AFRL-AFOSR-VA-TR-2023-0258

Online Modeling of Heterogeneous Autonomy

Alfredo Garcia
TEXAS ENGINEERING EXPERIMENT STATION COLLEGE STATION
1470 WILLIAM D FITCH PKY
COLLEGE STATION, TX,
US

01/19/2023
Final Technical Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
Air Force Office of Scientific Research
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE 20230119	**2. REPORT TYPE** Final	**3. DATES COVERED** **START DATE** 20190801	**END DATE** 20220731

4. TITLE AND SUBTITLE
Online Modeling of Heterogeneous Autonomy

5a. CONTRACT NUMBER	**5b. GRANT NUMBER** FA9550-19-1-0347	**5c. PROGRAM ELEMENT NUMBER** 61102F
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**

6. AUTHOR(S)
Alfredo Garcia

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) TEXAS ENGINEERING EXPERIMENT STATION COLLEGE STATION 1470 WILLIAM D FITCH PKY COLLEGE STATION, TX US	**8. PERFORMING ORGANIZATION REPORT NUMBER**

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Office of Scientific Research 875 N. Randolph St. Room 3112 Arlington, VA 22203	**10. SPONSOR/MONITOR'S ACRONYM(S)** AFRL/AFOSR RTA2	**11. SPONSOR/MONITOR'S REPORT NUMBER(S)** AFRL-AFOSR-VA-TR-2023-0258

12. DISTRIBUTION/AVAILABILITY STATEMENT
A Distribution Unlimited: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
In our recent publication [1] the task of estimating a model of dynamic decisions by a single human agent based upon the history of implemented actions with hidden (or partially observable) states. Under the assumption of complete state observability (for both the
agent and the modeler), this problem has been widely studied in two strands of the literature where it is referred to as structural estimation of Markov decision processes (MDP) [5] or alternatively as inverse reinforcement learning (IRL) [6]. In this paper we consider the case in which the agent makes decisions under partial observability of the relevant state variableand the modeler (which also only partially observes the state) uses a POMDP (Partially Observable MDP) model. We analyze the structural properties of the model and specify conditions under which the model is ident able without knowledge of the state dynamics. We consider a soft policy gradient algorithm to compute a maximum likelihood estimator and provide anite-time characterization of convergence to a stationary point. We test the model with data from engine replacement dynamic decisions. First, we use synthetic data to highlight the robustness of the proposed methodology and characterize the potential for mis-spec cation when partial state observability is ignored. We then apply the model to a subset of the dataset in [5] on bus engine replacement decisions. The results show that the proposed model can signi icantly improve modelt as measured by the log-likelihood function by 17:7%. More interestingly, the model reveals a feature of bus route assignment behavior in the dataset which was hitherto ignored, i.e. buses with engines believed to be in worse condition exhibit less utilization (mileage) and higher maintenance costs.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT** UU	**18. NUMBER OF PAGES** 5
a. REPORT U	**b. ABSTRACT** U	**c. THIS PAGE** U		

19a. NAME OF RESPONSIBLE PERSON WARREN ADAMS	**19b. PHONE NUMBER** *(Include area code)* 00000000

Final Report: Online Modeling of Heterogeneous Autonomy

PI: Alfredo Garcia (email: alfredo.garcia@tamu.edu)
Co-PI: Mingyi Hong (email: mhong@umn.edu)

Agency:	**Air Force Office of Scientific Research (AFOSR)**
Program:	**Mathematical Optimization**
Grant:	**FA9550-19-1-0347**
Reporting Period:	**August, 2019 - July 30, 2022**

1 Project Overview

This is the final report on the AFOSR project, "Online Modeling of Heterogeneous Autonomy" by A. Garcia (Texas A&M University) and Mingyi Hong (U. of Minnesota). The present final report covers the activities from August, 2019 to July 30, 2022.

Predicting an autonomous agent's actions (human or artificial) is an important task for autonomy research. For example, accurate modeling of a human user's dynamic decision making rationale facilitates coordination with artificially intelligent agents. However, model development entails solving *high-dimensional, non-convex* optimization problems. Thus, there is need for fast, scalable algorithms for real-time learning in mission critical applications with extensive human-machine interaction.

In this project, we *(i)* study fast distributed asynchronous stochastic gradient approaches to train a neural network representation of to predict an agent's dynamic decisions and *(ii)* apply these techniques for inverse reinforcement learning with partially observable states.

2 Summary of Research Activities

Below, we summarize three research directions we have pursued during the entire three-year project.

Decentralized Riemannian Gradient Descent on the Stiefel Manifold: High dimensional machine learning models (e.g. deep learning) can take weeks to train on a single GPU-equipped machine. Hence, distributed implementations of the stochastic gradient method are often used to speed-up the training time. In a *federated* learning architecture model updates are executed in a *centralized* (parameter) server based upon gradient estimates obtained from local nodes. However, the federated learning architecture can be slowed down by the communication overhead associated which exchanging high dimensional local gradient estimates with the centralized parameter server. An alternative *networked* architecture for distributed machine learning is depicted on Figure 1 in a ring configuration. In this architecture there is no *central* node and nodes only exchange model parameters *locally* to compute average models.

