RESEARCH STATEMENT

Jiajin Li
Stanford University

My research interests lie in mathematical optimization and its applications in data-driven decision making, machine learning, and data science. My goal is to develop novel theoretical frameworks to study and analyze the convergence and statistical behaviors of optimization algorithms. By leveraging these insights from theoretical development, I design efficient algorithms tailored to data-driven optimization problems in these areas, such as graph learning, finance, experimental design, etc.

1 Research Overview & Philosophy

Most of my research has contributed to structure-driven algorithm design and the development of its convergence analysis through error bound theory, see Figure 1 for a broad overview.

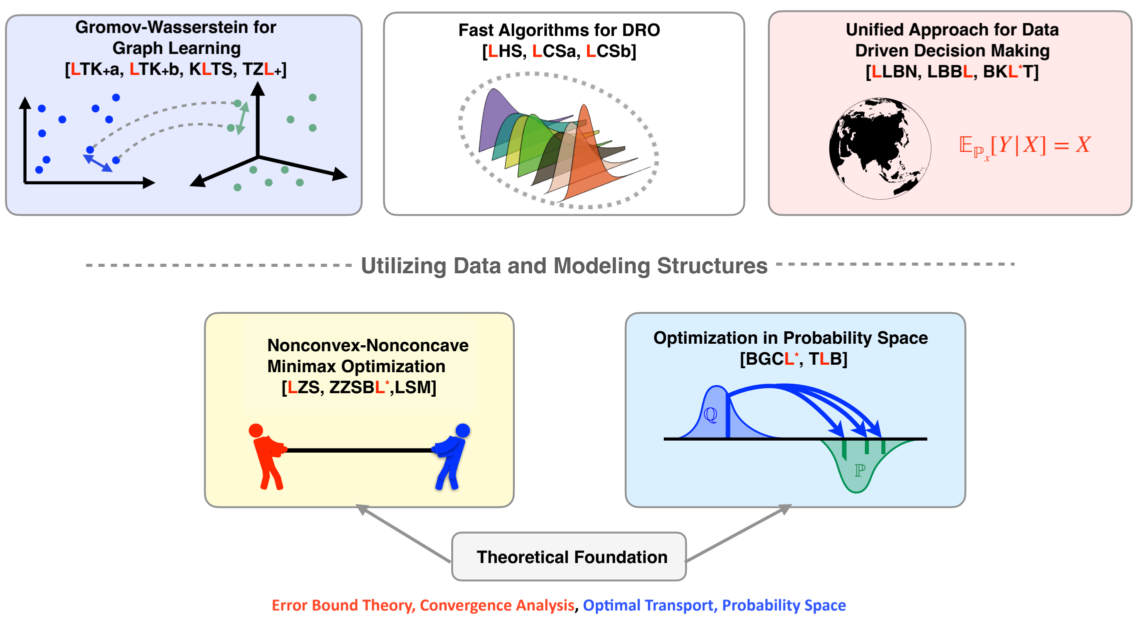

Figure 1: Optimization-centric Research

Now, I will discuss my research pipeline and philosophy, which motivate me to complete these lines of work.

The explosion of data and computing resources has led to the surge of automatic decision-making tools and modern machine learning methodologies. This data-driven revolution is underpinned by ever-evolving research in data-driven modeling and large-scale optimization. At the start of my PhD study in optimization, I was frequently frustrated by the inefficiency of classic optimization algorithms, such as gradient descent and subgradient methods, when applied to data-driven structured optimization problems. I gradually realized that this is the price of their generality, as classic optimization algorithms are typically designed to work for general problems (e.g., gradient Lipschitz functions), and thus do not take into account data and modeling structures. To ensure practical efficiency, I have needed to design optimization algorithms tailored to these particular structures.

Taking a step further, can we conduct convergence analyses for the proposed algorithms? Otherwise, we cannot make any statements about their worst-case performance. In my view, two general regularity conditions are essential for establishing sharp convergence rates of optimization algorithms: `smoothness` and `error bound conditions` (e.g., quantifying the perturbation quantity of the neighborhood around the optimal solution). Smoothness condition for a given problem is usually easy to verify, as intuitively it corresponds to whether gradient information is available. However, error bound analysis is typically treated to be independent of algorithm design, which makes algorithm design and convergence analysis two separate tasks. This can lead to suboptimal rates. Therefore, there is no theoretical guideline for designing optimization algorithms with strong performance guarantees. To bridge this gap, my research is committed to incorporating error bound analysis with algorithm design, so that convergence analysis and algorithm design can reinforce each other.

