Research Statement

Nghia Nguyen-Trung

Ph.D. Candidate, Statistics and Operations Research, UNC–Chapel Hill · nghiant@unc.edu · nghiant01.github.io · *Updated November 2026*

1 Overview

My research lies at the intersection of optimization, machine learning, and decision-making: I study first-order methods for *generalized equations*, also called composite inclusions, a model that covers minimization, minimax problems, variational inequalities, and fixed-point problems.

A standard optimization problem minimizes one objective function, but many problems in machine learning and decision-making cannot be stated this way, because they involve competing objectives, or conditions that come from no single objective. Training a model that is robust to adversarial attacks or data shifts is a game between a learner and an adversary; firms competing in a market, or devices sharing a power grid, are described by an equilibrium of a game; evaluating a policy in reinforcement learning means solving a Bellman equation. For many such problems, the optimality or equilibrium conditions can be written in a common form called a generalized equation:

$$\text{find } x^\star \text{ such that} \quad 0 \in Gx^\star + Tx^\star.$$

Here G describes how the decision variables "want to move" (for minimization, G is the gradient), and T encodes constraints or nonsmooth regularizers; because the model is general, one method and one analysis can serve many of these problems at once.

Why these problems are hard. First, they often lack *monotonicity*, the convexity-type structure on which the classical theory relies; without it, natural methods can fail, and in a simple two-player game, letting each player follow its own gradient can cycle or diverge. Second, the problems are large: G is typically an average over millions of data points, the variable can have millions of coordinates, and the data may sit on machines or devices connected by a network. An algorithm then cannot afford, or is not allowed, to compute the exact, current value of G at every step; it must use noisy samples, part of the variables, outdated values, or local information.

My research goal. I develop first-order methods for generalized equations that are fast, reliable, and practical on today's data and computing systems, where each step often has only incomplete information, with a focus on *accelerated* methods, whose error decreases faster than that of standard methods, and on guarantees for the *last iterate*, the point that an algorithm actually returns. Because a generalized equation usually has no objective to decrease, my main analytical tool is a Lyapunov (energy) function built for each method. Such guarantees tell practitioners which method to use, how to set its parameters, and how much computation and communication are needed (Figure 1).

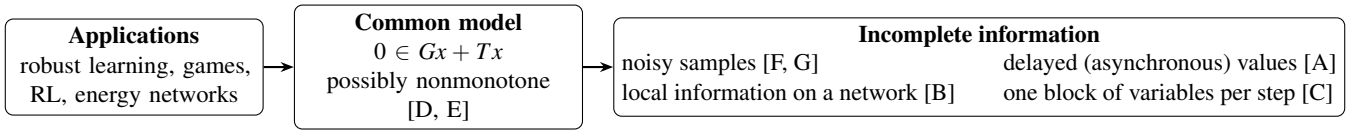


Figure 1: Many applications, one common model, and the incomplete-information settings I address ([A–C]: first author).

2 Research to date

Background. At Hanoi University of Science and Technology, I worked with Nguyen Thi Thu Thuy on self-adaptive projection methods for split feasibility and split variational inequality problems, with applications to optimal control and sparse recovery; this work led to six journal papers (see CV).

Ph.D. research. At UNC–Chapel Hill, I have written seven papers: one is published, one is accepted, two are under revision, and three are preprints. My collaboration with my advisor, Quoc Tran-Dinh, evolved from advisor-initiated projects [D, E, F, G], where I worked on the development and refinement of the convergence proofs and carried out all numerical experiments, to first-author projects [A, B, C], where I led the problem formulation, algorithm design, convergence analysis, and implementation.

2.1 Fast methods without monotonicity

The extragradient method fixes the cycling of simple gradient methods in games by taking a "look-ahead" step before each update; its many variants were usually analyzed one at a time and mostly for monotone problems. In [D], we built unified frameworks that contain many existing extragradient-type and operator-splitting methods, together with

new ones, and analyzed them under two conditions that allow a controlled amount of nonmonotonicity. In [E], we developed accelerated versions, based on anchoring and Nesterov's technique, that provably converge faster than the basic extragradient method. Since one analysis covers a whole family, users can choose a method by its cost without losing guarantees.

2.2 Noisy information: stochastic methods with variance reduction

When G is an average over a large dataset or an expectation, each step can only use a sample-based estimate of G, whose noise is hard to control without monotonicity; variance reduction reuses information from past samples so that this error shrinks as the method converges. In [F], we combined acceleration with variance reduction for possibly nonmonotone problems; one condition on the estimator covers several popular estimators, and the methods converge faster than non-accelerated stochastic methods, with guarantees that hold on essentially every run, not only on average. In [G], we used, to our knowledge for the first time for this problem class, biased variance-reduced estimators, analyzed how many evaluations of G they need, and tested the methods on AUC maximization for imbalanced data and on policy evaluation in reinforcement learning.

