

# Bokun Wang

(+1) 319-512-3088 | bokun-wang@tamu.edu | <https://bokun-wang.github.io> | Bokun Wang

## Education

### Texas A&M University

PH.D. STUDENT IN COMPUTER SCIENCE, ADVISED BY PROF. TIANBAO YANG

College Station, TX, USA

August 2022 -

My research focuses on stochastic optimization for machine learning and AI, with a particular emphasis on learning from imbalanced data, self-supervised representation learning, and distributionally robust optimization (DRO).

### The University of Iowa

PH.D. STUDENT IN COMPUTER SCIENCE (TRANSFERRED OUT WITH MY PH.D. ADVISOR)

Iowa, IA, USA

August 2021 - July 2022

### University of California, Davis

M.S. IN COMPUTER SCIENCE

Davis, CA, USA

Sept. 2018 - June 2020

### University of Electronic Science and Technology of China

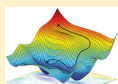
B.S. IN COMPUTER SCIENCE AND ENGINEERING (YINGCAI HONORS COLLEGE)

Chengdu, Sichuan, China

Sept. 2014 - June 2018

## Publications

### Theory for Modern Machine Learning and AI with Non-Decomposable Losses



#### Optimization Theory

Simple, Efficient, Minibatch Algorithms with Convergence Guarantees

Optimal Convergence Rates and Lower Complexity Bounds

Towards Fewer Assumptions on Lipschitz Continuity and Smoothness



Exploiting Data  
Parallelism and  
Collaboration



#### Distributed and Federated Learning

Adaptive Quantization for Faster  
Computation and Communication

Personalization for clients

Providing a  
Statistical  
Framework



#### Discriminative Probabilistic Modeling

Unified Framework for Supervised  
and Self-Supervised Learning

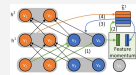
Generalization Analysis

[ICML 22, ICLR 22, NeurIPS 22, JMLR 23, ArXiv 23, ArXiv 24]

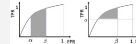
### Applications



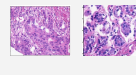
Contrastive Learning



Deep Learning on Graph



Learning with Imbalanced Data



Learning with Multi-Instance  
Medical Data



Group Robustness/Fairness



End-to-End 8-bit Minifloat  
Federated Learning



Differentiable Attack on Graph

[ICML 22, ICML 23, KDD-W 24, WWW 24]

### ► On Discriminative Probabilistic Modeling for Self-Supervised Representation Learning

Under Submission

BOKUN WANG, YUNWEN LEI, YIMING YING, TIANBAO YANG

2024

- **tl;dr:** We study the discriminative probabilistic modeling problem over a continuous domain for (multimodal) self-supervised representation learning, yielding a new contrastive objective for self-supervised representation learning.

### ► ALEXR: An Optimal Single-Loop Algorithm for Convex Finite-Sum Coupled Compositional Stochastic Optimization

Under Submission

BOKUN WANG AND TIANBAO YANG

2023

- **tl;dr:** We revisit a class of convex Finite-Sum Coupled Compositional Stochastic Optimization (cFCCO) problems and introduce an efficient single-loop primal-dual block-coordinate proximal algorithm with optimal oracle complexity.

### ► Towards Federated Learning with On-device Training and Communication in 8-bit Floating Point

FedKDD Workshop at 2024 ACM

SIGKDD Conference

BOKUN WANG, AXEL BERG, AND DURMUS ALP EMRE ACAR, AND CHUTENG ZHOU

2024

- **tl;dr:** In this work, we investigate the use of FP8 training in a federated learning context.

▷ **Everything Perturbed All at Once: Enabling Differentiable Graph Attacks**

*International World Wide Web  
Conference (TheWebConf)*

HAORAN LIU, **BOKUN WANG**, JIANLING WANG, XIANGJUE DONG, TIANBAO YANG, AND JAMES CAVERLEE

2023

- **tl;dr:** By leveraging continuous relaxation and parameterization of the graph structure, we propose a novel attack method to efficiently generate effective attacks and meanwhile eliminate the need for costly retraining.

