Computational Complexity of Statistical Inference

Guy Bresler (MIT)
Samuel B. Hopkins (MIT)
Tselil Schramm (Stanford)
Luca Trevisan (Bocconi)

February 26–March 1

1 Overview of the Field

The subject of this workshop is the computational complexity of statistical inference tasks in a multitude of contexts. This is a relatively new and rapidly developing research area. The fields of mathematical statistics and computational complexity have for a century existed largely independently of one another: the former has traditionally studied statistical or information limits, while the latter has focused mostly on combinatorial problems with worst-case (adversarially) chosen inputs that do not accurately reflect the reality of data problems. It is only in the last decade that a research community has emerged dedicated to addressing the fundamental questions at the interface. We briefly give context for why the new perspective is needed.

The two basic lines of inquiry in statistical inference have long been: (i) to determine fundamental statistical (i.e., information-theoretic) limits; and (ii) to find efficient algorithms achieving these limits. However, for many structured inference problems, it is not clear if statistical optimality is compatible with efficient computation. Statistically optimal estimators often entail an infeasible exhaustive search over possible structures. Conversely, for many settings the computationally efficient algorithms we know are statistically suboptimal, requiring higher signal strength or more data than is information-theoretically necessary. This phenomenon is both fascinating and startling. It suggests that the information-theoretic limit on the signal-to-noise ratio (or the amount of data) for these problems, as studied since the beginning of mathematical statistics, is not the practically relevant benchmark in modern high-dimensional settings. Instead, the practically relevant benchmark is the fundamental statistical limit for *computationally efficient* algorithms.

When efficient algorithms fail to achieve the statistical limit, a problem is said to have a *statistical-computational gap*. In many observed cases, the gap can be sizable, so that efficient algorithms require orders of magnitude more data than is information-theoretically necessary. Awareness of statistical-computational gaps is not new, with early work showing such gaps in artificially constructed learning problems [10, 19, 20] and more recent work focusing on algorithms that trade off between statistical and computational efficiency [21, 20, 8, 9].

By now dozens of important high-dimensional statistical estimation problems are conjectured to have different computational and statistical limits. These problems (for example, sparse linear regression or sparse phase retrieval [24, 7, 11, 17]) are ubiquitous in practice and well-studied theoretically, yet the central mysteries remain: What are the fundamental data limits for computationally efficient algorithms? How do we find optimal efficient algorithms? At a more basic level, are these statistical-computational gaps in distinct problems appearing for a common reason? Is there hope of building a widely applicable theory describing and explaining statistical-computational trade-offs?

1.1 Current State of the Field

In the context of statistical inference, computational complexity is necessarily broadly construed. The goal is to characterize the amount of computational resources needed to solve statistical problems in terms of the amount of data. Computational resources include not only running time, but also memory, communication, and more. Deepening our understanding of fundamental limits of efficient algorithms for these problems has a positive impact on algorithm design, as it allows us to understand which algorithms (or classes of algorithms) are likely to be optimal for different families of problems. Several approaches are predominant.

Reductions. A reduction from one inference task A to another task B shows that any algorithm for B can be transformed into an algorithm for A. If A is believed to be computationally intractable, then a reduction shows that B is hard as well. Reductions form the foundation of the classical theory of NP hardness, but historically they have mainly been applied to worst-case problems; in statistical contexts, it is challenging to design reductions which respect the distributional structure of data. A recent line of work has begun to overcome this challenge, providing reductions from the long-studied *planted clique* problem to a range of basic inference tasks (e.g. [3, 5, 4, 6]). While these works represent a tremendous start, in contrast with the worst-case setting we are still far from being able to reduce between most pairs of hard inference problems, and more mathematical tools are needed to build a satisfying theory.

Restricted Models of Computation. Unconditional evidence of hardness comes in the form of lower bounds against specific restricted models of computation. Popular restricted computational models include: sum-of-squares and other convex programming hierarchies, statistical query algorithms, local- and gradient-based algorithms, memory-bounded algorithms, and various classes of circuits, among others. Information-computation gaps are often consistent across all of these models, but our understanding of this phenomenon is limited. For each statistics problem that emerges, researchers prove lower bounds against each restricted model individually, each results in a disparate piece of evidence for an information-computation gap that must then be reconciled with the others. A more coherent understanding is needed, and this workshop aims to bring together researchers working on this topic to address questions such as: is there a hierarchy theorem for computational models in the context of statistics? Are some of these algorithmic methods better suited to certain types of problems?

Energy Landscapes. In Bayesian estimation problems, one typically views samples Y from a distribution P_θ parameterized by some $\theta \in \Theta$, and the goal is to estimate θ as faithfully as possible. A powerful perspective on estimation is to think of the distribution P_θ as creating an "energy landscape" over the space of possible parameters Θ, where the height of each $\theta' \in \Theta$ is related to the posterior $P(\theta' \mid Y)$. Geometric properties of this landscape are known to be related to the success or failure of local algorithms such as gradient descent, Langevin dynamics and Markov Chain Monte Carlo: does the landscape slope up to one large peak around θ, or are there deep valleys and other peaks as well? This powerful perspective originates in statistical physics, and has only recently begun to be rigorously related to computational complexity.

