Optimization for Data-Driven Learning and Control

By USMAN A. KHAN, *Senior Member IEEE*
Guest Editor

WAHEED U. BAJWA, *Senior Member IEEE*
Guest Editor

ANGELIA NEDIĆ, *Member IEEE*
Guest Editor

MICHAEL G. RABBAT, *Senior Member IEEE*
Guest Editor

ALI H. SAYED, *Fellow IEEE*
Guest Editor

There is little doubt that we are witnessing a data revolution, and this revolution is expected to impact almost all aspects of our lives with the passage of time. In particular, as the size and complexity of the data collected as part of this revolution grow at unprecedented rates, so does our reliance on the collected data for autonomous decision making in myriad applications. Examples of such applications include image and video classification, speech processing, natural language processing, weather forecasting, and robot path planning, to name a few. Already, many complex inference problems within these applications that seemed impossible to handle not too long ago can now be solved with the help of modern signal processing, control, and machine learning algorithms.

> This special issue provides a comprehensive overview of modern optimization tools and methods for the purposes of data-driven learning and control.

There are several key aspects that are common to such data-driven inference problems in countless applications, including massive data, streaming data, heterogeneous/multimodal data, communication and computation constraints, and privacy concerns, as well as the networked, autonomous, and/or mobile nature of the underlying data collection modules that include many different types of sensors, household devices, drones/robots, and social media platforms. The result is an ever-increasing integration of sensing, autonomy, and networking in many applications, with information exchange often carried out over shared and resource-limited media, such as the internet and the radio frequency spectrum. This integration, in turn, is giving rise to fundamentally new classes of data-driven learning and control problems that are of significant interest to the scientific community in general and to the IEEE community in particular.

Although the new classes of data-driven learning and control problems necessitate progress in several areas, advances in optimization tools and methods have emerged as one of the more effective means of addressing the emerging challenges in these problems. It is in this regard that several scientific communities have started to explore various aspects of optimization theory and algorithms for the purposes of computationally efficient data-driven learning and control in their respective disciplines.

Digital Object Identifier 10.1109/JPROC.2020.3031225

The purpose of this special issue is to summarize the implications of several of these efforts which are scattered across the scientific literature, in one location. To this end, the special issue brings together world-renowned experts from the areas of signal processing, control, optimization, and machine learning, who have contributed a total of 12 articles to this issue. These articles, which should be accessible to readers with different technical backgrounds, summarize the state-of-the-art theoretical and algorithmic advances in optimization for (large-scale) data-driven learning and control, and they elaborate on the implications of these advances in many real-world applications.

I. THEMATIC OVERVIEW OF THE SPECIAL ISSUE

In order to provide a comprehensive treatment to the special issue topic, "Optimization for Data-Driven Learning and Control," we have assembled a set of articles that provide broad coverage of the subject matter. These articles, which run the gamut from the state-of-the-art centralized and distributed methods in convex and nonconvex optimization to stochastic optimization and game theoretic methods for learning and control, not only focus on technical aspects of the underlying algorithms that are being widely used within the scientific community but also emphasize the intuition behind the analyses and results through simple examples. Another highlight of these articles is their ability to connect their technical results to real-world applications for the benefit of the diverse readership of this special issue. Thematically, these articles correspond to one or more of the following nonexclusive categories.

