

## **Chapter 7**

### **Afterword**

Dear Readers:

Congratulations on making it this far in the book. Even if you haven't read every chapter in full, I hope you've found parts of it useful and inspiring. If you are a practitioner, I hope this book convince you that theory can, at times, be genuinely useful in practice.

Before concluding this book, I would like to reflect on my journey into compositional optimization for advanced machine learning, which began in 2019. Before that, I was primarily focused on traditional stochastic optimization theory. During that time, we developed a stochastic algorithm for solving non-convex minimax optimization problems. In 2019, I spent a year in industry, where conversations with young professionals made me realize the importance of practicability. After coming back from the leave, I started to think about how to make the theory more practical. The first project was to apply our non-convex minimax optimization to AUC maximization for learning deep neural networks, leading us to achieve first place in the Stanford CheXpert competition for classifying X-ray images organized by Andrew Ng's ML group in 2020. In late 2020, my friend Shuiwang Ji at Texas A&M University introduced me to the MIT AICures challenge, which aimed to identify few molecules with properties suitable for COVID-19 drug development among many. Motivated by this challenge, I formulated the optimization problem of maximizing the empirical estimator of areas under precision-recall curves, known as average precision. This led me to define the novel finite-sum coupled compositional optimization (FCCO) framework. We developed the first algorithm for FCCO in 2021, which ultimately helped us win the MIT AICures Challenge.

As I explored further, I discovered broad applications of this framework in ML and AI, specifically in addressing the computational challenges inherent in contrastive learning, learning to rank, discriminative learning and continual learning. This series of work eventually led to the development of the LibAUC library for empirical X-risk minimization, which has since been downloaded over 100,000 times by researchers and developers across more than 85 countries. It also helped us to develop CLIP models better than OpenAI's models with 15 times less compute budget.

---

As I reflect on the journey that led to this book, I am reminded of the principle of ‘知行合一’ (Zhi Xing He Yi) by Wang, Yangming (a legendary sage of ancient China), cited in the preface. It is often translated as ‘unity of knowledge and action’, which I interpret as the idea that theory should guide practice, and practice can, in turn, inspire theory.

For decades, the field of machine learning has largely been framed through the lens of Empirical Risk Minimization (ERM). This book argues that such a view is increasingly insufficient for modern AI systems. As we have seen throughout these chapters, the “X” in EXM represents a diverse class of often non-decomposable objectives, such as AUC, ranking measures, cross-entropy loss with expensive normalization, and contrastive losses, which define the frontier of modern AI. The development of the LibAUC library and the success of the EXM framework in AI challenges have shown us that when we move beyond standard stochastic optimization, we unlock new levels of performance in critical domains like medical imaging and drug discovery. Yet, despite these successes, the systematic study of EXM is just beginning.

During my years as a graduate student, I immersed myself in many books on optimization and machine learning, which were instrumental in shaping my mathematical foundation. Now, after more than a decade of study and research, I am humbled to synthesize these insights—together with my own findings—into this book. I hope that the methods and theories presented here are not viewed as a final destination, but rather as a starting point. My hope is that this work encourages researchers to look beyond standard training loops and to explore new forms of “X-risks” that better capture the complexity of modern learning systems and the nuances of human intelligence and societal needs. Ultimately, I look forward to seeing how the next generation of researchers will build upon these ideas to bridge elegant mathematical theory with transformative real-world applications.

Finally, I should note that this book may contain typographical errors and may inevitably omit some important related works. I would be deeply grateful to readers who are willing to share corrections, suggestions, or feedback to help improve future revisions.