Final Program Report
Theoretical Foundations of Big Data Analysis (Fall 2013)

Michael Jordan (Organizing Chair)

The phenomenon of "Big Data" has become a central focus in a wide range of fields, including the physical sciences, the social sciences, the biological sciences and the inferential and computational sciences. The goal of the Simons Institute program on Big Data was to carve out a theoretical perspective on the topic. In what ways do the core ideas of complexity theory and the theory of algorithms need to be augmented or changed so as to address the inferential problems inherent in discussions of Big Data? How can computational ideas be integrated into the inferential theories of frequentist and Bayesian statistics? What role do theoretical ideas from mathematical disciplines such as optimization, combinatorics, signal processing, information theory and numerical linear algebra have to play?

A distinctive aspect of the program was the inclusion of leading researchers from each of these disciplines in the context of an overall focus on theoretical computer science. Indeed, while there have been many workshops bridging pairs of fields, the Simons program was perhaps the first in which major contingents of theoretical computer scientists, mathematical statisticians and optimization theorists have been brought together into a single program.

The Simons program was able to build on a decade or more of research in which ideas such as sampling, sketching, streaming and sparsity have been used to develop data analysis methodology that is increasingly scalable and robust. This backdrop provided many natural paths along which to proceed. For example, in what ways are sampling, sketching, streaming and sparsity complementary, so that their combination can be particularly useful for problems at large scale? How do we take into account emerging architectural constraints, particularly those associated with large-scale distributed computing? What kind of criteria does one use to assess efficiency in the setting of large-scale data analysis? In particular, how can we combine notions of time and space efficiency with statistical notions of risk and sample complexity?

In the next three sections of this report, we review some of the main themes that the program addressed, from the perspectives of the three major constituencies present in the program: theoretical computer science, statistics and optimization. We then turn to a review of selected concrete research outcomes to emerge from the program, focusing on collaborative efforts that arose during the program itself.

Theoretical Computer Science

Here are some of the key observations coming out of the program that are interesting from the perspective of theoretical computer science.

It is very important to study communication as a first-class resource (like space and time). Here are two settings where communications issues come up in quite different ways. In CPU-intensive linear algebraic computations that form the bedrock of scientific computing, minimizing communication between processing units serves two purposes: it allows computation to scale more gracefully with the number of processors, and it directly affects the energy output of the computation. While this latter aspect has not been explored traditionally, it is becoming more and more important as we hit millions of processing elements on a super computer and energy demands skyrocket.

The second setting is in data analysis. Here, the communication bottleneck arises from the

increase in data size (rather than the number of computing elements). If we have data distributed across multiple locations and wish to learn some model over the data, it is not practical to ship all the data to a single location. But most learning algorithms require that data be local for easy access. The theory of how to learn with distributed data in a communication-efficient manner is an important new development.

Statistical estimation introduces another resource to optimize. Traditional quality guarantees in algorithm design are in the form of approximations to an optimum for a given cost function. But for many statistical problems, the actual cost function is a proxy designed to help estimate a set of parameters for a model. But being close to the (proxy) cost function does not necessarily mean that we are close to the underlying model (where "close" is now measured statistically in terms of estimation error). If we now introduce statistical estimation error as a resource, there are results that suggest that obtaining a *weaker* approximation guarantee yields a better estimation error. We are still in the early days of understanding this phenomenon, but it is a new twist on traditional resource-constrained algorithm analysis.

Continuing the connection between optimization, learning and statistics, there are a number of first-order (and second-order) methods for getting local minima to optimizations associated with learning problems. Many of these methods involve solving linear systems, inverting matrices, estimating eigenstructures and so on. In the algorithms community, an impressive toolkit has been developed for getting approximation solutions to large linear algebra problems via sketching, sampling and approximations. We are now seeing some of these ideas percolate into numerical optimizations, with exact computations being replaced by provably approximate ones that maintain (approximate) guarantees on convergence while speeding up algorithms tremendously.

Statistics

The preceding two decades have seen much work in the statistical literature on high-dimensional statistics, where the number of observed variables is large relative to the number of data points. Ideas such as sparsity, convexity and concentration of measure have been exploited to develop statistical procedures that can perform well in high dimensions. The Big Data setting brings a number of new challenges to the fore—there is the computational challenge of achieving control on runtime and/or achieving reasonable approximation accuracy under subsampling, and there are a variety of issues associated with the heterogeneity and lack of control on the sampling of the data that is likely in Big Data settings. Furthermore, the problems of achieving control on error rates associated with large numbers of hypotheses are likely to be exacerbated in the Big Data setting, as the large amount of data is likely to lead researchers to wanting to be ambitious in terms of discovering new phenomena.

