Bihao SUN, Jinhui HU, Dawen XIA, Huaqing LI, 2021. A distributed stochastic optimization algorithm with gradient-tracking and distributed heavy-ball acceleration. *Frontiers of Information Technology & Electronic Engineering*, 22(11):1463-1476. https://doi.org/10.1631/FITEE.2000615

A distributed stochastic optimization algorithm with gradient-tracking and distributed heavy-ball acceleration

Key words: Distributed optimization; High-performance algorithm; Multi-agent system; Machine-learning problem; Stochastic gradient

Corresponding author: Huaqing LI
E-mail: huaqingli@swu.edu.cn
ORCID: https://orcid.org/0000-0001-6310-8965

Motivation

1. In distributed settings, distributed algorithms with exact gradient need massive calculation.
2. To increase the convergence rate of existing distributed stochastic first-order gradient methods, a momentum term is combined with a gradient-tracking technique.
3. Deep learning methods show great potential in computer-aided diagnosis.

Main idea

1. The momentum term is combined with a gradient-tracking technique to achieve an increased convergence rate over undirected networks.

2. The random average gradient technology is introduced in the proposed algorithm, to reduce the cost of calculating the random gradient at each iteration.

Method

Local update of solution:

$$\boldsymbol{x}_{k+1}^{i}=\sum_{j=1}^{m} w_{ij}\boldsymbol{x}_{k}^{j}-\alpha\boldsymbol{y}_{k}^{i}+\beta\left(\boldsymbol{x}_{k}^{j}-\boldsymbol{x}_{k-1}^{j}\right).$$

Local update of random gradient:

$$\boldsymbol{g}_{k+1}^{i}=\nabla f_{i}^{t_{k+1}^{i}}\left(\boldsymbol{x}_{k+1}^{i}\right)-\nabla f_{i}^{t_{k+1}^{i}}\left(\boldsymbol{v}_{k+1}^{i,t_{k+1}^{i}}\right)+\frac{1}{p_{i}}\sum\nolimits_{j=1}^{p_{i}}\nabla f_{i}^{j}\left(\boldsymbol{v}_{k+1}^{i,j}\right).$$

Local update of gradient tracking:

$$\boldsymbol{y}_{k+1}^{i}=\sum_{j=1}^{m} w_{ij}\boldsymbol{y}_{k}^{j}+\boldsymbol{g}_{k+1}^{i}-\boldsymbol{g}_{k}^{i}.$$

α is the step size of the algorithm, β is the momentum parameter, and w_{ij} is the element on row i and column j of the double random weight matrix $\boldsymbol{W}$.

Method (Cont'd)

Linear convergence condition:

$$\alpha < \bar{\alpha} \triangleq \min\left\{ \frac{\left(1-\sigma^2\right)^2}{380}\frac{\sqrt{p}}{\sqrt{PQ}}, \frac{1-\sigma^2}{12\sqrt{2087}}\frac{\sqrt{p}}{\sqrt{PQL_f}}, \frac{\sqrt{31}\left(1-\sigma^2\right)}{380\sqrt{3}L_f}, \frac{m}{QL_f}\frac{p}{164P} \right\},$$

$$\beta < \bar{\beta} \triangleq \min\left\{ \frac{3\sqrt{47}\left(1-\sigma^2\right)}{56\sqrt{34}}\frac{\sqrt{p}}{\sqrt{P}}, \frac{\sqrt{1259}\left(1-\sigma^2\right)}{280\sqrt{17}}\frac{\sqrt{p}}{\sqrt{P}}, \frac{\sqrt{PQ}}{\sqrt{p}}\frac{\sqrt{59}}{\sqrt{1122}}, \frac{\left(1-\sigma^2\right)^2\sqrt{69}}{560\sqrt{68}}\frac{p}{PQ} \right\},$$

where $P := \max_i \{p_i\}$, $p := \min_i \{p_i\}$, L_f is a Lipschitz constant, and μ is a strong convexity parameter.

$$Q = L_f / \mu, \qquad \sigma = \left\| W - \frac{1}{m}\mathbf{1}_m\mathbf{1}_m^{\mathrm{T}} \right\| < 1.$$