The relative performance of a *federated* vs. *networked* architectures for learning depends upon the interplay between computation vs. communication times. If the data across nodes is *independent* and *homogeneously* distributed and the expected time to communicate a newly obtained gradient estimate *exceeds* the expected time to compute such estimate, distributed networked implementations provide linear speedup with increasing number of

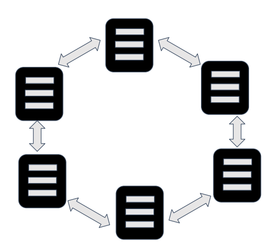

Figure 1: *Networked* architecture

nodes [3, 4] while federated learning does not scale up
well. Nonetheless, communication time can be traded-off for increased computational time by constraining the search for model parameters to a lower-dimensional Riemannian manifold. In our recent work [2], we have shown that a decentralized networked implementation of Riemannian Stochastic Gradient Descent (DRSGD) also guarantees linear speed up (see Figure 2) when computational time exceeds communication time.

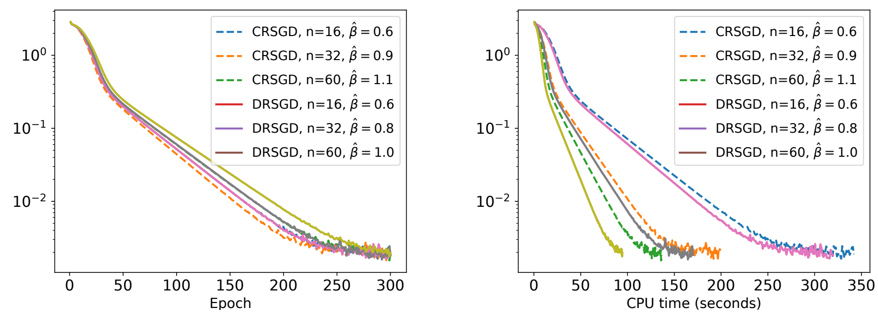

Figure 2: *Federated* (CRSGD) vs *Networked* (DRSGD) PCA on MNIST with $n \in \{16, 32, 30\}$ nodes.

Structural Estimation of Partially Observable Markov Decision Processes:

Dynamic discrete choice models are used to estimate the inter-temporal preferences of an agent as described by a reward function based upon observable histories of states and implemented actions. However, in many applications, such as reliability and healthcare, the system state is partially observable or hidden (e.g., the level of deterioration of an engine, the condition of a disease), and the decision maker only has access to information imperfectly correlated with the true value of the hidden state.

In our publication [1] the task of estimating a model of dynamic decisions by a single human agent based upon the history of implemented actions with hidden (or partially observable) states. Under the assumption of complete state observability (for both the agent and the modeler), this problem has been widely studied in two strands of the literature where it is referred to as *structural estimation* of Markov decision processes (MDP) [5] or alternatively as *inverse reinforcement learning* (IRL) [6]. In this paper we consider the case in which the agent makes decisions under *partial* observability of the relevant state variable and the modeler (which also only partially observes the state) uses a POMDP (Partially Observable MDP) model. We analyze the structural properties of the model and specify conditions under which the model is identifiable without knowledge of the state dynamics. We consider a *soft* policy gradient algorithm to compute a maximum likelihood estimator and provide a finite-time characterization of convergence to a stationary point. We test the model with data from engine replacement dynamic decisions. First, we use synthetic data to highlight the robustness of the proposed methodology and characterize the potential for misspecification when partial state observability is ignored. We then apply the model to a subset of the dataset in [5] on bus engine replacement decisions. The results show that the proposed model can significantly improve model fit as measured by the log-likelihood function by 17.7%. More interestingly, the model reveals a feature of bus route assignment behavior in the dataset which was hitherto ignored, i.e. buses with engines *believed* to be in worse condition exhibit less utilization (mileage) and higher maintenance costs.

Distributed Projected Subgradient Method for Weakly Convex Optimization:

The stochastic subgradient method is a widely-used algorithm for solving large-scale optimization problems arising in machine learning. Often these problems are neither smooth nor convex. Recently, Davis [1] characterized the convergence of the stochastic subgradient method for the weakly convex case, which encompasses many important applications (e.g., robust phase retrieval, blind deconvolution, biconvex compressive sensing, and dictionary learning). In practice, distributed implementations of the projected stochastic subgradient method (stoDPSM) are used to speed-up risk minimization. In this publication [**?**], we propose a distributed implementation of the stochastic subgradient method with a theoretical guarantee. Specifically, we show the global convergence of stoDPSM using the Moreau envelope stationarity measure. Furthermore, under a so-called sharpness condition, we show that deterministic DPSM (with a proper initialization) converges linearly to the sharp minima, using geometrically diminishing step-size. We provide numerical experiments to support our theoretical analysis.

