

# 有向网络中基于动量方法的加速行随机优化

任冬梅

河北工业大学理学院, 天津

收稿日期: 2023年2月13日; 录用日期: 2023年3月8日; 发布日期: 2023年3月15日

## 摘要

本文研究了有向网络下的分布式凸优化问题, 并在现有分布式算法的基础上提出了一种新颖的分布式动量加速算法, 叫做ARNH。该算法采用行随机矩阵和异构步长, 有效克服网络不平衡性的同时提高了网络灵活性。此外, 为了实现更快的收敛速率, ARNH采用Nesterov梯度法和Heavy-Ball法相结合的双加速机制。在局部目标函数可微且强凸的假设下, 本文证明通过选取合适的步长和动量参数, 算法可使节点状态渐近收敛到全局最优解。最后, 在仿真实验中将ARNH与相关算法进行性能比较, 验证了新算法的优越性。

## 关键词

分布式优化, 有向网络, 异构步长, 线性收敛

# Accelerated Row-Stochastic Optimization over Directed Networks Based on Momentum Methods

Dongmei Ren

College of Science, Hebei University of Technology, Tianjin

Received: Feb. 13<sup>th</sup>, 2023; accepted: Mar. 8<sup>th</sup>, 2023; published: Mar. 15<sup>th</sup>, 2023

## Abstract

In this paper, we study the distributed convex optimization problems over directed networks, and propose a novel distributed momentum acceleration algorithm called ARNH based on the existing distributed algorithms. ARNH uses row-stochastic matrix and heterogeneous step size, which effectively overcomes the network imbalance and improves the network flexibility. Furthermore, in order to achieve faster convergence, ARNH employs a double acceleration mechanism combining

**Nesterov gradient method and heavy-ball method. Under the assumption that the local objective function is differentiable and strongly convex, it is proved that the node state can be asymptotically converged to the global optimal solution by choosing appropriate step-sizes and momentum parameters. Finally, the superiority of the new algorithm is verified by comparing the performance of ARNH with related algorithms in simulation experiments.**

## Keywords

**Distributed Optimization, Directed Network, Heterogeneous Step Size, Linear Convergence**

Copyright © 2023 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

近年来,分布式优化受到了越来越多的关注,与传统的集中控制优化策略不同,分布式优化不需要集中的数据控制,可以有效减轻通信负担,节省通信成本,增强网络可扩展性。得益于上述优势,分布式优化算法被广泛应用于统计估计[1]、机器学习[2] [3] [4]、能源调度[5] [6]、网络化多智能体系统的控制[7]等领域。

分布式优化的研究早期是在无向网络上进行的, Nedić 等在文献[8]中为解决分布式无约束优化问题突破性地提出了分布式次梯度算法(DGD),但该算法受衰减步长影响存在收敛速度较慢的缺陷。为了达到更快的收敛速度, Shi 等在文献[9]中通过对前一次迭代信息的再次使用提出了即使采用固定步长也可实现线性收敛的 EXTRA 算法。随后, Qu 等人通过结合梯度追踪技术和不精确梯度法,在文献[10]中提出了适用于时变通信网络的 DIGing 算法,当局部目标函数强凸且光滑,该算法可使节点渐近收敛到全局最优解。此后, Jakovetić 等在文献[11]中揭示了 EXTRA 和 DIGing 之间的差异在于原始误差对对偶误差的影响不同,并通过推导出一种新的方法对上述两种方法进行统一。

上述算法要求节点间的信息传输是对称且双向的,这在实际网络系统中很难实现,因此研究有向网络更具有实用性。针对有向网络下的分布式无约束凸优化问题, Xi 等通过运用 push-sum 协议[12]设计了一种基于广播通信的次梯度下降算法[13],该算法采用列随机权重矩阵和衰减步长,虽然有效克服了有向网络的不平衡性,但只能达到次线性收敛。为解决收敛速度过慢的问题, Kempe 等通过将 push-sum 协议与 EXTRA 算法相结合,在文献[14]中提出了可以实现线性收敛的 DEXTRA 算法。进一步地,文献[15]考虑了更一般的时变网络,提出了一种带有异构步长的有向图算法(push-DIGing),该算法不仅为步长提供较广阔的选择范围,而且能实现更快收敛。此外,文献[16] [17]提出一类算法可同时使用了行和列随机权重矩阵来实现在强连通有向图中的线性收敛。

上述方法虽能有效应用于有向网络,但列随机矩阵的使用需要节点知道其自身出度,这在基于广播通信的网络中是不符合实际的。为了突破这一严苛条件,文献[18] [19]中提出了一类只使用行随机权重矩阵的分布式算法,此类算法通过引入辅助变量来消除网络不平衡性,当目标函数强凸且具有 Lipschitz 连续梯度时,可实现线性收敛。在此基础上, Xin 等人在文献[20]中进一步放宽了对步长的限制,设计出采用异构步长的行随机优化算法(FROST)。随后,文献[21]基于 Nesterov 加速机制提出了 FROST 算法的加速版本,文献[22]将 FROST 与 heavy-ball 和 Nesterov 动量项相结合,以实现更快收敛。除此之外,还有一些单使用行随机权重矩阵的原始对偶算法[23] [24],与现有的有向网络下的分布式算法相比,此类方法

在没有显式使用梯度跟踪技术(GT)的情况下也具有线性收敛性。

基于上述研究, 本文在文献[11]的基础上, 结合 Push-sum 协议提出了一种可适用于有向网络的加速行随机分布式优化算法(ARNH), 并通过采用异构步长和双加速机制, 在提高网络灵活性和保证节点隐私性的同时加快了收敛速度。后续证明在一定的假设条件下, ARNH 可以实现线性收敛, 最后, 根据数值实验结果验证了该算法相较于现有的有向图算法有更好的收敛性能。

## 2. 预备知识

### 2.1. 问题描述

在一个由  $n$  个节点组成的网络系统中, 节点间的信息交换可用图  $\mathcal{G}=(\mathcal{V}, \mathcal{E})$  进行描述, 其中  $\mathcal{V}=\{1, \dots, n\}$  是有限的非空点集合,  $\mathcal{E}=\{(i, j): i, j \in \mathcal{V}\}$  为边集合。若有  $(i, j) \in \mathcal{E}$ , 那么称节点  $j$  是节点  $i$  的邻居, 对节点  $i$ , 我们将其入度邻居定义为  $N_i^{\text{in}}=\{j \in \mathcal{V} \mid (j, i) \in \mathcal{E}\}$ , 相应的其出度邻居用  $N_i^{\text{out}}=\{j \in \mathcal{V} \mid (i, j) \in \mathcal{E}\}$  表示, 其中出度邻居的个数被定义为节点  $i$  的出度。若对任意边  $(i, j) \in \mathcal{V}$  必有相应的边  $(j, i) \in \mathcal{V}$  存在, 该网络拓扑图为无向图, 反之则为有向图。若图中任意两个节点都存在一条有向路径, 那么称此图是强连通的。

分布式优化即指上述网络中的所有节点共同解决以下分布式优化问题:

$$\min_{\tilde{s} \in \mathbb{R}^p} f(\tilde{s}) = \frac{1}{n} \sum_{i=1}^n f_i(\tilde{s}) \quad (1)$$

其中  $\tilde{s} \in \mathbb{R}^p$  为全局决策变量,  $f_i(\tilde{s}): \mathbb{R}^p \rightarrow \mathbb{R}$  为节点  $i$  私有的局部目标函数。若以分布形式来解决问题(1), 需要引入一个变量来模拟节点状态, 则上述优化问题可转化为:

$$\begin{aligned} \min_{s \in \mathbb{R}^{np}} f(s) &= \frac{1}{n} \sum_{i=1}^n f_i(s_i) \\ \text{s.t. } s_i &= s_j \quad \forall i, j = 1, 2, \dots, n \end{aligned} \quad (2)$$