Major results

1. Experimental results over logistic regression

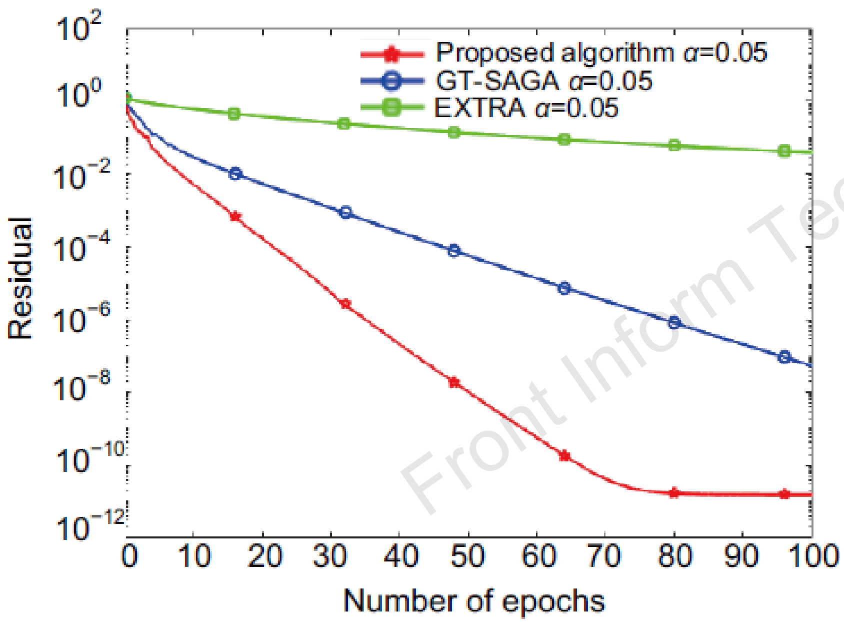

Fig. 2 Logistic regression over an undirected network

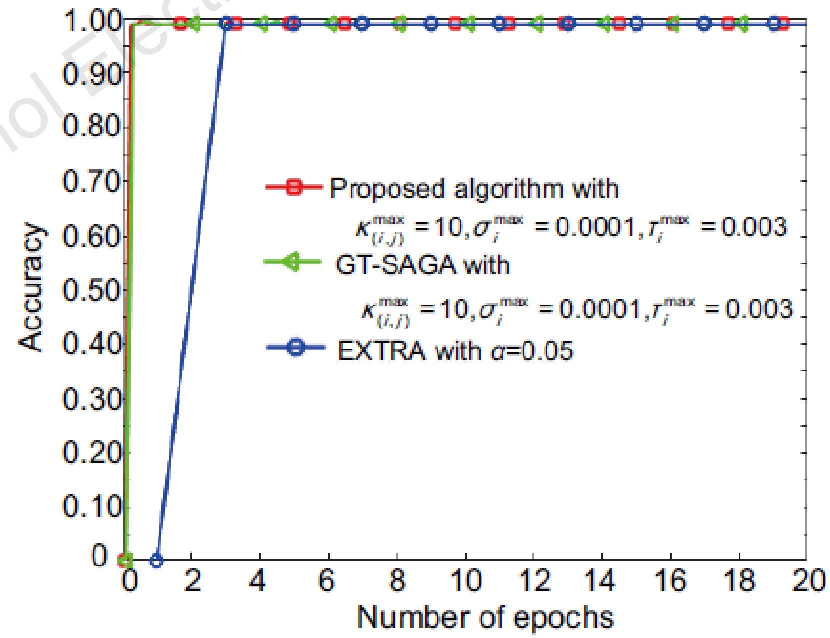

Fig. 3 Comparison of accuracy performance

Major results (Cont'd)

2. Experimental results over the least-square method of distributed signal and distributed quadratic programming processing