Later, when I started my postdoctoral study with Prof. Jose Blanchet at Stanford, I expanded my research area to data-driven stochastic modeling. This collaboration inspired me to rethink my research goals. I realized that the regularity conditions used to conduct convergence analysis of optimization algorithms can also provide strong practical guidelines and insights on how to design computationally amenable data-driven optimization models. All in all, past research experience and understanding on optimization and data-driven modeling have gradually formed my research pipeline as follows:

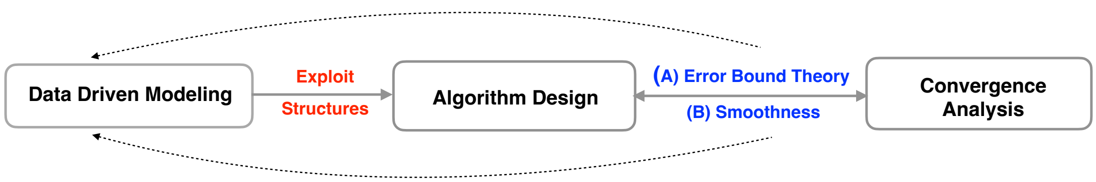

Figure 2: My Research Pipeline

2 Summary of Past and Ongoing Research

In this section, I will provide a detailed introduction to five lines of work illustrated in Figure 1, demonstrating the effectiveness of my research pipeline.

2.1 Nonconvex-Nonconcave Minimax Optimization

Nonconvex-nonconcave minimax optimization has attracted significant attention in machine learning and data science, particularly due to its close relation to robust training of neural networks and generative adversarial networks (GANs). My keen interest in this topic dates back to my PhD studies, when I began a line of research projects on designing efficient first-order algorithms for distributionally robust optimization (DRO) problems [LHS19, LCS23, LCS20, Li21]. In general, all DRO models can be formulated as minimax games, where a decision-maker (primal) engages in a game against a fictitious adversary (dual) to analyze the potential consequences of the worst-case attack.

The key to minimax optimization is balancing between the primal and dual updates. The optimality of this balance directly impacts the convergence rate. In my recent work [LZS23], I introduce a new concept called the primal-dual error bound, which characterizes "the degree" of this balance using tailored error bound theory. This fresh perspective provides a new algorithm design principle: optimal primal-dual balancing. That is, we should pay more attention to the player with the worse growth condition. The best convergence rate we can obtain is determined by the slower of the primal and dual updates. Moreover, this primal-dual error bound serves as a milestone for developing a unified convergence analysis framework for a broad class of nonconvex-nonconcave minimax optimization problems, culminating in the first universally applicable algorithm.

I believe that my works [LZS23, ZZS+23] make significant contributions to the minimax optimization literature. Before these works, the development of minimax optimization algorithms and their convergence analyses was still in its infancy stage. I summarize three major breakthroughs:

1. My works identify checkable regularity conditions for the largest class of nonsmooth nonconvex-nonconcave optimization problems with optimal convergence rates. Existing optimal methods only apply to smooth problems with gradient information [LSM20] or rely on other uncheckable conditions.

2. My works propose the first universally applicable algorithm for minimax optimization. This universality is crucial in practice as we often do not know which player has better properties, such as convexity or concavity.

3. My works also establish the first algorithm-independent quantitative relationships between ϵ-game stationarity and ϵ-optimization stationarity concepts for any $\epsilon \in (0, 1)$. These results demystify various notions of stationarity in the context of minimax optimization.

2.2 Optimization in Probability Space

Optimization of infinite-dimensional functionals of probability measures naturally arises in a wide range of problems, such as GANs, sequential decision making, and optimal control. In my recent work [BGKL23], a key observation is that we can leverage both Wasserstein geometry and the strong duality results recently developed in DRO to execute the gradient step in probability space. It turns out that the developed modified Frank-Wolfe step (gradient step) in probability space can be reduced to finite-dimensional convex optimization problems, leading to the first practically implementable algorithm to optimize the probability measure in the literature. On the theoretical side, our work gives the first non-asymptotic convergence rate under error bound type conditions, which matches the analogous results of finite-dimensional optimization problems in Euclidean geometry with similar regularity conditions. Later on, my collaborators and I extended [BGKL23] to weak the smoothness assumption by exploiting Bregman geometry [BLT23].

2.3 Fast First-Order Algorithms for DRO

As I mentioned, the first line of my research aims to realize the benefits of DRO in practice.

Many distributionally robust learning models (e.g., logistic regression, support vector machine) can be equivalently reformulated as tractable conic programs when the loss function is convex. These conic programs can then be solved by general-purpose off-the-shelf solvers, such as MOSEK and Gurobi. However, these general-purpose solvers do not scale well to large problems and can be slow in practice, as they do not exploit any useful latent structures in the problem.