3 Products

In this section, we list the main products generated by this project during the reporting period, as well as providing a short summary of each of the work.

1. "On the Divergence of Decentralized Non-Convex Optimization", Mingyi Hong, Siliang Zeng, Junyu Zhang, and Haoran Sun, **SIAM Journal on Optimization** (2022), DOI: 10.1137/20M1353149

   Description: In this work, we study a generic class of decentralized algorithms in which N agents jointly optimize the non-convex objective function $f(u) := 1/N \sum_{i=1}^{N} f_i(u)$, while only communicating with their neighbors. However, most of the existing decentralized algorithms require that the local function gradients ∇f_i's as well as the average function gradient ∇f are Lipschitz, that is, the *local* Lipschitz conditions (LLC) and global Lipschitz condition (GLC) are satisfied. In this work, we demonstrate the importance of the above Lipschitzness assumptions on the state-of-the-art decentralized algorithms. First, by constructing a series of examples, we show that when the *local* Lipschitz conditions (LLC) on the local function gradient ∇f_i's are not satisfied, a number of state-of-the-art decentralized algorithms diverge, even if the global

[1] "Stochastic model-based minimization of weakly convex functions," SIAM Journal on Optimization, vol. 29, no. 1, pp. 207-239, 2019

Lipschitz condition (GLC) still holds. This observation brings out a fundamental theoretical issue of the existing decentralized algorithms – their convergence conditions are *strictly stronger* than centralized algorithms such as the gradient descent (GD), which only requires the GLC. Second, we design two first-order algorithms, which are capable of computing stationary solutions of the original problem with neither the LLC nor the GLC condition. To our knowledge, this is the first attempt that studies decentralized non-convex optimization problems with neither the LLC nor the GLC.

2. "Maximum Likelihood Inverse Reinforcement Learning with Finite-Time Guarantees", S. Zeng, Ch. Li, A. Garcia and M., Hong, **Proceedings of Neural Information Processing Systems (NeurIPS)** (2022)

   Description: Inverse reinforcement learning (IRL) aims to recover the reward function and the associated optimal policy that best fits observed sequences of states and actions implemented by an expert. Many algorithms for IRL have an inherently nested structure: the inner loop finds the optimal policy given parametrized rewards while the outer loop updates the estimates towards optimizing a measure of fit. For high dimensional environments such nested-loop structure entails a significant computational burden. In this paper we develop a novel *single-loop* algorithm for IRL that does not compromise reward estimation accuracy. In the proposed algorithm, each policy improvement step is followed by a stochastic gradient step for likelihood maximization. We show that the proposed algorithm provably converges to a stationary solution with a finite-time guarantee. If the reward is parameterized linearly, we show the identified solution corresponds to the solution of the maximum entropy IRL problem. Finally, by using robotics control problems in MuJoCo and their transfer settings, we show that the proposed algorithm achieves superior performance compared with other IRL and imitation learning benchmarks.

3. "Structural Estimation of Partially Observable Markov Decision Processes", Y. Chang, A. Garcia, Z. Wang and L. Sun, **IEEE Transactions on Automatic Control** (2022), DOI: 10.1109/TAC.2022.3217908

   Description: Partially Observable Markov Decision Processes (POMDPs) is a well-developed framework for sequential decision making under uncertainty and partial information. This paper considers the (inverse) structural estimation of the primitives of a POMDP based upon data in the form of sequences of observables and implemented actions. We analyze the structural properties of an entropy regularized POMDP and specify conditions under which the model is identifiable without knowledge of the state dynamics. We consider a soft policy gradient algorithm to compute a maximum likelihood estimator, and illustrate the approach with an equipment replacement problem.

4. "Learning to Coordinate in Multi-Agent Systems: A Coordinated Actor-Critic Algorithm and Finite-Time Guarantees", S. Zeng, T. Chen, A. Garcia and Hong, M.

Proceedings of International Conference on Machine Learning Research vol 168, pp. 1-45 (2022)

Description: Multi-agent reinforcement learning (MARL) has attracted much research attention recently. However, unlike its single-agent counterpart, many theoretical and algorithmic aspects of MARL have not been well-understood. In this paper, we study the emergence of coordinated behavior by autonomous agents using an actor-critic (AC) algorithm. Specifically, we propose and analyze a class of coordinated actor-critic (CAC) algorithms in which individually parametrized policies have a *shared* part (which is jointly optimized among all agents) and a *personalized* part (which is only locally optimized). Such a kind of *partially personalized* policy allows agents to coordinate by leveraging peers' experience and adapt to individual tasks. The flexibility in our design allows the proposed CAC algorithm to be used in a *fully decentralized* setting, where the agents can only communicate with their neighbors, as well as in a *federated* setting, where the agents occasionally communicate with a server while optimizing their (partially personalized) local models. Theoretically, we show that under some standard regularity assumptions, the proposed CAC algorithm requires $\mathcal{O}(\epsilon^{-\frac{5}{2}})$ samples to achieve an ϵ-stationary solution (defined as the solution whose squared norm of the gradient of the objective function is less than ϵ). To the best of our knowledge, this work provides the first finite-sample guarantee for decentralized AC algorithm with partially personalized policies.