2 Recent Developments and Open Problems

In the last several years, exciting progress has been made in each of the core categories identified in Section 1.1, as well as in making new connections to other research areas. We describe some of the highlights in the last few years as well as where opportunities to connect with other areas are emerging. The workshop had emphasis on both.

2.1 Progress Highlights in Core Topics

Relating the Power of Different Classes of Algorithms Some of the algorithms under consideration are of algebraic nature, such as low-degree polynomial algorithms (motivated by the Sum-of-Squares hierarchy), and some of them of geometric nature, such as MCMC or free-energy based methods (motivated by statistical physics). Notably, the algebraic and geometric approaches can sometimes diverge in their predictions. For

example, in the well-studied tensor PCA model introduced by Richard and Montanari in 2014, there is a regime of parameters where simple low-degree (in fact, spectral) methods work, yet free-energy approaches suggest computational hardness.

Work by Bandeira, El Alaoui, Hopkins, Schramm, Wein, and Zadik, builds a link between the two points of view. They proposed the *Franz-Parisi* (FP) criterion, a new free-energy based criterion for hardness inspired by the seminal work of Franz and Parisi in statistical physics. They showed that for a large class of models, including the tensor PCA example, the prediction of the FP criterion matches exactly the optimal performance among low-degree methods. By leveraging this rigorous connection they were able to also establish that for various models the *algebraic* failure of low-degree methods implies the *geometric* failure of local MCMC algorithms. Finally, they showed that the FP criterion provides a useful tool purely to prove low-degree hardness, yielding new low-degree lower bounds for the sparse linear regression model.

Energy Landscapes of Random Optimization Problems *Random optimization problems* are high-dimensional optimization problems whose objectives are generated from random data. These problems are well-studied in the probability, computer science, and statistical physics communities, including models like the mean field spin glass, random perceptron, and random (max)-k-SAT. Practically, they model realistic tasks such as training a neural network. A central objective is to identify the *algorithmic threshold*, the largest objective attainable by an efficient algorithm. This question has until now remained poorly understood, without even a coherent way to *predict* the algorithmic threshold, let alone to find an algorithm and prove computational hardness above the threshold. (We note that the main distinction between random optimization problems and the other sorts of statistical inference problems studied in the program are that the latter are *planted* problems with a *signal* that is to be estimated. Both types of problems are of fundamental importance.)

Brice Huang and Mark Sellke (BIRS workshop participants) developed a unifying theory for random optimization problems which gives exact predictions of the algorithmic threshold. Their main insight is a geometric description of the algorithmic threshold ALG as the largest value E whose super-level set $S(E)$ typically contains an "everywhere-branching" ultrametric tree. This claim can be explained by the following intuitions. On the algorithmic side, efficient algorithms are capable of descending an everywhere-branching ultrametric tree via local updates and therefore can reach ALG. For lower bounds, Gamarnik and Sudan's Overlap Gap Property [14] shows that suitably stable algorithms fail to attain value E if $S(E)$ does not contain certain constellations, such as a pair of solutions a medium distance apart; Huang and Sellke realized that running this argument with an everywhere-branching ultrametric tree as the constellation shows a lower bound at ALG, which holds for a class of *overlap concentrated* algorithms.

Huang and Sellke have rigorously verified this prediction for the mean field spin glass, spherical multi-species spin glass, and random perceptron. The works [22, 18, 1] produce algorithms for the mean field spin glasses achieving value ALG. In [15], Huang and Sellke show that overlap concentrated algorithms cannot surpass ALG in spin glass models. In recent work [16], they also develop matching algorithms and lower bounds attaining ALG for the spherical multi-species spin glass and the random perceptron. It is worth emphasizing that without their framework, it is not even clear how to think about what an optimal algorithm should look like, let alone identify it and prove its optimality within a broad class.

2.2 Opportunities for Connecting with Other Fields

Machine Learning and Neural Networks In recent years, the machine learning community has devoted intense effort to understanding the capabilities and the limitations of deep learning. From the perspective of computational complexity and statistical tradeoffs, we can hope to develop a rigorous understanding of what structure of data is necessary and sufficient for neural networks to learn when trained with stochastic gradient descent (SGD) and different numbers of samples.