1) *Fundamental Optimization Methods:* We have enlisted a significant coverage of fundamental techniques that form the building blocks of modern optimization algorithms. These techniques, which include zeroth-order, first-order, second-order, submodular, and primal-dual methods, are covered in great depth by six articles (Xin *et al.*, Sahu and Kar, Mokhtari and Ribeiro, Jakovetić *et al.*, França and Bento, and Jaleel and Shamma) with an emphasis on decentralized and distributed solutions. In addition to their coverage of fundamental optimization methods, these articles also describe their utility for learning and control in the context of real-world applications such as training of machine learning and deep learning models, search-engine advertisements, smart grids, social networks, sensor localization, swarms, and robotics.
2) *Stochastic Optimization Methods:* Statistical and empirical risk minimization problems in which only imperfect/noisy gradients can be utilized for optimization purposes are highly relevant to many modern-day learning and control problems. Articles by Xin *et al.*, Sahu and Kar, Mokhtari and Ribeiro, Gower *et al.*, Assran *et al.*, and Li *et al.* in this special issue pay particular attention to these problems within the context of explicit real-world applications.
3) *Nonconvex Optimization Methods:* Convex problems, owing to their nice geometries, have been ubiquitous in early works on optimization methods for data-driven learning and control. However, several of the recently emerging data-driven problems in applications such as representation learning, deep learning, reinforcement learning, and tensor factorization tend to have nonconvex geometry. Keeping this in mind, a significant focus of several articles in the special issue is to cover nonconvex optimization methods and their applicability to learning and control problems; these include articles by Xin *et al.*, Mokhtari and Ribeiro, Nokleby *et al.*, Assran *et al.*, and Li *et al.*
4) *Practical Aspects and Applications:* The last five articles in the special issue focus on several practical aspects related to optimization methods in learning and control problems. These include compute- and/or communication-limited systems, time-varying problems and networks, asynchronous communications, topology learning, and acceleration techniques. In particular: Nokleby *et al.* focus on the interplay between solution accuracy, communication capabilities, and computational resources in distributed systems that carry out learning using streaming data; Assran *et al.* discuss asynchronous methods over decentralized networks of nodes where the communication and computation across the nodes need not be synchronized; Simonetto *et al.* bring out the challenges in solving online problems where the objective functions are dynamic; Matta *et al.* discuss how optimization algorithms over a graph can lead to inverse problems for topology learning; and, Li *et al.* talk about accelerated methods for stochastic and possibly nonconvex optimization problems.

II. SUMMARY OF ARTICLES

We now provide a short summary of each of the 12 articles appearing in this special issue.

A General Framework for Decentralized Optimization With First-Order Methods

by R. Xin, S. Pu, A. Nedić, and U. A. Khan

This article focuses on decentralized optimization methods for the problem of minimizing a finite sum of convex functions, which has been of interest in control, signal processing and estimation, and more recently in large-scale data science and machine learning problems. Decentralized and distributed first-order methods are discussed, which have found success in many emerging machine learning problems where they serve as the first method of choice for many complex inference and training tasks. A general framework for distributed first-order methods over undirected and directed graphs is presented, together with its stochastic variants that rely on stochastic gradient evaluations at each node in the graph. The effectiveness of the methods is demonstrated on real-world, complex training problems implemented in large-scale distributed environments.

Decentralized Zeroth-Order Constrained Stochastic Optimization Algorithms: Frank–Wolfe and Variants With Applications to Black-Box Adversarial Attacks
by A. K. Sahu and S. Kar

In many practical optimization scenarios, the objective functions may not be known in a closed form or the gradient computation may be too expensive. In such cases, zeroth-order or gradient-free methods, which do not directly access gradients of the objective function, are an attractive alternative to gradient-based approaches. This article presents an overview of the recent work in the area of distributed zeroth-order optimization, focusing on constrained optimization settings and algorithms built around the Frank-Wolfe framework. In particular, the article reviews different types of architectures, from master–worker-based decentralized to fully distributed and describe appropriate zeroth-order projection free schemes for solving constrained stochastic optimization problems catered to these architectures. Limitations of this framework are also discussed in terms of dimension dependence and the ideas are illustrated with the help of adversarial attacks on deep learning models.

Stochastic Quasi-Newton Methods
by A. Mokhtari and A. Ribeiro

Model training in machine learning is often formulated as the solution of empirical risk minimization, which are optimization programs whose complexity scales with the number of elements in a data set. Training over massive data sets thus induces a significant computational burden, and such challenges are typically addressed with the help of stochastic optimization methods. A relevant approach in this direction is to consider stochastic variants of the quasi-Newton method, which approximate the curvature (second-order information) of the objective function using stochastic gradients. This article discusses recent developments to accelerate the convergence of such stochastic methods and describes their convergence aspects. Applications are discussed in the context of predicting the click-through rate of an advertisement displayed in response to a specific search engine query.