To resolve this issue, my research has developed a first-order algorithmic framework to efficiently solve a class of distributionally robust generalized linear classification models [LHS19, LCS23, LCS20, Li21]. The key novelty of my work is the careful identification and exploitation of useful latent structures in the problem. This results in the first practical algorithm to achieve faster convergence rates in theory and practice. Indeed, our wall-clock time is up to **1000+** times faster than the standard off-the-shelf solver, with the performance gap growing with problem size.

2.4 A Unified Approach for DRO

During my PhD study, as an optimization researcher, I typically overlooked modeling design and statistical properties of data-driven optimization problems, since my only objective was to achieve the global optimum or stationary point of a fixed problem more quickly. However, after I started my Postdoc study, I was inspired to use my mathematical optimization expertise to develop new data-driven optimization models with computational tractability in mind.

My recent work [BKLT23] introduces a novel approach that unifies most of existing DRO models into a single framework using optimal transport (OT) with martingale constraints, while maintaining computational tractability. Moreover, this approach allows us to incorporate distributional uncertainty sets to address misspecification for both likelihoods and actual outcomes. Also, we find out that the incorporation of martingale constraints in conventional DRO models has far-reaching implications in the supervise learning [LLBN22] and dynamic learning setting [LBBL23].

2.5 Gromov Wasserstein for Graph Learning

A significant portion of my research involves close collaboration and discussion with researchers in applied areas such as graph learning [LTK+23a, LTK+23b, KLJS23, TLGL22, SLS22], NLP [TZL+23], finance [BLPZ23] and experimental design [LLYB23]. In this process, I use my research experience in optimization to solve real-world problems.

Below, I will use my line of research on the Gromov-Wasserstein (GW) distance for graph learning to highlight this research experience. The GW distance provides a flexible way to compare and couple probability distributions supported on different metric spaces. It has been applied to various structural data analysis tasks, such as cross-lingual knowledge graph (KG) alignment in NLP for machine translation, social network analysis, and shape correspondence in computer graphics.

My recent work [LTK+23a] provides the first provable single-loop algorithm — BAPG for approximately computing the Gromov-Wasserstein (GW) distance, which achieves state-of-the-art (SOTA) performance in various tasks, including graph alignment and partition. Inspired by error bound condition satisfied by GW problem, I introduce a novel relaxation technique to balance accuracy and computational efficiency. BAPG will be included in the `POT: Python Optimal Transport` soon. Based on [LTK+23a], my coauthors and I developed an unsupervised graph alignment framework that jointly performs structure learning and GW-based graph alignment. This method achieves SOTA performance on the DBP15K KG alignment benchmark dataset[1]. We also fully investigated a robust version of the GW distance to account for outliers in real-world applications [KLJS23], which achieves SOTA performance on several social network analysis datasets, including Douban Online-Offline.

[1]DBP15k contains four language-specific KGs that are respectively extracted from English, Chinese, French and Japanese, which is the most popular dataset for the evaluation of entity alignment tasks.

3 Future Research Agenda

My long-term research goal is to make far-reaching impacts in both theory and practice. Building on the research pipeline I introduced in the first section, the research direction I would like to expand is to consider both smoothness and error bound conditions of optimization problems into data-driven modeling design. This direction has become increasingly pressing and relevant to practitioners as more structured data-driven decision-making models emerge.

To be specific, what I have discovered to be both challenging and intriguing in this field is to balance computational effectiveness and statistical efficiency across various optimization algorithms and statistical (data-driven) models. Optimizers typically focus on developing effective algorithms for a specific abstract optimization problem at hand. By contrast, as statisticians often prioritize the statistical properties of the global optimal solution, the resulting optimization problem may lack certain desirable properties such as smoothness and error bound conditions. Consequently, existing iterative methods may face challenges such as slow convergence rates and heavy computational burden. My work will offer a comprehensive perspective and answer on how to optimally close this gap for different data-driven decision-making problems.

Additionally, as an optimizer, a natural direction to further explore is general coupled nonconvex-nonconcave minimax optimization problems, which can cover a lot of operation management and machine learning problems, such as mechanism design, inventory sharing, meta learning, game theory, etc. A long-standing open problem is to find the largest problem class for which we can develop a universally applicable first-order algorithm to achieve the polynomial rate to any types of stationary points. Besides that, I am also keenly interested in demystifying the different stationary concepts. On the practical side, my long-term goal is to produce an open-source optimization package for minimax optimization problems, which could have a high impact among operations research, machine learning, and data science communities.

With my rich experience in both optimization and data-driven decision making, I am the right person to pursue them.

References

[BGKL23] Jose Blanchet, Peter W Glynn, Carson Kent, and Jiajin Li. Modified Frank-Wolfe in probability space. *To be submitted to Mathematical Programming. Preliminary version appears in NeurIPS 2021.*, 2023.