其中  $s = [s_1^T, s_2^T, \dots, s_n^T] \in \mathbb{R}^{np}$ , 因为问题(1)和(2)等价, 则问题(2)的最优解可写为  $s^* = 1 \otimes \tilde{s}^* \in \mathbb{R}^{np}$ 。

### 2.2. 相关定义和假设

**定义 2.1** 若矩阵  $A = [a_{ij}]$  每一行(列)求和都为 1, 称矩阵为行(列)随机矩阵。若该矩阵同时满足行随机与列随机, 那么称矩阵  $A = [a_{ij}]$  为双随机矩阵。

**定义 2.2** 若一个连续可微函数  $f: \mathbb{R}^p \rightarrow \mathbb{R}$ , 对任意的  $\tilde{s}_1, \tilde{s}_2 \in \mathbb{R}^p$  存在  $\mu$  使得:

$$f(\tilde{s}_1) \geq f(\tilde{s}_2) + \nabla f(\tilde{s}_2)^T (\tilde{s}_1 - \tilde{s}_2) + \frac{\mu}{2} \|\tilde{s}_1 - \tilde{s}_2\|^2 \quad (3)$$

成立, 则称函数  $f$  是  $\mu$ -强凸, 其中  $\mu$  是强凸系数。

**定义 2.3** 若一个连续可微函数  $f: \mathbb{R}^p \rightarrow \mathbb{R}$ , 对任意的  $\tilde{s}_1, \tilde{s}_2 \in \mathbb{R}^p$  存在  $l$  使得:

$$\|\nabla f(\tilde{s}_2) - \nabla f(\tilde{s}_1)\| \leq l \|\tilde{s}_2 - \tilde{s}_1\| \quad (4)$$

成立, 则称函数  $f$  是  $l$ -光滑或简称光滑, 其中  $l$  是 Lipschitz 常数。

**假设 2.1** 网络拓扑图  $\mathcal{G}=(\mathcal{V}, \mathcal{E})$  强连通。

**假设 2.2** 局部目标函数  $f_i: \mathbb{R}^p \rightarrow \mathbb{R}$  强凸且光滑。

### 2.3. 算法发展

1) Heavy-ball 方法

Polyak 为解决一般优化问题在文献[25]中提出了 Heavy-ball 方法:

$$x_{k+1} = x_k - \alpha \nabla f(x_k) + \eta(x_k - x_{k-1})$$

其中  $\eta(x_k - x_{k-1})$  为重球动量项, 文中指出当步长  $\alpha$  和动量参数  $\eta$  的选取满足约束条件时, 该算法可达到比一般梯度下降法更快的收敛速度, 因此后续有学者将该方法引入分布式优化算法[23] [26]中以实现更快收敛。

2) Nesterov 梯度法

文献[26]对一阶梯度法使用历史信息提出了 Nesterov 梯度方法:

$$\begin{aligned} x_{k+1} &= s_k - \alpha \nabla f(s_k) \\ s_{k+1} &= x_{k+1} + \eta(x_{k+1} - x_k) \end{aligned}$$

其中  $\beta(x_{k+1} - x_k)$  为 Nesterov 动量项, 当目标函数满足强凸且光滑条件,  $\alpha = \frac{1}{l}$  且  $\beta = \frac{\sqrt{l} - \sqrt{\mu}}{\sqrt{l} + \sqrt{\mu}}$  时, 该算法可以达到  $O\left(\left(1 - \sqrt{\alpha\mu}\right)^k\right)$  的收敛速率, 这比相同情况下的梯度下降法更快, 因此后续有学者将其应用到分布式算法[22] [23]中。

受上述算法的启发, 我们将动量项与文献[11]中的算法相结合, 设计出一种行随机加速优化算法 (ARNH)。所提出的算法中, 对任意第  $k \in N$  次迭代, 每个节点  $i \in \mathcal{V}$  持有四个变量,  $x_k^i$ 、 $s_k^i$ 、 $v_k^i$  和  $u_k^i$ , 根据如下规则进行更新

$$x_{k+1}^i = \sum_{j=1}^n r_{ij} s_k^j - \alpha_i \left( \frac{\nabla f_i(s_k^i)}{[v_k^i]_i} + u_k^i \right) + \eta_i (x_k^i - x_{k-1}^i) \tag{5}$$

$$s_{k+1}^i = x_{k+1}^i + \eta_i (x_{k+1}^i - x_k^i) \tag{6}$$

$$u_{k+1}^i = u_k^i - \sum_{j=1}^n l_{ij} \left( \frac{\nabla f_j(s_k^j)}{[v_k^j]_j} + u_k^j - \sum_{l=1}^n b_{jl} s_k^l \right) \tag{7}$$

$$v_{k+1}^i = \sum_{j=1}^n r_{ij} v_k^j \tag{8}$$

算法初始化设定任意  $x_0^i = s_0^i \in \mathbb{R}^p$ ,  $u_k^i = \mathbf{0}_p$  和  $v_0^i = e_i$ , 权重矩阵  $R = [r_{ij}]$  为行随机矩阵,  $L = [l_{ij}] = I_n - R$ , 矩阵  $B = [b_{ij}]$  为  $n \times n$  维的行随机矩阵, 并且满足有  $b \in \mathbb{R}$  存在使得  $B \mathbf{1}_n = b \mathbf{1}_n$  成立。  $\alpha_i$  和  $\eta_i$  分别对应节点  $i$  的步长和动量参数。为了更好地分析算法收敛性, 我们将 ARNH 改写成矩阵形式前需要引入必要符号:

$$\alpha = [\alpha_1, \alpha_2, \dots, \alpha_n]^T, \quad D_\alpha = \text{diag}\{\alpha\} \otimes I_p, \quad V_k = [v_k^1, v_k^2, \dots, v_k^n]^T, \quad \bar{B} = B \otimes I_p, \quad \bar{L} = L \otimes I_p,$$

$$\eta = [\eta_1, \eta_2, \dots, \eta_n]^T, \quad D_\eta = \text{diag}\{\eta\} \otimes I_p, \quad \bar{V}_k = \text{diag}\{V_k\} \otimes I_p, \quad \bar{R} = R \otimes I_p.$$

那么 ARNH 算法可转换为如下矩阵形式:

$$x_{k+1} = \bar{R} s_k - D_\alpha (\bar{V}_k^{-1} \nabla F(s_k) + u_k) + D_\eta (x_k - x_{k-1}) \tag{9}$$

$$s_{k+1} = x_{k+1} - D_\eta (x_{k+1} - x_k) \tag{10}$$

$$u_{k+1} = u_k - \bar{L} (\bar{V}_k^{-1} \nabla F(s_k) + u_k - \bar{B} s_k) \tag{11}$$

$$V_{k+1} = \bar{R}V_k \quad (12)$$

其中  $\nabla F(s_k)$ 、 $x_k$ 、 $s_k$  和  $u_k$  被分别定义为  $\nabla F(x_k) = [\nabla f_1(x_k^1)^T, \nabla f_2(x_k^2)^T, \dots, \nabla f_n(x_k^n)^T]^T$ 、  
 $x_k = [(x_k^1)^T, (x_k^2)^T, \dots, (x_k^n)^T]^T$ 、 $s_k = [(s_k^1)^T, (s_k^2)^T, \dots, (s_k^n)^T]^T$  和  $u_k = [(u_k^1)^T, (u_k^2)^T, \dots, (u_k^n)^T]^T$ 。

