r\|_2^2)g}} \right\}$ 并且 $0 \leq \beta < 1$, 则下列不等式成立

$$\Gamma_k \leq G_{\alpha, \beta} \Gamma_{k-1}, \tag{50}$$

其中 $\Gamma_k \in \mathbb{R}^5$ 和 $G_{\alpha, \beta} \in \mathbb{R}^{5 \times 5}$ 定义如下

$$\Gamma_k = \begin{pmatrix} \mathbb{E}[\|\mathbf{s}_k - \mathbf{1}_n \hat{\mathbf{s}}_k\|_{\boldsymbol{\pi}_r}^2] \\ \mathbb{E}[\|\hat{\mathbf{s}}_k - \mathbf{x}^*\|_2^2] \\ \mathbb{E}[n \|\mathbf{s}_k - \mathbf{s}_{k-1}\|_2^2] \\ \mathbb{E}[\mathbf{v}_k] \\ \mathbb{E}[\|\mathbf{y}_k - \mathcal{B}_\infty \mathbf{y}_k\|_{\boldsymbol{\pi}_c}^2] \end{pmatrix}$$

$$G_{\alpha,\beta} = \begin{pmatrix} \frac{1+\sigma_A^2}{2} + 9npp_1p_2p_5\alpha^2 + \beta & 5np_1p_2\alpha^2 & ((1 + \sigma_A^2)\bar{\pi}_r + 9np_1p_2p_5\alpha^2)\beta & 2p_1p_2\alpha^2 & 10\bar{\pi}_r p_2\alpha^2 \\ 5pp_3\alpha & 1 & 5p_3\alpha\beta^2 & 2.5n(\pi_r^T \pi_c)^2 \alpha^2 & p_4\alpha \\ 16p + 9npp_1p_5\alpha^2 & 5np_1p_5\alpha^2 & (4h + 9np_1p_5\alpha^2)\beta^2 & 2p_1\alpha^2 & 10\bar{\pi}_c\alpha^2 \\ \frac{4p_5}{m\bar{\pi}_r} & \frac{2p_5}{m} & \frac{4p_5}{m}\beta^2 & 1 - \frac{1}{m} & 0 \\ 973pp_5p_6 & 169p_5p_6 & 401hp_5p_6\beta & 22p_5 & \frac{3+\sigma^2}{4} \end{pmatrix}$$

在上述矩阵中 $p = \frac{1}{\pi_r}, p_1 = 10\|\pi_c\|_2^2, p_2 = \frac{\bar{\pi}_r}{1-\sigma_A^2}, p_3 = \frac{\pi_r^T \pi_c L^2}{\mu}, p_4 = \frac{5\|\pi_r\|_2^2 \bar{\pi}_c}{\mu \pi_r^T \pi_c}, p_5 = L^2, p_6 = \frac{2}{\pi_c(1-\sigma_B^2)}$.

接下来根据性质 1 结果, 推导使得 $\rho(G_{\alpha,\beta}) < 1$ 的步长和动量系数的取值范围.

引理12. 对 *ABN-LSVRG* 算法, 当步长 α 满足

$$0 < \alpha \leq \min\left\{ \sqrt{\frac{(1 - \sigma_A^2)\delta_1}{2(9npp_1p_2p_5\delta_1 + 5np_1p_2\delta_2 + 2p_1p_2\delta_4 + 10\bar{\pi}_r\delta_5)}}, \right. \\ \left. \frac{\delta_2}{2(5pp_3\delta_1 + p_4\delta_5)}, \sqrt{\frac{5\delta_2}{2n(\pi_r^T \pi_c)^2\delta_4}}, \right. \\ \left. \sqrt{\frac{\delta_3 - 16p\delta_1}{9npp_1p_5\delta_1 + 5np_1p_5\delta_2 + 2p_1\delta_4 + 10\bar{\pi}_c\delta_5}}, \right. \\ \left. \frac{1}{4L\sqrt{10n\|\pi_c\|_2^2(h + \beta^2) + 2n^2(\pi_r^T \pi_c)^2}}, \right. \\ \left. \frac{1}{5n\kappa L\pi_r^T \pi_c}, \frac{(1 - \sigma_B^2)^2}{16L\sqrt{(15(h + \beta^2) + n\|\pi_r\|_2^2)g}} \right\}. \tag{51}$$

