Research Statement Tengyuan Liang, Assistant Professor, University of Chicago

My research interests lie broadly in learning theory, especially developing new statistical and computational paradigms to understand how complex over-parametrized Machine Learning (ML) models extract information effectively from data. I begin by describing my research in *Three Phases* of efforts to advance the frontier of learning theory and mathematical statistics, and then lay out my future research agenda. All numbered references below refer to items on my vitae.

Phase I: Classic Empirical Risk Minimization (ERM) Principle. Uniform Law of Large Numbers (ULLN) laid the probabilistic foundation for classical learning theory and mathematical statistics. When the model class is convex with controlled complexity, the ERM principle ensures that learning from empirical data is possible and often optimal. The complexity control is typically achieved by explicitly regularizing the problem. In this phase of research, I studied the extent to which extensive analysis based on classic ULLN and ERM principle can provide mathematical guarantees for non-convex models, including deep neural networks.

For learning functions, [16, *Econometrica*] studied optimal high probability bounds on estimation with deep neural networks and downstream semi-parametric inference. The analysis leverages careful localization arguments for the non-convex class (to improve upon naive ULLN) and recent approximation theory developments. [3, *COLT*] discovered a new localization approach by proposing "offset Rademacher complexity" to derive rates for general non-convex and unbounded function spaces. For learning probability distributions, [18] formulated the first statistical framework to understand generative adversarial networks. The framework explained the curious regularization effect of the generator-discriminator-pair tradeoffs.

Phase II: New Minimum-Norm Interpolation (MNI) Principle. According to conventional wisdom, explicit regularization should be added to the model to prevent "interpolating" the training data. Curiously, abundant empirical evidence suggests that modern ML methods perform well statistically even at interpolation, without "explicit regularization." ULLN and ERM principle struggle to explain learning in this new interpolation regime. One likely hypothesis is that methods or algorithms favor a certain "minimal" way of interpolating the data, typically measured by certain norms induced by algorithms. In this phase of research, I investigated the MNI principle for practical ML models in the interpolation regime, and discovered new phenomena in high dimensions.

[13, *Ann. Stat.*] studied interpolation with kernel ridgeless regression and identified that, surprisingly, high dimensionality and non-linear inner-product kernel generate an "implicit regularization" effect. Hence, the MNI principle generalizes well with favorable geometric properties on the design. [15, *COLT*] extended the above linear growth regime (dimension proportional to sample-size) to a broader polynomial growth regime, and discovered a curious "multiple-descent shape" of the risk curve as the dimension grows. A similar analysis holds for wide neural networks with random weights. [21] studied the mathematical role of non-linear activation in deep neural networks and how it affects interpolation and memorization capacity. [20] examined the MNI principle for classification problems and established a precise high dimensional asymptotic theory for boosting. At the heart of the analysis lies an exact characterization of the max-min margin and the generalization error of MNI induced by AdaBoost, which significantly enriches the prior understanding of boosting.

Phase III: Algorithmic Component of Learning Theory. One distinctive advantage of modern ML models over classic non-parametric models is the flexible yet tractable algorithmic component. In this phase of research, I examined the algorithmic component of learning theory, and unveiled the behavior of stochastic-gradient-type algorithms for learning non-convex and over-parametrized ML models.

[8, *COLT*] and [12, *J. Royal Stat. Soc. B*] considered the Langevin dynamics for training general learning models with possible non-convex landscapes. Using stochastic analysis, we showed that for a particular local optimum, with arbitrary initialization, the algorithm trajectory either lands outside the neighborhood within a "short recurrence time," or enters the neighborhood and remains within for an exponential "long escape time." In the latter meta-stable regime, we proposed and analyzed new stochastic-gradient-type algorithms for local inference using Langevin dynamics. [14, *J. Am. Stat. Assoc.*] considered gradient flow training of two-layer neural networks, and explored provable benefits of the adaptive representation in neural networks compared to the prespecified fixed-basis representation in classical non-parametrics.

Future Research. A major open question in deep neural networks is the advantage over other non-parametric approaches such as kernel machines. To answer this question, I plan to study with practical algorithms what structures of the function are easy to learn, continuing my work [13, 14, 15, 21] on the connection and difference between (adaptive) kernel machines and neural networks. My future research aims to understand when and why modern ML methods work well on data and to devise the next generation non-parametric statistical models with computation in mind.