### 3. 算法收敛性证明

$$\begin{aligned} V_\infty &= \lim_{k \rightarrow \infty} V_k, & v &= \sup_{k \geq 0} \|V_k\|, & \bar{\eta} &= \max_i \{\eta_i\}, & \tau &= \|\bar{R} - I_{np}\|, & \lambda &= \|\bar{R} + I_{np}\|, \\ \bar{V}_k &= \text{diag}(V_k), & \hat{v} &= \sup_{k \geq 0} \|\bar{V}_k^{-1}\|, & \bar{\alpha} &= \max_i \{\alpha_i\}, & \gamma &= \|I_{np} - \bar{V}_\infty\|. \end{aligned}$$

#### 3.1. 辅助引理

**引理 3.1** [24] 考虑由行随机矩阵  $R$  生成的  $\bar{V}_k$ ，对于任意的  $k \in N$ ，有下列不等式成立

$$\|\bar{V}_k - \bar{V}_\infty\| \leq k_1 \sigma^k$$

其中  $k_1$  为大于 0 的常数， $\sigma \in (0,1)$  为收缩因子。

**引理 3.2** [22] 考虑由行随机矩阵  $R$  生成的  $\bar{V}_k$ ，对于所有的  $k \in N$ ，有下列不等式成立

$$\|\bar{V}_k^{-1} - \bar{V}_\infty^{-1}\| \leq k_1 \sqrt{n} \hat{v}^2 \sigma^k$$

其中  $\sigma \in (0,1)$  为收缩因子在引理 1 中被定义，其中  $k_1$  为大于 0 的常数。

**引理 3.3** [21] 给定  $0 < \alpha < \frac{2}{nl}$ ，令  $\omega = \max\{|1 - \alpha n \mu|, |1 - \alpha n l|\}$ ，如果假设 2.2 成立，那么对于  $\forall x \in \mathbf{R}^{np}$ ，我们有

$$\|x - \alpha \nabla F(x) - x^*\| \leq \omega \|x - x^*\|_2$$

引理 3.2 建立了  $\bar{V}_k^{-1}$  的线性收敛性。引理 3.3 指出，在集中式梯度法中，当前迭代值与最优解之间的距离在每次迭代中收缩。

**引理 3.4** [21] 对于一个非负矩阵  $X \in \mathbf{R}^{n \times n}$  和正向量  $x \in \mathbf{R}^n$ ，如果有  $Xx < \zeta x$ ，那么  $\rho(X) < \zeta$ 。

#### 3.2. 收敛关系

被提出的算法收敛关系将通过推导如下四项误差的上界得到

- 1) 对偶误差:  $\|u_{k+1} - u^*\|$ ;
- 2) 最优误差:  $\|\hat{x}_{k+1} - s^*\|$ ;
- 3) 状态误差:  $\|x_k - x_{k-1}\|$ ;
- 4) 一致性误差:  $\|x_{k+1} - \hat{x}_{k+1}\|$ 。

其中  $\hat{x}_{k+1} = \bar{V}_\infty^{-1} x_{k+1}$ 。

**引理 3.5** 若假设 2.1-2.2 成立，对任意  $k \geq 0$  有下列不等式成立

$$\begin{aligned} \|u_{k+1} - u^*\| &\leq \sigma \|u_k - u^*\| + (\|B\| + \hat{v}l) \|L\| \|x_k - \hat{x}_k\| + (\|B\| + \hat{v}l) \|L\| \|\hat{x}_k - s^*\| \\ &\quad + \bar{\eta} \|L\| (\|B\| + \hat{v}l) \|x_k - x_{k-1}\| + \hat{v}^2 \sigma^k k_1 \|L\| \|\nabla F(s_k)\| \end{aligned}$$

证明 考虑  $k$  趋近于无穷时算法收敛于  $(s^*, u^*)$  根据式子(11)，我们有

$$u^* = u^* - \bar{L}(\bar{V}_\infty^{-1} \nabla F(s^*) + u^* - \bar{B}s^*) \quad (13)$$

这意味着  $u^* = -\bar{V}_\infty^{-1}\nabla F(s^*)$ ，则等式(13)可变换为下列形式

$$\begin{aligned} u_{k+1} - u^* &= u_k - u^* - \bar{L}(\bar{V}_k^{-1}\nabla F(s_k) - \bar{V}_\infty^{-1}\nabla F(s^*) + \bar{V}_\infty^{-1}\nabla F(s^*) + u_k) + \bar{L}\bar{B}s_k \\ &= \bar{R}(u_k - u^*) + \bar{L}(\bar{V}_k^{-1}\nabla F(s_k) - \bar{V}_\infty^{-1}\nabla F(s^*)) + \bar{L}\bar{B}s_k, \end{aligned} \tag{14}$$

根据初始化条件  $u_0 = \mathbf{0}_{np}$  可证得  $\bar{V}_\infty u_{k+1} = \bar{V}_\infty u_k = \mathbf{0}_{np}$ ，由于  $\bar{L}\bar{B}s^* = b\bar{L}s^* = \mathbf{0}_{np}$ ，(14)可改写为

$$u_{k+1} - u^* = (\bar{R} - \bar{V}_\infty)(u_k - u^*) + \bar{L}(\bar{V}_k^{-1}\nabla F(s_k) - \bar{V}_\infty^{-1}\nabla F(s^*)) + \bar{L}\bar{B}(s_k - s^*) \tag{15}$$

根据引理 2.2~2.3 以及假设 2，等式(15)右边的第二项可改写为

$$\begin{aligned} \|\bar{V}_k^{-1}\nabla F(s_k) - \bar{V}_\infty^{-1}\nabla F(s^*)\| &\leq \|\bar{V}_k^{-1}\nabla F(s_k) - \bar{V}_\infty^{-1}\nabla F(s_k)\| + \|\bar{V}_\infty^{-1}\nabla F(s_k) - \bar{V}_\infty^{-1}\nabla F(s^*)\| \\ &\leq \hat{\nu}l\|x_k - \hat{x}_k\| + \hat{\nu}l\|\hat{x}_k - s^*\| + \bar{\eta}\hat{\nu}l\|x_k - x_{k-1}\| + \hat{\nu}^2\sigma^k k_1\|\nabla F(s_k)\| \end{aligned} \tag{16}$$

将等式(16)带入等式(15)，并对两边取范数可得

$$\begin{aligned} \|u_{k+1} - u^*\| &\leq \sigma\|u_k - u^*\| + (\|\bar{B}\| + \hat{\nu}l)\|\bar{L}\| \|x_k - \hat{x}_k\| + (\|\bar{B}\| + \hat{\nu}l)\|\bar{L}\| \|\hat{x}_k - s^*\| \\ &\quad + \bar{\eta}\|\bar{L}\|(\|\bar{B}\| + \hat{\nu}l)\|x_k - x_{k-1}\| + \hat{\nu}^2\sigma^k k_1\|\bar{L}\| \|\nabla F(s_k)\| \end{aligned} \tag{17}$$

**引理 3.6** 若假设 1~3 成立，对  $k \geq 0$  下面不等式成立：

$$\begin{aligned} \|x_{k+1} - \hat{x}_{k+1}\| &\leq \bar{\alpha}\gamma\|u_k - u^*\| + (\sigma + \bar{\alpha}\gamma\hat{\nu}l)\|x_k - \hat{x}_k\| + \bar{\alpha}\gamma\hat{\nu}l\|\hat{x}_k - s^*\| \\ &\quad + (\lambda\gamma\bar{\eta} + \bar{\alpha}\gamma\bar{\eta}\hat{\nu}l)\|x_k - x_{k-1}\| + \bar{\alpha}\gamma\hat{\nu}^2\sigma^k k_1\|\nabla F(s_k)\| \end{aligned}$$