动量参数 β 满足

$$0 \leq \beta < \min\left\{ \frac{(1 - \sigma_A^2)\delta_1 - 2(9npp_1p_2p_5\delta_1 + 5np_1p_2\delta_2 + 2p_1p_2\delta_4 + 10\bar{\pi}_r\delta_5)\alpha^2}{2(\delta_1 + (1 + \sigma_A^2)\bar{\pi}_r + 9np_1p_2\alpha^2)}, \right. \\ \left. \sqrt{\frac{\delta_2 - (5pp_3\delta_1 + p_4\delta_5)\alpha - 2.5n(\pi_r^T \pi_c)^2\delta_4\alpha^2}{5p_3\alpha\delta_3}}, \right. \\ \left. \sqrt{\frac{\delta_3 - 16p\delta_1 - (9npp_1p_5\delta_1 + 5np_1p_5\delta_2 + 2p_1\delta_4 + 10\bar{\pi}_c\delta_5)\alpha^2}{(4h + 9np_1p_2\alpha^2)\delta_3}}, \right. \\ \left. \sqrt{\frac{\delta_4 - 4pp_5\delta_1 - 2p_5\delta_2}{4p_5\delta_3}}, \right. \\ \left. \frac{(1 - \sigma_B^2)\delta_5 - 4(973pp_5p_6\delta_1 - 169p_5p_6\delta_2 - 22p_5\delta_4)}{401hp_5p_6\delta_3} \right\}, \tag{52}$$

则有 $\rho(G_{\alpha,\beta}) < 1$. 故 *ABN-LSVRG* 线性收敛到最优解 \mathbf{x}^* .

证明: 根据引理 5, 推导步长 α 和动量系数 β 的上界和正向量 $\varepsilon = [\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5]^T$ 使 $G_{\alpha, \beta} \varepsilon < \varepsilon$. 上式等价的写为

$$\begin{aligned}
 (\delta_1 + (1 + \sigma_A^2)\bar{\pi}_r + 9np_1p_2\alpha^2)\beta &< \frac{1 - \sigma_A^2}{2}\delta_1 - (9npp_1p_2p_5\delta_1 + 5np_1p_2\delta_2 \\
 &+ 2p_1p_2\delta_4 + 10\bar{\pi}_r\delta_5)\alpha^2, \\
 5p_3\alpha\delta_3\beta^2 &< \delta_2 - (5pp_3\delta_1 + p_4\delta_5)\alpha - 2.5n(\bar{\pi}_r^T \bar{\pi}_c)^2\delta_4\alpha^2, \\
 (4h + 9np_1p_2\alpha^2)\delta_3\beta^2 &< \delta_3 - 16p\delta_1 - (9npp_1p_5\delta_1 + 5np_1p_5\delta_2 \\
 &+ 2p_1\delta_4 + 10\bar{\pi}_c\delta_5)\alpha^2, \\
 \frac{4p_5}{m}\delta_3\beta^2 &< \frac{1}{m}\delta_4 - \frac{4pp_5}{m}\delta_1 - \frac{2p_5}{m}\delta_2, \\
 401hp_5p_6\delta_3\beta &< \frac{1}{4}\delta_5 - 973pp_5p_6\delta_1 - 169p_5p_6\delta_2 - 22p_5\delta_4.
 \end{aligned} \tag{53}$$

注意到不等式 (53) 的右边为正, 因此有

$$\begin{aligned}
 \alpha &< \sqrt{\frac{(1 - \sigma_A^2)\delta_1}{2(9npp_1p_2p_5\delta_1 + 5np_1p_2\delta_2 + 2p_1p_2\delta_4 + 10\bar{\pi}_r\delta_5)}} \\
 \alpha &< \frac{\delta_2}{2(5pp_3\delta_1 + p_4\delta_5)}, \\
 \alpha &< \sqrt{\frac{5\delta_2}{2n(\bar{\pi}_r^T \bar{\pi}_c)^2\delta_4}}, \\
 \alpha &< \sqrt{\frac{\delta_3 - 16p\delta_1}{9npp_1p_5\delta_1 + 5np_1p_5\delta_2 + 2p_1\delta_4 + 10\bar{\pi}_c\delta_5}}.
 \end{aligned} \tag{54}$$

