

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 inaccurate 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]. Integrating the strength of algorithms and deep learning, my primary research objective is to develop intelligent systems with learning and logical abilities, able to perceive, think, and explore the world beyond human-level capabilities.

Algorithms were my idea of fun in high school. Participating in the Chinese Olympiad of Informatics, 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, encountering techniques like the Fast Fourier Transform, the Miller-Rabin test, and Euler’s iterative method. These experiences revealed 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 theoretical foundations.

As I learned more about machine learning (ML), I found it fascinating that, without assurances of worst-case performance, ML models deliver impressive outcomes in a wide range of real-world applications. Their success in the game of Go and protein structure prediction well surpass that of classic algorithms. This led me to a realization that is very 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 approach these problems? Motivated by this question, I entered the world of deep learning through an algorithmic lens, seeking to bring theoretical guarantees to the practical models.

Learning-Augmented Sorting One of my projects came about using inaccurate advice to speed up sorting, perhaps the most fundamental algorithmic task. This followed the line of work on learning-augmented algorithms, started by [5], which seeks to design algorithms with theoretical guarantees that can leverage predictions from ML models. Previous works have proposed approximated sorting algorithms with noisy signals, but no one has invented provably correct sorting algorithms to cope with erroneous advice. Working with Prof. Christian Coester at Oxford, I proposed a setting 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 clean comparisons, if given dirty comparisons are mostly correct. Subsequently, I proved the optimality of my algorithm and demonstrated its effectiveness in experiments.

This work [2] will appear at **NeurIPS 2023**. The proof of my algorithm is elegant enough that I was told by an attendee of a workshop that he was considering teaching it at his university. I have also discussed with applied drug discovery researchers, who suggest that my results can be applied to reduce the number of comparative experiments needed to select molecular structures 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, finding the right theoretical tools for analyzing the worst-case behaviour is still an open challenge. Working with Prof. Jan Obloj at Oxford, we used optimal transport to lift the robustness problem into infinite-dimensional space, then leveraged Distributional Robust Optimization (DRO) sensitivity analysis to evaluate robustness. As the only team member with an ML background, I worked on translating mathematical formulas into algorithms. Specifically, I proposed a new loss function, Rectified DLR, along with an adversarial attack algorithm with certified accuracy bound. Further, I employed off-policy learning techniques from RL to mitigate the minimax optimization cost, and proposed a new robust training algorithm. This work [3] will appear at **NeurIPS 2023**.

Neuro-symbolic control on diffusion models Diffusion models have all the strengths of “System 1 thinking”¹ for image generation tasks: 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 diffusion models, I developed a two-stage pipeline using bounding boxes as a bridge and decomposed image generation into the generation of individual objects and relations. This approach achieved state-of-the-art performance in generating multiple objects and multiple relations on synthetic datasets.

Future Directions My research goal is to develop learning systems that combine the reliability of classic algorithms with the seemingly unreasonable power of machine learning. Such systems would be more interpretable and composable than traditional ML models while surpassing classic algorithms in learning abilities.

To start with, several questions immediately pique my interest:

- **How to invent algorithms to leverage “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 tradeoffs. My goal is to develop algorithms with provable guarantees that can effectively utilize this spectrum of dirty predictions to produce provable outputs.
- **How to leverage algorithmic techniques to enhance ML models?** Classic algorithms offer a wealth of powerful tools – recursion, divide-and-conquer, dynamic programming, etc. Integrating these into learning systems can significantly enhance their logical reasoning capabilities. As a preliminary approach, I plan to use implicit neural networks to embed algorithmic principles within learning systems. The primary challenge involves finding appropriate settings and robustly training these sophisticated networks.

Oxford is an optimal place for me to pursue my research ambitions. It is a place I am already familiar with, and I have already had research interactions with many researchers here. I am particularly excited to continue working with **Professor Christian Coester**. Building upon our prior work in sorting, I plan to explore other settings where dirty subroutines can be leveraged. I am also keen on learning the Mirror Descent technique and applying it to devise algorithms for settings with uncertainty, such as the explorable heap selection. Further, I am keen to work with him on developing more efficient algorithms for NP-hard problems, utilizing ML models to inform our approach. I would also like to continue to work with **Professor Christian Rupprecht**, my thesis supervisor. With him, I would work on building machine learning models, possibly in the field of Computer Vision, endowed with nice theoretical properties. Specifically, I aim to further investigate fixed-point implicit neural networks and explore ways to merge algorithmic insights into learning systems.

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

References

- [1] **Xingjian Bai** and Christian Coester, “Sorting with Predictions,” in *Conference on Neural Information Processing Systems (NeurIPS)*, 2023, Available: <https://arxiv.org/abs/2311.00749>.
- [2] **Xingjian Bai**, Guangyi He, Yifan Jiang, and Jan Obloj, “Wasserstein Distributional Robustness of Neural Networks,” in *Conference on Neural Information Processing Systems (NeurIPS)*, 2023. Available: <https://arxiv.org/abs/2306.09844>.
- [3] **Xingjian Bai** and Luke Melas-Kyriazi, “Fixed Point Diffusion Models” Under review.
- [4] Jacek Karwowski, Oliver Hayman, **Xingjian Bai**, Klaus Kiendlhofer, Charlie Griffin, and Joar Skalse, “Goodhart’s Law in Reinforcement Learning,” Under review. Available: <https://arxiv.org/abs/2310.09144>.
- [5] Thodoris Lykouris and Sergei Vassilvitskii, “Competitive Caching with Machine Learned Advice,” in *Proc. of the 35th Int. Conf. on Machine Learning (ICML 2018)*, vol. 80, pp. 3302–3311, 2018.