Statement of Research

Jingbo Liu

My postdoctoral research centers around statistical inference, while my PhD work lies mainly in information theory. Although historically the two fields evolved through the pursuits of different engineering problems (i.e. data analysis and prediction vs. communication engineering), both fields share the element of the analysis of probabilistic models, through which we gain insights into the fundamental limitations of system designs, or touchstones for comparing algorithms. Below, my previous research topics are summarized, and some proposals for future research are sketched.

1 Statistical Inference

1.1 Memory Constraints in Message Passing Algorithms

Message passing is an inference algorithm for graphical models, with wide applications such as parity check codes, satisfiability, community detection, and linear regression. The message passing algorithm proceeds by recursively updating values associated with the vertices in the graph, based on the values of the neighbors. The algorithm enjoys low complexity and high accuracy for sparse (e.g. locally tree-like) graphs, and, in particular, is known to compute the exact posterior for trees. For $d$-regular trees with edge-disagreement probability $\epsilon < 1/2$, the Kesten-Stigum (KS) threshold states that asymptotically recovering the root from the leaves via a recursive algorithm is possible if and only if $\epsilon > (1 - d^{-1/2})/2$. In realistic applications, however, the processing units can only store values in a finite number of bits. In 2000, Evans, Kenyon, Peres, and Schulman conjectured that the KS threshold no longer holds in this memory-bounded setting [18, Section 9]:

\[\text{Any recursive reconstruction algorithm that stores only a bounded number of bits at each node, must fail asymptotically above a certain noise threshold which is strictly lower than } \epsilon > (1 - d^{-1/2})/2.\]

This conjecture was motivated by earlier partial results in [21], and was also discussed later in [22]. At the first sight, it is not even clear whether this problem is combinatorial in nature, or should be solved by calculus. Indeed, standard data processing arguments only proves the 1-bit case, whereas some interesting combinatorial constructions of recursive algorithms seem to refute the conjecture beyond the 1-bit case, as we report in [4].

**Main result:** We resolve the EKPS conjecture in the affirmative [4] by establishing the following novel asymptotic normality result:

\[\text{For any algorithm close to optimal, the posterior probability must be close to the Gaussian distribution (in the Wasserstein-2 distance).}\]

This normality result does not following immediately from CLT since the degree of the tree is not growing. Moreover, by a fixed-point style argument, we show that a simple “belief quantization” algorithm improves a previous “periodic majority” algorithm, and is, in fact, order-wise optimal.
1.2 Communication Complexity and Statistical Risk Tradeoff

Communication complexity is a topic at the intersection of information and computer science, with traditional applications such as circuit complexity and designs of data structures. Today, the ever-increasing amount of data and the demand on privacy motivate the study of statistical inference in distributed settings, where a communication complexity constraint is usually imposed. In models where the data are i.i.d. from a common parameter family, the communication-risk tradeoff has been well-studied in the literature, where the lower bound uses the strong data processing inequality (or similar techniques such as contraction of $\chi^2$-information or the Fisher information).

In [3][2], we study another very natural yet rather different type of distributed estimation problem, where the joint data distribution comes from a parameter family, and furthermore, the parameter is estimated by each terminal, instead of a central processor, after interactive communications. Here, our lower-bound technique is no longer the strong data processing inequality, but a so-called “symmetric strong data processing inequality” (introduced in my previous work with coauthors [11] in other contexts). Combining the result with novel achievability schemes, we derive the risk-communication tradeoff for the Gaussian and binary families, exact up to first-order approximations. Alternative proofs via optimal transportation theory were also reported in [2] and the arXiv version of [3].

Curiously, the probabilistic setup we consider also implies the result for a combinatorial problem, the Gap-Hamming problem in computer science. Previously, the solution to the Gap-Hamming problem uses an extremal inequality of Borell, which appears applicable only to highly structured spaces such as the Hamming cube. In contrast, our new bound via the symmetric strong data processing inequality can be computed, in principle, for all finite-alphabet distributions.