选择适当的 $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5$ 使得步长 α 的取值为正, 则

$$\begin{aligned}
 \delta_1 &> 0, \\
 \delta_2 &> 0, \\
 \delta_3 &> 16p\delta_1, \\
 \delta_4 &> 4pp_5\delta_1 + 2p_5\delta_2, \\
 \delta_5 &> 4(973pp_5p_6\delta_1 - 169p_5p_6\delta_2 - 22p_5\delta_4).
 \end{aligned} \tag{55}$$

进一步, 根据 (53) 得到动量系数 β 的上界

$$\begin{aligned}
 \beta &< \frac{(1 - \sigma_A^2)\delta_1 - 2(9npp_1p_2p_5\delta_1 + 5np_1p_2\delta_2 + 2p_1p_2\delta_4 + 10\bar{\pi}_r\delta_5)\alpha^2}{2(\delta_1 + (1 + \sigma_A^2)\bar{\pi}_r + 9np_1p_2\alpha^2)}, \\
 \beta &< \sqrt{\frac{\delta_2 - (5pp_3\delta_1 + p_4\delta_5)\alpha - 2.5n(\pi_r^T \pi_c)^2\delta_4\alpha^2}{5p_3\alpha\delta_3}}, \\
 \beta &< \sqrt{\frac{\delta_3 - 16p\delta_1 - (9npp_1p_5\delta_1 + 5np_1p_5\delta_2 + 2p_1\delta_4 + 10\bar{\pi}_c\delta_5)\alpha^2}{(4h + 9np_1p_2\alpha^2)\delta_3}}, \\
 \beta &< \sqrt{\frac{\delta_4 - 4pp_5\delta_1 - 2p_5\delta_2}{4p_5\delta_3}}, \\
 \beta &< \frac{(1 - \sigma_B^2)\delta_5 - 4(973pp_5p_6\delta_1 - 169p_5p_6\delta_2 - 22p_5\delta_4)}{401hp_5p_6\delta_3}.
 \end{aligned} \tag{56}$$

证明完成.

5. 数值仿真

本节通过一些数值实验来验证所提出算法的有效性. 考虑如下分布式逻辑回归问题:

$$\min_{\mathbf{x} \in \mathbb{R}^p} F(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \frac{1}{m} \sum_{j=1}^{m_i} \log[1 + \exp(-(C_{ij}^T x)y_{ij})] + \frac{\lambda}{2} \|\mathbf{x}\|_2^2,$$

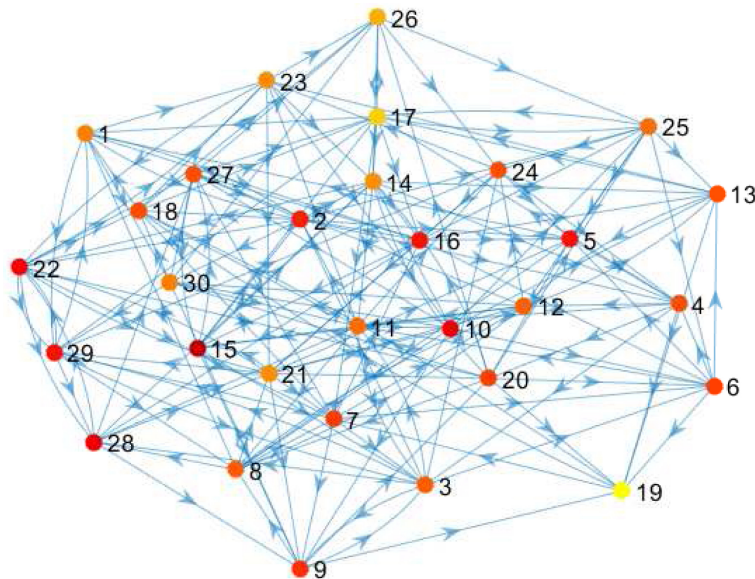


Figure 1. Directed network with 30 agents

图 1. 30 个智能体的有向网络

其中 $f_i(\mathbf{x}) = \frac{1}{m} \sum_{j=1}^{m_i} \log[1 + \exp(-(C_{ij}^T x)y_{ij})] + \frac{\lambda}{2} \|\mathbf{x}\|_2^2$, $C_{ij} \in \mathbb{R}^p$ 是特征向量, $y_{ij} \in \{+1, -1\}$ 是对应

