NHAN H. PHAM

nhanph.github.io ⬦ nhp@ibm.com ⬦ (775)-501-2570

RESEARCH INTERESTS

- ⬦ Stochastic methods for machine learning, deep learning, reinforcement learning, and federated learning
- ⬦ LLM-based solutions for enterprise data management applications

EDUCATION

Ph.D. in Operations Research   2017–2021
Department of Statistics and Operations Research
University of North Carolina at Chapel Hill · Chapel Hill, NC, USA

Graduate Study in Computer Engineering   2015–2017
Department of Computer Science and Engineering
University of Nevada, Reno · Reno, NV, USA

Bachelor of Engineering (Honor Program) in Computer Engineering   2008–2013
Department of Computer Science and Engineering
Ho Chi Minh City University of Technology · Ho Chi Minh City, Vietnam

INDUSTRY EXPERIENCES

Staff Research Scientist   2024–Present
IBM Research, Thomas J. Watson Research Center · Yorktown Heights, NY

Research Scientist   2022–2024
IBM Research, Thomas J. Watson Research Center · Yorktown Heights, NY

Summer Machine Learning Intern   2021
Blue River Technology Inc. · Sunnyvale, CA

Summer Research Intern   2020
IBM Research, Thomas J. Watson Research Center · Yorktown Heights, NY

RESEARCH EXPERIENCES

Automation for Enterprise Data Management   *Dec. 2023–Present*

- ⬦ Build an end-to-end system for text-to-SQL using Large Language Models (LLMs).
- ⬦ Experiments with both in-context learning and finetuning LLMs for individual tasks such as schema linking and SQL generation.
- ⬦ Curate training and evaluation datasets from open-source and enterprise usecases.
- ⬦ Develop reasoning capability for text-to-SQL models.
- ⬦ Achieved top rannking on BIRD leaderboard.

Table metadata to business glossaries mapping   *2023–Jan. 2024*

- ⬦ Leverage Large Language Models (LLMs) to design generic matching methods that do not require manual tuning and can identify complex relations between column names and glossaries.
- ⬦ Propose methods that utilize LLMs in two ways: generating additional context for column names that can aid with matching and using LLMs to directly infer if there is a relation between column names and glossary description.

An Automated System for Tuning and Alignment of Foundation Models with Applications in Data Management
Mar. 2022–Mar. 2023

- ⬦ Build a configurable search system that allows automated rapid experimentation for tuning foundation models on data management tasks.
- ⬦ Experimenting with different reward functions for reinforcement learning (RL) fine-tuning for Column-to-Concept mapping use-case.

Automated Decision Optimization   *Jan. 2022–Mar. 2023*

- ⬦ Design the application framework and system architecture for data and knowledge-driven Automated Decision Optimization ((AutoDO)).
- ⬦ Demonstrate, benchmark, and experiment for effectiveness and solution quality from AutoDO.
- ⬦ AutoDO is available on IBM API Hub portal.

Evaluating Robustness of Cooperative MARL: A Model-based approach *Jul. 2021–May 2023*
Joint work with: Dr. Lam M. Nguyen, Dr. Jie Chen, Dr. Hoang Thanh Lam, Dr. Subhro Das, and Dr. Tsui-Wei Weng.
Accepted for the 2023 IEEE International Conference on Data Mining (ICDM), eprint.

⋄ Propose the first model-based adversarial attacks, called cMBA, for cooperative multi-agent reinforcement learning by solving a constrained nonconvex optimization problem at every timestep.
⋄ Propose a new victim agent selection strategy which has not been considered in previous works.
⋄ Conduct experiments on multi-agent MuJoCo environments.

Federated Learning with Randomized Douglas-Rachford Splitting Methods *Aug. 2020–Jun. 2021*
Graduate Research Assistant, Supervisor: Dr. Quoc Tran-Dinh, Dr. Lam M. Nguyen.
Accepted for the 35th Conference on Neural Information Processing Systems, eprint.

⋄ Propose two new algorithms, FedDR and asyncFedDR, to solve finite-sum nonconvex problems in federated learning by combining Douglas-Rachford splitting, randomized strategy, and asynchronous update.
⋄ Achieve best-known communication complexity and handle data heterogeneity.
⋄ Conduct experiments on federated learning examples using synthetic and real datasets.