1.3 Sparsity, Regression, and False Discovery Rate Control

In many data science applications, for example, finding those genes that are related to a certain disease, one would like to efficiently throw out those predictive variables (e.g. gene expressions) that are not really related to a response variable. The number of predictors are usually comparable or larger than the number of samples, hence the false discovery rate (FDR) control is arguably more reasonable than the family-wise error rate control (i.e. demanding no false discovery at all). The “knockoff filter” introduced recently by Baber and Candès offers provable guarantees on the false discovery control in wide settings. Intuitively speaking, the knockoff filter constructs knockoff (fake) variables, so that the algorithm can leverage them to choose a data-dependent threshold, by which the solution to the LASSO algorithm can choose predictors with guaranteed FDR. However, previous methods for choosing the generating laws of these knockoff variables, such as solving semidefinite programs (SDP), were generally guided by heuristics and computation efficiency considerations. It remained an open problem how to choose among a universe of possible mechanisms for generating the knockoffs for data with a given covariance structure.

In [14], we study FDR control in the regime where the number of predictors is proportional to the number of samples. The predictors are assumed to be Gaussian with a general covariance matrix, but satisfying a certain distributional limit condition introduced by Javanmard and Montanari [19]. We introduce effective signal deficiency (ESD) as the equivalence class of functions of the covariance matrix that “metrize” consistency of a given variable selection algorithm. By leveraging the replica results on LASSO in [19], we determine the ESD for the knockoff filter with a generic knockoff mechanism, which is the right cost function that an engineer should optimize when choosing among knockoff mechanisms.

Furthermore, in the case of tree graphical models (or generally, walk-summable graphical mod-
els), we propose a new knockoff mechanism with an explicit analytic formula, hence faster computation. On synthetic data on various tree models, the statistical error slightly improves the best among previous knockoff mechanisms.

1.4 Topics of Future Interest

- **Learning with structures.** Often, exploiting the underlying structures in the data can lead to more efficient algorithms with good statistical performance (cf. tree graphical model examples in Section 1.1 and 1.3). Many interesting open problems (e.g. computational-statistical gaps) remain, when one moves from graphical models with binary labels to the case of multiple labels (a.k.a. Potts model). Beyond graphical models, other emerging modeling assumptions include nonparametric Bayes and negative dependence. Robustness of algorithms with respect to these model assumptions is of practical importance, and the Wasserstein approximation technique in Section 1.1 brings a promising approach.

- **Learning under system constraints.** Besides communication complexity (cf. Section 1.1, 1.2), other forms of system constraints that can be incorporated in an estimation framework include: differential privacy, for which a theory usually exists in parallel to communication constraints; query complexity in optimization, for which existing results have not answered important cases, for example, parallel query for certain natural function classes.

- **Methods from optimal transport.** Optimal transportation sometimes arises unexpected in the statistical lower-bound proofs of some inference problems (e.g. Section 1.1, 1.2). In addition, optimal transportation is inherent to some learning problems, such as transfer learning and generative adversarial networks. Interestingly, while the optimal transportation solution is known to have various equivalent analytic formulations, they may lead to different performances in numerical computations (e.g. [20]).

- **Kernel regression and neural network.** Linear regression can be studied by some nice analytic tools (e.g. replica method, cf. Section 1.3). Recently, some authors showed that in the infinite width limit, the gradient descent in a neural network becomes a kernel regression. Can we make a bridge to the standard finite width setting, so as to extend theories of kernel regression to neural networks? In particular, can the implicit regularization theory for kernel regression be extended, so as to explain the generalization mystery of neural networks? I have recently started looking at these problems with Philippe Rigollet.