证明 结合(9)和(10)可得

$$x_{k+1} = \bar{R}x_k - D_\alpha(\bar{V}_k^{-1}\nabla F(s_k) + u_k) + (\bar{R} + I_{np})D_\eta(x_k - x_{k-1}) \tag{18}$$

在上式两端乘以  $(I_{np} - \bar{V}_\infty)$  并使用  $\bar{R}\bar{V}_\infty = \bar{V}_\infty$ ，我们有

$$\begin{aligned} x_{k+1} - \hat{x}_{k+1} &= (\bar{R} - \bar{V}_\infty)x_k + (I_{np} - \bar{V}_\infty)(\bar{R} + I_{np})D_\eta(x_k - x_{k-1}) \\ &\quad - (I_{np} - \bar{V}_\infty)D_\alpha(\bar{V}_k^{-1}\nabla F(s_k) + u_k), \end{aligned} \tag{19}$$

根据  $(\bar{R} - \bar{V}_\infty)x_k = (\bar{R} - \bar{V}_\infty)(x_k - \hat{x}_k)$ ，进一步可得

$$\begin{aligned} x_{k+1} - \hat{x}_{k+1} &= (\bar{R} - \bar{V}_\infty)(x_k - \hat{x}_k) + (I_{np} - \bar{V}_\infty)(\bar{R} + I_{np})D_\eta(x_k - x_{k-1}) \\ &\quad - (I_{np} - \bar{V}_\infty)D_\alpha(\bar{V}_k^{-1}\nabla F(s_k) - \bar{V}_\infty^{-1}\nabla F(s^*) + \bar{V}_\infty^{-1}\nabla F(s^*) + u_k) \\ &= (\bar{R} - \bar{V}_\infty)(x_k - \hat{x}_k) + (I_{np} - \bar{V}_\infty)(\bar{R} + I_{np})D_\eta(x_k - x_{k-1}) \\ &\quad - (I_{np} - \bar{V}_\infty)D_\alpha(\bar{V}_k^{-1}\nabla F(s_k) - \bar{V}_\infty^{-1}\nabla F(s^*)) \\ &\quad - (I_{np} - \bar{V}_\infty)D_\alpha(\bar{V}_\infty^{-1}\nabla F(s^*) + u_k), \end{aligned} \tag{20}$$

对式子(20)取范数并将式子(16)带入，我们可得

$$\begin{aligned} \|x_{k+1} - \hat{x}_{k+1}\| &\leq \bar{\alpha}\gamma\|u_k - u^*\| + (\sigma + \bar{\alpha}\gamma\hat{\nu}l)\|x_k - \hat{x}_k\| + \bar{\alpha}\gamma\hat{\nu}l\|\hat{x}_k - s^*\| \\ &\quad + (\lambda\gamma\bar{\eta} + \bar{\alpha}\gamma\bar{\eta}\hat{\nu}l)\|x_k - x_{k-1}\| + \bar{\alpha}\gamma\hat{\nu}^2\sigma^k k_1\|\nabla F(s_k)\| \end{aligned}$$

**引理 3.7** 若假设 1~3 成立，对  $k \geq 0$  下面不等式成立：

$$\begin{aligned} \|x_{k+1} - x_k\| &\leq (\tau + \bar{\alpha}\hat{v}l)\|x_k - \hat{x}_k\| + (\bar{\alpha}\bar{\eta}\hat{v}l + \bar{\eta}\lambda)\|x_k - x_{k-1}\| \\ &\quad + \bar{\alpha}\hat{v}l\|\hat{x}_k - s^*\| + \bar{\alpha}\hat{v}^2\sigma^k k_1 \|\nabla F(s_k)\| + \bar{\alpha}\|u_k - u^*\| \end{aligned}$$

在等式(9)两边同时减去  $x_k$  并运用(10)可得

$$x_{k+1} - x_k = (\bar{R} - I_{\text{np}})x_k - D_\alpha(\bar{V}_k^{-1}\nabla F(s_k) + u_k) + (\bar{R} + I_{\text{np}})D_\eta(x_k - x_{k-1})$$

根据  $\bar{R}\hat{x}_k = \hat{x}_k$ , 我们有下式成立

$$\begin{aligned} x_{k+1} - x_k &= (\bar{R} - I_{\text{np}})(x_k - \hat{x}_k) + (\bar{R} + I_{\text{np}})D_\eta(x_k - x_{k-1}) \\ &\quad + D_\alpha(\bar{V}_\infty^{-1}\nabla F(s^*) + u_k + \bar{V}_k^{-1}\nabla F(s_k) - \bar{V}_\infty^{-1}\nabla F(s^*)), \end{aligned} \quad (21)$$

对(21)两边同时取范数, 我们有

$$\begin{aligned} \|x_{k+1} - x_k\| &\leq (\tau + \bar{\alpha}\hat{v}l)\|x_k - \hat{x}_k\| + (\bar{\eta}\bar{\alpha}\hat{v}l + \lambda\bar{\eta})\|x_k - x_{k-1}\| \\ &\quad + \bar{\alpha}\hat{v}l\|\hat{x}_k - s^*\| + \bar{\alpha}\hat{v}^2\sigma^k k_1 \|\nabla F(s_k)\| + \bar{\alpha}\|u_k - u^*\| \end{aligned}$$

引理 3.7 证明完毕。

**引理 3.8** 若假设 1~3 成立, 对  $k \geq 0$  下列不等式成立

$$\begin{aligned} \|\hat{x}_{k+1} - s^*\| &\leq (\omega + v\hat{v}l\Delta_\alpha)\|\hat{x}_k - s^*\| + (\bar{\alpha} + \Delta_\alpha)v\hat{v}l\|x_k - \hat{x}_k\| + \Delta_\alpha v\|u_{k+1} - u^*\| \\ &\quad + (\bar{\alpha}\hat{v}l + \Delta_\alpha\hat{v}l + \lambda)v\bar{\eta}\|x_k - x_{k-1}\| + (\bar{\alpha} + \Delta_\alpha)v\hat{v}^2\sigma^k k_1 \|\nabla F(s_k)\| \end{aligned}$$

其中  $\Delta_\alpha = \bar{\alpha} - \alpha_{\min}$ 。

根据  $\bar{V}_\infty \bar{R} = \bar{V}_\infty$  在等式(9)两边同时乘  $\bar{V}_\infty$  并运用(10)可得

$$\bar{V}_\infty x_{k+1} = \bar{V}_\infty x_k + \bar{V}_\infty (\bar{R} + I_{\text{np}})D_\eta(x_k - x_{k-1}) - \bar{V}_\infty D_\alpha(\bar{V}_k^{-1}\nabla F(s_k) + u_k) \quad (22)$$

根据  $\bar{V}_\infty u_{k+1} = \bar{V}_\infty u_k = \mathbf{0}_{\text{np}}$ , 在(22)右边加减  $\bar{\alpha}\bar{V}_\infty \bar{V}_\infty^{-1}\nabla F(\hat{x}_k)$ , 我们可得

$$\begin{aligned} \bar{V}_\infty x_{k+1} &= \bar{V}_\infty x_k - \bar{\alpha}\bar{V}_\infty \bar{V}_\infty^{-1}\nabla F(\hat{x}_k) + \bar{\alpha}\bar{V}_\infty (\bar{V}_\infty^{-1}\nabla F(\hat{x}_k) - \bar{V}_k^{-1}\nabla F(s_k)) \\ &\quad + \bar{V}_\infty (\bar{\alpha}I_{\text{np}} - D_\alpha)(\bar{V}_k^{-1}\nabla F(s_k) - \bar{V}_\infty^{-1}\nabla F(s^*)) \\ &\quad + \bar{V}_\infty (\bar{\alpha}I_{\text{np}} - D_\alpha)(u_{k+1} - u^*) + \bar{V}_\infty (\bar{R} + I_{\text{np}})D_\eta(x_k - x_{k-1}) \end{aligned}$$