One productive approach to combine statistical perspectives with computational and architectural perspectives is to augment standard analyses of minimax risk with constraints. For example, one can ask what rate of convergence of minimax risk is achieved (both in terms of lower bounds and upper bounds that are achieved by specific procedures) in the distributed setting under communication constraints. One can also treat privacy as a constraint and compute minimax risks when the channel between the observed data and the statistician is constrained to respect differential privacy at level α.

The analysis of networks (e.g., social networks) continues to inspire much statistical research. Goals include obtaining models that are realistic (for example, capturing the emergence of large cliques at scale, and capturing power laws), that permit efficient statistical estimation and are tractable computationally. Much of the work conducted during the program revolved around ideas emerging from graph limits, exchangeability theorems and subsampling-based bootstrap.

Optimization

Ideas from optimization theory played a central, linking role in the program. Many statistical inference problems are naturally posed as optimization problems, and perspectives from combinatorial and constrained optimization have long driven research in theoretical computer science. Moreover, in aiming for optimality results in the form of lower bounds, the notions of oracle complexity from optimization theory have proved very useful. Here the space of procedures is much simplified relative to classical complexity theory and useful lower bounds (with matching upper bounds) are often within reach.

Much of the optimization-based work in the program focused on first-order methods, with stochastic gradient descent and stochastic coordinate descent notable areas of focus. Here "stochastic" refers to the notion that optimization functionals of interest in data analysis typically involve sums over data points, and it is natural to approximate such sums (and their gradients) with samples. In the "Big Data" setting such sums involve huge numbers of summands and the advantages of stochastic approaches are manifest.

The class of methods known as "conditional gradient," or "Frank-Wolfe," algorithms were promoted as instances of this overall focus on first-order methods; these methods make it possible to treat constrained optimization problems efficiently in the case of optimization problems involving sums of large number of data points.

Interactions between optimization and sparsity were also explored in the program. In particular, Big Data problems often involve vectors in which many components are equal to zero (e.g., in representations of linguistic data). In this setting it is natural to consider methods that have a second-order (while still stochastic) character, such that step sizes are tuned to focus on informative coordinates.

Lastly, much of the work in the optimization area was focused on taking account of the fact that statistical errors tend to scale as $1/\sqrt{n}$, where n is the number of data points, such that optimization accuracy beyond this range is of little value for inference (and indeed may be counter-productive). Focusing on methods that quickly arrive at $1/\sqrt{n}$ yields a different perspective from the (classical) focus on obtaining arbitrary accuracy.

Some Research Highlights

In this section we overview five major concrete results that arose out of the Big Data program. All of these results tackle aspects of the fundamental theoretical problem posed by Big Data, namely that of managing computational resources such as time, space, communication and privacy while maintaining control over statistical risk, and doing so in the context of large, high-dimensional data sets. All of these results have since appeared in leading peer-reviewed publications and are showing indications of significant impact. They were all based on collaborations that either formed *ab initio* during the program or were decisively catalyzed by the program.

Verifiable Stream Computation and Arthur-Merlin Communication

The *data streaming* paradigm considers settings in which one wants to compute a property of a massive input, but lacks the resources to store even a small fraction of the input. Streaming algorithms must therefore process the input in a single pass, using very limited memory (i.e., using space that is at most sublinear, and ideally polylogarithmic, in the size of the input). The model was formalized in seminal work of Alon, Matias and Szegedy in 1996, and since then it has received intense study. A collaboration between Simons program participants Amit Chakrabarti, Graham

Cormode, Andrew McGregor, Justin Thaler and Suresh Venkatasubramanian has yielded significant progress in this area.

Because the data streaming paradigm is so restrictive, it is impossible for streaming algorithms to obtain exactly correct answers for all but the simplest problems (without storing essentially the entire data stream). Given this reality, the goal becomes to develop small-space streaming algorithms that are guaranteed to return *approximately correct* answers with high probability. There are many important problems for which this is possible, such as computing the number of distinct elements, frequency moments, or frequent items in a data stream. However, there are many other problems that cannot even be approximated by streaming algorithms using sublinear space.