2 Information Theory

2.1 Information Theory and Functional Analysis

**Brascamp-Lieb inequality.** The Brascamp-Lieb inequality and its reverse generalize many fundamental inequalities; as such, they sit at the very top of the hierarchy of inequalities in Gardner’s influential survey paper “The Brunn-Minkowski inequality”. We propose a forward-reverse Brascamp-Lieb inequality which even generalizes the two [7][1]. This clarifies that there is no fundamental distinction between the Brascamp-Lieb inequality and its reverse, in the sense that they are merely instances of the new inequality with different parameters. For the new inequality, we derive the information-theoretic characterizations, Gaussian optimality, and finiteness and extremisability conditions.
Reverse hypercontractivity approach to converses [15][5]. We introduce a novel approach to information-theoretic converses based on two ideas: 1. Use Fenchel’s duality to find a functional functional inequality corresponding to a generic information theoretic expression. 2. Apply reverse hypercontractivity to the functional inequality. This approach is at least as general as the famous concentration-of-measure (blowing-up lemma) approach, but often gives a sharper bound. In particular, this answers an open problem about the second-order asymptotics in “side information problems” [23]. Furthermore, this approach also leads to simpler proofs and tighter bounds for the relay channel (Cover’s open problem) than the measure concentration bound [13]. The work was awarded the Thomas Cover Dissertation Award of the information theory society and the Bede Liu Dissertation Award of Princeton University, and my solo-author paper [5] was selected in the TPC choice session at ISIT.

2.2 Miscellaneous Topics

• Interactive communications. As alluded in Section 1.2, we introduced some converse methods for interactive communication complexity problems in [11], in the context of common randomness generation and secret key generation. The work resolved a conjecture of Himanshu Tyagi, and received some attention in the computer science community as well.

• Nonasymptotic information theory.
  – Resolvability, strong soft-covering lemmas. We extend the channel resolvability theory to a new approximation metric ($E_\gamma$-resolvability) [12], and derive strong soft-covering lemmas via McDiarmid’s inequality. The latter result forms the basis of the solution of the wiretap-II channel capacity and exact channel resolvability exponents by other authors.
  – Covering lemmas. Novel proofs of mutual covering lemma via resolvability or Talagrand’s concentration inequality [12][17]. Conference version presented as semi-plenary talk at ISIT.
  – “Robust” version of information theoretic inequalities. Information spectrum bounds for the Brascamp-Lieb inequality under perturbation in the total variation distance, leading to single-shot bounds for distributed source coding problems [8].
  – Sampling techniques. Asymptotically tight exponential bounds for rejection sampling techniques via novel achievability schemes [16].

• Information-theoretic secrecy.
  – Wiretap channels. Tradeoff between communication rate and the eavesdropper list-decoding size via $E_\gamma$-resolvability [12].

• Polar codes. The encoding/decoding complexities of the polar codes are shown to be a polynomial of the gap to the fundamental limit, under certain settings [6].

• Novel decoding metrics. $\alpha$-decodability is proposed as an interpolation of the convention maximum error criterion ($\alpha = -\infty$) and average error criterion ($\alpha = 1$); the critical threshold $\alpha = 0$ is identified for the strong Fano’s inequality to hold [10].
2.3 Topics of Future Interest

- **Converse techniques.** Applications of information-theoretic converse techniques in impact areas beyond communication engineerings, e.g. inference, generalization bounds, bandits. Immediate projects include applying our reverse hypercontractivity techniques to group testing, or developing non i.i.d. results using available reverse hypercontractivity bounds for weakly correlated processes.

- **Many-user information theory.** The “5-G” and “IoT” era expects a dramatic increase in the number of connections in the network. While classical information theory focuses on the setting of fixed number of users and increasing blocklength, non-asymptotic information theory offers an abundance of techniques that may be adapted to the scenario where the blocklength grows proportionally to the network size.

- **Information flow in neural networks.** Tishby’s information bottleneck theory offers a way to potentially look under the hood of the neural networks. How to efficiently estimate the strength of the information flow is an interesting question where statistics and information theory meet. Moreover, unlike Shannon theory where the mutual information is absent from the problem formulation but arises naturally as the solution, the information bottleneck theory lacks justifications of the use of the mutual information metric or comparisons with other information metrics.

3 References from my publications


4 Other references