根据假设 2 和引理 2.3, (23)式的第二项满足下列不等式

$$\begin{aligned} \|\bar{V}_k^{-1}\nabla F(s_k) - \bar{V}_\infty^{-1}\nabla F(\hat{x}_k)\| &\leq \|\bar{V}_k^{-1}\nabla F(s_k) - \bar{V}_\infty^{-1}\nabla F(s_k)\| + \|\bar{V}_\infty^{-1}\nabla F(s_k) - \bar{V}_\infty^{-1}\nabla F(\hat{x}_k)\| \\ &\leq \hat{v}^2\sigma^k k_1 \|\nabla F(s_k)\| + \hat{v}l\|x_k - \hat{x}_k\| + \bar{\eta}\hat{v}l\|x_k - x_{k-1}\| + \bar{\eta}\hat{v}l\|x_k - x_{k-1}\| \end{aligned} \quad (23)$$

将(23)和(16)带入(22), 并对等式两边减去  $s^*$  同时取范数可得

$$\begin{aligned} \|\hat{x}_{k+1} - s^*\| &\leq (\omega + v\hat{v}l\Delta_\alpha)\|\hat{x}_k - s^*\| + (\bar{\alpha} + \Delta_\alpha)v\hat{v}l\|x_k - \hat{x}_k\| + \Delta_\alpha v\|u_{k+1} - u^*\| \\ &\quad + (\bar{\alpha}\hat{v}l + \Delta_\alpha\hat{v}l + \lambda)v\bar{\eta}\|x_k - x_{k-1}\| + (\bar{\alpha} + \Delta_\alpha)v\hat{v}^2\sigma^k k_1 \|\nabla F(s_k)\| \end{aligned}$$

### 3.3. 主要结果

在呈现主要结果之前, 根据以上所提及的结果和符号, 我们进行如下定义:

$$t_k = \begin{bmatrix} \|u_k - u^*\| \\ \|x_k - \hat{x}_k\| \\ \|\hat{x}_k - s^*\| \\ \|x_k - x_{k-1}\| \end{bmatrix}, H_k = \sigma^k \begin{bmatrix} \hat{v}^2 \sigma^k k_1 \|\bar{L}\| & 0 & 0 & 0 \\ \bar{\alpha} \gamma \hat{v}^2 k_1 & 0 & 0 & 0 \\ v \hat{v}^2 k_1 (\Delta_\alpha + \bar{\alpha}) & 0 & 0 & 0 \\ \bar{\alpha} \hat{v}^2 k_1 & 0 & 0 & 0 \end{bmatrix}, s_k = \begin{bmatrix} \|\nabla F(s_k)\| \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$J_{(\alpha, \eta)} = \begin{bmatrix} \sigma & w_1 & w_1 & \bar{\eta} w_1 \\ \bar{\alpha} \gamma & \lambda + \bar{\alpha} w_2 & \bar{\alpha} w_2 & \lambda \gamma \bar{\eta} + \bar{\alpha} \bar{\eta} w_2 \\ v \Delta_\alpha & (\Delta_\alpha + \bar{\alpha}) w_3 & \omega + \Delta_\alpha w_3 & \bar{\eta} (\bar{\alpha} w_3 + \Delta_\alpha w_3 + \lambda v) \\ \bar{\alpha} & \tau + \bar{\alpha} w_4 & \bar{\alpha} w_4 & \bar{\eta} (\bar{\alpha} w_4 + \lambda) \end{bmatrix}, \begin{cases} w_1 = (\|\bar{B}\| + \hat{v} l) \|\bar{L}\|, \\ w_2 = \gamma \hat{v} l, \\ w_3 = v \hat{v} l, \\ w_4 = \hat{v} l. \end{cases}$$

**定理 3.1** 若假设 1~2 成立, 根据引理 3.5~3.8 有下列线性不等式成立

$$t_{k+1} \leq J_{(\alpha, \eta)} t_k + H_k s_k$$

当最大步长  $\bar{\alpha}$  满足:

$$0 \leq \bar{\alpha} < \min \left\{ \frac{c_2 - \lambda(\bar{\eta} c_4 + c_2)}{\gamma c_1 + w_2 c_2 + w_2 c_3 + w_2 \bar{\eta} c_4}, \frac{c_4 - \tau c_2}{c_1 + w_4 c_2 + w_4 c_3 + \bar{\eta} w_4 c_4}, \frac{1}{nl} \right\} \tag{24}$$

动量参数  $\bar{\eta}$  满足:

$$0 \leq \bar{\eta} < \min \left\{ \frac{(1-\lambda)c_2 - \bar{\alpha}(\gamma c_1 + w_2 c_2 + w_2 c_3)}{\lambda c_4 + \bar{\alpha} w_2 c_4}, \frac{c_4 - \tau c_2 - \bar{\alpha}(c_1 + w_4 c_2 + w_4 c_3)}{\lambda c_4 + \bar{\alpha} w_4 c_4}, \frac{\bar{\alpha}(n\mu c_3 - w_3 c_2) - \Delta_\alpha(v c_1 + w_3 c_2 + w_3 c_3)}{n\mu c_3 + \bar{\alpha} w_3 c_4 + w_3 \Delta_\alpha c_4} \right\}. \tag{25}$$

其中  $c_1, c_2, c_3, c_4$  为随机常数, 取值满足:

$$0 < c_2, \tau c_2 < c_4, \frac{w_1(c_2 + c_3) + \bar{\eta} w_1 c_4}{1 - \sigma} < c_1, 0 < c_3 < \frac{n\mu c_3 - \bar{\eta} w_3 c_4}{w_3}$$

最大步长和最小步长之间的差距满足:

$$0 < \Delta_\alpha < \frac{\bar{\alpha}(n\mu c_3 - w_3 c_2 - \bar{\eta} w_3 c_4)}{v c_1 + w_3 c_2 + w_3 c_3 + w_3 \bar{\eta} c_4}$$

则矩阵  $J_{(\alpha, \eta)}$  的谱半径小于 1, 即  $\rho(J_{(\alpha, \eta)}) < 1$ 。

证明 由于  $l > \mu$  所以  $\omega = 1 - \bar{\alpha} n \mu$ , 根据引理 4 可知存在正向量  $c = [c_1, c_2, c_3, c_4]^T \in \mathbb{R}_+^4$  使得  $J_{(\alpha, \eta)} c < c$ , 即下列不等式组成立

$$\begin{cases} w_1(c_2 + c_3) + \bar{\eta} w_1 c_4 < (1 - \sigma) c_1, \\ \lambda \gamma \bar{\eta} c_4 < (1 - \lambda) c_2 - \bar{\alpha}(\gamma c_1 + w_2 c_2 + w_2 c_3 + \bar{\eta} w_2 c_4), \\ v \lambda \bar{\eta} c_4 < \bar{\alpha}(n\mu c_3 - w_3 c_2 - \bar{\eta} w_3 c_4) - \Delta_\alpha(v c_1 + w_3 c_2 + w_3 c_3 + w_3 \bar{\eta} c_4), \\ \bar{\eta} \lambda c_4 < c_4 - \tau c_2 - \bar{\alpha}(c_1 + w_4 c_2 + w_4 c_3 + \bar{\eta} w_4 c_4). \end{cases} \tag{26}$$