Cryptography The field of cryptography builds various useful procedures and primitives, essentially all of which rely in some way on computational hardness. The ideal computationally hard problems, upon which all of modern cryptography is based, are *average case* hard hypothesis testing problems. The average-case nature of the hardness means that one does not have to go far to find a hard instance – indeed, a random one works. These are precisely the sort of problem studied by the computational complexity for statistical inference community. For somewhat nuanced reasons, however, the problems useful for cryptography cannot be *too*

hard, in a certain sense. This makes suitable problems somewhat rare, and the cryptography community is eager to expand their list of suitable problems. At the same time, cryptographers have developed sophisticated tools for reasoning about computational complexity of these problems, and CCSI can potentially benefit from these.

Combinatorics Many of the problems of interest to the CCSI community relate to the combinatorics of graphs or hypergraphs and there is significant opportunity for cross-pollination in both directions. As an example, delicate combinatorial properties of hypergraphs turn out to relate to existence of succinct refutation arguments, a core concept in CCSI. Conversely, techniques developed in the context of CCSI, such as algorithms based on spectra of matrices, may be useful for reasoning about combinatorial objects. Fourier analysis of Boolean functions plays a key role in the modern theory of both subjects.

3 Presentation Highlights

High-dimensional Statistics

1. In the presentation by Jiaming Xu, titled "Sharp Statistical Limits for Shuffled Linear Regression," the focus was on addressing the shuffled linear regression problem. This issue arises when the link between predictors and responses in a linear model is hidden due to a latent permutation of the samples, necessitating the simultaneous estimation of both the permutation and regression coefficients. Shuffled linear regression is relevant in various fields, including robotics, multi-target tracking, signal processing, and even in data integration and de-anonymization tasks. Despite significant interest and extensive research over the past decade, the exact statistical thresholds for successful recovery in this models had not been clearly defined. Xu discussed the team's breakthrough in pinpointing these thresholds for both perfect and near-perfect recovery of a hidden linear regressor, thanks to a detailed examination of the maximum likelihood estimator (MLE) across different error regimes. The computational aspect of MLE, tied to the NP-hard quadratic assignment problem, poses challenges, as no known polynomial-time algorithms achieve success near these statistical limits outside of noiseless or low-dimensional scenarios. This research marks a notable advancement in understanding the statistical and computational intricacies of shuffled linear regression.

2. In the presentation titled "Computational and Statistical Limits of Inference from Gaussian Fields," Frederic Koehler delved into recent advancements in learning and testing Gaussian graphical models, with a focus on significant examples such as Gaussian free fields on unknown graphs. The talk emphasized the close relationship between these models and the challenge of sparse linear regression, shedding light on potential computational-statistical gaps. Furthermore, Koehler explored the connections to sparse principal component analysis (PCA), providing further insight into the computational complexities of learning Gaussian Graphical Models.

3. In the talk titled "Fine-Grained Extensions of the Low-Degree Testing Framework," Alex Wein discussed the low-degree polynomial framework as an effective tool for exploring the computational complexity of statistical problems. This approach focuses on the capabilities and limits of using low-degree polynomials as algorithms, particularly in the context of hypothesis testing. The talk highlighted extensions of this framework that address more detailed questions than the low-degree method as so-far developed over the last five years. For example, Wein discussed the spiked Wigner model, aiming to find the precise optimal trade-off between type I and type II error rates achievable by polynomial-time algorithms. Using a strengthened variant of the "low-degree conjecture" which the field has been investigating for the last five years, Wein showed showed that spectral-based tests are the most effective among polynomial-time algorithms, although exponential-time non-spectral tests could achieve better results.

4. Kiril Bangachev delved into the topic of "On The Fourier Coefficients of High-Dimensional Random Geometric Graphs," focusing on the random geometric graph $RGG(n, S^{d-1}, p)$, which is generated by sampling n i.i.d. vectors uniformly on S^{d-1} and connecting vertices i and j if their dot product is greater than or equal to $\tau_p(d)$, where $\tau_p(d)$ ensures the graph's expected density is p. Bangachev's

research primarily investigates the low-degree Fourier coefficients of the $RGG(n, S^{d-1}, p)$ distribution. The importance of low-degree Fourier coefficients of such models has been made clear by the last five years of investigation into the "low-degree conjecture". However, methods to analyze low-degree Fourier coefficients of random graph models are thus-far largely limited to random graph models with simple product structures.

The main conceptual contribution of this work is a novel two-step approach for bounding Fourier coefficients in the RGG, which lacks a lot of the independence structure found in simpler random graph models. This new technique is potentially applicable to a broader study of latent space distributions. Initially, the method involves localizing the dependence among edges to a limited number of "fragile" edges. Following this, the space of latent vector configurations is partitioned based on the fragile edges, and within each partition, a noise operator is defined that acts independently on edges not adjacent to these fragile edges.

The implications of these bounds include: 1) Clarifying the low-degree polynomial complexity of differentiating spherical random geometric graphs from Erdös–Rényi graphs, both when observing a complete set of edges and under the non-adaptively chosen mask model introduced by prior work; 2) Demonstrating a statistical-computational gap for distinguishing RGG from the planted coloring model in conditions where RGG can be distinguished from Erdös–R; 3) Reestablishing known bounds on the second eigenvalue of random geometric graphs.