Primal–Dual Optimization Methods for Large-Scale and Distributed Convex Optimization and Data Analytics
by D. Jakovetić, D. Bajović, J. Xavier, and J. M. F. Moura

This article focuses on the augmented Lagrangian method (ALM), where a constrained optimization problem is solved with a series of unconstrained subproblems, with respect to the original (primal) variable, while the constraints are controlled via dual variables. In particular, the article provides a tutorial-style introduction to the ALM and its variant primal-dual methods for solving convex optimization problems in large-scale and distributed settings. Special emphasis is on control-theoretic tools for the design and analysis of the underlying methods, while discussion and insights are provided in the context of two emerging applications: federated learning and distributed-energy trading.

Distributed Optimization, Averaging via ADMM, and Network Topology
by G. França and J. Bento

Scalable optimization methods distribute the computation over a network of processors and are of interest due to the explosion in the size of data sets and model complexity in modern machine learning applications. This article reviews recent research quantifying the influence of the network topology on the convergence behavior of distributed methods. It compares different algorithms when applied to a canonical average consensus problem. It further explores the connections between the alternating direction method of multipliers (ADMM) and lifted Markov chains in addition to providing an explicit characterization of its convergence and optimal parameter tuning in terms of spectral properties of the underlying network. The connection between network topology and convergence rates are numerically validated on a real-world sensor localization problem.

Distributed Optimization for Robot Networks: From Real-Time Convex Optimization to Game-Theoretic Self-Organization
by H. Jaleel and J. S. Shamma

This article presents a collection of state-of-the-art results for distributed optimization problems arising in the context of robot networks, with a focus on two special classes of problems, namely real-time path planning for multirobot systems and self-organization in multirobot systems using game-theoretic approaches. For real-time multi-robot path planning, recent approaches based on approximately solving distributed optimization problems are discussed for both continuous- and discrete-action domains. For game-theoretic self-organization, some results for area coverage and real-time target assignment are presented, where the problems are formulated as games and online updating rules are designed to enable teams of robots to achieve the collective objective in a distributed manner.

Variance-Reduced Methods for Machine Learning
by R. M. Gower, M. Schmidt, F. Bach, and P. Richtárik

This article discusses stochastic optimization methods with a focus on variance reduction techniques. These variance reduced methods excel in settings with convex objectives when multiple passes through the training data is allowed, achieving faster convergence than the simple stochastic gradient method, both in theory and in practice. The article reviews the key principles and main developments behind variance reduced methods for optimization with a primary focus on convex problems and with pointers to the literature for nonconvex problems.

Scaling-Up Distributed Processing of Data Streams for Machine Learning

by M. Nokleby, H. Raja, and W. U. Bajwa

Emerging applications of machine learning in numerous areas—including online social networks, remote sensing, internet-of-things systems, smart grids, and more—involve a continuous gathering of and learning from streams of data samples. Real-time incorporation of streaming data into the learned machine learning models is essential for improved inference in these applications. It is in this regard that this article reviews recently developed methods that focus on distributed training of large-scale machine learning models from streaming data in the compute-limited and bandwidth-limited regimes, with an emphasis on convergence analysis that explicitly accounts for the mismatch between computation, communication and streaming rates, and that provides sufficient conditions for order-optimum convergence. In particular, it focuses on methods that solve: 1) distributed stochastic convex problems and 2) distributed principal component analysis, which is a nonconvex problem with a geometric structure that permits global convergence.

Advances in Asynchronous Parallel and Distributed Optimization

by M. Assran, A. Aytekin, H. R. Feyzmahdavian, M. Johansson, and M. G. Rabbat

The article focuses on asynchronous parallel and distributed methods for large-scale optimization problems in machine learning, where the processors may maintain an inconsistent view of the optimization variables. Such methods can more efficiently utilize the computational resources than synchronous methods, and they are not sensitive to issues such as stragglers and unreliable communication links. This article reviews recent developments in the design and analysis of asynchronous methods, including both centralized and decentralized methods. The analysis provides insights into how the degree of asynchrony impacts convergence rates, especially in stochastic methods.