Regression Optimization for System-level Production Control *Jun. 2020–Aug. 2020*
IBM Research Intern, Supervisor: Dr. Roman Vaculin, Dr. Dzung T. Phan, Dr. Lam M. Nguyen.
Accepted for the 2021 American Control Conference (ACC).

Stochastic Gauss-Newton Algorithms for Nonconvex Compositional Optimization *Sept. 2019–Feb. 2020*
Graduate Research Assistant, Supervisors: Dr. Quoc Tran-Dinh, Dr. Lam M. Nguyen.
Accepted for the 37th International Conference on Machine Learning, eprint.

⋄ Propose two new Stochastic Gauss-Newton algorithms to solve stochastic nonconvex compositional problems that use both classical stochastic and SARAH estimators for function values and Jacobian estimators.
⋄ Give first stochastic Gauss-Newton algorithm with global complexity analysis.
⋄ Conduct numerical experiments on two examples: stochastic nonlinear equations and asset allocation problem.

Regularization Techniques on Deep Learning *Sept. 2019–Dec. 2019*
SAMSI Research Fellow, Supervisor: Dr. Quoc Tran-Dinh.

⋄ Study the principle of different regularization techniques on training Deep Neural Networks (DNNs).
⋄ Conduct numerical experiments on different DNN models consisting two or more regularizers on both model parameters (e.g., ℓ_2-norm, max-norm constraint, etc.) and training process (dropout, batch normalization, etc.).

Hybrid Stochastic Policy Gradient Algorithm for Reinforcement Learning *Jul. 2019–Dec. 2019*
Graduate Research Assistant, Supervisors: Dr. Quoc Tran-Dinh, Dr. Lam M. Nguyen.
Accepted for the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS 2020), eprint.

⋄ Propose new biased policy gradient estimator from REINFORCE/GPOMDP and adopted SARAH estimators and use it to derive first algorithm that has convergence guarantee to solve a composite policy optimization problem in reinforcement learning.
⋄ Prove proposed algorithm achieves best-known convergence rate over existing methods and conduct experiments to verify the advantage using OpenAI gym environments.

Hybrid Optimization Framework for Composite Nonconvex Optimization *Feb. 2019–Aug. 2019*
Graduate Research Assistant, Supervisors: Dr. Quoc Tran-Dinh, Dr. Lam M. Nguyen.
Accepted for Mathematical Programming, eprint.

⋄ Introduce a new stochastic gradient estimator that combines SGD and SARAH estimators and use it to develop a new algorithm for composite nonconvex optimization problems which achieves best-known convergence rate.
⋄ Verify the effectiveness of the proposed algorithm via numerical experiments using Python and Tensorflow.

ProxSARAH: A Framework for Stochastic Composite Nonconvex Optimization *Aug. 2018–Feb. 2019*
Graduate Research Assistant, Supervisors: Dr. Quoc Tran-Dinh, Dr. Lam M. Nguyen.
Accepted for Journal of Machine Learning Research (JMLR), eprint.

⋄ Develop new stochastic algorithm for composite nonconvex optimization problems which utilizes existing SARAH estimator and achieves the best-known convergence rate.
⋄ Conduct numerical experiments to illustrate advantage of proposed algorithms on three examples: Non-negative PCA, classification with 3 nonconvex losses, and neural network training using Python and Tensorflow.

Autonomous Robots for Bridge Inspection *Aug. 2015–Feb. 2017*

Graduate Research Assistant, Supervisor: Dr. Hung M. La.
In Proceedings of the 54th Annual Allerton Conference on Communication, Control, and Computing, eprint.
In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (*ICRA*), eprint.
The 33rd International Symposium on Automation and Robotics in Construction and Mining (*ISARC*), eprint.

◇ Propose four-wheeled robot for steel bridge inspection with permanent magnets embedded inside each wheel equipped with different type of sensors: visual camera, 3D sensor, IMU for localization and mapping purposes.

◇ Build controller unit with minicomputer (Intel NUC) running Robot Operating System communicating with low-level controller (Arduino-based) for sensory data collection, implement sensor fusion and mapping algorithms.

PREPRINTS

1. G. Rossiello, **N. H. Pham**, M. Glass, J. Lee, D. Subramanian. Rationalization Models for Text-to-SQL. *arXiv:2502.06759*, 2025.

2. T. T. Doan, L. M. Nguyen, **N. H. Pham**, and J. Romberg. Convergence Rates of Accelerated Markov Gradient Descent with Applications in Reinforcement Learning. *arXiv:2002.02873*, 2020.