5. Brice Huang presented "A Constructive Proof of the Spherical Parisi Formula," shedding light on one of the most celebrated results in the theory of spin glasses—the Parisi formula for free energy. Huang introduced a new proof for the lower bound of the spherical mean-field model that stands out for its constructive nature. The proof is significantly more transparent than earlier proofs, which were widely viewed as opaque. By contrast, the new proof is modular, and it draws on algorithmic insights generated by the computer science community studying spin glasses in the last few years. Huang's talk demonstrates the influence of algorithmic developments from the CCSI community on problems in pure mathematics and probability theory.

Algorithms for Learning

1. Adam Klivans presented on the new learning framework "TDS Learning," addressing the challenge of learning with distribution shift. In learning with distribution shift, a learner is provided labeled samples from a training distribution D, alongside unlabeled samples from a test distribution D', with the goal of developing a classifier that minimizes test error. Traditional methods focus on bounding the classifier's loss by measuring some form of distance between D and D'. However, these distance measures are often complex and don't necessarily lead to practical algorithms. Moving beyond conventional strategies, Klivans introduced a novel approach termed testable learning with distribution shift (TDS learning). This model enables the creation of efficient algorithms that can certify a classifier's performance on the test distribution. Specifically, it allows for the output of a classifier that has low test error, provided the samples from D and D' pass a specific test, which must also accept if the marginal distributions of D and D' are identical. The presentation showcased several TDS algorithms learning well-known concept classes, including halfspaces, intersections of halfspaces, and decision trees. Previously, no efficient distribution-shift-robust algorithms existed for these categories without imposing strong assumptions on D'.

2. Jane Lange discussed the first agnostic, efficient, proper learning algorithm for monotone Boolean functions. The algorithm, given $2^{\tilde{O}(\sqrt{n}/\epsilon)}$ uniformly random examples of an unknown n-variate Boolean function f, outputs a hypothesis g that is monotone and $(\text{opt} + \epsilon)$-close to f, where opt is the distance from f to the nearest monotone function. The running time, as well as the size and evaluation time of the hypothesis, is also $2^{\tilde{O}(\sqrt{n}/\epsilon)}$, nearly aligning with the best known lower bound of Blais et al. Lange also presented an algorithm for estimating, up to an additive error ϵ, the distance of an unknown function f to monotone, with a runtime of $2^{\tilde{O}(\sqrt{n}/\epsilon)}$. Prior to this work, while sample-efficient algorithms for these tasks existed, they lacked runtime efficiency. Lange's algorithm builds on prior works,

which supply an improper learning algorithm. Given the output of the improper learner, Lange's approach is to corrects it to monotone using query-efficient local computation algorithms. This method is enhanced by incorporating a convex optimization step into the improper learner and correcting a real-valued function before rounding its values to Boolean.

3. Mitali Bafna introduced efficient algorithms for computing power sum decompositions of polynomials. The goal is to decompose an input polynomial $P(x) = \sum p_i(x)^d$ into its component polynomials p_i. In the classic case when the p_i's are linear, this is the extensively researched tensor decomposition problem. When the p_i are quadratic, the power-sum decomposition is a promising primitive for learning Gaussian mixtures from low-order moments. Unlike tensor decomposition, unique identifiability and algorithms for power sum decomposition have remained elusive. Bafna's algorithm effectively decomposes a sum of $m = O(n)$ generic quadratic p_i's for $d = 3$ and, more broadly, the dth power sum of $m = n^{O(d)}$ generic degree-K polynomials for any $K > 2$. The algorithm can accommodate inverse polynomial noise, assuming the p_i's are generic.

4. David Wu's presentation focused on the challenge of robust community recovery in sparse stochastic block models, when edges may be added and deleted from the graph by an adversary. Traditionally, efficient algorithms exist for community detection when the signal-to-noise ratio surpasses the Kesten–Stigum (KS) threshold, which is conjectured to be the computational threshold for this task. Wu's research sought to determine whether the KS threshold also represents the computational limit for robust community recovery. His findings confirm this hypothesis, introducing an algorithm capable of robust community recovery across arbitrary stochastic block models with any constant number of communities. This extends the prior work of Ding, d'Orsi, Nasser, and Steurer, which provided an efficient solution above the KS threshold for two-community models. The algorithm is spectral, making use of a specific matrix, called the Bethe Hessian. Wu was able to demonstrate that the spectrum of the Bethe Hessian is robust to perturbations, in the sense that any perturbation to a principal minor of the matrix either doesn't change the spectrum, or else produces delocalized eigenvectors which are easy to filter out.

