SCIENCE CHINA
Information Sciences

• LETTER •

September 2023, Vol. 66 199202:1–199202:2
https://doi.org/10.1007/s11432-022-3637-y

A zeroth-order algorithm for distributed optimization with stochastic stripe observations

Yinghui WANG[1], Xianlin ZENG[2*], Wenxiao ZHAO[3] & Yiguang HONG[4]

[1]*School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing* 100083*, China;*
[2]*Key Laboratory of Intelligent Control and Decision of Complex Systems, School of Automation, Beijing Institute of Technology, Beijing* 100081*, China;*
[3]*Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing* 100190*, China;*
[4]*Department of Control Science and Engineering, Tongji University, Shanghai* 201804*, China*

Received 3 May 2022/Revised 23 August 2022/Accepted 22 October 2022/Published online 14 April 2023

Citation Wang Y H, Zeng X L, Zhao W X, et al. A zeroth-order algorithm for distributed optimization with stochastic stripe observations. Sci China Inf Sci, 2023, 66(9): 199202, https://doi.org/10.1007/s11432-022-3637-y

Recent years have seen a rising interest in distributed optimization problems because of their widespread applications in power grids, multi-robot control, and regression learning. Over the last few decades, many distributed algorithms have been developed for tackling distributed optimization problems. In these algorithms, agents over the network only have access to their own local functions and exchange information with their neighbors.

Moreover, observations in image processing and economic dispatch are often captured with stochastic stripe observations. However, most of these mentioned studies are centralized. Thus far, solving optimization models with stochastic stripe observations for distributed networks remains an underexplored topic.

This motivates us to investigate the distributed optimization problem with stochastic stripe observations. Challenges of the problem come from the uncertain stripe observations of local functions because the observations are not equipped with exact analytic expressions. Therefore, we propose a distributed zeroth-order algorithm to determine approximate solutions of distributed optimization problems with stochastic stripe observations and perform an algorithm analysis.

Problem formulation and algorithm analysis. In this study, we investigate the following distributed problem:

$$\begin{aligned} &\min_{x} \ f(\boldsymbol{x}), \quad f(\boldsymbol{x})=\sum_{i=1}^{n} f_i(x_i), \\ &\text{s.t.} \quad x_i=x_j, \quad x\in X, \end{aligned} \tag{1}$$

where $f_i:\mathcal{R}^p\to\mathcal{R}$ is a convex function. In our problem setup, agents over the network only have access to observations of local functions with stochastic stripe observations $[Y_{L_i}(x_i),Y_{R_i}(x_i)]$ and global constraint X.

In our setting, each agent can observe the inexact local objective function $[Y_{L_i}(x_i),Y_{R_i}(x_i)]$ with stochastic stripe noises. $[Y_{L_i}(x_i),Y_{R_i}(x_i)]$ provides a stochastic inexact observation pair of local objective function $f_i(x_i)$ with $Y_{L_i}(x_i)\leqslant Y_{R_i}(x_i)$. For each point x_i, different stochastic stripe observations $[Y_{L_i}(x_i)_1,Y_{R_i}(x_i)_1]$, $[Y_{L_i}(x_i)_2,Y_{R_i}(x_i)_2]$, $[Y_{L_i}(x_i)_3,Y_{R_i}(x_i)_3]$, ... could be made. We only use the information of one stochastic observation pair $[Y_{L_i}(x_i),Y_{R_i}(x_i)]$ in this study.

The following is an approximate expression of problem (1):

$$\begin{aligned} &\min_{x} \ g(\boldsymbol{x},\boldsymbol{\lambda}), \quad g(\boldsymbol{x},\boldsymbol{\lambda})=\sum_{i=1}^{n} g_i(x_i,\lambda_i), \\ &\text{s.t.} \quad x_i=x_j, \quad x\in X, \quad \forall i\in\mathcal{N}, \\ &\lambda_i=\lambda_j, \end{aligned} \tag{2}$$

where $\boldsymbol{\lambda}=[\lambda_1,\ldots,\lambda_n]^{\mathrm{T}}$ with $\lambda_i\in(0,1)$. $g(\boldsymbol{x},\boldsymbol{\lambda})\triangleq\sum_{i=1}^{n} g_i(x_i,\lambda_i)$ and $g_i(x_i,\lambda_i)\triangleq\lambda_i f_{L_i}(x_i)+(1-\lambda_i)f_{R_i}(x_i)$.