[BKLT23] Jose Blanchet, Daniel Kuhn, Jiajin Li, and Bahar Taskesen. Unifying distributionally robust optimization via optimal transport theory. *To be submitted to SIAM Journal on Optimization*, 2023.

[BLPZ23] Jose Blanchet, Jiajin Li, Markus Pelger, and Greg Zanotti. Optimistic optimal transport automatic outlier removal mechanism. *Woking paper*, 2023.

[BLT23] Jose Blanchet, Jiajin Li, and Lezhi Tan. Generalized Frank-Wolfe method for optimization in probability space via Bregman proximal mapping. *In preparation for Mathematics of Operation Research*, 2023.

[KLJS23] Lemin Kong, Jiajin Li, Tang Jianheng, and Anthony Man-Cho So. Outlier-robust Gromov-Wasserstein in graph data. *Advances in Neural Information Processing Systems (Spotlight)*, 2023.

[LBBL23] Kyriakos Lotidi, Nicholas Bambos, Jose Blanchet, and Jiajin Li. Wasserstein distributionally robust linear-quadratic estimation under martingale constraints. *The 25th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2023.

[LCS20] Jiajin Li, Caihua Chen, and Anthony Man-Cho So. Fast epigraphical projection-based incremental algorithms for wasserstein distributionally robust support vector machine. *Advances in Neural Information Processing Systems*, 33:4029–4039, 2020.

[LCS23] Jiajin Li, Caihua Chen, and Anthony Man-Cho So. Towards a first-order algorithmic framework for distributionally robust optimization. *To be submitted to Mathematical Programming*, 2023.

[LHS19] Jiajin Li, Sen Huang, and Anthony Man-Cho So. A first-order algorithmic framework for distributionally robust logistic regression. *Advances in Neural Information Processing Systems*, 32, 2019.

[Li21] Jiajin Li. Efficient and provable algorithms for wasserstein distributionally robust optimization in machine learning. *Ph.D. Thesis*, 2021.

[LLBN22] Jiajin Li, Sirui Lin, Jose Blanchet, and Viet Anh Nguyen. Tikhonov regularization is optimal transport robust under martingale constraints. *Advances in Neural Information Processing Systems*, 2022.

[LLYB23] Yiping Lu, Jiajin Li, Lexing Ying, and Jose Blanchet. Synthetic principle component design: Fast experimental design with synthetic controls. *NeurIPS 2022 Workshop on Causality for Real-world Impact*, 2023.

[LSM20] Jiajin Li, Anthony Man-Cho So, and Wing-Kin Ma. Understanding notions of stationarity in nonsmooth optimization: A guided tour of various constructions of subdifferential for nonsmooth functions. *IEEE Signal Processing Magazine*, 37(5):18–31, 2020.

[LTK+23a] Jiajin Li, Jianheng Tang, Lemin Kong, Huikang Liu, Jia Li, Anthony Man-Cho So, and Jose Blanchet. A convergent single-loop algorithm for relaxation of gromov-wasserstein in graph data. *International Conference on Learning Representation*, 2023.

[LTK+23b] Jiajin Li, Jianheng Tang, Lemin Kong, Huikang Liu, Jia Li, Anthony Man-Cho So, and Jose Blanchet. Trade-off among infeasibility, efficiency and accuracy for gromov-wasserstein computation. *To be submitted to Journal of Machine Learning Research*, 2023.

[LZS23] Jiajin Li, Linglingzhi Zhu, and Anthony Man-Cho So. Nonsmooth nonconvex-nonconcave minimax optimization: Primal-dual balancing and iteration complexity analysis. *Under review at Mathematical Programming*, 2023.

[SLS22] Oded Stein, Jiajin Li, and Justin Solomon. A splitting scheme for flip-free distortion energies. *SIAM Journal on Imaging Sciences*, 15(2):925–959, 2022.

[TLGL22] Jianheng Tang, Jiajin Li, Ziqi Gao, and Jia Li. Rethinking graph neural networks for anomaly detection. *International Conference on Machine Learning*, 2022.

[TZL+23] Jianheng Tang, Weiqi Zhang, Jiajin Li, Kangfei Zhao, Fugee Tsung, and Jia Li. Robust attributed graph alignment via joint structure learning and optimal transport. *IEEE International Conference on Data Engineering (ICDE)*, 2023.

[ZZS+23] Taoli Zheng, Linglingzhi Zhu, Anthony Man-Cho So, Jose Blanchet, and Jiajin Li. Doubly smoothed GDA: Global convergent algorithm for constrained nonconvex-nonconcave minimax optimization. *Advances in Neural Information Processing Systems*, 2023.