Statement of Purpose

Xingjian Bai *Ph.D. Applicant*

My research interests lie at the intersection of algorithms, mathematics, and deep learning – a domain where elegant theories meet practical models. I have explored these intersections through several research projects. These include adapting classic algorithms to leverage learned predictions [1], and enhancing machine learning models through rigorous theoretical analysis, spanning the fields of adversarial robustness [2], generative modeling [3], and reinforcement learning [4]. Applying algorithmic insights and mathematical theories to deep learning, my primary research objective is to develop intelligent systems with both learning and logical abilities, able to perceive, think, and explore the world beyond human-level capabilities.

Algorithms were my passion in high school. Participating in the Chinese Olympiad, advancing to the USACO Camp, and winning 1st Place in the Canadian Computing Olympiad National Team Selection, I enjoyed learning algorithms with friends worldwide. Later, I delved deeper into algorithms with theoretical foundations, such as the Fourier Transform, the Miller-Rabin test, and Euler's iterative method. These algorithms reveal a complex world behind codes, of analytical and algebraic structures I could not yet understand. I majored in mathematics and computer science at Oxford to further explore these theories.

Studying more about machine learning (ML), I was captivated by its strong performance across various applications, despite the absence of worst-case guarantees. Its performance in tasks such as the game of Go and protein structure prediction well surpasses that of classic algorithms. This led me to a realization similar to what Prof. Sanjeev Arora described in his blog, *Off the Convex Path*: Beyond the idealized world of problems addressed by classic algorithms, there lies a broader set of challenges – erroneous, stochastic, and worst-case intractable by nature – yet they are prevalent in real-world scenarios. What, then, is the suitable paradigm to tackle these problems? Is it possible to develop solutions that are both practically effective and theoretically sound? Driven by these questions, I entered the world of deep learning, aiming to develop theory-inspired learning systems that are able to tackle real-world problems with reliability.

Learning-Augmented Sorting One of my projects is about using inaccurate advice from black-box machine learning models to speed up sorting, perhaps the most fundamental algorithmic task. This project followed the line of work on learning-augmented algorithms, initiated by Lykouris and Vassilvitskii [5], which seeks to craft algorithms with theoretical guarantees that can also leverage the power of ML models. While prior works have introduced approximated sorting algorithms utilizing noisy signals, no one has invented provable sorting algorithms to cope with erroneous advice. Working with Prof. Christian Coester at Oxford, I developed a framework where a quick-and-dirty comparison function is available, besides the original exact (but slow) comparator. I drew insights from self-balanced data structures and randomized algorithms, and designed a novel algorithm that can precisely sort with a small number of exact comparisons, if given dirty comparisons are mostly correct. Subsequently, I established the optimality of my algorithm and validated its efficacy through experiments.

I presented this work [1] at NeurIPS 2023. The proof of my algorithm is elegant enough that it draws the attention of many theoretical researchers. I have also discussed with researchers in applied drug discovery, who suggested that my findings could be used to reduce the comparative experiments needed for selecting molecular structures for drugs among thousands of candidates.

Wasserstein Distributional Adversarial Robustness of Neural Networks Unlike algorithms with provable guarantees, neural networks are vulnerable to worst-case attacks, making them unsafe for real-world applications. Despite the extensive literature on practical attacks and defenses [6,7], finding the right theoretical tools for analyzing neural networks' worst-case behavior is still an open challenge. Working with Prof. Jan Obloj at Oxford, I used Optimal Transport to lift the robustness problem into infinite-dimensional space, then applied tools in Distributional Robust Optimization (DRO) sensitivity analysis. As the only team member with an ML background, I worked on translating mathematical concepts into functional algorithms. I proposed a new loss function, Rectified DLR, along with an attack algorithm with certified bounds. Further, I integrated off-policy techniques from Reinforcement Learning to mitigate the computational cost of minimax optimization and proposed a robust training algorithm. This work [2] has appeared at NeurIPS 2023.

Neurosymbolic Diffusion Models Generative diffusion models have the strengths of "System 1 thinking"[1]: they excel at rendering emotions, styles, and scene diversity. However, they fall short in compositionality, consistency, and generalizability, the hallmarks of a logical system. To address these limitations, I worked with Prof. Jiajun Wu at Stanford to improve the relational compositionality of diffusion models. Diverging from traditional end-to-end approaches, I developed a two-stage pipeline using bounding boxes as a bridge and decomposed image generation into the generation of individual objects and relations. On CLEVR dataset, my model achieved the best performance in generating images with multiple objects and relations.