▷ **Memory-Based Optimization Methods for Model-Agnostic Meta-Learning and Personalized Federated Learning**

*Journal of Machine Learning  
Research (JMLR), 24(145):1–46*

**BOKUN WANG**, ZHUONING YUAN, YIMING YING, TIANBAO YANG

2023

- **tl;dr:** We propose memory-based stochastic algorithms for MAML that converge with vanishing error. The proposed algorithms require sampling a constant number of tasks and data samples per iteration, making them suitable for the continual learning scenario. Moreover, we also extend the proposed algorithms to personalized federated learning settings.

▷ **Provable Multi-instance Deep AUC Maximization with Stochastic Pooling**

*40th International Conference  
on Machine Learning (ICML)*

DIXIAN ZHU, **BOKUN WANG**, ZHI CHEN, YAXING WANG, MILAN SONKA, XIAODONG WU, AND TIANBAO YANG

2023

- **tl;dr:** We consider a novel application of deep AUC maximization (DAM) for multi-instance learning (MIL), in which a single class label is assigned to a bag of instances (eg, multiple 2D slices of a CT scan for a patient). We propose an algorithm called MIDAM to address a neglected yet non-negligible computational challenge of MIL in the context of DAM: bag size is too large to be loaded into GPU memory for backpropagation.

▷ **Theoretically Better and Numerically Faster Distributed Optimization with Smoothness-Aware Quantization Techniques**

*36th Conference on Neural  
Information Processing Systems  
(NeurIPS)*

**BOKUN WANG**, MHER SAFARYAN, AND PETER RICHTÁRIK

2022

- **tl;dr:** In this work, we extend the smoothness-aware compression strategy in Safaryan et al (2021) to arbitrary unbiased compression operators and then instantiate the theory by quantization operators.

▷ **When AUC meets DRO: Optimizing Partial AUC for Deep Learning with Non-Convex Convergence Guarantee**

*39th International Conference  
on Machine Learning (ICML)*

DIXIAN ZHU, GANG LI, **BOKUN WANG**, XIAODONG WU, AND TIANBAO YANG

2022

- **tl;dr:** We propose new formulations of partial area under the ROC Curve (pAUC) by using the distributionally robust optimization (CVaR or KL divergence based) to define the loss for each individual positive data. For one-way/two-way pAUC maximization, we propose two provably convergent algorithms and numerically verify the effectiveness on various image and medical datasets.

▷ **Finite-Sum Coupled Compositional Stochastic Optimization: Theory and Applications**

*39th International Conference  
on Machine Learning (ICML)*

**BOKUN WANG** AND TIANBAO YANG

2022

- **tl;dr:** This paper studies stochastic optimization for a sum of compositional functions, where the inner-level function of each summand is coupled with the corresponding summation index. We provide convergence analysis and numerical verification of a simple algorithm.
- The proposed algorithm is **a workhorse of the LibAUC library**.

▷ **GraphFM: Improving Large-Scale GNN Training via Feature Momentum**

*39th International Conference  
on Machine Learning (ICML)*

HAIYANG YU\*, LIMEI WANG\*, **BOKUN WANG\***, MENG LIU, TIANBAO YANG, AND SHUIWANG JI (\*EQUAL CONTRIBUTION)

2022

- **tl;dr:** We propose a new way of incorporating historical embeddings when updating feature representations in sampling-based and memory-based graph neural networks (GNNs).

▷ **Optimal Algorithms for Stochastic Multi-Level Compositional Optimization**

*39th International Conference  
on Machine Learning (ICML)*

WEI JIANG, BOKUN WANG, YIBO WANG, LIJUN ZHANG, AND TIANBAO YANG

2022

- **tl;dr:** we propose a Stochastic Multi-level Variance Reduction method (SMVR), which achieves the optimal sample complexity of  $O(1/\epsilon^3)$  to find an  $\epsilon$ -stationary point for non-convex objectives.