2.3 Delayed, local, and partial information (first-author work)

Delayed information. Asynchronous parallel methods avoid waiting for the slowest processor by using values of G computed at older iterates; in [A], I developed accelerated fixed-point methods that work with delayed and inexact information. One error condition covers several asynchronous architectures, and the methods converge faster than classical and asynchronous fixed-point methods, with a cost that grows only linearly with the maximum delay; the analysis also gives stochastic, incremental, and shuffling variants, tested on games and neural networks.

Information spread over a network. Often each agent (a machine, a device, or an energy resource) knows only its own part of the problem and talks only to its neighbors. In [B], with Ion Necoara (Politehnica Bucharest) and my advisor, I developed, to our knowledge, the first accelerated decentralized methods for composite monotone inclusions. In one method, each agent chooses its own step size without global information about the network; both converge at accelerated rates for the error of the original problem at the network-average iterate and for the disagreement between agents. Experiments include distributed games and a simulated virtual power plant, in which distributed energy resources coordinate to act as one power plant.

Partial updates. When data and variables are both numerous, even one full update is expensive; in [C], with the same collaborators, I combined variance reduction with updates of one randomly chosen block of variables per step, and analyzed the methods without monotonicity. A sampling rule that favors important blocks exploits large differences between blocks; experiments include a quadratic minimax problem with a nonconvex regularizer.

These projects show that acceleration can be kept when each step sees only delayed, local, or partial information; they are the starting point of my independent program.

3 Research plan

I want fast methods for generalized equations to work on realistic computing systems, which are asynchronous, networked, and run without hand-tuned parameters, and to support decisions made from uncertain data.

3.1 Years 1–2: building on my first-author work

Asynchronous methods with noisy data. My delayed-oracle methods [A] handle noise only under a bounded-variance assumption, without variance reduction; for monotone inclusions, asynchronous [2] and variance-reduced [13] methods have been developed separately; their combination has been studied for minimization [5, 6]. I plan to combine the variance-reduced estimators of [F, C] with the delayed oracles of [A], while keeping accelerated rates. A main difficulty is that variance-reduced estimators already reuse old information, so their staleness interacts with the communication delay; a second goal is one model with noise, delay, and block updates, and a bound showing the cost of each.

Directed networks. The analysis in [B] assumes two-way communication and a symmetric mixing matrix, but many networks have one-directional links. I plan to extend the accelerated methods of [B] to directed networks by adding push-sum or tracking-type corrections; both projects use tools I already have and are natural first papers in a new position.

3.2 Years 3–5: broader settings

Realistic delays and dynamic networks. In [A], the cost depends on the worst-case delay, which can far exceed a typical one, and acceleration requires co-coercivity, which fails for many minimax problems. For asynchronous SGD in minimization, recent bounds depend on the average delay instead [3, 4]; I want accelerated rates for generalized equations with such delay-sensitive costs, beyond co-coercivity. In sensor, federated, and energy networks, links also change and resources join or leave; methods over time-varying or directed networks [7, 8] exist for (strongly) monotone problems but lack the accelerated guarantees of [B], and I plan to study such networks, including nonmonotone local operators.

Decision-making under uncertainty. Many robust and distributionally robust decision models can be formulated as minimax problems, in which a decision maker hedges against the worst case over plausible data distributions; with large or nonconvex models, these are large-scale and may lack monotonicity. I plan to study how the techniques developed in my work, such as variance reduction [F, G], randomized coordinate updates [C], and decentralized computation [B], can be adapted to such models, so that robust decisions can be computed reliably from sampled or distributed data.

Adaptive step sizes. My step sizes depend on constants, such as Lipschitz constants, that are often unknown or conservative; adaptive methods [1, 11] estimate them locally but are deterministic and non-accelerated, and I plan adaptive versions of my stochastic and decentralized methods, with random or agent-local estimates [12].

Reinforcement learning. Value iteration is a fixed-point method for the Bellman operator and is slow when future rewards are discounted only slightly; anchoring gives a near-optimal rate here [9], with stochastic versions for Q-learning [10]. I want to know whether such acceleration survives asynchronous, partial, or decentralized updates, which requires extending my Lyapunov tools to the maximum norm, in which the Bellman operator contracts.