Agents have different observations in different distributed stochastic settings. In the first-order settings [1, 2], agents over the network have access to inexact (sub)gradient information $\nabla f_i(x_i)+\epsilon_i(x_i)$, whereas, in distributed zeroth-order settings [3–6], agents over the network obtain information of $f_i(x_i)+\epsilon_i$. In these earlier settings, $\epsilon_i(x_i)$s are the martingale-difference noise with zeroth means and bounded variances. In our setting, agents have access to stochastic stripe observations $[Y_{L_i}(x_i),Y_{R_i}(x_i)]([f_i(x_i)+\epsilon_i,f_i(x_i)+\iota_i])$.

To solve problem (1) with stochastic stripe observations $[Y_{L_i}(x),Y_{R_i}(x)]$, zeroth-order designs, which only make use of function information over the network, are considered in this study.

Distributed zeroth-order algorithm is given in Algorithm 1.

$\{\triangle_i^q(k)\}_{k\geqslant 0}$, $q=1,2,\ldots,p$, $k=1,2,\ldots$ is a sequence of mutually independent and identically distributed ran-

* Corresponding author (email: xianlin.zeng@bit.edu.cn)

Algorithm 1 Distributed zeroth-order algorithm

Input: Total numbers of iteration T, step-size sequence $\{\iota(k)\}$, parameter sequence $\{c(k)\}$, and random variables $\{\triangle_i(k)\}$.
Initialize: $x_i(0) \in X$ for all $i = 1, 2, \ldots n$.
while $k = 0, \ldots T$, for $i \in \mathcal{N}$ **do**
Average of random differences $x_i(k)$: $\xi_i(k) = \sum_{j=1}^{n} w_{ij}(k)x_j(k)$;
Calculation of local measurement $d_i(k)$: $d_i(k) = \dfrac{[y_i^+(k) - y_i^-(k)] \triangle_i^-(k)}{2c(k)}$;
Descent step: $x_i(k) = P_{\xi_i(k)\in X}(\xi_i(k) - \iota(k)d_i(k))$;
Average of local observations $\lambda_i(k)$: $\lambda_i(k+1) = \sum_{j=1}^{n} w_{ij}(k)\lambda_i(k)$;
end while
return $(\boldsymbol{x}(T), \boldsymbol{\lambda}(T))$.

dom variables with zero mean, $\triangle_i(k) = [\triangle_i^1(k), \triangle_i^2(k), \ldots, \triangle_i^p(k)]^{\mathrm{T}}$, and $\triangle_i^-(k) = [\frac{1}{\triangle_i^1(k)}, \frac{1}{\triangle_i^2(k)}, \ldots, \frac{1}{\triangle_i^p(k)}]^{\mathrm{T}}$.

$y_i^+(k)$ and $y_i^-(k)$ are given as

$$\begin{aligned} y_i^+(k) &= \lambda_i(k)Y_{L_i}\Big(\xi_i(k) + c(k)\triangle_i(k)\Big) \\ &\quad + \big(1-\lambda_i(k)\big)R_{L_i}\Big(\xi_i(k) + c(k)\triangle_i(k)\Big), \\ y_i^-(k) &= \lambda_i(k)Y_{L_i}\Big(\xi_i(k) - c(k)\triangle_i(k)\Big) \\ &\quad + \big(1-\lambda_i(k)\big)R_{L_i}\Big(\xi_i(k) - c(k)\triangle_i(k)\Big). \end{aligned}$$

$\{c(k)\}$ and $\{\iota(k)\}$ are two nonnegative step-size sequences that gradually trend to zero with

$$\iota(k) = \frac{1}{k^{1-\epsilon}}, \quad c(k) = \frac{1}{k^{\delta}}, \quad 0 \leqslant \epsilon < \frac{1}{4}, \quad \frac{1}{2} - \epsilon > \delta > \epsilon.$$

Results of Algorithm 1. We first show the almost-sure convergence and convergence rate of $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ to problem (2). For problem (1), we provide the asymptotic behavior of the approximate solution $\boldsymbol{x}^*$ of problem (1). The assumptions and parameter selection of Algorithm 1 are given in Appendix A.

Let $\lambda(0) = \frac{1}{n}\sum_{i=1}^{n}\lambda_i(0)$. The almost-sure convergence and convergence rate of $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ to problem (2) are given in Theorems 1 and 2, whose proofs are similar to [4] and are omitted in this study.