的标签且 λ 是正则化参数. 实验中所用的两个数据集 a9a 和 w8a 是从 UCI 机器学习官方网站下载. 本文考虑由 30 个节点构成的不平衡有向网络, 如图 1 所示. 与网络拓扑相关的行随机和列随机权重矩阵 $A = [a_{ij}]$ 和 $B = [b_{ij}]$ 根据一致权重策略生成, 即 $a_{ij} = \frac{1}{N_i^{in}} \forall i$ 和 $b_{ij} = \frac{1}{N_j^{out}} \forall j$. 假设数据集上的数据被均匀的分到每个智能体上. 在所有的数值实验中, 我们使用 $\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_k^i - \mathbf{x}^*\|_2$ 表示平均残差.

当步长确定时, 我们研究动量参数 β 对 ABN-LSVRG 算法的影响. 从图 2 中可以发现, 适当的增加动量系数 β 的值可以加快收敛. 在 a9a 和 w8a 两个数据集中, 当 $\beta = 0.9$ 时, ABN-LSVRG 算法表现最好. 因此, 在以下的实验中我们设置 $\beta = 0.9$.

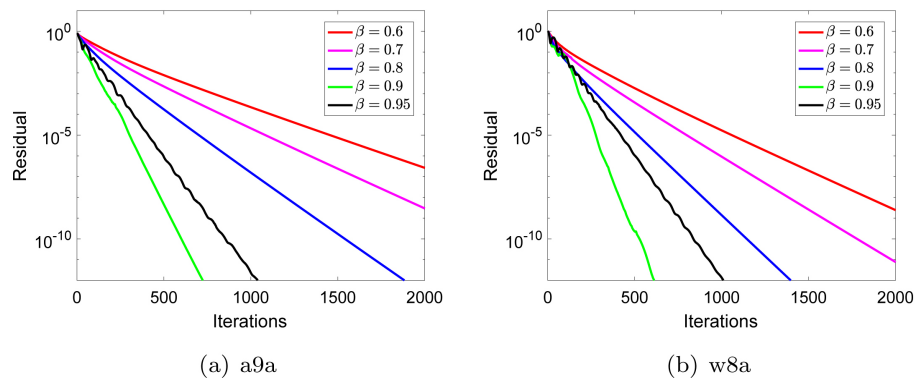


Figure 2. The impact of the momentum coefficient β for ABN-LSVRG

图 2. 动量系数 β 对 ABN-LSVRG 算法的影响

进一步, 为了说明所提出算法的有效性, 本文将 ABN-LSVRG 算法与现有的分布式随机算法进行比较, 例如 SADDOPT, S-AB, Push-SAGA, Push-LSVRG, AB-SAGA. 数值实验中, 所有算法的步长已通过手动调整至最优. 从图 3 可以看出, 本文提出的算法比存在的一些分布式随机优化算法收敛的更快. 反映了本文所提出的算法 ABN-LSVRG 相较于其他算法更具有优势.