Cryptography

1. In his presentation, Andrej Bogdanov explored the nuanced relationship between search and decision problems within statistical and cryptographic contexts. In cryptography, the interest often lies in understanding the hardness of both of these problems. Hardness of search problems in statistics is akin to hardness of inverting one-way functions in cryptography, while hardness of distinguishing problems is akin to pseudorandomness.

   Bogdanov explored the possibility of formal equivalences between search and decision hardness in CCSI, drawing on insights from cryptography which relate hardness of inverting one-way functions and the existence of pseudorandom generators.

2. In her presentation, Rachel Lin explored the Learning Sparse Parity with Noise (LPN) problem, where the (public) coefficient matrix is k-sparse (i.e., each column has Hamming weight k). These problems have been extensively studied within average-case complexity literature, and have also been widely studied as canonical hard problems in learning theory for at least the last three decades. Lin's talk expanded on LPN by generalizing to larger fields $\mathbb{F}_q$. This generalization has already been used in cryptography.

   Rachel discussed the computational hardness of sparse LPN problems and focused especially on how the sparsity parameter impacts their hardness. The amount of sparsity also plays a central role in developing cryptographic protocols such as multi-party homomorphic secret sharing and secure multiparty computation. These protocols are notable for their efficient communication, requiring sublinear bandwidth relative to the computation's complexity. This work, a collaboration with Yuval Ishai, Aayush Jain, and Dao Quang, marks a significant contribution to understanding how aspects of hard computational problems such as LPN impact their application in designing advanced cryptographic protocols.

3. Aayush Jain's talk centered on a new code-based cryptographic assumption known as *Dense-Sparse LPN*. This innovation comes in response to the challenges of constructing advanced cryptographic primitives from other code-based assumptions. In particular, the only known constructions were based on Learning Parity with Noise (LPN) with an extremely low noise rate, which does not have sufficient security as it is solvable in quasi-polynomial time.

   The Dense-Sparse LPN assumption is inspired by McEliece's cryptosystem and the random k-XOR problem studied in average-case complexity. As its main benefit, it is plausibly secure against subexponential time attacks. The assumption involves a dense matrix, a random sparse matrix, and a sparse noise vector, which are combined in a nontrivial way to construct cryptographic primitives that were previously thought to be hard or impossible to achieve with code-based assumptions.

   Leveraging this new assumption, Jain discussed the construction of *lossy trapdoor functions*, a cornerstone in cryptography that enables the creation of a wide array of cryptographic primitives, both lossy and non-lossy. This development presents the first post-quantum alternative to lattice-based constructions.

4. Yuval Ishai's presentation highlighted the cryptographic potential of conjectures regarding the difficulty of identifying planted graphs within random ambient graphs, generalizing the planted clique problem which is one of the foundational problems in CCSI. Yuval's talk showed how natural generalizations of the planted clique assumption can lead to new secure constructions of cryptographic primitives, including new secret-sharing schemes and secure computation protocols.

   Ishai's discussion opened the door to several intriguing questions about the feasibility of concealing a specific planted graph within a small, potentially arbitrarily distributed ambient graph. Yuval's talk was an invitation to the cryptography and CCSI communities to delve deeper into the complexities of graph-based security assumptions.

Combinatorics

1. Pravesh Kothari and Peter Manohar presented on the "Kikuchi Matrix Method" in two exciting talks. They introduced a novel technique that allows addressing combinatorial problems by instead proving the unsatisfiability of k-sparse linear equations mod 2, or k-XOR formulas, with minimal randomness. The conceptual highlight is reducing arbitrary instances to random ones—or structured ones—and in either case useful mathematical properties exist that can be exploited.

   From a technical standpoint, the main ideas rest on spectral properties of "Kikuchi" graphs, induced subgraphs of Cayley graphs on the hypercube or its variations, such as products of hypercubes, which are formed by the equations' coefficient vectors. The connection between the spectral characteristics of these graphs and the combinatorial attributes of the input formulas can be directly related. They outlined the method's impact through several applications: 1) Proof of Feige's conjecture on the hypergraph Moore bound, 2) Showing a cubic lower bound for 3-query locally decodable codes, advancing beyond the quadratic bound established by Kerenedis and de Wolf (2004); 3) Establishing an exponential lower bound for linear 3-query locally correctable codes, thereby showing a separation between what is achievable with 3-query LCCs and LDCs, with the latter previously achieving sub-exponential length constructions. This marks the first instance of a super-polynomial lower bound for local codes with queries exceeding two. All three of these applications represent major advances in combinatorics.

   The method itself draws heavily on insights from CCSI. Indeed, the Kikuchi matrices were first examined by Ahmed El Alaoui, Cris Moore, and Alex Wein in the context of tensor PCA, a fundamental problem in the complexity of statistical inference, underscoring the theme of the workshop. Indeed, Cris reported during Pravesh's talk that the inception of the method occurred at a prior workshop at BIRS! While Ahmed, Cris, and Alex used Kikuchi matrices to study statistical inference problems, Pravesh and Peter showed how spectra of these Kikuchi matrices are highly consequential in combinatorics.

