REPORT DOCUMENTATION PAGE	Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY) 03-12-2022	2. REPORT TYPE Final Report	3. DATES COVERED (From - To) 15-Jun-2018 - 14-Feb-2021

4. TITLE AND SUBTITLE Final Report: Optimization for Distributed Machine Learning	5a. CONTRACT NUMBER W911NF-18-1-0223
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER 611102
6. AUTHORS	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES Georgia Tech Research Corporation 505 Tenth Street NW Atlanta, GA 30332 -0420	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211	10. SPONSOR/MONITOR'S ACRONYM(S) ARO
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) 72672-MI.1

12. DISTRIBUTION AVAILIBILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			George Lan
UU	UU	UU	UU		19b. TELEPHONE NUMBER 404-385-5402

Standard Form 298 (Rev 8/98)
Prescribed by ANSI Std. Z39.18

RPPR Final Report
as of 07-Dec-2022

Agency Code: 21XD

Proposal Number: 72672MI

Agreement Number: W911NF-18-1-0223

INVESTIGATOR(S):

Name: George Lan
Email: gl68@gatech.edu
Phone Number: 4043855402
Name: George Lan
Email: gl68@gatech.edu
Phone Number: 4043855402
Principal: Y

Organization: **Georgia Tech Research Corporation**
Address: 505 Tenth Street NW, Atlanta, GA 303320420
Country: USA
DUNS Number: 097394084
EIN: 580603146
Report Date: 14-May-2021
Date Received: 03-Dec-2022
Final Report for Period Beginning 15-Jun-2018 and Ending 14-Feb-2021
Title: Optimization for Distributed Machine Learning
Begin Performance Period: 15-Jun-2018
End Performance Period: 14-Feb-2021
Report Term: 0-Other
Submitted By: George Lan
Email: gl68@gatech.edu
Phone: (404) 385-5402
Distribution Statement: 1-Approved for public release; distribution is unlimited.

STEM Degrees:

STEM Participants:

Major Goals: The major goals of this project consist of the following aspect.
(a) Design communication-efficient algorithms over general complex network to enable decentralize machine learning.
(b) Design asynchronous and randomized methods over a star network to enable federated learning.
(c) Design efficient algorithms to enable robust learning.

Army relevance: to learn from data collected by agents (e.g., airfighters, autonomous vehicles) distributed over networks.

To achieve these goals, we pursued the following approaches.
(a) Propose algorithms with optimal communication complexity for decentralize/federated learning.
(b) Propose algorithms with graph topology independent gradient and sampling complexity.
(c) Propose novel algorithms for solving minmax and general variational inequalities from robust learning.

The Scientific Barriers Exist in the lack of understanding about the interaction among communication of agents, computation performed by agents, data (samples) collected by agents, and the impact to solution accuracy. Removing these barriers will enable the design of efficient algorithms for learning over networks in terms of all three aspects, i.e., communication, computation and sampling.

Accomplishments: Our accomplishments are listed as follows.
(a) Developed novel decentralized communication sliding methods that can judiciously skip communication.
(b) Developed graph topology independent method for decentralized and stochastic optimization.
(c) Developed accelerated stochastic algorithm for nonconvex distributed finite-sum and multiblock optimization.
(d) Developed new optimal methods for robust optimization and machine learning.

The major conclusions include:
(a) Communication complexity for stochastic optimization can be as small as for deterministic problems.
(b) Gradient and sampling complexity can be independent of graph topology.

Significance of our findings:

(a) Developed deep understanding about the interaction of communication, computation and sampling for distributed machine learning.
(b) Proposed novel and theoretically optimal algorithms for distributed machine learning.
(c) Lead to more accurate and faster solution methods for online learning, optimization and control for distributed systems.

Training Opportunities: This project provide partial support to six Ph.D. students and one master student.

Results Dissemination: Research results have been disseminated through numerous invited seminars as listed in the final report.

Honors and Awards: PI Lan gave a plenary talk titled “Stochastic Optimization Algorithms for Machine Learning” at the XV International Conference on Stochastic Programming (ICSP) in 2019. ICSP is the flag-ship conference for stochastic programming society and is held triennially.