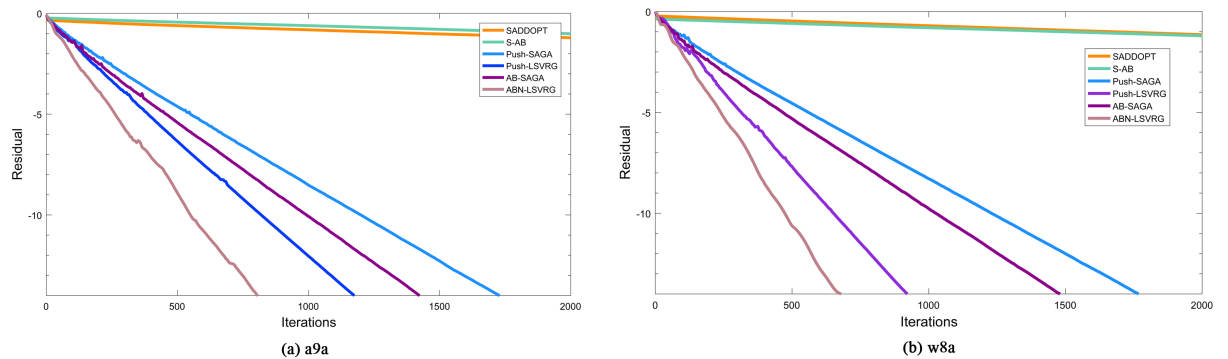


Figure 3. Comparison of ABN-LSVRG with the existing methods

图 3. ABN-LSVRG 与现有的分布式随机算法的对比

6. 结语

基于单循环的分布式方差缩减技术和 Nesterov 动量加速技巧, 本文提出了有向网络中一种快速

的分布式随机算法. 通过同时使用行随机和列随机权重矩阵, 避免了由特征向量计算所产生的额外的计算和通信. 分布式单循环的方差缩减技术消除了随机梯度产生的方差并且不需要额外的存储空间, 减少了存储成本. Nesterov 动量项的引入加快了算法的收敛. 理论分析证明本文所提出的算法可以线性收敛到最优解. 基于分布式逻辑回归问题, 数值结果验证了所提出的算法比现有一些分布式算法表现好.

参考文献

- [1] Pu, S., Olshevsky, A. and Paschalidis, I. (2020) Asymptotic Network Independence in Distributed Stochastic Optimization for Machine Learning: Examining Distributed and Centralized Stochastic Gradient Descent. *IEEE Signal Processing Magazine*, **37**, 114-122. <https://doi.org/10.1109/MSP.2020.2975212>
- [2] 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>
- [3] 王龙, 卢开红, 关永强. 分布式优化的多智能体方法[J]. 控制理论与应用, 2019, 36(11): 1820-1833.
- [4] Lü, Q., Liao, X., Li, H., *et al.* (2020) Achieving Acceleration for Distributed Economic Dispatch in Smart Grids over Directed Networks. *IEEE Transactions on Network Science and Engineering*, **7**, 1988-1999. <https://doi.org/10.1109/TNSE.2020.2965999>
- [5] 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>
- [6] Nedic, A., Olshevsky, A. and Shi, W. (2017) Achieving Geometric Convergence for Distributed Optimization over Time-Varying Graphs. *SIAM Journal on Optimization*, **27**, 2597-2633. <https://doi.org/10.1137/16M1084316>
- [7] Gao, J., Liu, X., Dai, Y., *et al.* (2022) Achieving Geometric Convergence for Distributed Optimization with Barzilai-Borwein Step Sizes. *Science China Information Sciences*, **65**, 149-204.
- [8] Ghaderyan, D., Aybat, N., Aguiar, A., *et al.* (2023) A Fast Row-Stochastic Decentralized Method for Distributed Optimization over Directed Graphs. *IEEE Transactions on Automatic Control*, **69**, 275-289. <https://doi.org/10.1109/TAC.2023.3275927>
- [9] Lian, X., Zhang, C., Zhang, H., *et al.* (2017) Can Decentralized Algorithms Outperform Centralized Algorithms? A Case Study for Decentralized Parallel Stochastic Gradient Descent. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, CA, 4-9 December 2017, 5336-5346.
- [10] Lee, S. and Zavlanos, M.M. (2017) Approximate Projection Methods for Decentralized Optimization with Functional Constraints. *IEEE Transactions on Automatic Control*, **63**, 3248-3260. <https://doi.org/10.1109/TAC.2017.2778696>
- [11] Qu, G. and Li, N. (2017) Harnessing Smoothness to Accelerate Distributed Optimization. *IEEE Transactions on Control of Network Systems*, **5**, 1245-1260. <https://doi.org/10.1109/TCNS.2017.2698261>