5. "On Distributed Nonconvex Optimization: Projected Subgradient Method for Weakly Convex Problems in Networks", Shixiang Chen, Alfredo Garcia, Shahin Shahrampour **IEEE Transactions on Automatic Control** (2022), DOI: 10.1109/TAC.2021.3056535

   Description: The stochastic subgradient method is a widely used algorithm for solving large-scale optimization problems arising in machine learning. Often, these problems are neither smooth nor convex. Recently, Davis et al., 2018 characterized the convergence of the stochastic sub-gradient method for the weakly convex case, which encompasses many important applications (e.g., robust phase retrieval, blind deconvolution, biconvex compressive sensing, and dictionary learning). In practice, distributed implementations of the projected stochastic subgradient method (stoDPSM) are used to speed up risk minimization. In this article, we propose a distributed implementation of the stochastic subgradient method with a theoretical guarantee. Specifically, we show the global convergence of stoDPSM using the Moreau envelope stationarity measure. Furthermore, under a so-called sharpness condition, we show that deterministic DPSM (with a proper initialization) converges linearly to the sharp minima, using geometrically diminishing step size. We provide numerical experiments to support our theoretical analysis.

   Description:

6. "Distributed Networked Learning with Correlated Data", L. Hong, A. Garcia and C. Eksin, **Automatica** (2022), Vol. 137, pp 110-134 DOI: 10.1016/j.automatica.2021.110134

Description: We consider a distributed estimation method in a setting with heterogeneous streams of correlated data distributed across nodes in a network. In the considered approach, linear models are estimated locally (i.e., with only local data) subject to a network regularization term that penalizes a local model that differs from neighboring models. We analyze computation dynamics (associated with stochastic gradient updates) and information exchange (associated with exchanging current models with neighboring nodes). We provide a finite-time characterization of convergence of the weighted ensemble average estimate and compare this result to federated learning, an alternative approach to estimation wherein a single model is updated by locally generated gradient updates. This comparison highlights the trade-off between speed vs precision: while model updates take place at a faster rate in federated learning, the proposed networked approach to estimation enables the identification of models with higher precision. We illustrate the method's general applicability in two examples: estimating a Markov random field using wireless sensor networks and modeling prey escape behavior of flocking birds based on a publicly available dataset.

7. "Decentralized Riemannian Gradient Descent on the Stiefel Manifold" S. Chen, A.Garcia. M. Hong and S. Shahrampour, **Proceedings of International Conference on Machine Learning ICML** (2021)

   Description: We consider a distributed non-convex optimization where a network of agents aims at minimizing a global function over the Stiefel manifold. The global function is represented as a finite sum of smooth local functions, where each local function is associated with one agent and agents communicate with each other over an undirected connected graph. The problem is non-convex as local functions are possibly non-convex (but smooth) and the Steifel manifold is a non-convex set. We present a decentralized Riemannian stochastic gradient method (DRSGD) with the convergence rate of $\mathcal{O}(1/\sqrt{K})$ to a stationary point. To have exact convergence with constant stepsize, we also propose a decentralized Riemannian gradient tracking algorithm (DRGTA) with the convergence rate of $\mathcal{O}(1/K)$ to a stationary point. We use multi-step consensus to preserve the iteration in the local (consensus) region. DRGTA is the first decentralized algorithm with exact convergence for distributed optimization on Stiefel manifold.

8. "Linearized ADMM Converges to Second-Order Stationary Points for Non-Convex Problems", S. Lu, J. D. Lee, M. Razaviyayn, and M. Hong, **IEEE Transactions on Signal Processing** (2021), DOI: 10.1109/TSP.2021.3100976

   Description: In this work, a gradient-based primal-dual method of multipliers is proposed for solving a class of linearly constrained non-convex problems. We show that with random initialization of the primal and dual variables, the algorithm is able to compute second-order stationary points (SOSPs) with probability one. Further, we present applications of the proposed method in popular signal processing and machine learning problems such as decentralized matrix factorization and decentralized training of over-parameterized neural networks. One of the key steps in the analysis is to construct a new loss function for these problems such that the required convergence conditions (especially the gradient Lipschitz conditions) can be satisfied without changing the global optimal points.