Streaming interactive proofs (SIPs) are a recent rethinking of the data streaming paradigm in light of the increasingly ubiquitous practice of outsourcing the processing of massive inputs to commercial data centers. In an SIP, a computationally limited client wants to perform some analysis of a massive input, but cannot do so because it is forced to operate within the restrictive data streaming paradigm. The client therefore accesses a powerful but untrusted service provider (e.g., a commercial cloud computing service) that *can* store and process the input. However, the client is unwilling to blindly trust answers returned by this service. Thus, the service cannot simply supply the desired answer; it must also convince the client of its correctness via a short interaction after the stream has been seen.

Several recent papers have introduced and studied SIPs and their variants. These papers have begun to reveal a rich theory, leveraging algebraic techniques developed in the classical theory of interactive proofs to obtain efficient verification protocols for a variety of problems that require linear space in the standard streaming model.

In work performed during the Big Data program, Chakrabarti *et al.* [36] studied "barely interactive" SIPs. Specifically, they showed that two or three rounds of interaction suffice to *exactly* solve several basic problems—including Median, Nearest Neighbor Search, Pattern Matching and Range Counting—with polylogarithmic space and communication costs. Such efficiency with a constant number of rounds of interaction was thought to be impossible based on previous work of Klauck and Prakash (*ITCS*, 2013); Chakrabarti *et al.* circumvent this lower bound claim by identifying and exploiting an implicit assumption in the earlier proof.

As a particular example of this SIP framework, here is an informal version of one of their results for Nearest Neighbor Search:

Result 1: For data sets consisting of points from $[n]^d$ under a reasonable metric, such as the Manhattan distance ℓ_1^d or the Euclidean distance ℓ_2^d, there is a three-round SIP with space and communication costs bounded by $\text{poly}(d, \log n)$ allowing exact nearest neighbor queries to the data set. The client's runtime per stream update is also $\text{poly}(d, \log n)$.

An important feature of this result is that the costs of the SIP for the client depend only polynomially on the dimension d of the dataset. In contrast, all known (even non-streaming) algorithms that exactly solve nearest neighbor search suffer from a "curse of dimensionality": they either require a data structure of superpolynomial size, or they force a query time that is either linear in the number of input data points, or exponential in the dimension d.

Their work also initiated a formal study of the *limitations* of contest-round SIPs, showing that our algorithms already approach the limits of existing algorithmic techniques. This study reveals several important phenomena exhibited by streaming interactive proofs that are not exhibited by their classical (non-streaming) counterparts.

Preserving Validity in Adaptive Data Analysis

There is a growing recognition throughout the scientific community that claims of statistical significance in published research are frequently invalid. There has been a great deal of effort to understand and propose mitigations for this problem, largely focusing on statistical approaches to controlling the false discovery rate in multiple hypothesis testing. However, the statistical inference theory surrounding this body of work assumes a fixed procedure to be performed, selected independently of the data. In contrast, the practice of data analysis in scientific research is by its nature an adaptive process, in which new analyses emerge on the basis of data exploration and previous analyses of the same data.

At the Big Data workshop on "Differential Privacy," program participants Cynthia Dwork, Vitaly Feldman, Moritz Hardt, Toniann Pitassi, Omer Reingold and Aaron Roth, inspired in large part by an influential tutorial on "Selective Inference and False Discovery Rate" presented at the workshop by Yoav Benjamini of Tel Aviv University, initiated a collaboration on understanding connections between the problems of preventing false discovery and preserving privacy in statistical data analysis. Recently, in the *ACM STOC* Conference [48] and in the journal *Science* [49], they have published some of their findings demonstrating that the challenges of adaptivity can, in fact, be addressed using insights from differential privacy—a notion of privacy preservation in data analysis introduced in computer science about ten years ago by Dwork *et al.*. Roughly speaking, differential privacy ensures that the probability of observing any outcome from an analysis is "essentially unchanged" by modifying any single data point. Such a condition is often called a "stability" guarantee. An important line of work establishes connections between the stability of a learning algorithm and its ability to *generalize* from the sample to the underlying distribution. It is known that certain stability notions are necessary and sufficient for generalization. Unfortunately, the stability notions considered in these prior works do not compose in the sense that running multiple stable algorithms sequentially and adaptively may result in a procedure that is not stable. Differential privacy is stronger than these previously studied notions of stability, and in particular enjoys strong adaptive composition guarantees.