4 Meeting Activities

4.1 Icebreakers

On Monday morning, we ran an icebreaker activity called "find your pair." Each participant was given a notecard with a thematic phrase on it, and the participants mingled in order to find the participant with the most closely related notecard. For example, one notecard might have a large integer on it, and the second might have its prime factors. The activity lasted for about 20 minutes, leading naturally into continued discussion and interaction. Participants noted that this icebreaker activity was a fun opportunity to meet others and begin informal discussions.

4.2 Lightning talks

On Monday evening we had a session of open-min 5-minute lightning talks. This allowed many participants to present short research snippets and introduce what they were currently thinking about. This had the dual benefit of allowing non-speakers to present some of their work, and also allowing speakers to broadcast their research interests on Monday, so that they could connect with other participants of similar interest early on in the workshop.

4.3 Visioning Session

On Tuesday morning, we assigned each participant to one of four discussion groups, each with a specific topic: (1) connections between solution geometry and other notions of algorithmic intractability, (2) making use of new average-case hard problems in cryptography, (3) deep learning, and (4) formulating questions beyond testing and estimation. Each discussion group met over dinner, with the goal of formulating an agenda for the research direction including both concrete open problems, as well as a 5-year "vision" for where the field should go. Then, in the evening session, we came together, each group made a report, and we had a robust discussion. The highlights of the questions generated are given below.

Solution Geometry This group was interested in clarifying the connection between optimization landscapes – that is, the complexity, in terms of local minima and maxima of a random "energy surface" which arises in e.g. maximum likelihood estimation, and computational complexity. This connection can be considered in both "planted" settings, where there is a hidden signal one aims to recover (e.g. planted clique), or "null" settings, where the landscape is purely noise (e.g. the Sherrington-Kirkpatrick model). Various versions of the theme "clustering phenomena in the landscape implies computational hardness" are now known:

- Optimization in random models: (various versions of) Overlap-Gap property imply failure of stable algorithms ([14] and subsequent works)
- Sampling in random models: non-monotonicity of Franz-Parisi potential implies failure of stable samplers [1].
- Hypothesis testing in planted models: there is a connection between annealed Franz-Parisi potential (a "landscape" view) and low degree hardness [2].

These results are so far fairly disjointed – an important goal is to develop a unifying picture.

This group also discussed possible connections between landscapes and computational complexity for worst case problems. A concrete instance of this question is as follows. It is now known ([13, 23]) that in a sparse Erdös-Rényi graph, the largest independent set found by e.g. low degree polynomials has size half the optimum, and this hardness is witnessed by an OGP in the landscape of more-than-half-optimal independent sets. In worst case graphs, the ratio between the sizes of the largest independent set and the largest algorithmically-findable independent set can be arbitrarily large. What does this say about what would happen if we run a random graph through a gap amplification gadget?

Cryptography Cryptography relies fundamentally on hard problems to build useful primitives. While cryptography has reaped excellent rewards working with only a handful hardness assumptions (such as Learning with Errors and Groups), there is a great shortage of assumptions for realizing post-quantum cryptography, important feasibility goals and for efficient constructions. Since cryptography relies so fundamentally on new useful sources of hardness, it is important for the advancement of the field that we must continue to look for new and usable sources of hardness. This is where the fields of statistical inference and average-case complexity provides a great prospect. These areas could serve not only as a supply of large variety of problems for cryptography, but have systematic mechanisms to give evidence of hardness for these problems in form of lower-bounds in different restricted algorithmic models (such as SoS, low-degree tests and other models).

Towards this broad vision, in the workshop the following questions were brought up (in no specific order of relevance):

- **Public-Key Encryption from a New Source** While most average-case statistical inference problems readily gives rise to one-way functions (the most basic building block from cryptography), currently, we only have a handful sources of hardness implying public-key encryption or other advanced cryptographic applications. A new *structure* yielding public-key encryption (and hopefully more!) is a very worthy goal. This is also interesting from a complexity theoretic standpoint. Most of the current set of assumptions (such as LWE) implying encryption schemes can be shown to be in relatively structured complexity classes such as $\mathsf{NP} \cap \mathsf{coNP}$ and in particular have short proofs of unsatisfiability. This is not true for most of the statistical inference problems in their computationally hard regime. To understand this better, perhaps, considering proofs of unsatisfiability via interactive proofs would be a good idea to consider for statistical inference problems. Further, this might have to do with finding "efficient trapdoors" within statistical inference problems. This will perhaps require fundamentally new algorithmic results on these problems.

- **Low-Degree Heuristic** The next question that was discussed was regarding low-degree hardness conjecture. As cryptographers are considering new assumptions in recent works, hardness in the low-degree tests framework could provide a good first evidence of hardness for new assumptions. What is not so very well-formulated are the types of problems where it might be safe to assume security solely based on low-degree lower-bounds. More understanding of limitations and strengths of low-degree tests and its relationship with other algorithms would be good.