9. "On the Local Linear Rate of Consensus on the Stiefel Manifold", S. Chen, A. Garcia, M. Hong, S. Shahrampour, revise and resubmit, **IEEE Transactions on Automatic Control** (2021)

   Description: Coordinated group behavior arising from purely local interactions has been successfully modeled with distributed consensus-seeking dynamics, where the local behavior is aimed at minimizing the disagreement with neighboring peers. However, it has been recently shown that when constrained by a manifold geometry, distributed consensus-seeking dynamics may ultimately fail to converge to a global consensus state. In this paper, we study discrete-time consensus-seeking dynamics on the Stiefel manifold and identify conditions on the network topology to ensure convergence to a global consensus state. We further prove a (local) linear convergence rate to the consensus state that is on par with the well-known rate in the Euclidean space. These results have implications for consensus applications constrained by manifold geometry, such as synchronization and collective motion, and they can be used for convergence analysis of decentralized Riemannian optimization on the Stiefel Manifold.

10. "The Effects of Mental Fatigue on Effort Allocation: Modeling and Estimation", Z. Wang, Y. Chang, B.J. Schmeichel, A. Garcia, **Psychology Review** (2021), DOI: 10.1037/rev0000365

    Description: Mental fatigue is usually accompanied by drops in task performance and reduced willingness for further exertion. A value-based theoretical account may help to explain such negative effects. In this view, mental fatigue influences perceived costs and rewards of exerting effort. However, no formal mathematical framework has yet been proposed to model and quantitatively estimate the effects of mental fatigue on subjective evaluations of effort expenditure, under possibly imperfect self-perceptions of internal fatigue states. We proposed a mathematical framework to model human cognitive effort allocations, assuming mental fatigue states are partially observable with semi-Markov dynamics. We modeled effort allocation decisions as a means to the goal of maximizing cumulative subjective values over a given time horizon. We developed an estimation method to identify subjective values and the hidden dynamics of mental fatigue, which can in future work be applied to self-reports, psychophysiological indices, and behavioral outcomes associated with fatigue. The modeling and estimation method was tested using a simulated n-back task under a free-choice paradigm, with model parameters fine-tuned from past studies. The proposed approach was able to recapitulate task performance and task engagement patterns observed under mental fatigue. This work advances a reward/cost trade-off account for explaining the exertion of mental effort and suggests new avenues for both theoretically and empirically relevant understandings of how cognitive operations are affected by mental fatigue.

11. “Distributed Networked Real-time Learning,” A. Garcia, L. Wang, J. Huang and L. Hong, **IEEE Transactions on Control of Network Systems**, (2020) DOI: 10.1109/TCNS.2020.3029992

    Description: Many machine learning algorithms have been developed under the assumption that data sets are already available in batch form. Yet in many application domains data is only available sequentially overtime via compute nodes in different geographic locations. In this paper, we consider the problem of learning a model when streaming data cannot be transferred to a single location in a timely fashion. In such cases, a distributed architecture for learning relying on a network of interconnected ”local” nodes is required. We propose a distributed scheme in which every local node implements stochastic gradient updates based upon a local data stream. To ensure robust estimation, a network regularization penalty is used to maintain a measure of cohesion in the ensemble of models. We show the ensemble average approximates a stationary point and characterize the degree to which individual models differ from the ensemble average. We compare the results with federated learning to conclude the proposed approach is more robust to heterogeneity in data streams (data rates and estimation quality). We illustrate the results with an application to image classification with a deep learning model based upon convolutional neural networks.

12. “Min-Max Optimization without Gradients: Convergence and Applications to Black-Box Evasion and Poisoning Attacks”, Sijia Liu, Songtao Lu, Xiangyi Chen, Yao Feng, Kaidi Xu, Abdullah Al-Dujaili, Mingyi Hong, Una May O’Reilly, **Proceedings of Machine Learning Research** (2020)

    Description. In this paper, we study the problem of constrained min-max optimization in a black-box setting, where the desired optimizer cannot access the gradients of the objective function but may query its values. We present a principled optimization framework, integrating a zeroth-order (ZO) gradient estimator with an alternating projected stochastic gradient descent-ascent method, where the former only requires a small number of function queries and the later needs just one-step descent/ascent update. We show that the proposed framework, referred to as ZO-Min-Max, has a sublinear convergence rate under mild conditions and scales gracefully with problem size. We also explore a promising connection between black-box min-max optimization and black-box evasion and poisoning attacks in adversarial machine learning (ML). Our empirical evaluations on these use cases demonstrate the effectiveness of our approach and its scalability to dimensions that prohibit using recent black-box solvers.