As an application of this general connection, Dwork *et al.* propose a *reusable holdout method* that allows an analyst to safely reuse a holdout dataset many times to validate the results of adaptively chosen analyses. The analyst is given unfettered access to a training dataset, but can only access the holdout set via an algorithm that allows the analyst to validate statistics on the holdout set. Armed with such a tool, the analyst is free to explore the (training) data *ad libitum*, generating and computing statistics, validating them on the holdout, and repeating this procedure, as well as sharing outcomes with other analysts who may also use the same holdout set.

Simply put, the reusable holdout method is as follows: access the holdout set only via a differentially private algorithm. The intuition is that if we can learn about the data set in aggregate, while provably learning very little about any individual data element, then we can control the information leaked, and thus prevent overfitting.

Other researchers have already followed up on these ideas, including, for example, computational lower bounds obtained by Hardt and Ullman [68].

Hypothesis Testing for Large Networks

Extracting information from large graphs has become an important data-analysis problem, as network data is now common in various fields, including the social sciences, cellular biology and neuroscience. Network analysis has been pursued within both computer science and statistics but there is little work that bridges these disciplines.

From a statistical perspective the most basic type of questions are formalized through the notion of hypothesis testing. That is, given two competing generative models for a random network, one wants to decide which model is the most likely for the network at hand. In the last decade there has been an explosion of work on frameworks for realistic network models. Most of this work has focused on simply studying properties of networks rather than solving inferential problems such as hypothesis testing (a notable counter-example is the problem of community detection).

In a sequence of papers, Simons Institute visitors Sébastien Bubeck, Nati Linial, Elchanan Mossel and Miklós Rácz[1] introduced hypothesis-testing approaches for studying a range of problems in dimension reduction and growth. In particular, given a large graph, one wants to assess whether the observed connections result from a latent geometrical structure in the vertices, or are purely random. The paper [28] considered the most basic version of this problem in which the null hypothesis is that the observed graph on n vertices has been generated by the standard Erdös-Rényi random graph model $G(n,p)$, and where the alternative is a geometric model where each vertex is labeled with a point in some metric space, and an edge is present between two vertices if the distance between the corresponding labels is smaller than some prespecified threshold. They focus on the case where the underlying metric space is the Euclidean sphere, and the latent labels are i.i.d. uniform random vectors. This model is denoted by $G(n,p,d)$, where p is the probability of an edge between two vertices (p determines the threshold distance for connection). From a mathematical point of view the task is to understand the total variation distance between $G(n,p)$ and $G(n,p,d)$. This is a surprisingly rich question, which has led them to define a new basic statistic for networks—*signed triangles*—and which remains mysterious in the sparse regime where $p = c/n$. A particularly intriguing open question that emerged from this work is to understand the topological properties of the latent feature space that remain visible in the network data. For instance, can one detect if the sphere is replaced by a torus?

Most real-life networks are formed through a growth process. One dynamic growth model has become very popular in the recent years, namely the Preferential Attachment model $PA(n)$. The basic version of the PA model can be defined by induction as follows. First $PA(2)$ is the unique tree on two vertices. Then, given $PA(n)$, $PA(n+1)$ is formed from $PA(n)$ by adding a new vertex u and a new edge uv where v is selected at random among vertices in $PA(n)$ with probability proportional to their degrees. Note that we view $PA(n)$ as an unlabelled graph. One of the reason for the success of this model is that it captures the essence of the process behind the evolution of the Internet. However there is one potentially important aspect which is overlooked: in the early days of the Internet the PA model was probably very far from being a good approximation to the evolution of the network. It is perhaps reasonable to assume that after 1995 the network was evolving according to PA, but certainly from 1970 to 1995 the evolution followed fundamentally different rules. This observation suggests to study the PA model with a seed, that is we are now interested in $PA(n,T)$, where T is a finite seed tree (formally $PA(|T|,T) = T$ and $PA(n+1,T)$ is formed from $PA(n,T)$ as before). The basic hypothesis testing problem in this context is whether one can identify if the seed was S or T. It is not at all obvious that such a test is possible as the network becomes larger (the influence of the seed may vanish), but it turns out that the seed always has an influence in the following strong sense (the following result was conjectured and partially proved in [32]; the full proof is from Curien *et al.* slightly later): for any $S \neq T$ with at least three vertices,

$$\lim_{n\to\infty} TV(PA(n,S), PA(n,T)) > 0.$$

[1]Bubeck was a participant in the Big Data program, but the other three were primarily associated with the partner program on "Real Analysis in Computer Science"; thus this is a notable instance of cross-fertilization between the two programs.