Protocol Activity Status:

Technology Transfer: This project supports basic research and does not involve technology transfer

PARTICIPANTS:

Participant Type: Graduate Student (research assistant)
Participant: Zhe Zhang
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Georgios Kotalis
Person Months Worked: 12.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Yi Zhou
Person Months Worked: 10.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Alejandro Carderera
Person Months Worked: 10.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Cyrille Combettes
Person Months Worked: 10.00 **Funding Support:**
Project Contribution:

National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Yan Li
Person Months Worked: 5.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Yifei Luo
Person Months Worked: 3.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Partners

,

I certify that the information in the report is complete and accurate:
Signature: Guanghui Lan
Signature Date: 12/3/22 6:01PM

Report for ARO Project (W911NF-18-1-0223)
"Optimization for Distributed Machine Learning"

Guanghui (George) Lan and Sebastian Pokutta

We report our research progresses made from June 2017 to Feburary 2021 for the ARO project entitled "Optimization for Distributed Machine Learning".

1 Research objectives and problems

This project intends to significantly advance the design and analysis of optimization algorithms for distributed machine learning. As outlined in the research proposal, our research is being pursued along the following three thrusts.

- Decentralized learning over a general complex network. In this research thrust, we consider a general network topology and focus on addressing the following research problems:
  - P1. how to design efficient algorithms for asynchronous distributed machine learning?
  - P2. how to perform distributed stochastic optimization and learning over time varying networks?
- Federated learning over a distributed network. In this research thrust, we consider networks with a special but important structure, i.e., star network topology, and focus on the following research problem:
  - P3. How to design asynchronous and randomized methods for distributed and federated learning?
- Differentially private (distributed) and resilient machine learning. In this research thrust, we focus on addressing the privacy and robustness issues that are common for both decentralized and federated machine learning. We intend to answer the following research question:
  - P4. Can we define efficient differentially private and/or robust convex optimization algorithms for the distributed setting?

2 Research achievements

I. Research Thrust I - Decentralized Learning Year 1: Existing optimization theory mostly concerns the number of required numerical operations performed on CPUs. However, in a distributed setting, communication among agents can incur a much more significant latency. High performance CPUs in these days can read and write the memory at 10 - 100 GB per second whereas communication over TCP/IP is about 100 MB per second. Therefore, the gap between intra-node computation and inter-node communication is about 3 orders of magnitude. It is of paramount importance to study communication-efficient optimization methods for solving different types of machine learning problems over different types of network topologies. In our recent paper entitled "Communication-Efficient Algorithms for Decentralized and Stochas- tic Optimization", we present a new class of decentralized first-order methods for nonsmooth and stochastic optimization problems defined over multiagent networks. Considering that communication is a major bottleneck in decentralized optimization, our main goal is to develop algorithmic frameworks which can significantly reduce the number of inter-node communications. In comparison with existing results for decentralized nonsmooth and stochastic optimization, we can reduce the total number of inter-node communication rounds by orders of magnitude while still maintaining the optimal complexity bounds on intra-node stochastic subgradient evaluations. The bounds on the (stochastic) subgradient evaluations are actually comparable to those required for centralized nonsmooth and stochastic optimization under certain conditions on the target accuracy. This work served as our preliminary research for this project and has been formally accepted by Mathematical Programming (one of the very top journals in the optimization area) in 2018.

Year 2: In a more recent work entitled "Asynchronous decentralized accelerated stochastic gradient descent" (arXiv preprint:1809.09258), we introduced an asynchronous decentralized accelerated stochas-tic gradient descent type of method for decentralized stochastic optimization. We establish $\mathcal{O}(1/\epsilon)$ (resp., $\mathcal{O}(1/\sqrt{\epsilon})$ communication complexity and $\mathcal{O}(1/\epsilon^2)$ (resp., $\mathcal{O}(1/\epsilon)$) sampling complexity for solving general convex (resp., strongly convex) problems. This work generalizes our earlier work on accelerated stochastic gradient descent (or SGD with momentum) under the decentralized setting and shows that one can still save communication costs for this widely used algorithm. In addition, it helps to answer our Research Problem P1 regarding the design of efficient algorithms for asynchronous distributed machine learning. This paper has been accepted by IEEE Journal on Selected Areas in Information Theory.