-
- [12] Di Lorenzo, P. and Scutari, G. (2016) Next: In-Network Nonconvex Optimization. *IEEE Transactions on Signal and Information Processing over Networks*, **2**, 120-136.
<https://doi.org/10.1109/TSIPN.2016.2524588>
- [13] Pu, S. and Nedic, A. (2021) Distributed Stochastic Gradient Tracking Methods. *Mathematical Programming*, **187**, 409-457. <https://doi.org/10.1007/s10107-020-01487-0>
- [14] Defazio, A., Bach, F. and Lacoste-Julien, S. (2014) SAGA: A Fast Incremental Gradient Method with Support for Non-Strongly Convex Composite Objectives. *Proceedings of the 27th International Conference on Neural Information Processing Systems*, **1**, 1646-1654.
- [15] Tan, C., Ma, S., Dai, Y., *et al.* (2016) Barzilai-Borwein Step Size for Stochastic Gradient Descent. *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016*, 5-10 December 2016, Barcelona.
- [16] Nguyen, L., Liu, J., Scheinberg, K., *et al.* (2017) SARAH: A Novel Method for Machine Learning Problems Using Stochastic Recursive Gradient. *International Conference on Machine Learning*, **70**, 2613-2621.
- [17] Xin, R., Khan, U. and Kar, S. (2020) Variance-Reduced Decentralized Stochastic Optimization with Accelerated Convergence. *IEEE Transactions on Signal Processing*, **68**, 6255-6271.
<https://doi.org/10.1109/TSP.2020.3031071>
- [18] Kempe, D., Dobra, A. and Gehrke, J. (2003) Gossip-Based Computation of Aggregate Information. *44th Annual IEEE Symposium on Foundations of Computer Science*, Cambridge, MA, 11-14 October 2003, 482-491.
- [19] Nedic, A. and Olshevsky, A. (2016) Stochastic Gradient-Push for Strongly Convex Functions on Time-Varying Directed Graphs. *IEEE Transactions on Automatic Control*, **61**, 3936-3947.
<https://doi.org/10.1109/TAC.2016.2529285>
- [20] Spiridonoff, A., Olshevsky, A. and Paschalidis, I.C. (2020) Robust Asynchronous Stochastic Gradient-Push: Asymptotically Optimal and Network-Independent Performance for Strongly Convex Functions. *Journal of Machine Learning Research*, **21**, 58.
- [21] Qureshi, M.I., *et al.* (2020) S-ADDOPT: Decentralized Stochastic First-Order Optimization over Directed Graphs. *IEEE Control Systems Letters*, **5**, 953-958.
<https://doi.org/10.1109/LCSYS.2020.3006420>
- [22] Xin, R., Sahu, A., Khan, U., *et al.* (2019) Distributed Stochastic Optimization with Gradient Tracking over Strongly-Connected Networks. *2019 IEEE 58th Conference on Decision and Control (CDC)*, Nice, 11-13 December 2019, 8353-8358.
- [23] Qureshi, M., Xin, R., Kar, S., *et al.* (2021) Push-SAGA: A Decentralized Stochastic Algorithm with Variance Reduction over Directed Graphs. *IEEE Control Systems Letters*, **6**, 1202-1207.
<https://doi.org/10.1109/LCSYS.2021.3090652>
- [24] Qureshi, M., Xin, R., Kar, S., *et al.* (2022) Variance Reduced Stochastic Optimization over Directed Graphs with Row and Column Stochastic Weights. arXiv Preprint arXiv:2202.03346
- [25] Qureshi, M. and Khan, U.A. (2022) Stochastic First-Order Methods over Distributed Data. *2022 IEEE 12th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Trondheim, 20-23 June 2022, 405-409.
<https://doi.org/10.1109/SAM53842.2022.9827792>

- [26] Hu, J., Chen, G., Li, H., *et al.* (2022) Push-LSVRG-UP: Distributed Stochastic Optimization over Unbalanced Directed Networks with Uncoordinated Triggered Probabilities. *IEEE Transactions on Network Science and Engineering*, **10**, 934-950. <https://doi.org/10.1109/TNSE.2022.3225229>
- [27] Kovalev, D., Horváth, S. and Richtárik, P. (2020) Don't Jump Through Hoops and Remove Those Loops: SVRG and Katyusha Are Better Without The Outer Loop. *Proceedings of the 31st International Conference on Algorithmic Learning Theory*, **117**, 451-467.
- [28] Sun, B., Hu, J., Xia, D., *et al.* (2021) A Distributed Stochastic Optimization Algorithm with Gradient-Tracking and Distributed Heavy-Ball Acceleration. *Frontiers of Information Technology & Electronic Engineering*, **22**, 1463-1476. <https://doi.org/10.1631/FITEE.2000615>
- [29] Hu, J., Xia, D., Cheng, H., *et al.* (2022) A Decentralized Nesterov Gradient Method for Stochastic Optimization over Unbalanced Directed Networks. *Asian Journal of Control*, **24**, 576-593. <https://doi.org/10.1002/asjc.2483>