Finally, the paper by Bubeck and Linial [31] investigated the extent to which large graphs can be characterized by their "k-profiles", i.e., the sets of small subgraphs that occur locally within them. This is an important and largely unexplored question that is very relevant to the compression of large graphical data sets: if small profiles provide a good characterization, then a large graph can be represented succinctly by its profile. Bubeck and Linial considered the interesting special case where the graphs are trees, and showed that this case is very different from that of general graphs: in particular, surprisingly, the limit set of k-profiles of trees is convex. Their work opens up a number of intriguing conjectures and questions that are sure to occupy mathematicians and computer scientists in the coming years.

Non-Oblivious Dimension Reduction

Big Data are often high-dimensional in nature and an important pre-processing tool in classical data analysis involves mapping a high-dimensional dataset to a lower-dimensional one. This makes it possible to apply more sophisticated algorithms to the data, while controlling the overall processing time. Dimension reduction methods are ubiquitous in the theory and practice of data analysis.

But it is also important to emphasize that there is a notable gap between the theory and the practice of dimension reduction methods. In theory, the dimension reduction methods, such as the classic Johnson-Lindenstrauss Lemma from 1984, are oblivious. These methods use a function that maps any high-dimensional vector into a lower dimensional one. Most importantly, it reduces the dimension without knowledge of the entire dataset, while providing strong guarantees of correctness and performance. Such methods have been used heavily for solving problems, such as Nearest Neighbor Search (NNS), with provable guarantees.

In contrast, in practice, there are many useful dimension reduction methods that are non-oblivious. These methods aim to optimize the dimension-reducing map by using knowledge of the entire dataset. One example of such a method is Principal Component Analysis (PCA), which extracts the maximal variance direction(s) of the dataset. Non-oblivious methods tend to outperform oblivious ones empirically. On the other hand, non-oblivious methods usually come with no guarantees, for either correctness or performance.

Program participants Alex Andoni and Ravi Kannan investigated the theoretical aspects of non-oblivious dimension reduction methods for the nearest neighbor search (NNS) problem. In the NNS problem, we are to preprocess a dataset of objects (such as images) so that, for a given new query object, we can quickly report similar objects in the dataset. Such objects are usually represented by high-dimensional vectors: e.g., a 20x20 image can be represented as a 400-dimensional vector, one coordinate per pixel. Andoni and Kannan, with collaborators Amirali Abdullah and Robert Krauthgamer [3], introduce a natural data model for such settings and show that, while PCA does not lead to a correct algorithm, it is possible to design more careful spectral algorithms that are correct and perform well. In particular they demonstrate algorithms that are based on iterative PCA and on a PCA-tree algorithm.

This work is the first to provide theoretical guarantees for NNS algorithms based on spectral non-oblivious dimension reduction methods that are popular in practice.

Communication Efficient Distributed Optimization for Machine Learning

The size and scale of machine learning applications has grown rapidly in recent years, giving rise to a major research focus on the parallelization of machine learning algorithms. Much of this work has focused on the shared-memory paradigm, but in real-world applications it is common to see fully-distributed platforms encompassing many thousands or ten of thousands of computers and many

terabytes of data. In such settings, communication costs can dominate data processing costs. Thus there is a major need for new theoretical work that can guide the design of communication-efficient machine learning.

During the Big Data program, participants Martin Jaggi, Michael Jordan and Virginia Smith initiated a new line of research aimed at fully-distributed optimization-based machine learning [75, 91]. Within a primal-dual setting, their framework uses local computation on dual variables to dramatically reduce the amount of communication necessary for a variety of standard machine learning algorithms. Local slaves are able to operate only on their own data in a block-parallel framework, with local results combined by a master via a simple averaging or adding operation. The primal-dual structure of the framework can be exploited to obtain theoretical convergence rates, even if the subproblems are only solved approximately. Moreover, the structure makes possible a form of meta-analysis in which local convergence rates can be combined to obtain a global convergence rate under generic assumptions about the local solvers.

A practical advantage of this framework is that it provides reusability of existing and well-established single machine solvers. Moreover, the framework is adaptive with respect to communication bandwidth.

Jaggi et al. have implemented these ideas in an open-source library, referred to as CoCoA (COmmunication efficient dual COordinate Ascent), which runs on the Apache Spark platform. They have shown empirically that CoCoA surpasses the state of the art on a a variety of distributed machine learning benchmarks.