  Another question that was asked if we can better models that better capture search assumptions and distinguishing (weak detection) assumptions since cryptographic assumptions are of this flavor as opposed to say being of refutation flavor.

- **Large Alphabet CSPs** This appears to be a more technical but nevertheless important question is around understanding large alphabet CSPs better. The idea is that in context of Boolean CSPs we can show tight upper bounds and lower bounds in various models (such as SoS, and low-degree tests). Unfortunately, these bounds are no longer tight when we consider CSPs over large alphabet with a large gap between the easy regime and provably hard regime. It is believed that large alphabet CSPs could remain secure for a larger number of samples than Boolean CSPs. As assumptions such as sparse LPN over large fields (which corresponds to large alphabet CSP) are gaining importance in cryptography, tighter bounds will directly translate to better parameters for cryptographic schemes.

Deep learning The deep learning theory discussion group primarily discussed three general research directions. The first is to identify richer data models that go beyond the typical Gaussianity assumptions that are pervasive in theory for supervised learning. For instance, recent works on training dynamics for neural networks have posited various forms of causal structure, and it would be interesting to explore what kinds of distributional assumptions of this flavor strike the right balance between not making the original learning problem too easy that it obviates the need for deep learning, but also capturing useful properties of real-world data.

The second research direction proposed was to revisit problems which practitioners have moved on from due to the empirical difficulties they pose, but which could benefit from provable guarantees and serious theoretical study. The prime example of this is adversarial examples, which have become all the more relevant given recent jailbreaking attacks on large language models.

Finally, it was proposed to consider alternative learning guarantees beyond the usual statistical ones studied in the theory literature. For instance, can one revisit classic distribution learning questions from the perspective of *computational* indistinguishability? Such questions have already been studied in the supervised setting, by leveraging connections to boosting. It would be interesting if one could develop analogous tools for distribution learning, or better yet, to show that existing algorithms in practice like GANs or diffusion models achieve this weaker learning guarantee.

Verification of statistical inference This group discussed open problems pertaining to whether statistical inference becomes easier in the presence of an untrusted but computationally very powerful prover. Untrusted provers have been conceptually crucial in the development of worst-case complexity theory, so it is natural to wonder whether they can shed light on CCSI as well.

One inspiration for this question is the work of Feige, Kim, and Ofek [12], which shows that a random 3CNF formula with n variables and $n^{1.4}$ clauses admits a poly-size and poly-time verifiable witnesses for unsatisfiability with high probability. By contrast, *generating* such a witness in polynomial time is believed to require $n^{1.5}$ random clauses.

This group asked: could this phenomenon extend beyond random constraint satisfaction problems? For example, can an untrusted prover help distinguish $G(n, 1/2)$ from $G(n, 1/2)$ with a planted clique, below the $\sqrt{n}$ threshold where the problem becomes computationally easy? In other words, is there a range of clique sizes where the distinguishing problem is hard, but a prover can give a certificate that is efficiently verifiable? Certainly the answer is yes in the clique case — the clique itself is an efficiently verifiable proof — but it is not clear in the non-clique case. We don't expect to find a refutation that exists for every clique-free graph — this would imply $coNP \subseteq NP$ – but rather to exploit randomness of $G(n, 1/2)$.

We also consider ways to relax the "poly-size witness" requirement in ways that are still make good use of an untrusted prover. For example, is there additional power in having multiple rounds of communication? Are there problems which don't have poly-size certificates (representing a single round of communication), but can be distinguished by efficient multi-round interaction with an untrusted prover? (In worst-case complexity these are referred to as "Arthur-Merlin protocols".) The practical applications of such a protocol would be similar to the practical applications of a one-round protocol; both allow you to verify the output of a black-box predictive model, for example.

4.4 Afternoon collaboration time

Afternoons were for the most part left open to allow for collaboration and discussion. Participants commented on how much they appreciated this opportunity to follow up on interesting questions that arose in the talks.

5 Outcomes of the meeting

Even at this early stage, a few concrete outcomes of the meeting are clear.

The first is that many of the junior participants were introduced to the broader community at this event. We received a lot of positive feedback from junior participant about how easy it was to meet and talk with senior researchers. An arXiv preprint posted in the weeks after the workshop by a postdoctoral scholar participant acknowledges a senior participant for a helpful discussion during the workshop.

The second is new questions and directions for collaborations at the intersection of cryptography and computational complexity of statistical inference. The cryptographer participants identified a variety of average-case computational problems as being worthy of further study by the CCSI community. They also surveyed a larger variety of cryptographic primitives, beyond public-key encryption, that could benefit from new computationally complex average-case problems, including secret sharing and multiparty computation. In the other direction, the CCSI participants were able to give an exposition of the variety of techniques for predicting computational intractability of an average case problem, and to better understand cryptographers' concerns on this front.