3. T. T. Doan, L. M. Nguyen, **N. H. Pham**, and J. Romberg. Finite-Time Analysis of Stochastic Gradient Descent under Markov Randomness. *arXiv:2003.10973*, 2020.

PUBLICATIONS

1. Q. Xiao, D. Bhattacharjya, B. Ganesan, R. Marinescu, K. Mirylenka, **N. H. Pham**, M. Glass, and J. Lee. The Consistency Hypothesis in Uncertainty Quantification for Large Language Models. *Proceedings of the Forty-First Conference on Uncertainty in Artificial Intelligence*, 2025.

2. **N. H. Pham**, L. M. Nguyen, J. Chen, H. T. Lam, S. Das, T. W. Weng. Evaluating Robustness of Cooperative MARL: A Model-based Approach. *2023 IEEE International Conference on Data Mining* (ICDM), pp. 1271-1276, Shanghai, China, 2023.

3. Q. Tran-Dinh, **N. H. Pham**, D. T. Phan, and L. M. Nguyen. FedDR–Randomized Douglas-Rachford Splitting Algorithms for Nonconvex Federated Composite Optimization. *The 35th Conference on Neural Information Processing Systems*, 2021.

4. D. T. Phan, L. M. Nguyen, P. Murali, **N. H. Pham**, H. Liu, and J. R. Kalagnanam. Regression Optimization for System-level Production Control. *American Control Conference* (*ACC*), 2021.

5. Q. Tran-Dinh, **N. H. Pham**, D. T. Phan, and L. M. Nguyen. A Hybrid Stochastic Optimization Framework for Composite Nonconvex Optimization. *Mathematical Programming*, 2021.

6. Q. Tran-Dinh, **N. H. Pham**, and L. M. Nguyen. Stochastic Gauss-Newton Algorithms for Nonconvex Compositional Optimization. *Proceedings of the 37th International Conference on Machine Learning*, PMLR 119:9572-9582, 2020.

7. **N. H. Pham**, L. M. Nguyen, D. T. Phan, and Q. Tran-Dinh. ProxSARAH: An Efficient Algorithmic Framework for Stochastic Composite Nonconvex Optimization. *Journal of Machine Learning Research*, 2020.

8. **N. H. Pham**, L. M. Nguyen, D. T. Phan, P. H. Nguyen, M. van Dijk, and Q. Tran-Dinh. A Hybrid Stochastic Policy Gradient Algorithm for Reinforcement Learning. *The 23rd International Conference on Artificial Intelligence and Statistics* (*AISTATS*), 2020, Palermo, Italy.

9. H. M. La, T. H. Dinh, **N. H. Pham**, Q. P. Ha, and A. Q. Pham. Automated Robotic Monitoring and Inspection of Steel Structures and Bridges. *Robotica*, Cambridge University Press, 1-21, 2018.

10. T. D. Le, S. Gibb, **N. H. Pham**, H. M. La, L. Falk, and T. Berendsen. Autonomous Robotic System using Non-Destructive Evaluation methods for Bridge Deck Inspection. In *Proceedings of the 2017 IEEE International Conference on Robotics and Automation* (*ICRA*), May 29-June 3, 2017, Singapore.

11. **N. H. Pham** and H. M. La. Design and Implementation of an Autonomous Robot for Steel Bridge Inspection. In *Proceedings of the 54th Annual Allerton Conference on Communication, Control, and Computing*, pages 1-8, Sept. 27-30, 2016, Urbana-Champaign, Illinois, USA.

12. **N. H. Pham**, H. M. La, Q. P. Ha, S. N. Dang, A. H. Vo, and Q. H. Dinh. Visual and 3D Mapping for Steel Bridge Inspection Using a Climbing Robot. *The 33rd International Symposium on Automation and Robotics in Construction and Mining* (*ISARC*), pages 1-8, July 18-21, 2016, Auburn, Alabama, USA.