Research Monograph

Finally we mention an additional project that has emerged from the program and will, it is hoped, serve to extend its impact. In an attempt to bridge the differences in viewpoint between the main constituencies of this diverse field, three program participants—Han Liu (Princeton; Statistics), Peter Richtarik (Univ of Edinburgh; Optimization) and Suresh Venkatasubramanian (Univ of Utah; CS Theory)—are in the final stages of writing a monograph on open problems in the theory of Big Data. This project was initiated towards the end of the semester-long program, and has continued largely through remote interaction since then; the three authors were invited back to the Institute during June 2015 to assist with the completion of the writing. We expect that the monograph will be published in the Foundations & Trends series of NOW Publishers, as well as being posted on the Institute's website.

Programmatic Considerations

The formal program of the Theoretical Foundations of Big Data semester was based around a sequence of four week-long workshops. There was also a one-week "Boot Camp" at the outset of the program. In this section we briefly consider the "lessons learned" regarding this programmatic organization.

The Boot Camp was viewed as a major success. The heterogeneity of the participants necessitated some form of introduction to the major theoretical perspectives involved in Big Data analysis (theoretical computer science, statistics and optimization). Indeed, few researchers have worked across all of these areas, and fewer still have kept up with recent developments in these rapidly developing fields. Nonetheless, there are enough commonalities at a technical level across the fields (with ideas from allied fields such as probability theory, optimization theory, geometry and information theory providing points of contact), such that Boot Camp lecturers were able to convey technical ideas succinctly and rapidly. Basic technical topics such as dimension reduction,

streaming, sampling, linear optimization, random matrices and concentration theory were reviewed, and emerging perspectives and challenges presented by scale were emphasized. Many participants referred to the Boot Camp as a major stimulus for their subsequent research.

The four workshops were all also viewed as quite successful. Their format was similar, involving mid-length talks (e.g., 45 minutes) that allowed technical concepts to be developed, and involving a mix of disciplines. That said, while the first two workshops were heavy on talks and light on discussion time, it became clear (from comments and attendance drop-off) that enhanced focus on discussion time and free time for research was preferred by many participants. The fourth workshop (on differential privacy) involved a mix of talk lengths (some 30 minutes and some one hour) and interactive events such as rump sessions, and participants appeared to be particularly satisfied by the format. This is not to suggest that this is generically a preferred model. Indeed, early in the semester the more talk-heavy format may be necessary to engender dialog. However, if we were to re-run the program I suspect that we would focus on formats involving fewer talks.

Theoretical Foundations of Big Data Analysis, Fall 2013

[1] A. ABDULLAH, R. KUMAR, A. MCGREGOR, S. VASSILVITSKII, and S. VENKATASUBRAMANIAN. Sketching and Dimensionality Reduction for Information Spaces. Manuscript. 2015.

[2] A. ABDULLAH and S. VENKATASUBRAMANIAN. A Directed Isoperimetric Inequality with application to Bregman Near Neighbor Lower Bounds. ACM Symposium on the Theory of Computing. 2015.

[3] A. ABDULLAH, A. ANDONI, R. KANNAN, and R. KRAUTHGAMER. Spectral Approaches to Nearest Neighbor Search. In *55th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2014.

[4] N. AGARWAL, G. KINDLER, A. KOLLA, and L. TREVISAN. Unique games on the Hypercube. In *Chicago Journal of Theoretical Computer Science*, 2014.

[5] N. K. AHMED, N. DUFFIELD, J. NEVILLE, and R. KOMPELLA. Graph sample and hold: A framework for big-graph analytics. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1446–1455, ACM, 2014.

[6] H. M. AKTULGA, A. BULUÇ, S. WILLIAMS, and C. YANG. Optimizing sparse matrix-multiple vectors multiplication for nuclear configuration interaction calculations. In *Proceedings of the IPDPS*, 2014.

[7] A. E. ALAOUI and M. W. MAHONEY. Fast Randomized Kernel Methods With Statistical Guarantees. *arXiv preprint arXiv:1411.0306v1,* 2014. Submitted for publication.

[8] A. AMIR, A. LEVY, E. PORAT, and B. R. SHALOM. Dictionary Matching with One Gap. In *Combinatorial Pattern Matching*, pp. 11–20, Springer, 2014.

[9] A. ANDONI, R. KRAUTHGAMER, and D. P. WOODRUFF. The Sketching Complexity of Graph Cuts. *arXiv preprint arXiv:1403.7058v2,* 2014.