In the coming months and years, we expect that many of the problems and directions discussed at the meeting will come to fruition as scientific progress. We are very thankful to BIRS for hosting us and giving us this incredible opportunity to connect and collaborate.

References

[1] A. El Alaoui, A. Montanari, and M. Sellke. Optimization of mean-field spin glasses. *Ann. Prob.*, 49(6):2922–2960, 2021.

[2] Afonso S Bandeira, Ahmed El Alaoui, Samuel B Hopkins, Tselil Schramm, Alexander S Wein, and Ilias Zadik. The franz-parisi criterion and computational trade-offs in high dimensional statistics. *arXiv preprint arXiv:2205.09727*, 2022.

[3] Quentin Berthet and Philippe Rigollet. Optimal detection of sparse principal components in high dimension. *The Annals of Statistics*, 41(4):1780–1815, 2013.

[4] Matthew Brennan and Guy Bresler. Optimal average-case reductions to sparse PCA: From weak assumptions to strong hardness. In *Conference on Learning Theory (COLT)*, 2019.

[5] Matthew Brennan and Guy Bresler. Statistical-computational gaps for learning sparse mixtures, robust estimation and semirandom adversaries. *arXiv:1908.06130*, 2019.

[6] Matthew Brennan and Guy Bresler. Reducibility and statistical-computational gaps from secret leakage. In *Conference on Learning Theory*, pages 648–847. PMLR, 2020.

[7] T Tony Cai, Xiaodong Li, and Zongming Ma. Optimal rates of convergence for noisy sparse phase retrieval via thresholded Wirtinger flow. *The Annals of Statistics*, 44(5):2221–2251, 2016.

[8] Venkat Chandrasekaran and Michael I Jordan. Computational and statistical tradeoffs via convex relaxation. *Proceedings of the National Academy of Sciences*, 110(13):E1181–E1190, 2013.

[9] Yudong Chen and Jiaming Xu. Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *Journal of Machine Learning Research*, 17(27):1–57, 2016.

[10] Scott E Decatur, Oded Goldreich, and Dana Ron. Computational sample complexity. *SIAM Journal on Computing*, 29(3):854–879, 2000.

[11] Jianqing Fan, Han Liu, Zhaoran Wang, and Zhuoran Yang. Curse of heterogeneity: Computational barriers in sparse mixture models and phase retrieval. *arXiv preprint arXiv:1808.06996*, 2018.

[12] Uriel Feige, Jeong Han Kim, and Eran Ofek. Witnesses for non-satisfiability of dense random 3cnf formulas. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pages 497–508. IEEE, 2006.

[13] David Gamarnik, Aukosh Jagannath, and Alexander S Wein. Circuit lower bounds for the p-spin optimization problem. *arXiv preprint arXiv:2109.01342*, 2021.

[14] David Gamarnik and Madhu Sudan. Limits of local algorithms over sparse random graphs. In *Proceedings of the 5th conference on Innovations in theoretical computer science*, pages 369–376. ACM, 2014.

[15] B. Huang and M. Sellke. Tight Lipschitz hardness for optimizing mean field spin glasses. In *63rd FOCS, to appear*, 2022.

[16] Brice Huang and Mark Sellke. Algorithmic threshold for multi-species spherical spin glasses. *arXiv preprint arXiv:2303.12172*, 2023.

[17] Jonathan A Kelner, Frederic Koehler, Raghu Meka, and Dhruv Rohatgi. On the power of preconditioning in sparse linear regression. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 550–561. IEEE, 2022.

[18] A. Montanari. Optimization of the Sherrington-Kirkpatrick Hamiltonian. *SIAM J. Comp.*, 2021.

[19] Rocco A Servedio. Computational sample complexity and attribute-efficient learning. In *Proceedings of the thirty-first annual ACM symposium on Theory of computing*, pages 701–710. ACM, 1999.

[20] Shai Shalev-Shwartz, Ohad Shamir, and Eran Tromer. Using more data to speed-up training time. In *Artificial Intelligence and Statistics (AISTATS)*, pages 1019–1027, 2012.

[21] Shai Shalev-Shwartz and Nathan Srebro. SVM optimization: inverse dependence on training set size. In *Proceedings of the 25th international conference on Machine learning*, pages 928–935. ACM, 2008.

[22] E. Subag. Following the ground states of full-RSB spherical spin glasses. *CPAM*, 74(5):1021–1044, 2021.

[23] A. Wein. Optimal low-degree hardness of maximum independent set. *Math. Stat. & Learn.*, 4(3):221–251, 2022.

[24] Yuchen Zhang, Martin J Wainwright, and Michael I Jordan. Lower bounds on the performance of polynomial-time algorithms for sparse linear regression. In *Conference on Learning Theory (COLT)*, pages 921–948, 2014.