13. “Distributed Learning in the Non-Convex World: From Batch to Streaming Data, and Beyond”, Tsung-Hui Chang, Mingyi Hong, Hoi-To Wai, Xinwei Zhang and Songtao Lu, **IEEE Signal Processing Magazine** (2020), DOI: 10.1109/MSP.2020.2970170

    Description. Distributed learning has become a critical enabler of the massively connected world envisioned by many. This article discusses four key elements of scalable distributed processing and real-time intelligence — problems, data, communication and computation. Our aim is to provide a fresh and unique perspective about how these elements should work together in an effective and coherent manner. In particular, we provide a selective review about the recent techniques developed for optimizing non-convex models (i.e., problem classes), processing batch and streaming data (i.e., data types), over the networks in a distributed manner (i.e., communication and computation paradigm). We describe the intuitions and connections behind a core set of popular distributed algorithms, emphasizing how to trade off between computation and communication costs. Practical issues and future research directions will also be discussed.

14. "Variance Reduced Policy Evaluation with Smooth Function Approximation", Hoi-To Wai, Mingyi Hong, Zhuoran Yang, Zhaoran Wang, Kexin Tang, **Proceeding of Neural Information Processing** (2019)

    Description. Policy evaluation with smooth and nonlinear function approximation has shown great potential for reinforcement learning. Compared to linear function approxi- mation, it allows for using a richer class of approximation functions such as the neural networks. Traditional algorithms are based on two timescales stochastic approximation whose convergence rate is often slow. This paper focuses on an offline setting where a trajectory of m state-action pairs are observed. We formulate the policy evaluation problem as a non-convex primal-dual, finite-sum optimization problem, whose primal sub-problem is non-convex and dual sub-problem is strongly concave. We suggest a single-timescale primal-dual gradient algorithm with variance reduction, and show that it converges to an ϵ-stationary point using $O(m/\epsilon)$ calls (in expectation) to a gradient oracle.

15. "Improving the Sample and Communication Complexity for Decentralized Non-Convex Optimization: Joint Gradient Estimation and Tracking", Haoran Sun, Songtao Lu, Mingyi Hong, **Proceedings of the International Conference on Machine Learning** (2020)

    Description. Many modern large-scale machine learning problems benefit from decentralized and stochastic optimization. Recent works have shown that utilizing both decentralized computing and local stochastic gradient estimates can outperform state-of-the-art centralized algorithms, in applications involving highly non-convex problems, such as training deep neural networks.

    In this work, we propose a decentralized stochastic algorithm to deal with certain smooth non-convex problems where there are m nodes in the system, and each node has a large number of samples (denoted as n). Differently from the majority of the existing decentralized learning algorithms for either stochastic or finite-sum problems, our focus is given to *both* reducing the total communication rounds among the nodes, while accessing the minimum number of local data samples. In particular, we propose an algorithm named D-GET (decentralized gradient estimation and tracking), which jointly performs decentralized gradient estimation (which estimates the local gradient using a subset of local samples) *and* gradient tracking (which tracks the global full gradient using local estimates). We show that, to achieve certain ϵ stationary solution of the deterministic finite sum problem, the proposed algorithm achieves an $\mathcal{O}(mn^{1/2}\epsilon^{-1})$ sample complexity and an $\mathcal{O}(\epsilon^{-1})$ communication complexity. These bounds significantly improve upon the best existing bounds of $\mathcal{O}(mn\epsilon^{-1})$ and $\mathcal{O}(\epsilon^{-1})$, respectively. Similarly, for online problems, the proposed method achieves an $\mathcal{O}(m\epsilon^{-3/2})$ sample complexity and an $\mathcal{O}(\epsilon^{-1})$ communication complexity, while the best existing bounds are $\mathcal{O}(m\epsilon^{-2})$ and $\mathcal{O}(\epsilon^{-2})$.

16. "Distributed Networked Learning with Correlated Data", Lingzhou Hong, Alfredo Garcia and Ceyhun Eksin, **The Proceedings of IEEE CDC** (2020)

    Description. This is the conference version of the paper published in **Automatica** which is summarized above.

4 Students Funded

This project has partially supported the following graduate students

1. Lingzhou Hong, Texas A&M University (PhD)
2. Wei Ran, Texas A&M University (PhD)
3. Zhide Wang, Texas A&M University (PhD)
4. Siliang Zeng, U. of Minnesota (PhD)

5 Other Activities Supported

This project has supported the PIs through partial summer support.

References

[1] Yanling Chang, Alfredo Garcia, Zhide Wang, and Lu Sun. Structural estimation of partially observable markov decision processes. *IEEE Transactions on Automatic Control*, pages 1–8, 2022. 2

[2] Shixiang Chen, Alfredo Garcia, Mingyi Hong, and Shahin Shahrampour. Decentralized riemannian gradient descent on the stiefel manifold. *ICML 2021*, 2021. 2

[3] X. Lian, C. Zhang, H. Zhang, C. Hsieh, W. Zhang, and J. Liu. Can decentralized algorithms outperform centralized algorithms? a case study for decentralized parallel stochastic gradient descent. In *Advances in Neural Information Processing Systems*. 2017. 2

[4] Shi Pu and Alfredo Garcia. Swarming for faster convergence in stochastic optimization. *SIAM Journal on Control and Optimization*, 56(4):2997–3020, 2018. 2