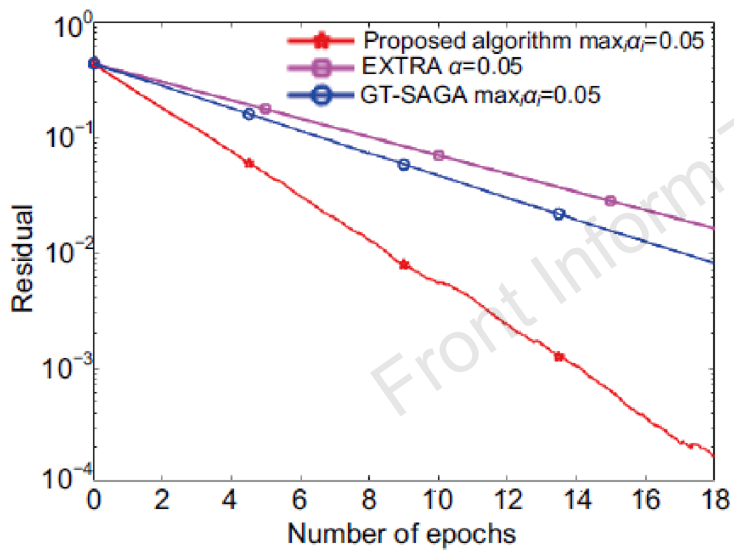

Fig. 4 Performance comparison over the least-squares method of distributed signal processing

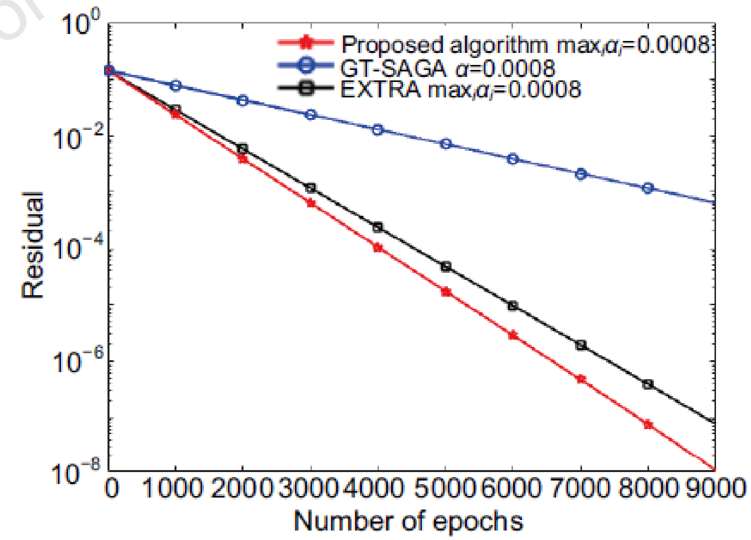

Fig. 5 Performance comparison over an undirected network when condition number $Q = 500$

Major results (Cont'd)

3. Test results of different experimental accuracy

Table 1 Convergence performance comparison over logistic regression

Algorithm	Number of epochs		
	Accuracy=10^{-2}	10^{-4}	10^{-6}
Proposed algorithm	8	22	35
GT-SAGA	16	46	79
EXTRA	193	549	935

Table 2 Convergence performance comparison over least-squares

Algorithm	Number of epochs	
	Accuracy=10^{-1}	10^{-2}
Proposed algorithm	3	18
GT-SAGA	6	8
EXTRA	7	20

Table 3 Convergence performance comparison over the distributed quadratic programming

Algorithm	Number of epochs	
	Accuracy=10^{-1}	10^{-2}
Proposed algorithm	1472	2747
GT-SAGA	4412	8237
EXTRA	1657	3097

Conclusions

1. A distributed stochastic algorithm was presented which is capable of solving large-scale optimization problems over undirected networks.

2. It was shown that the proposed algorithm achieves accelerated linear convergence with a constant step size.

3. Extensive experiments on real-world datasets illustrated that the performance of the proposed algorithm is superior to those of other comparable algorithms.

Huaqing LI received his BS degree in Information and Computing Science in 2009 from Chongqing University of Posts and Telecommunications, Chongqing, China, and his PhD degree in Computer Science and Technology in 2013 from Chongqing University. From Sept. 2014 to Sept. 2015, he was a postdoctoral researcher at the School of Electrical and Information Engineering, The University of Sydney, Australia. From Nov. 2015 to Nov. 2016, he was a postdoctoral researcher at the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. He is currently a professor at the College of Electronic and Information Engineering, Southwest University, Chongqing, China. His main research interests include nonlinear dynamics and control, multi-agent system, and distributed optimization. Prof. LI currently serves as a regional editor for *Neur Comput Appl*, an editorial board member for *IEEE Access*, and a corresponding expert for *Front Inform Technol Electron Eng*.