[10] P. AWASTHI, A. BLUM, O. SHEFFET, and A. VIJAYARAGHAVAN. Learning mixtures of ranking models. In *Advances in Neural Information Processing Systems*, pp. 2609–2617, 2014.

[11] T. BANSAL, C. BHATTACHARYYA, and R. KANNAN. A provable SVD-based algorithm for learning topics in dominant admixture corpus. *arXiv preprint arXiv:1410.6991,* 2014.

[12] L. BARENBOIM. Deterministic (Delta + 1)-coloring in sublinear (in Delta) time in static, dynamic and faulty networks. 2015. Submitted to PODC'15.

[13] L. BARENBOIM. Efficient processing of dynamic data inspired by distributed computing. In preparation. 2015.

[14] L. BARENBOIM, M. ELKIN, and C. GAVOILLE. A fast network-decomposition algorithm and its appliacations to constant-time distributed computation. 2015. Submitted to PODC'15.

[15] A. R. BENSON, J. D. LEE, B. RAJWA, and D. F. GLEICH. Scalable methods for nonnegative matrix factorizations of near-separable tall-and-skinny matrices. In *Advances in Neural Information Processing Systems*, pp. 945–953, 2014.

[16] I. BEZÁKOVÁ and Z. LANGLEY. Minimum Planar Multi-sink Cuts with Connectivity Priors. In *Mathematical Foundations of Computer Science*, pp. 94–105, 2014.

[17] U. BHASKAR, K. LIGETT, and L. J. SCHULMAN. Network Improvement for Equilibrium Routing. In *Integer Programming and Combinatorial Optimization (IPCO)*, 138–149. Springer, 2014.

[18] U. BHASKAR, K. LIGETT, L. J. SCHULMAN, and C. SWAMY. Achieving Target Equilibria in Network Routing Games without Knowing the Latency Functions. In *Proceedings of the 55th IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 31–40, 2014.

[19] N. BHATNAGAR, A. SLY, and P. TETALI. Decay of Correlations for the Hardcore Model on the $ d $-regular Random Graph. *arXiv preprint arXiv:1405.6160,* 2014.

[20] S. BHATTACHARYYA and P. J. BICKEL. Community detection in networks using graph distance. *arXiv preprint arXiv:1401.3915,* 2014.

[21] S. BHATTACHARYYA and P. J. BICKEL. Subsampling Bootstrap of Count Features of Networks. *arXiv preprint arXiv:1312.2645,* 2013.

[22] P. BICKEL and P. SARKAR. Hypothesis testing for Automated Community Detection. JRSS B. 2015. To appear.

[23] A. BLUM and M. HARDT. The Ladder: A Reliable Leaderboard for Machine Learning Competitions. In *ICML*, 2015.

[24] C. BOUTSIDIS, P. DRINEAS, and M. MAGDON-ISMAIL. Rich Coresets for Constrained Linear Regression. *IEEE Transactions on Information Theory,* 59, 10, pp. 6880–6892, 2014.

[25] C. BOUTSIDIS, P. DRINEAS, and M. MAGDON-ISMAIL. Near-optimal column-based matrix reconstruction. *SIAM Journal on Computing,* 43, 2, pp. 687–717, 2014.