为了确保该不等式的左项严格大于 0, 我们需要下列不等式组成立

$$\begin{cases} 0 < (1 - \lambda) c_2 - \bar{\alpha}(\gamma c_1 + w_2 c_2 + w_2 c_3 + \bar{\eta} w_2 c_4), \\ 0 < \bar{\alpha}(n\mu c_3 - w_3 c_2 - \bar{\eta} w_3 c_4) - \Delta_\alpha(v c_1 + w_3 c_2 + w_3 c_3 + w_3 \bar{\eta} c_4), \\ 0 < c_4 - \tau c_2 - \bar{\alpha}(c_1 + w_4 c_2 + w_4 c_3 + \bar{\eta} w_4 c_4). \end{cases} \tag{27}$$



通过计算可知要使不等式(26)第一项成立, 最大步长  $\bar{\alpha}$  需要满足:

$$\bar{\alpha} < \frac{c_2 - \lambda(\bar{\eta}c_4 + c_2)}{\gamma c_1 + w_2 c_2 + w_2 c_3 + w_2 \bar{\eta} c_4} \quad (28)$$

根据式子(27)的第二项, 参数  $c_2$  和  $\bar{\alpha}$  必须满足下列条件:

$$0 < c_3 < \frac{n\mu c_3 - \bar{\eta} w_3 c_4}{w_3}, \quad 0 < \Delta_\alpha < \frac{\bar{\alpha}(n\mu c_3 - w_3 c_2 - \bar{\eta} w_3 c_4)}{v c_1 + w_3 c_2 + w_3 c_3 + w_3 \bar{\eta} c_4} \quad (29)$$

根据式子(27)的第三项, 可得

$$c_4 > \tau c_2, \quad \bar{\alpha} < \frac{c_4 - \tau c_2}{c_1 + w_4 c_2 + w_4 c_3 + \bar{\eta} w_4 c_4} \quad (30)$$

为了确保式子(26)成立, 必须有

$$\frac{w_1(c_2 + c_3) + \bar{\eta} w_1 c_4}{1 - \sigma} < c_1 \quad (31)$$

首先选择合适的参数  $c_2$  之后, 根据式子(29)选择  $c_3$ , 根据式子(30)选择  $c_4$ , 最后选择满足(31)的  $c_1$ , 则可以找出符合要求的正向量  $c$ 。由于步长的选择需要满足原始条件  $\bar{\alpha} < \frac{2}{nl}$ , 结合(28)和(30)将式子进行整理得到步长的最终取值范围:

$$0 \leq \bar{\alpha} < \min \left\{ \frac{c_2 - \lambda(\bar{\eta}c_4 + c_2)}{\gamma c_1 + w_2 c_2 + w_2 c_3 + w_2 \bar{\eta} c_4}, \frac{c_4 - \tau c_2}{c_1 + w_4 c_2 + w_4 c_3 + \bar{\eta} w_4 c_4}, \frac{1}{nl} \right\} \quad (32)$$

根据(26)可以求得最大动量参数  $\bar{\eta}$  的取值范围如下

$$0 \leq \bar{\eta} < \min \left\{ \frac{(1-\lambda)c_2 - \bar{\alpha}(\gamma c_1 + w_2 c_2 + w_2 c_3)}{\lambda c_4 + \bar{\alpha} w_2 c_4}, \frac{c_4 - \tau c_2 - \bar{\alpha}(c_1 + w_4 c_2 + w_4 c_3)}{\lambda c_4 + \bar{\alpha} w_4 c_4}, \frac{\bar{\alpha}(n\mu c_3 - w_3 c_2) - \Delta_\alpha(v c_1 + w_3 c_2 + w_3 c_3)}{n\mu c_3 + \bar{\alpha} w_3 c_4 + w_3 \Delta_\alpha c_4} \right\} \quad (33)$$

当步长和参数选取满足约束时, 我们有  $\rho(J_{(\alpha, \bar{\eta})}) < 1$ 。

**引理 3.9** [15] 假设所选步长满足(32), 则对于所有  $k \geq 0$ , 下列不等式均成立:

$$\begin{aligned} \|J_{(\alpha, \bar{\eta})}^k\| &\leq \Gamma_1 \theta_1^k \quad 0 < \theta_1 < 1, 0 < \Gamma_1 < \infty, \\ \|H_r\| &\leq \Gamma_2 \theta_2^k \quad 0 < \theta_2 < 1, 0 < \Gamma_2 < \infty, \\ \|J_{(\alpha, \bar{\eta})}^{k-r-1} H_r\| &\leq \Gamma \theta^k \quad 0 \leq r \leq k-1, \end{aligned}$$

其中  $\theta = \max\{\theta_1, \theta_2\}$ ,  $\Gamma = \Gamma_1 \Gamma_2 / \theta$ 。

**定理 3.2** 假定假设 1~2 成立, 当步长和动量项满足定理 3.1 时, 由所提出的算法生成的序列  $\{x_k\}$  线性收敛于全局最优解  $s^*$ , 即存在常数  $p > 0$  使得:

$$\|x_k - s^*\| \leq 2p(\theta + \xi)^k$$

其中  $\theta = \max\{\theta_1, \theta_2\}$  取自引理 3.9,  $\xi$  是满足  $0 < \xi + \theta < 1$  的足够小的正常数。

证明 利用数学归纳法对  $t_{k+1} \leq J_{(\alpha, \bar{\eta})} t_k + H_k s_k$  进行不断递归, 整理后可得到

$$t_{k+1} \leq J_{(\alpha, \bar{\eta})} t_0 + \sum_{r=0}^{k-1} J_{(\alpha, \bar{\eta})}^{k-r-1} H_r s_r \quad (34)$$

对(34)两边同时取范数并利用引理 3.9 我们有

$$\|t_k\| \leq \|J_{(\alpha,\eta)}\| \|t_0\| + \sum_{r=0}^{k-1} \|J_{(\alpha,\eta)}^{k-r-1} H_r\| \|s_r\| \leq \Gamma \omega^k \left( \|t_0\| + \sum_{j=0}^{k-1} \|s_r\| \right) \tag{35}$$

进一步，利用三角不等式和假设 3 对  $\|s_r\|$  进行定界

$$\begin{aligned} \|s_k\| &\leq \|\nabla F(V_\infty x_k) - \nabla F(s^*)\| + \|\nabla F(s^*)\| \\ &\leq L \|x_k - V_\infty x_{k-1}\| + l \|V_\infty x_k - s^*\| + \|\nabla F(s^*)\| \\ &\leq 2L \|t_k\| + \|\nabla F(s^*)\| \end{aligned} \tag{36}$$

将(36)代入(35)可得下式成立：

$$\|t_k\| \leq \left( \|t_0\| + 2l \sum_{r=0}^{k-1} \|t_r\| + k \|\nabla F(s^*)\| \right) \Gamma \theta^k \tag{37}$$

对  $k \geq 0$ ，定义  $a_k \triangleq \sum_{r=0}^{k-1} \|t_r\|$ ， $b_k = 0$ ， $c_k = l\Gamma\theta^k$ ， $d_k \triangleq \theta^k (\Gamma_2 \|t_0\| + k\Gamma \|\nabla F(s^*)\|)$  那么(37)可重写为：

$$\|t_k\| = a_{k+1} - a_k \leq c_k a_{k+1} + d_k, \quad \forall k > 0$$

根据引理 3.9 我们可知  $a_k$  收敛且有界，则对于任意  $0 < \omega < \xi < 1$ ，可得

$$\lim_{k \rightarrow \infty} \frac{\|t_k\|}{(\theta + \xi)^k} \leq \frac{c_k a_{k+1} + d_k}{(\theta + \xi)^k} = 0$$