[5] J. Rust. Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica*, pages 999–1033, 1987. 2

[6] B. D. Ziebart, A. Maas, J. A. Bagnell, and A. K. Dey. Maximum entropy inverse reinforcement learning. *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence*, pages 1433–1438, 2008. 2

Final Report: Online Modeling of Heterogeneous Autonomy

PI: Alfredo Garcia (email: alfredo.garcia@tamu.edu)
Co-PI: Mingyi Hong (email: mhong@umn.edu)

Agency:	**Air Force Office of Scientific Research (AFOSR)**
Program:	**Mathematical Optimization**
Grant:	**FA9550-19-1-0347**
Reporting Period:	**August, 2019 - July 30, 2022**

1 Project Overview

This is the final report on the AFOSR project, "Online Modeling of Heterogeneous Autonomy" by A. Garcia (Texas A&M University) and Mingyi Hong (U. of Minnesota). The present final report covers the activities from August, 2019 to July 30, 2022.

Predicting an autonomous agent's actions (human or artificial) is an important task for autonomy research. For example, accurate modeling of a human user's dynamic decision making rationale facilitates coordination with artificially intelligent agents. However, model development entails solving *high-dimensional, non-convex* optimization problems. Thus, there is need for fast, scalable algorithms for real-time learning in mission critical applications with extensive human-machine interaction.

In this project, we *(i)* study fast distributed asynchronous stochastic gradient approaches to train a neural network representation of to predict an agent's dynamic decisions and *(ii)* apply these techniques for inverse reinforcement learning with partially observable states.

2 Summary of Research Activities

In this second year we highlight two recent lines of work:

Decentralized Riemannian Gradient Descent on the Stiefel Manifold: High dimensional machine learning models (e.g. deep learning) can take weeks to train on a single GPU-equipped machine. Hence, distributed implementations of the stochastic gradient method are often used to speed-up the training time. In a *federated* learning architecture model updates are executed in a *centralized* (parameter) server based upon gradient estimates obtained from local nodes. However, the federated learning architecture can be slowed down by the communication overhead associated which exchanging high dimensional local gradient estimates with the centralized parameter server. An alternative *networked* architecture for distributed machine learning is depicted on Figure 1 in a ring configuration. In this architecture there is no *central* node and nodes only exchange model parameters *locally* to compute average models.

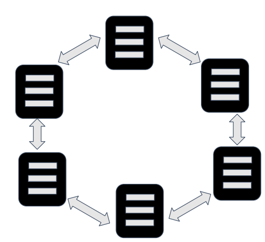

Figure 1: *Networked* architecture

The relative performance of a *federated* vs. *networked* architectures for learning depends upon the interplay between computation vs. communication times. If the data across nodes is *independent* and *homogeneously* distributed and the expected time to communicate a newly obtained gradient estimate *exceeds* the expected time to compute such estimate, distributed networked implementations provide linear speedup with increasing number of nodes [3, 4] while federated learning does not scale up well. Nonetheless, communication time can be traded-off for increased computational time by constraining the search for model parameters to a lower-dimensional Riemannian manifold. In our recent work [2], we have shown that a decentralized networked implementation of Riemannian Stochastic Gradient Descent (DRSGD) also guarantees linear speed up (see Figure 2) when computational time exceeds communication time.

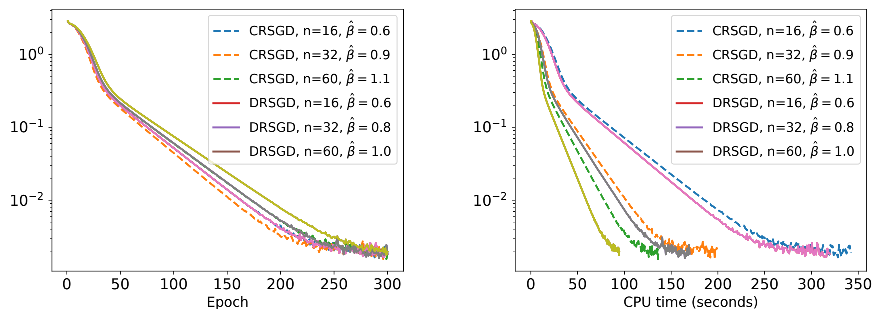

Figure 2: *Federated* (CRSGD) vs *Networked* (DRSGD) PCA on MNIST with $n \in \{16, 32, 30\}$ nodes.