Fixed Point Diffusion Models The heavy memory and computational demand of diffusion models present a significant bottleneck for their deployment in production environments. In my fourth-year thesis at the Visual Geometry Group Oxford, I tackled this issue by embedding an implicit fixed-point solving layer within the denoising network of diffusion models. This approach transformed the diffusion process into a sequence of closely-related fixed point problems, resulting in reduced model size and memory usage, and enabling information sharing across diffusion timesteps. My empirical evaluations on four datasets demonstrated that this model surpasses state-of-the-art diffusion models in both performance and efficiency. This work is under review for **CVPR 2024**.

Future Directions My research goal is to develop theory-inspired learning systems that combine the theoretical grounding of classic algorithms and mathematics with the seemingly unreasonable power of machine learning.

To start with, several questions immediately pique my interest:

- **How to design algorithms to leverage a spectrum of "dirty" predictions?** In multiple algorithmic and practical scenarios, such as sorting, we often rely on a potentially costly evaluator (e.g., comparator, attacker, or scoring system). ML models could be trained to approximate such evaluations, albeit with errors, resulting in a range of "dirty" predictors with varying accuracy and cost. I aim to develop algorithms that can utilize this spectrum of dirty predictions to produce provably efficient outputs.
- **How to effectively integrate learning components into classic algorithmic pipelines?** Previous work in learning-augmented algorithms predominantly used discrete signals to transfer information from learning components to classic algorithms, mainly for simplicity in complexity analysis. However, this approach fails to leverage the full potential of neural networks. It not only oversimplifies the signals that neural networks can produce, but also makes these neural networks unaware of the classic algorithms receiving their signals. To address this, I aim to design continuous, interactive, bi-directional interfaces between learning components and classic algorithms. These interfaces would enable dynamic and lossless interaction, leading to an integrated system with both learning capabilities and provable guarantees.
- **In statistical inference, integer programming, and optimization, what are some problems where ML can be applied to inform classic algorithms?** Traditionally focused on worst-case scenarios, these fields stand on the cusp of a paradigm shift with ML's ability to inform and refine algorithms. I aim to identify and tackle problems, such as mixed-integer programming and submodular function maximization, where the integration of ML can go beyond traditional worst-case analysis and unlock more efficient solutions.
- **How to leverage algorithmic techniques to enhance learning systems?** ML models, while proficient in a wide range of tasks, often lack consistency, interpretability, and reasoning capabilities. By integrating algorithmic tools such as recursion, dynamic programming, and bipartite matching into learning systems, we can introduce a logical component to elevate their logical abilities. The effectiveness of this approach is backed by the recent development of neural-symbolic models [10, 11]. As an initial attempt, I plan to develop implicit neural networks [8, 9] to incorporate algorithmic principles directly into learning systems. This approach not only aligns neural network training with algorithmic processes but also facilitates a natural synergy between networks and algorithms.

[1] as described in the book *Thinking, Fast and Slow*

References

[1] **Xingjian Bai** and Christian Coester. Sorting with Predictions. *NeurIPS 2023*. [arXiv]

[2] **Xingjian Bai**, Guangyi He, Yifan Jiang, and Jan Obloj. Wasserstein Distributional Robustness of Neural Networks. *NeurIPS 2023*. [arXiv]

[3] **Xingjian Bai** and Luke Melas-Kyriazi. Fixed Point Diffusion Models. *Under review at CVPR 2024.*

[4] Jacek Karwowski, Oliver Hayman, **Xingjian Bai**, Klaus Kiendlhofer, Charlie Griffin, and Joar Skalse. Goodhart's Law in Reinforcement Learning. *Under review at ICLR 2024.* [arXiv]

[5] Thodiris Lykouris and Sergei Vassilvitskii. Competitive Caching with Machine Learned Advice. *ICML 2018*.

[6] Francesco Croce and Matthias Hein. Reliable evaluation of adversarial robustness with an ensemble of diverse parameter-free attacks. *ICML 2020*.

[7] Francesco Croce, Maksym Andriushchenko, Vikash Sehwag, Edoardo Debenedetti, Nicolas Flammarion, Mung Chiang, Prateek Mittal, and Matthias Hein. RobustBench: a standardized adversarial robustness benchmark. *NeurIPS Datasets and Benchmarks 2021.*

[8] Shaojie Bai, J. Zico Kolter, Vladlen Koltun. Deep Equilibrium Models. *NeurIPS 2019.*

[9] Fangda Gu, Heng Chang, Wenwu Zhu, Somayeh Sojoudi, and Laurent El Ghaoui. Implicit Graph Neural Networks. *NeurIPS 2020.*

[10] Tanmay Gupta and Aniruddha Kembhavi. Visual Programming: Compositional visual reasoning without training. *CVPR 2023.*

[11] Dídac Surís, Sachit Menon, and Carl Vondrick. ViperGPT: Visual Inference via Python Execution for Reasoning. arXiv preprint arXiv:2303.08128.