因此存在  $p > 0$  使得下式成立

$$\|t_k\| \leq p(\theta + \xi)^k, \quad \forall k \geq 0$$

由此，我们可得到以下结果

$$\|x_k - s^*\| \leq \|x_k - V_\infty x_k\| + \|V_\infty x_k - s^*\| \leq 2p(\theta + \xi)^k$$

即得到定理 3.2 的结果，证明结束。

#### 4. 数值实验

我们考虑有向图上的分布式逻辑回归问题：

$$\tilde{s}^* = \arg \min_{\tilde{s} \in \mathbb{R}^p} \frac{\delta}{2} \|\tilde{s}\|^2 + \sum_{i=1}^n \sum_{j=1}^{m_i} \ln [1 + \exp(-y_{ij} c_{ij}^T \tilde{s})]$$

其中  $(c_{ij}, y_{ij}) \in \mathbb{R}^p \times \{-1, +1\}$ ， $c_{ij}$  为节点  $i$  的第  $j$  个数据样本特征向量，其相应的标签用  $y_{ij}$  表示， $n$  个节点共同处理  $N = \sum_{i=1}^n m_i$  个样本，其中  $m_i$  对应节点  $i$  所需要处理的样本数。若任意节点所需要处理的样本数相同，则上述问题便转化为分布式无约束优化问题：

$$f_i(\tilde{s}) = \frac{\delta}{2n} \|\tilde{s}\|^2 + \sum_{j=1}^{m_i} \ln [1 + \exp(-y_{ij} c_{ij}^T \tilde{s})]$$

其中正则项  $\frac{\delta}{2n} \|\tilde{s}\|^2$  是用以规避数据过拟合。

实验中的初始强连通有向网络由十个节点随机生成，应用规则  $R = I - \frac{1}{2d_{\max}^{\text{in}}} L_R$  生成行随机矩阵，其

中  $d_{\max}^{\text{in}}$  为最大入度,  $L_R$  是有关于矩阵  $R$  的拉普拉斯矩阵。类似地,  $C = I - \frac{1}{2d_{\max}^{\text{out}}} L_C$  用以列随机矩阵的生成。此外, 对于所有的分布式算法, 我们使用  $\{r(x)\}_{k \geq 0}$  来刻画残差序列, 其中  $r(k) = \frac{\|\bar{x}_k - \bar{x}^*\|}{\|\bar{x}_0 - \bar{x}^*\|}$ 。

根据定理 3.1 和 3.2 可看出 ARNH 的收敛速率与步长和参数有关, 对于动量参数对收敛速率的影响, 我们设定步长  $\alpha = 0.004$ , 参数  $b = 130$ , 动量参数的选取为  $\eta \in [0, 0.5]$ , 算法在不同动量参数下第 300 次迭代时的残差变化曲线见图 1, 从中可以看出最优动量参数为 0.3。

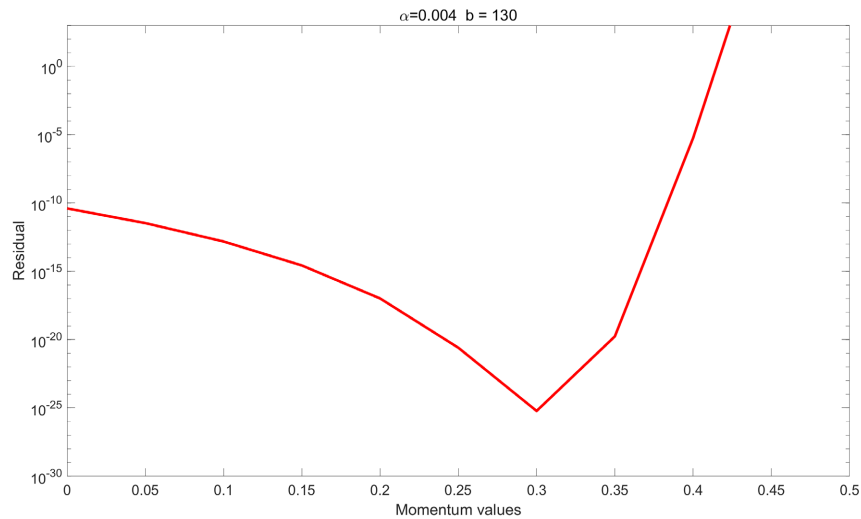


Figure 1. Performance of ARNH algorithm with different momentum parameters  $\eta$

图 1. 不同参数  $\eta$  下 ARNH 算法性能

实验 2 测试参数  $b$  对算法性能的影响, 选取异构步长  $\alpha = 0.004$ , 动量参数  $\eta = 0.3$ , 将带有不同参数  $b$  的算法进行性能比较, 从实验结果图 2 可以观察到 ARSD 在  $b = 130$  时达到最佳收敛。

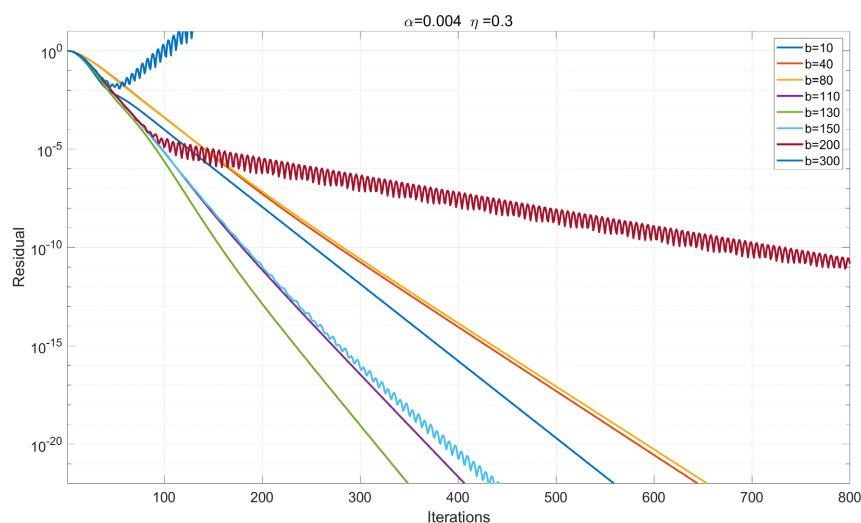


Figure 2. Performance of ARNH algorithm with different parameters  $b$

图 2. 不同参数  $b$  下 ARNH 算法性能

实验 3 设定所提出算法的非协作步长  $\alpha_i \in [0.0035, 0.0045]$ ，动量参数  $\eta_i \in [0.3, 0.35]$ ，参数  $b = 130$ ，并 ARNH 与前沿算法进行比较，所有相关算法的参数已调至最优可获得最佳性能表现，由图 3 可以看出 ARNH 相较于现有的分布式一阶算法有较快的收敛速度。