Structural Estimation of Partially Observable Markov Decision Processes:

In our recent publication [1] the task of estimating a model of dynamic decisions by a single human agent based upon the history of implemented actions with hidden (or partially observable) states. Under the assumption of complete state observability (for both the agent and the modeler), this problem has been widely studied in two strands of the literature where it is referred to as *structural estimation* of Markov decision processes (MDP) [5] or alternatively as *inverse reinforcement learning* (IRL) [6]. In this paper we consider the case in which the agent makes decisions under *partial* observability of the relevant state variable and the modeler (which also only partially observes the state) uses a POMDP (Partially Observable MDP) model. We analyze the structural properties of the model and specify conditions under which the model is identifiable without knowledge of the state dynamics. We consider a *soft* policy gradient algorithm to compute a maximum likelihood estimator and provide a finite-time characterization of convergence to a stationary point. We test the model with data from engine replacement dynamic decisions. First, we use synthetic data to highlight the robustness of the proposed methodology and characterize the potential for misspecification when partial state observability is ignored. We then apply the model to a subset of the dataset in [5] on bus engine replacement decisions. The results show that the proposed model can significantly improve model fit as measured by the log-likelihood function by 17.7%. More interestingly, the model reveals a feature of bus route assignment behavior in the dataset which was hitherto ignored, i.e. buses with engines *believed* to be in worse condition exhibit less utilization (mileage) and higher maintenance costs.

3 Products

In this section, we list the main products generated by this project during the reporting period:

1. "On the Divergence of Decentralized Non-Convex Optimization", Mingyi Hong, Siliang Zeng, Junyu Zhang, and Haoran Sun, accepted by **SIAM Journal on Optimization**, 2022

2. "Maximum Likelihood Inverse Reinforcement Learning with Finite-Time Guarantees", S. Zeng, Ch. Li, A. Garcia and Hong, M. **Proceedings of Neural Information Processing Systems (NeurIPS)** (2022)

3. "Structural Estimation of Partially Observable Markov Decision Processes", Y. Chang, A. Garcia, Z. Wang and L. Sun (2022), **IEEE Transactions on Automatic Control** , DOI: 10.1109/TAC.2022.3217908

4. "Learning to Coordinate in Multi-Agent Systems: A Coordinated Actor-Critic Algorithm and Finite-Time Guarantees", S. Zeng, T. Chen, A. Garcia and Hong, M. **Proceedings of Machine Learning Research** (2022) vol 168, pp. 1-45

5. "Decentralized Riemannian Gradient Descent on the Stiefel Manifold" S. Chen, A.Garcia. M. Hong and S. Shahrampour, **Proceedings of International Conference on Machine Learning ICML** (2021)

6. "On Distributed Non-convex Optimization: Projected Subgradient Method For Weakly Convex Problems in Networks ", S. Chen, A. Garcia and S. Shahrampour, (2022) **IEEE Transactions on Automatic Control**, Vol. 67 No.2, pp. 662-675

7. "Linearized ADMM Converges to Second-Order Stationary Points for Non-Convex Problems", **IEEE Transactions on Signal Processing**, S. Lu, J. D. Lee, M. Razaviyayn, and M. Hong, DOI: 10.1109/TSP.2021.3100976

8. "Distributed Networked Real-time Learning," A. Garcia, L. Wang, J. Huang and L. Hong, (2021) **IEEE Transactions on Control of Network Systems**,Vol. 8, No. 1, pp. 28-38

9. "Min-Max Optimization without Gradients: Convergence and Applications to Black-Box Evasion and Poisoning Attacks", Sijia Liu, Songtao Lu, Xiangyi Chen, Yao Feng, Kaidi Xu, Abdullah Al-Dujaili, Mingyi Hong, Una May O'Reilly, **Proceedings of Machine Learning Research** (2020)

4 Students Funded

This project has partially supported the following graduate students

1. Lingzhou Hong, Texas A&M University (PhD).
2. Zhide Wang, Texas A&M University (PhD)
3. Siliang Zeng, U. of Minnesota (PhD)

5 Other Activities Supported

This project has supported the PIs through partial summer support.

References

[1] Yanling Chang, Alfredo Garcia, Zhide Wang, and Lu Sun. Structural estimation of partially observable markov decision processes. *IEEE Transactions on Automatic Control*, pages 1–8, 2022. 2

[2] Shixiang Chen, Alfredo Garcia, Mingyi Hong, and Shahin Shahrampour. Decentralized riemannian gradient descent on the stiefel manifold. *ICML 2021*, 2021. 2

[3] X. Lian, C. Zhang, H. Zhang, C. Hsieh, W. Zhang, and J. Liu. Can decentralized algorithms outperform centralized algorithms? a case study for decentralized parallel stochastic gradient descent. In *Advances in Neural Information Processing Systems*. 2017. 2

[4] Shi Pu and Alfredo Garcia. Swarming for faster convergence in stochastic optimization. *SIAM Journal on Control and Optimization*, 56(4):2997–3020, 2018. 2

[5] J. Rust. Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica*, pages 999–1033, 1987. 2

[6] B. D. Ziebart, A. Maas, J. A. Bagnell, and A. K. Dey. Maximum entropy inverse reinforcement learning. *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence*, pages 1433–1438, 2008. 2