Time-Varying Convex Optimization: Time-Structured Algorithms and Applications

by A. Simonetto, E. Dall'Anese, S. Paternain, G. Leus, and G. B. Giannakis

Driven by modern infrastructure and social network platforms, recent years have witnessed a major focus on dynamic, time-varying optimization problems and applications. This article reviews a broad class of algorithms for time-varying optimization with an emphasis on both algorithmic development and performance analysis. It provides a comprehensive overview of available tools and methods and unveils open challenges in application domains of broad interest. Real-world examples are presented on smart power systems, robotics, machine learning, and data analytics, highlighting domain-specific challenges and solutions.

Graph Learning Under Partial Observability

by V. Matta, A. Santos, and A. H. Sayed

Many optimization, inference, and learning tasks can be accomplished efficiently by means of decentralized processing algorithms where the network topology (i.e., the graph) plays a critical role in enabling the interactions among neighboring nodes. There is a large body of literature examining the effect of the graph structure on the performance of decentralized processing strategies. In this article, the authors examine the inverse problem and consider the reverse question: How much information can one glean about the underlying graph topology by observing the behavior at the nodes of the graph? For large-scale networks, the difficulty in addressing such inverse problems is compounded by the fact that usually only a limited fraction of the nodes can be probed, giving rise to a second important question: Despite the presence of unobserved nodes, can partial observations still be sufficient to discover the graph linking the probed nodes? The article surveys recent advances on this important inverse problem and related questions.

Accelerated First-Order Optimization Algorithms for Machine Learning

by H. Li, C. Fang, and Z. Lin

To meet the demand of big data applications, considerable effort has been expended to design theoretically and practically fast algorithms. This article provides a comprehensive survey of accelerated first-order methods with a particular focus on stochastic algorithms. The article reviews the basic accelerated algorithms for deterministic convex optimization problems. Subsequently, it concentrates on the extensions of the deterministic convex problems to their stochastic counterparts and, at last, it introduces some recent developments on accelerated methods for nonconvex optimization problems.

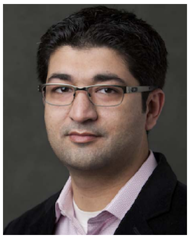
III. CONCLUSION

This special issue assembles a collection of articles on the topic of optimization for data-driven learning and control. We hope the contents of this issue will be useful to the many readers of Proceedings of the IEEE. The articles in this special issue provide a tutorial-style content aimed at nonexpert, as well as reviews of latest advances that we expect to be valuable to researchers in optimization and machine learning who are interested in an overview of recent developments and future directions. Throughout the process of preparing this special issue, we have striven to balance the discussion of advances in theory with implications for practitioners.

The field of optimization for machine learning and control is broad and growing. The topics covered in this special issue reflect what we believe to be important at this time. Due to restrictions on space and the number of articles in this special issue, however, we are aware that there are some topics which may not be covered here, and we apologize for any omissions.

We would like to thank all of the contributing authors. This special issue would not have been possible without their hard work preparing and revising these articles. We also sincerely thank the reviewers for contributing their time and knowledge to improve the contents of each article. Finally, we would like to offer our gratitude to the editorial staff of Proceedings of the IEEE and the Editorial Board, who provided valuable feedback on the special issue proposal. We would like to express special thanks to Jo Sun, Senior Publications Editor, Vaishali Damle, Managing Editor, and Gianluca Setti, Editor-in-Chief, for their support, feedback, and guidance throughout the process of preparing this issue. ■

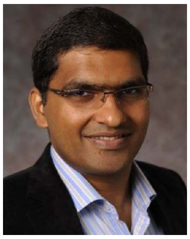
ABOUT THE GUEST EDITORS



Usman A. Khan (Senior Member, IEEE) received the B.S. degree from the University of Engineering and Technology, Lahore, Pakistan, in 2002, the M.S. degree from the University of Wisconsin–Madison, Madison, WI, USA, in 2004, and the Ph.D. degree from Carnegie Mellon University, Pittsburgh, PA, USA, in 2009, all in electrical and computer engineering.