Year 3: We investigate save communication costs and local computation for smooth and/or strongly convex stochastic optimization under the decentralized setting. We proposed a novel decentralized algorithm that can require the least possible communication rounds, while using the smallest possible number of stochastic gradient computation. Another distinctive feature of this algorithm is that it allows the agents to collect local stochastic samples (or observations) and to communicate with their neighbors over the network simultaneously. The research results are summarized in the paper entitled "Graph topology invariant gradient and sampling complexity for decentralized and stochastic optimization", currently under revision for SIAM Journal on Optimization, a very top journal in the area of optimization.

The investigation of these questions helped us to address Research Problems P1 and P2.

II. Research Thrust II - Federated Learning.

In federated learning, we consider a class of finite-sum optimization problems defined over a distributed multiagent network with m agents connected to a central server. In particular, the objective function consists of the average of $m(\geq 1)$ smooth components associated with each network agent together with some regularization terms.

Year 1: In our work entitled "Random gradient extrapolation for distributed and stochastic optimization", we developed a new randomized incremental gradient algorithm, namely random gradient extrapolation method (RGEM), which does not require any exact gradient evaluation even for the initial point, but can achieve the optimal $\mathcal{O}(\log(1/\epsilon))$ complexity bound in terms of the total number of gradient evaluations of component functions to solve the finite-sum problems. Furthermore, we demonstrate that for stochastic finite-sum optimization problems, RGEM maintains the optimal $\mathcal{O}(1/\epsilon)$ complexity (up to a certain logarithmic factor) in terms of the number of stochastic gradient computations, but attains an $\mathcal{O}(\log(1/\epsilon))$ complexity in terms of communication rounds (each round involves only one agent). It is worth noting that the former bound is independent of the number of agents m, while the latter one only linearly depends on m. To the best of our knowledge, this is the first time that these complexity bounds have been obtained for distributed and stochastic optimization problems. This work, which also served as a preliminary research for this project, has been accepted by SIAM Journal on Optimization in 2018.

Following this work, we studied new stochastic methods for solving two important classes of nonconvex optimization problems. We first introduced a randomized accelerated proximal gradient (RapGrad) method for solving a class of nonconvex distributed optimization problems and show that it can significantly reduce the number of gradient computations and hence communication costs especially when the condition number L/μ (i.e., the ratio between the Lipschitz constant and negative curvature) is large. Inspired by RapGrad, we also developed a new randomized accelerated proximal dual (RapDual) method for solving a class of multi-block nonconvex optimization problems coupled with linear constraints and some special structural properties. We demonstrated that RapDual can also save up to a factor of $\mathcal{O}(\sqrt{m})$ block updates than its batch counterpart, where m denotes the number of blocks. To the best of our knowledge, all these complexity results associated with RapGrad and RapDual seem to be new in the literature. Our paper entitled “Accelerated Stochastic Algorithms for Nonconvex Finite-sum and Multi-block Optimization”, which summarizes the aforementioned contributions, has been accpted by SIAM Journal on Optimization.

Year 2: With the support of this ARO grant, we also explored the design and analysis for stochastic second methods that can further save communication costs for distributed learning. In our paper entitled “Stochastic Variance-Reduced Cubic Regularization for Nonconvex Optimization”, we established the best-known so far rate of convergence for stochastic cubic reguarized Newton method. This work has been accepted by AISTATS in 2018. Moreover, in “A note on inexact gradient and Hessian conditions for cubic regularized Newton’s method”, accepted by Operations Research Letters in 2019, we resolved an important issue regarding how to adaptively implement this type of method.

For the convex setting, we designed a unified optimal variance-reduced accelerated gradient method that is not only optimal for strongly convex problems, but also for convex problems without strongly convexity. This work addressed a few open problems in the area of variance reduced methods for distributed optimization and federated learning. Our paper entitled ”A unified variance-reduced accelerated gradient method for convex optimization” has been accepted to NeuIPS 2019, a very top conference in the area of machine learning.

The investigation of these questions helps us to address Research Problems P3.