3.3 Students, collaboration, and funding

This agenda offers projects at every level: Ph.D. students can develop the theory for directed networks, adaptive stochastic methods, or robust decision models; master's and undergraduate students can implement the methods, run simulations, and test them on data, for example on robust portfolio models such as the one my two undergraduate mentees are now building. I will continue my current collaborations and would welcome work with colleagues in statistics, machine learning, and engineering. These directions fit the NSF Division of Mathematical Sciences and the DOE Office of Science, whose Applied Mathematics program supports numerical optimization and methods that use asynchrony or randomness, including their early-career programs; the networked-energy applications offer a basis for joint proposals with engineering colleagues.

References

[A] **N. Nguyen-Trung** and Q. Tran-Dinh. A class of accelerated fixed-point-based methods with delayed inexact oracles and its applications. arXiv:2512.13547, 2025.

[B] **N. Nguyen-Trung**, I. Necoara, and Q. Tran-Dinh. Distributed fast fixed-point algorithms for composite monotone inclusions over networks. arXiv:2609.14953, 2026.

[C] **N. Nguyen-Trung**, I. Necoara, and Q. Tran-Dinh. Optimistic gradient algorithms with randomized coordinate and variance reduction strategies for composite inclusions. Preprint, 2026. DOI:10.13140/RG.2.2.26234.86726.

[D] Q. Tran-Dinh and **N. Nguyen-Trung**. Revisiting extragradient-type methods – Part 1: Generalizations and sublinear convergence rates. *Comput. Optim. Appl.*, 94(1):141–210, 2026.

[E] Q. Tran-Dinh and **N. Nguyen-Trung**. Revisiting extragradient-type methods – Part 2: Generalization and unified convergence analysis of anchored extragradient-type variants. *Comput. Optim. Appl.*, to appear, 2026. DOI:10.1007/s10589-026-00851-8.

[F] Q. Tran-Dinh and **N. Nguyen-Trung**. New accelerated past-extragradient methods with variance reduction for generalized equations. arXiv:2508.16791, 2025. Under revision at *Math. Oper. Res.*

[G] Q. Tran-Dinh and **N. Nguyen-Trung**. Unbiased and biased variance-reduced forward-reflected-backward splitting methods for stochastic composite inclusions. arXiv:2603.15576, 2026. Under revision at *Comput. Optim. Appl.*

[1] A. Alacaoglu, A. Böhm, and Y. Malitsky. Beyond the golden ratio for variational inequality algorithms. *J. Mach. Learn. Res.*, 24(172):1–33, 2023.

[2] P. L. Combettes and J. Eckstein. Asynchronous block-iterative primal-dual decomposition methods for monotone inclusions. *Math. Program.*, 168(1–2):645–672, 2018.

[3] H. R. Feyzmahdavian and M. Johansson. Asynchronous iterations in optimization: New sequence results and sharper algorithmic guarantees. arXiv:2109.04522, 2023.

[4] A. Koloskova, S. U. Stich, and M. Jaggi. Sharper convergence guarantees for asynchronous SGD for distributed and federated learning. *NeurIPS*, 2022.

[5] S. J. Reddi, A. Hefny, S. Sra, et al. On variance reduction in stochastic gradient descent and its asynchronous variants. *NeurIPS*, 2015.

[6] R. Leblond, F. Pedregosa, and S. Lacoste-Julien. Improved asynchronous parallel optimization analysis for stochastic incremental methods. *J. Mach. Learn. Res.*, 19(81):1–68, 2018.

[7] D. Kovalev, A. Beznosikov, A. Sadiev, et al. Optimal algorithms for decentralized stochastic variational inequalities. *NeurIPS*, 2022.

[8] M. Bianchi and S. Grammatico. Fully distributed Nash equilibrium seeking over time-varying communication networks with linear convergence rate. *IEEE Control Syst. Lett.*, 5(2):499–504, 2021.

[9] J. Lee and E. K. Ryu. Accelerating value iteration with anchoring. *NeurIPS*, 2023.

[10] M. Bravo and J. P. Contreras. Stochastic Halpern iteration in normed spaces and applications to reinforcement learning. *Math. Program.*, 2026.

[11] L. Shen and F. Kılınç-Karzan. Parameter-free non-ergodic extragradient algorithms for solving monotone variational inequalities. arXiv:2604.07662, 2026.

[12] D. Ghaderyan and S. Werner. AdGT: Decentralized gradient tracking with adaptive per-agent stepsizes. *IEEE Trans. Signal Process.*, 74:4097–4113, 2026.

[13] X. Zhang, W. B. Haskell, and Z. Ye. A unifying framework for variance-reduced algorithms for finding zeroes of monotone operators. *J. Mach. Learn. Res.*, 23(60):1–44, 2022.