He held a postdoctoral position at the GRASP Laboratory, University of Pennsylvania, Philadelphia, PA, USA. In 2011, he joined Tufts University as an Assistant Professor. In Spring 2015, he was a Visiting Professor with KTH, Stockholm, Sweden. He is currently an Associate Professor of electrical and computer engineering (ECE) with Tufts University, Medford, MA, USA, where he is also an Adjunct Professor of computer science. His research interests include signal processing, machine learning, control, and optimization. He has published extensively in these topics with more than 100 papers in journals and conference proceedings and holds multiple patents. Recognition of his work includes the prestigious National Science Foundation (NSF) Career Award, several NSF REU awards, an IEEE journal cover, three best student paper awards in IEEE conferences, and several news articles, including two in *IEEE Spectrum.*

Dr. Khan was an Associate Member of the Sensor Array and Multichannel Technical Committee, IEEE Signal Processing Society, from 2010 to 2019, where he has been an elected full member since 2019. He was an elected full member of the IEEE Big Data Special Interest Group from 2017 to 2019 and has served on the IEEE Young Professionals Committee and the IEEE Technical Activities Board. He is also the Technical Area Chair of the Networks track at the IEEE 2020 Asilomar Conference on Signals Systems and Computers. He has served on the technical program committee of several IEEE conferences and has organized/chaired several IEEE workshops and sessions. He served as an Editor for IEEE Transactions on Smart Grid from 2014 to 2017. He is also serving as an Associate Editor for IEEE Control System Letters, IEEE Transactions on Signal and Information Processing over Networks, and IEEE Open Journal of Signal Processing. He also served as a Guest Associate Editor for IEEE Control System Letters—Special Issue on Learning and Control both to appear in 2020.



Waheed U. Bajwa (Senior Member, IEEE) received the B.E. degree (Honors) in electrical engineering from the National University of Sciences and Technology, Islamabad, Pakistan, in 2001, and the M.S. and Ph.D. degrees in electrical engineering from the University of Wisconsin–Madison, Madison, WI, USA, in 2005 and 2009, respectively.

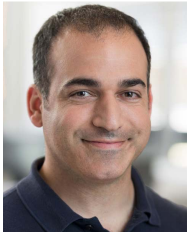
He was a Postdoctoral Research Associate of the Program in Applied and Computational Mathematics with Princeton University, Princeton, NJ, USA, from 2009 to 2010, and a Research Scientist with the Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA, from 2010 to 2011. He has been with Rutgers University, Piscataway, NJ, USA, since 2011, where he is currently an Associate Professor with the Department of Electrical and Computer Engineering and an Associate Member of the graduate faculty of the Department of Statistics. His research interests include statistical signal processing, high-dimensional statistics, machine learning, inverse problems, and networked systems.

Dr. Bajwa received the Best in Academics Gold Medal and President's Gold Medal in Electrical Engineering from the National University of Sciences and Technology in 2001, the Morgridge Distinguished Graduate Fellowship from the University of Wisconsin–Madison in 2003, the Army Research Office Young Investigator Award in 2014, the National Science Foundation CAREER Award in 2015, the Rutgers University's Presidential Merit Award in 2016, the Rutgers University's Presidential Fellowship for Teaching Excellence in 2017, and the Rutgers Engineering Governing Council ECE Professor of the Year Award in 2016, 2017, and 2019. He is a co-investigator on a work that received the Cancer Institute of New Jersey's Gallo Award for Scientific Excellence in 2017, a coauthor on papers that received best student paper awards at IEEE IVMSP 2016 and IEEE CAMSAP 2017 workshops, and a member of the Class of 2015 National Academy of Engineering Frontiers of Engineering Education Symposium. He served as the Lead Guest Editor for *IEEE Signal Processing Magazine*—Special Issue on Distributed, Streaming Machine Learning in 2020, the Technical Co-Chair for the IEEE SPAWC 2018 Workshop, the Technical Area Chair of the 2018 Asilomar Conference on Signals, Systems, and Computers, the General Chair for the 2017 DIMACS Workshop on Distributed Optimization, Information Processing, and Learning, and an Associate Editor for IEEE Signal Processing Letters from 2014 to 2017. He is also serving as a Senior Area Editor for IEEE Signal Processing Letters and an Associate Editor for IEEE Transactions on Signal and Information Processing over Networks.