III. Research Thrust III - Robustness.

Privacy and robustness issues will be studied, naturally after we presented the basic schemes for decentralized and federated learning. Therefore, the Research Problem P4 will be one of the focuses of our studies in Year 2 of this project. More studies on privacy issue were expected to be carried out by Dr. Sebastian Pokutta, who was initially a co-PI of this project. Since he moved to Germany, our focus has been shifted more to robustness and adversary issues in machine learning.

Year 2: Regarding the robustness issue, we investigated a class of distributionally robust optimization problems where distribution of data is only partially known (or the adversary may maliciously manipulate the distribution of data). We are investigating the design of efficient algorithms that can be used to solve this type of problem, especially under a distributed setting. We designed a few efficient algorithms, including sequential dual method and sequential smoothing level method, that can be used to solve this type of problem, especially under a distributed setting. Our paper entitled "Efficient Algorithms for Distributionally Robust Stochastic Optimization with Discrete Scenario Support" has been accepted by SIAM Journal on Optimization.

Year 3: We further studied general variational inequality problems that play a significant role in robustness and adversary concerns. We first present a novel operator extrapolation (OE) method for solving deterministic variational inequality (VI) problems. Similar to the gradient (operator) projection method, OE updates one single search sequence by solving a single projection subproblem in each iteration. We show that OE can achieve the optimal rate of convergence for solving a variety of VI problems in a much simpler way than existing approaches. We then introduce the stochastic operator extrapolation (SOE) method and establish its optimal convergence behavior for solving different stochastic VI problems. In particular, SOE achieves the optimal complexity for solving a fundamental problem, i.e., stochastic smooth and strongly monotone VI, for the first time in the literature. We also present a stochastic block operator extrapolations (SBOE) method to further reduce the iteration cost for the OE method applied to large-scale deterministic VIs with a certain block structure. Numerical experiments have been conducted to demonstrate the potential advantages of the proposed algorithms. In fact, all these algorithms are applied to solve generalized monotone variational inequality (GMVI) problems whose operator is not necessarily monotone. This paper has been accepted by SIAM Journal on Optimization.

3 Research Training

Ph.D. students Alejandro Carderera, Cyrille Combettes, Georgios Kotalis, Yan Li, Zhe Zhang, and Yi Zhou, and one master student Yifei Luo are involved in this project.

4 Dissemination

PI Lan has finished a monograph titled "First-order and stochastic optimization methods for machine learning", published by Springer-Nature in May 2020. Chapter 4 and chapter 8 of this book discussed distributed and decentralized optimization learning, both related to this ARO project.

Numerous invited seminars have been given to a broad audience. These seminars are listed below.

Year 1: a) G. Lan, invited speaker, “Random Gradient Extrapolation for Distributed and Stochastic Op- timization”, International Symposium on Mathematical Programming, Bordeaux, France, July 3, 2018. b) S. Pokutta, invited speaker, “Blended Conditional Gradients”, Oberwolfach, November, 2018 Conference presentations. c) S. Pokutta, plenary speaker, “Smooth Constraint Convex Minimization via Conditional Gradients”, INFORMS Computing Society Conference, January 2019 d) G. Lan, invited speaker, “Accelerated Stochastic Algorithms for Nonconvex Finite-sum Optimization”, Workshop on Bridging Mathematical Optimization, Information Theory, and Data Science, Princeton University, May 15, 2018. e) G. Lan, minicourse speaker (3 hours), “Stochastic Optimization Algorithms for Machine Learn- ing”, CIMI Workshop on Optimization and Learning, Toulouse, France, September 10-13, 2018. f) G. Lan, invited speaker, “Accelerated Stochastic Algorithms for

Nonconvex Finite-sum and Multi-block Optimization", Workshop on theoretical Foundation of Deep Learning, Georgia Institute of Technology, October 7, 2018. g) G. Lan, invited speaker, "Stochastic Optimization for Learning over Networks", Department of Industrial Engineering and Management Sciences, Northwestern University, November 29, 2018. h) G. Lan invited speaker, East Coast Optimization Meeting 2019, April 4-5, 2019 Department of Mathematical Sciences, George Mason University, Fairfax, Virginia.