## ▷ IntSGD: Adaptive Floatless Compression of Stochastic Gradients

10th International Conference on  
Learning Representations (ICLR)

KONSTANTIN MISHCHENKO, **BOKUN WANG**, DMITRY KOVALEV, AND PETER RICHTÁRIK

2022

- **tl;dr:** We propose a family of adaptive integer compression operators for distributed Stochastic Gradient Descent (SGD) that do not communicate a single float, which is provably convergent and computationally cheap. IntSGD has the same iteration complexity as that of FP32 SGD while only communicating 25% bits. We numerically compare our algorithm to baselines on image classification and language modelling tasks.

## ▷ Riemannian Stochastic Proximal Gradient Methods for Nonsmooth Optimization over the Stiefel Manifold

Journal of Machine Learning  
Research (JMLR), 23(106), 1-33

**BOKUN WANG**, SHIQIAN MA, AND LINGZHOU XUE

2022

- **tl;dr:** We present two Riemannian stochastic proximal gradient methods for nonsmooth optimization over the Stiefel manifold.

## ▷ Towards Fair Deep Clustering With Multi-State Protected Variables

ICML Workshop on the Security  
and Privacy of Machine Learning

BOKUN WANG AND IAN DAVIDSON

2019

- **tl;dr:** In this work, we consider a general setting where there can be many protected groups and propose a deep fair clustering algorithms, which learns a discriminative but fair cluster assignment function.

## Honors & Awards

---

- 2022 **Scholar Award**, 36th Conference on Neural Information Processing Systems (NeurIPS)
- 2022 **Travel Award**, 39th International Conference on Machine Learning (ICML)
- 2022 **Spotlight Presentation**, 10th International Conference on Learning Representations (ICLR)
- 2018 **Outstanding Graduate**, University of Electronic Science and Technology of China (UESTC)
- 2016 **National Scholarship**, Ministry of Education of the People's Republic of China
- 2016 **People's Scholarship (Special Class)**, University of Electronic Science and Technology of China (UESTC)

## Experience

---

### Central Technology / Machine Learning Group, ARM

Waltham, Massachusetts

RESEARCH INTERN ADVISED BY DR. CHUTENG ZHOU AND DR. DURMUS ALP EMRE ACAR,

May 2023 - August 2023

- Topic: Propose and implement a new framework for efficient federated learning with neural network quantization and communication quantization, based on the new FP8 data type (see the whitepaper by NVIDIA, ARM and Intel).
- Results lead to a patent application and a KDD 2024 workshop paper.

### OptML Lab, King Abdullah University of Science and Technology (KAUST)

Thuwal, KSA

RESEARCH INTERN ADVISED BY PROF. PETER RICHTÁRIK,

Sept. 2020 - August 2021

- Topic: efficient optimization algorithms for distributed ML. I contributed to theoretical and numerical parts: building problem formulation, establishing convergence analysis, and writing Python (including Pytorch) code for numerical verifications. In particular, I have experience using King Abdullah University of Science and Technology's large-scale Shaheen supercomputer to conduct experiments on 16 GPU nodes.
- Results lead to an ICLR 2022 paper and a NeurIPS 2022 paper.

## Services

---

- 2024 **Co-organizer of a session titled "New Min-max Optimization Algorithms and Their Applications in Machine Learning and AI"**, 2024 INFORMS Annual Meeting, Seattle
- 2024 **Reviewer**, The Forty-first International Conference on Machine Learning (ICML)
- 2023 **Reviewer**, The Eleventh International Conference on Learning Representations (ICLR)
- 2023 **Reviewer**, The 26th International Conference on Artificial Intelligence and Statistics (AISTATS)
- 2023 **Reviewer**, The Thirty-seventh Conference on Neural Information Processing Systems (NeurIPS)
- 2023 **Reviewer**, The Forty International Conference on Machine Learning (ICML)
- 2021 **Reviewer**, The Thirty-fifth Conference on Neural Information Processing Systems (NeurIPS)

Since 2021 **Reviewer**, Neurocomputing

Since 2023 **Reviewer**, IJSE Transactions

Since 2024 **Reviewer**, Mathematical Programming