Angelia Nedić (Member, IEEE) received the Ph.D. degree in computational mathematics and mathematical physics from Moscow State University, Moscow, Russia, in 1994, and the Ph.D. degree in electrical and computer science engineering from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2002.

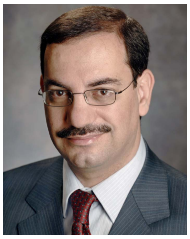
She has worked as a Senior Engineer with BAE Systems North America, Arlington, VA, USA, and the Advanced Information Technology Division, Burlington, MA, USA. She has been a Willard Scholar Faculty Member with the University of Illinois at Urbana–Champaign, Champaign, IL, USA. She is currently a Faculty Member with the School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ, USA. Her general research interests include optimization, large-scale complex systems dynamics, variational inequalities, and games.

Dr. Nedić was a recipient (jointly with her coauthors) of the best paper awards at the Winter Simulation Conference 2013 and the International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt) 2015.



Michael G. Rabbat (Senior Member, IEEE) received the B.Sc. degree from the University of Illinois at Urbana–Champaign, Champaign, IL, USA, in 2001, the M.Sc. degree from Rice University, Houston, TX, USA, in 2003, and the Ph.D. degree from the University of Wisconsin–Madison, Madison, WI, USA, in 2006, all in electrical engineering.

He is currently a Research Scientist with the Facebook Artificial Intelligence Research Group (FAIR), Montreal, QC, Canada. From 2007 to 2018, he was a Professor with the Department of Electrical and Computer Engineering, McGill University, Montreal, QC, Canada. From 2013 to 2014, he held visiting positions at Télécom Bretegne, Brest, France, the Inria Bretagne-Atlantique Research Centre, Rennes, France, and the KTH Royal Institute of Technology, Stockholm, Sweden. His research interests include optimization, distributed algorithms, graph signal processing, and machine learning.



Ali H. Sayed (Fellow, IEEE) is the Dean of Engineering with EPFL, Lausanne, Switzerland, where he also leads the Adaptive Systems Laboratory. He served as a Distinguished Professor and the former Chairman of electrical engineering at the University of California at Los Angeles, Los Angeles, CA, USA. He is a member of the U.S. National Academy of Engineering and is recognized as a Highly Cited Researcher by Thomson Reuters and Clarivate Analytics. He is an author of over 570 publications and six books. His research interests include adaptation and learning theories, data and network sciences, statistical inference, optimization, and biologically inspired designs.

Prof. Sayed is a Fellow of EURASIP and the American Association for the Advancement of Science (AAAS). His work has been recognized with several awards, including the 2015 Education Award from the IEEE Signal Processing Society, the 2014 Papoulis Award from the European Association for Signal Processing, the 2013 Meritorious Service Award and the 2012 Technical Achievement Award from the IEEE Signal Processing Society, the 2005 Terman Award from the American Society for Engineering Education, the 2005 Distinguished Lecturer from the IEEE Signal Processing Society, the 2003 Kuwait Prize in Basic Sciences, and the 1996 IEEE Fink Prize. He received several best paper awards from the IEEE in 2002, 2005, 2012, and 2014 and EURASIP in 2015. He served as the President of the IEEE Signal Processing Society from 2018 to 2019.