Year 2: a) Plenary speaker (1 hour), "Stochastic Optimization Algorithms for Machine Learning", the XV International Conference on Stochastic Programming (ICSP), Trondheim, July 31, 2019. b) Invited speaker, "Proximal Point Methods for Optimization with Nonconvex Functional Constraints", c) International Conference on Continuous Optimization (ICCOPT) August 5, 2019. d) Invited speaker, "Optimization and Learning with Nonconvex Functional Constraints", SAMSI Deep Learning Workshop August 14, 2019. e) Invited speaker, "Stochastic First-order Methods for Convex Functional Constrained Optimization", Workshop: New Ideas in Quantitative Finance Stony Brook University, Stony Brook, NY, November 5, 2019. f) Invited speaker, "Projection-free Methods and Their Applications", Operations Research Reaching Thousands Miles Away Online Forum, organized by the Operations Research Society of China, July 12, 2020.

Year 3:

a) Invited seminar (virtual), "Projection-free Methods and Their Applications", Boris Polyak's Russian Seminar on Optimization, Moscow Institute of Physics and Technology, July 29, 2020 (online). b) Invited seminar (virtual), "Complexity of Stochastic Dual Dynamic Programming", Department of Mathematics, Shanghai University, October 24, 2020 (online). c) Invited seminar (virtual), "Decentralized Stochastic Gradient Descent and Beyond", Department of Applied Mathematics and Statistics, Johns Hopkins University, January 13, 2021. d) Invited seminar (virtual), "Advancing Stochastic Optimization for Reinforcement Learning", Daniel J. Epstein Seminar Series, University of Southern California, February 16, 2021.

Publication

1. G. Lan and Y. Yang, Accelerated stochastic algorithms for nonconvex finite-sum and multiblock optimization, SIAM Journal on Optimization 29 (4), 2753-2784, 2019.

2. G. Lan, Z. Li, Y. Zhou, A unified variance-reduced accelerated gradient method for convex optimization, Advances in Neural Information Processing Systems (NeurIPS), 10462-10472, 2019.

3. G. Lan, First-order and Stochastic Optimization Methods for Machine Learning, Springer Nature, 2020.

4. G. Lan, S. Lee and Y. Zhou, Communication-efficient algorithms for decentralized and stochastic optimization, Mathematical Programming 180 (1), 237-284, 2020.

5. G. Lan and Y. Zhou, Asynchronous decentralized accelerated stochastic gradient descent, IEEE Journal on Selected Areas in Information Theory, under review, 2020.

6. Z. Zhang, S. Ahmed, G. Lan, Efficient algorithms for distributionally robust stochastic optimization with discrete scenario support, SIAM Journal on Optimization, accepted for publication, 2021.

7. Z. Zhang and G. Lan, Optimal algorithms for nested stochastic composite optimization, submitted to Mathematical Programming, 2020.
8. G. Lan, Y. Ouyang and Y. Zhou, Graph topology invariant gradient and sampling complexity for decentralized and stochastic optimization, SIAM Journal on Optimization, under revision, 2022.
9. G. Kotalis, G. Lan and T. Li, Simple and optimal methods for stochastic variational inequalities, I: operator extrapolation, SIAM Journal on Optimization, 2021.
10. Z. Wang, Y. Zhou, Y. Liang, and **G. Lan**, A note on inexact gradient and Hessian conditions for cubic regularized Newton's method, Operations Research Letters, Volume 47, Issue 2, March 2019, Pages 146-149.
11. Z. Wang, Y. Zhou, Y. Liang, and **G Lan**, "Sample Complexity of Stochastic Variance-Reduced Cubic Regularization for Nonconvex Optimization", AISTATS 2018, PMLR 80:5990-5999, 2018.

5 Industry interactions

We have had the following industrial interactions.

- Our Ph.D. student Yi Zhou joined IBM's Almaden Research Center as a Research Staff Member in the AI Platforms Research group after she graduated in Summer 2018. Part of her responsibility at IBM is to develop new products for decentralized/federated learning algorithms and platforms
- Our Ph.D. student Georgios Kotalis Joined Amazon AI department in Summer 2022.