Theorem 1. With Assumptions 1 and 2,

(a) the sequence $\{\lambda_i(k)\}$, $i \in \mathcal{N}$ generated by Algorithm 1 converges to the same point $\lambda(0)$;

(b) the sequence $\{x_i(k)\}$, $i \in \mathcal{N}$ generated by Algorithm 1 reaches a consensus almost surely, i.e., $\lim_{k\to\infty}\|x_i(k) - x_j(k)\| = 0$, almost surely;

(c) the sequence $\{x_i(k)\}$, $i \in \mathcal{N}$ generated by Algorithm 1 converges to the same point x^* almost surely.

Theorem 2. With Assumptions 1 and 2, we have $R(T) \sim O(\frac{1}{T^{\epsilon}})$.

Subsequently, we consider the relationship between the approximate solution $\big(x^*, \lambda(0)\big)$ given by Algorithm 1 and the exact solution $\hat{x}^*$ of problem (1) in terms of network scale.

Theorem 3. (a) With Assumptions 1–3 and 4(a), $\|g\big(\boldsymbol{x}^*, \boldsymbol{\lambda}(0)\big) - g\big(\hat{\boldsymbol{x}}^*, \boldsymbol{\lambda}_0\big)\| = 0$.

(b) With Assumptions 1–3 and 4(b), $\lim_{n\to\infty} P(\|g(\boldsymbol{x}^*, \boldsymbol{\lambda}(0)) - g\big(\hat{\boldsymbol{x}}^*, \boldsymbol{\lambda}_0\big)\| < \epsilon c_0(\hat{x}^*)) = 1$, where $c_0\big(\hat{x}^*\big) = \|f_L\big(\hat{x}^*\big) - f_R\big(\hat{x}^*\big)\|$.

(c) With Assumptions 1–3 and 4(c), $\lim_{n\to\infty} P(\|g\big(\boldsymbol{x}^*, \boldsymbol{\lambda}(0)\big) - g\big(\hat{\boldsymbol{x}}^*, \boldsymbol{\lambda}_0\big)\| \leqslant ac_0\big(\hat{x}^*\big)) \geqslant 2\phi(\frac{\sqrt{n}a}{\nu}) - 1$, and $\lim_{n\to\infty} P(\|g\big(\boldsymbol{x}^*, \boldsymbol{\lambda}(0)\big) - g\big(\hat{\boldsymbol{x}}^*, \boldsymbol{\lambda}_0\big)\| \geqslant bc_0\big(\hat{x}^*\big)) \leqslant 2 - 2\phi(\frac{\sqrt{n}b}{\nu})$, where $\phi(\cdot)$ is the cumulative distribution function of Gaussian distribution $\mathcal{N}(0,1)$.

Conclusion. This study investigated a distributed optimization problem with stochastic stripe observations subject to local convex constraints. A distributed stochastic zeroth-order algorithm was developed to determine an approximate solution to the proposed problem. Moreover, the almost-sure convergence and convergence rate were obtained for an approximate solution to the problem, and the asymptotic behavior of the approximate solution was further discussed.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 72101026, 61621063), State Key Laboratory of Intelligent Control and Decision of Complex Systems, and Fundamental Research Funds for the China Central Universities of USTB (Grant No. FRF-TP-22-141A1).

Supporting information Appendixes A and B. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Ram S S, Nedić A, Veeravalli V V. Distributed stochastic subgradient projection algorithms for convex optimization. J Optim Theor Appl, 2010, 147: 516–545
2 Wang Y H, Lin P, Hong Y G. Distributed regression estimation with incomplete data in multi-agent networks. Sci China Inf Sci, 2018, 61: 092202
3 Anit K S, Dusan J, Dragana B, et al. Distributed zeroth order optimization over random networks: a Kiefer-Wolfowitz stochastic approximation approach. In: Proceedings of IEEE Conference on Decision and Control (CDC), 2018
4 Wang Y, Zhao W, Hong Y, et al. Distributed subgradient-free stochastic optimization algorithm for nonsmooth convex functions over time-varying networks. SIAM J Control Optim, 2019, 57: 2821–2842
5 Yuan D, Ho D W C, Xu S. Zeroth-order method for distributed optimization with approximate projections. IEEE Trans Neural Netw Learn Syst, 2015, 27: 284–294
6 Yuan D M, Ho D W C. Randomized gradient-free method for multiagent optimization over time-varying networks. IEEE Trans Neural Netw Learn Syst, 2014, 26: 1342–1347