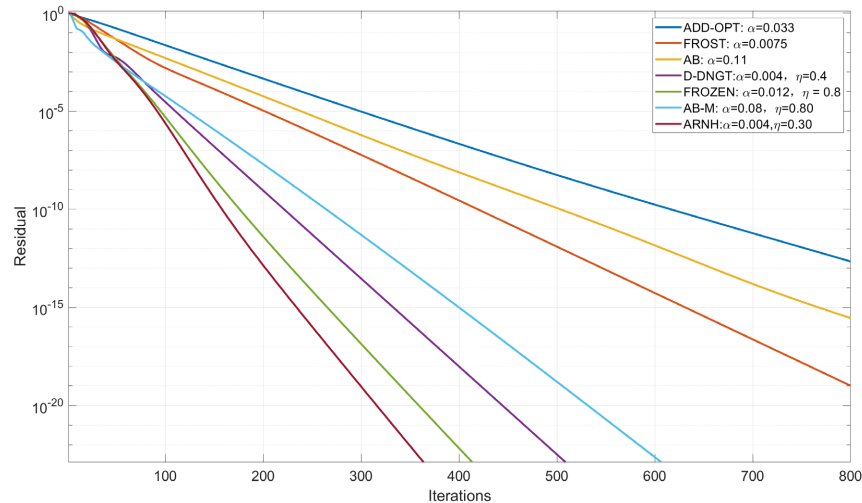


Figure 3. Performance comparison of different algorithms

图 3. 不同算法的性能比较

## 5. 总结

本文提出了一种基于行随机优化的分布式同步动量加速算法，称为 ARNH。该算法利用 Push-Sum 技术消除网络不平衡性，并运用双加速机制提高算法收敛速率。若目标函数强凸且具有 Lipschitz 连续梯度，本文证明当所选步长和动量参数满足取值范围，ARNH 能更快地线性收敛到全局最优解。最后，数值实验验证了理论分析的正确性以及算法的高效性。

## 参考文献

- [1] Meng, M. and Li, X. (2020) Distributed Nonlinear Estimation Over Unbalanced Directed Networks. *IEEE Transactions on Signal Processing*, **68**, 6212-6223. <https://doi.org/10.1109/TSP.2020.3033389>
- [2] Forero, P.A., Cano, A. and Giannakis, G.B. (2010) Consensus-Based Distributed Support Vector Machines. *Journal of Machine Learning Research*, **11**, 1663-1707. <https://doi.org/10.1145/1791212.1791218>
- [3] Raja, H. and Bajwa, W.U. (2016) Cloud K-SVD: A Collaborative Dictionary Learning Algorithm for Big, Distributed Data. *IEEE Transactions on Signal Processing*, **64**, 173-188. <https://doi.org/10.1109/TSP.2015.2472372>
- [4] Lorenzo, P.D. and Sayed, A.H. (2016) Sparse Distributed Learning Based on Diffusion Adaptation. *IEEE Transactions on Signal Processing*, **61**, 1419-1433. <https://doi.org/10.1109/TSP.2012.2232663>
- [5] Wu, D., Yang, T., Stoorvogel, A.A. and Stoustrup, J. (2017) Distributed Optimal Coordination for Distributed Energy Resources in Power Systems. *IEEE Transactions on Automation Science and Engineering*, **14**, 414-424. <https://doi.org/10.1109/TASE.2016.2627006>
- [6] Safavi, S., Khan, U.A., Kar, S. and Moura, J.M.F. (2018) Distributed Localization: A Linear Theory. *Proceedings of the IEEE*, **106**, 1204-1223. <https://doi.org/10.1109/JPROC.2018.2823638>
- [7] Mateos, G., Bazerque, J.A. and Giannakis, G.B. (2010) Distributed Sparse Linear Regression. *IEEE Transactions on Signal Processing*, **58**, 5262-5276. <https://doi.org/10.1109/TSP.2010.2055862>
- [8] Nedić, 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>
- [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. (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>
- [11] Jakovetić, D. (2019) A Unification and Generalization of Exact Distributed First-Order Methods. *IEEE Transactions on Signal and Information Processing over Networks*, **5**, 31-46. <https://doi.org/10.1109/TSIPN.2018.2846183>
- [12] Xi, C., Wu, Q. and Khan, U.A. (2017) On the Distributed Optimization over Directed Networks. *Neurocomputing*, **267**, 508-515. <https://doi.org/10.1016/j.neucom.2017.06.038>
- [13] Kempe, D., Dobra, A. and Gehrke, J. (2003) Gossip-Based Computation of Aggregate Information. *44th Annual IEEE Symposium on Foundations of Computer Science*, Cambridge, 11-14 October 2003, 482-491. <https://doi.org/10.1109/SFCS.2003.1238221>
- [14] Xi, C. and Khan, U.A. (2017) DEXTRA: A Fast Algorithm for Optimization over Directed Graphs. *IEEE Transactions on Signal Processing*, **62**, 4980-4993. <https://doi.org/10.1109/TAC.2017.2672698>
- [15] Nedić, A., Olshevsky, A. and Shi, A. (2017) Achieving Geometric Convergence for Distributed Optimization over Time-Varying Graphs. *SIAM Journal on Optimization*, **27**, 2597-2633. <https://doi.org/10.1137/16M1084316>
- [16] Pu, S., Shi, W., Xu, J. and Nedić, A. (2018) Push-Pull Gradient Methods for Distributed Optimization in Networks. *IEEE Transactions on Automatic Control*, **66**, 1-16. <https://doi.org/10.1109/TAC.2020.2972824>
- [17] Xin, R. and Khan, U.A. (2018) A Linear Algorithm for Optimization over Directed Graphs with Geometric Convergence. *IEEE Control Systems Letters*, **2**, 315-320. <https://doi.org/10.1109/LCSYS.2018.2834316>
- [18] Xi, C., Mai, V.S. and Abed, E.H. (2018) Linear Convergence in Optimization over Directed Graphs with Row-Stochastic Matrices. *IEEE Transactions on Automatic Control*, **63**, 3558-3565. <https://doi.org/10.1109/TAC.2018.2797164>
- [19] Mai, V.S. and Abed, E.H. (2016) Distributed Optimization over Weighted Directed Graphs Using Row Stochastic Matrix. *2016 American Control Conference (ACC)*, Boston, 6-8 July 2016, 7165-7170.
- [20] Xin, R., Xi, C. and Khan, U.A. (2019) FROST—Fast Row-Stochastic Optimization with Uncoordinated Step-Sizes. *EURASIP Journal on Advances in Signal Processing*, **2019**, Article No. 1. <https://doi.org/10.1186/s13634-018-0596-y>
- [21] Xin, R., Jakovetić, D. and Khan, U.A. (2019) Distributed Nesterov Gradient Methods over Arbitrary Graphs. *IEEE Signal Processing Letters*, **26**, 1247-1251. <https://doi.org/10.1109/LSP.2019.2925537>
- [22] Lu, Q., Liao, X., Li, H. and Huang, T. (2021) A Nesterov-Like Gradient Tracking Algorithm for Distributed Optimization over Directed Networks. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, **51**, 6258-6270. <https://doi.org/10.1109/TSMC.2019.2960770>
- [23] Li, H., Wang, J. and Wang, Z. (2019) Row-Stochastic Matrices Based Distributed Optimization Algorithm with Uncoordinated Step-Sizes. *2019 6th International Conference on Information, Cybernetics, and Computational Social Systems (ICCSS)*, Chongqing, 27-30 September 2019, 124-131. <https://doi.org/10.1109/ICCSS48103.2019.9115478>
- [24] Ghaderyan, D., Aybat, N.S., Aguiar, A.P. and Pereira, F.M.L. (2021) A Fast Row-Stochastic Decentralized Optimization Method over Directed Graphs.
- [25] Polyak, B. (1964) Some Methods of Speeding up the Convergence of Iteration Methods. *USSR Computational Mathematics and Mathematical Physics*, **4**, 1-17. [https://doi.org/10.1016/0041-5553\(64\)90137-5](https://doi.org/10.1016/0041-5553(64)90137-5)
- [26] Nesterov, Y. (2013) *Introductory Lectures on Convex Optimization: A Basic Course*. Springer Science & Business Media, Berlin, 87.