PATENT APPLICATIONS

1. M. Eyceoz, G. Rossiello, A. M. Gliozzo, M. R. Glass, N. Mihindukulasooriya, **N. H. Pham**, L. H. Vu, D. Subramanian, F. M. Chowdhury. Generating structured query language queries from natural language inputs with schema enrichment. *Filed on Aug 19, 2025.*

2. T. R. Dinger, A. M. Gliozzo, **N. H. Pham**, O. Hassanzadeh, D. Subramanian, L. Amini, G. Rossiello, M. F. M. Chowdhury, L. VU, T. Kaple, M. Glass. Database Querying Using Natural Language Processing. *Filed on March 26, 2025.*

3. E. Lobo, **N. H. Pham**, L. Vu, T. Mummert, and D. Subramanian. A novel system for metadata to glossary matching in data lakes using generative models. *Filed on August 22, 2024.*

4. L. Vu, **N. H. Pham**, D. Subramanian, T. Mummert. System and Method for Combining Data Selection and Reward Function for Tuning LLMs using Reinforcement Learning. *Filed on Jan. 11, 2024.*

5. T. L. Hoang, M. M. Galindo, G. Picco, M. Zayats, **N. H. Pham**, L. M. Nguyen, M. L. Sbodio, D. T. Phan, and V. L. Garcia. Evolution and regression generative models and sequence representation learning from multi sequence alignment and phylogenetic trees data. *Filed on Jun. 30, 2023.*

6. L. Vu, P. Kirchner, R. Marinescu, D. Subramanian, and **N. H. Pham**. A novel meta-hyperparameter tuning system for RL using sequence model. *Filed on Jun. 26, 2023.*

7. E. Lobo, **N. H. Pham**, D. Subramanian, and T. Pedapati. A novel meta-hyperparameter tuning system for RL using sequence model. *Filed on Jun. 23, 2023.*

8. **N. H. Pham**, L. M. Nguyen, J. Chen, T. L. Hoang, S. Das. A systematic approach for evaluating robustness of cooperative multi-agent reinforcement learning. *Filed on Sep. 28, 2022.*

9. D. T. Phan, **N. H. Pham**, L. M. Nguyen. Site-Wide Optimization for Mixed Regression Models and Mixed Control Variables. *Filed on May 25, 2021.*

TUTORIALS/LAB

AAAI 2023 Tutorial and Lab Organizer   *Feb. 2023*
Title: *Automated AI For Decision Optimization with Reinforcement Learning.*

INVITED TALKS

MIT-IBM Guest Seminar   *Jul. 2021*
Title: *Stochastic Recursive Gradient Algorithms for Stochastic Composite Nonconvex Optimization and Policy Optimization.*

INFORMS Annual Meeting 2020 Virtual   *Nov. 2020*
Title: *A Hybrid Stochastic Policy Gradient Algorithm for Reinforcement Learning.*

INFORMS Annual Meeting 2019 Seattle   *Oct. 2019*
Title: *ProxSARAH: An Efficient Algorithmic Framework for Stochastic Composite Nonconvex Optimization.*

SKILLS & QUALIFICATIONS

Technical   Python, Tensorflow, Keras, Pytorch, Scikit-learn, C/C++, Matlab, Data Structures & Algorithms
Other skills   Linux Development Environment, Robotics, Embedded Systems

HONORS & AWARDS

IBM Outstanding Technical Achievement Award   *2025*
IBM Granite Text-to-SQL: BIRD Leaderboard First Place

IBM Research Accomplishments   *2025*
IBM Granite fine-tuned Text-to-SQL models sweep top spots in BIRD Leaderboard (A-level)

IBM Research Pat Goldberg Memorial Best Paper   *2022*
IBM Research

PROFESSIONAL SERVICES

REVIEWER (PEER-REVIEWED CONFERENCES)

International Conference on Machine Learning (ICML)   *2020–present*

Conference on Neural Information Processing Systems (NIPS/NeurIPS)   *2020–present*

International Conference on Artificial Intelligence and Statistics (AISTATS)   *2021–present*

International Conference on Learning Representations (ICLR)	*2021–present*
AAAI Conference on Artificial Intelligence	*2022–present*
Conference on Uncertainty in Artificial Intelligence	*2022*

REVIEWER (PEER-REVIEWED JOURNALS)

Automatica	*2024*
Applied Intelligence	*2023–2024*
Transactions on Machine Learning Research (TMLR)	*2023–2024*
IEEE Transactions on Neural Networks and Learning Systems (IEEE TNNLS)	*2022–2023*
Machine Learning	*2021–2023*
Journal of Machine Learning Research (JMLR)	*2022–2023*
IEEE Transactions on Automatic Control (IEEE TAC)	*2022–2023*
IMA Journal of Numerical Analysis (IMAJNA)	*2022–2023*
Journal of Scientific Computing (JOMP)	*2022*
Neural Networks (NEUNET)	*2022*
Computational Optimization and Applications (COAP)	*2021–2022*