[26] C. BOUTSIDIS and D. P. WOODRUFF. Optimal CUR Matrix Decompositions. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, STOC '14, pp. 353–362, ACM, 2014.
[27] J. BRADIC and M. KOLAR. Inference for Regression Quantiles After Model Selection. Preprint. 2015.
[28] S. BUBECK, J. DING, R. ELDAN, and M. RACZ. Testing for high-dimensional geometry in random graphs. *arXiv preprint arXiv:1411.5713,* 2014.
[29] S. BUBECK, R. ELDAN, E. MOSSEL, and M. RACZ. From trees to seeds: On the inference of the seed from large trees in the uniform attachment model. *arXiv preprint arXiv:1409.7685,* 2014.
[30] S. BUBECK, G. LUGOSI, and L. DEVROYE. Finding Adam in random growing trees. *arXiv preprint arXiv: 1411.3317,* 2014.
[31] S. BUBECK and N. LINIAL. On the local profiles of trees. *arXiv preprint arXiv:1310.5396,* 2013. To appear in Journal of Graph Theory.
[32] S. BUBECK, E. MOSSEL, and M. Z. RÁCZ. On the influence of the seed graph in the preferential attachment model. *arXiv preprint arXiv:1401.4849,* 2014. To appear in IEEE Transactions on Network Science and Engineering.
[33] M. BUN and J. THALER. Hardness amplification and the approximate degree of constant-depth circuits. *arXiv preprint arXiv:1311.1616,* 2013.
[34] J.-Y. CAI, A. GALANIS, L. A. GOLDBERG, H. GUO, M. JERRUM, D. STEFANKOVIC, and E. VIGODA. #BIS-Hardness for 2-Spin Systems on Bipartite Bounded Degree Graphs in the Tree Nonuniqueness Region. *APPROX-RANDOM,* pp. 582–595, 2014.
[35] A. CHAKRABARTI, G. CORMODE, N. GOYAL, and J. THALER. Annotations for Sparse Data Streams. In *SODA*, pp. 687–706, SIAM, 2014.
[36] A. CHAKRABARTI, G. CORMODE, A. MCGREGOR, J. THALER, and S. VENKATASUBRAMANIAN. Verifiable Stream Computation and Arthur-Merlin Communication. *Computational Complexity Conference,* 2015.
[37] A. CHAKRABARTI and S. KALE. Submodular Maximization Meets Streaming: Matchings, Matroids, and More. In *Integer Programming and Combinatorial Optimization*, 210–221. Springer, 2014.
[38] X. CHEN and M. KOLAR. Fast Inference for Markov Random Fields. Working paper. 2015.
[39] Y. CHEN and R. J. SAMWORTH. Generalised additive and index models with shape constraints. *arXiv preprint arXiv:1404.2957v1,* 2014.
[40] P. CONSTANTINE and D. GLEICH. Computing Active Subspaces. *arXiv preprint arXiv:1408.0545v1,* 2014.
[41] G. CORMODE, S. MUTHUKRISHNAN, and J. YAN. First author advantage: citation labeling in research. In *Proceedings of the 2013 workshop on Computational scientometrics: theory & applications*, pp. 13–18, ACM, 2013.
[42] G. CORMODE, S. MUTHUKRISHNAN, and J. YAN. People like us: mining scholarly data for comparable researchers. In *Companion Publication of the 23rd International Conference on World Wide Web*, WWW Companion '14, pp. 1227–1232, International World Wide Web Conferences Steering Committee, 2014.
[43] P. DRINEAS and M. W. MAHONEY. Randomization in Numerical Linear Algebra (RandNLA). Invited to Communications of the ACM. 2015.
[44] J. DUCHI, M. JORDAN, and B. MCMAHAN. Estimation, optimization, and parallelism when data is sparse. In *Advances in Neural Information Processing Systems*, pp. 2832–2840, 2013.
[45] J. DUCHI, M. J. WAINWRIGHT, and M. JORDAN. Local privacy and minimax bounds: sharp rates for probability estimation. In *Advances in Neural Information Processing Systems*, pp. 1529–1537, 2013.
[46] J. C. DUCHI, M. I. JORDAN, M. J. WAINWRIGHT, and A. WIBISONO. Optimal rates for zero-order optimization: the power of two function evaluations. *arXiv preprint arXiv:1312.2139,* 2013. Accepted for publication at IEEE Transactions on Information Theory.
[47] J. C. DUCHI and M. J. WAINWRIGHT. Distance-based and continuum Fano inequalities with applications to statistical estimation. *arXiv preprint arXiv:1311.2669,* 2013.
[48] C. DWORK, V. FELDMAN, M. HARDT, T. PITASSI, O. REINGOLD, and A. ROTH. Preserving Statistical Validity in Adaptive Data Analysis. In *Proceedings of 47th Symposium on Theory of Computing (STOC)*, 2015.
[49] C. DWORK, V. FELDMAN, M. HARDT, T. PITASSI, O. REINGOLD, and A. ROTH. The reusable holdout: Preserving validity in adaptive data analysis. *Science,* 349, 6248, pp. 636-638, 2015.
[50] J. FAN, T. KE, H. LIU, and L. XIA. QUADRO: A Supervised Dimension Reduction Method via Rayleigh Quotient Optimization. *The Annals of Statistics,* 2015.
[51] E. X. FANG, B. HE, H. LIU, and X. YUAN. Generalized Alternating Direction Method of Multipliers: New Theoretical Insight and Application. *Mathematical Programming Computation,* 2015.

[52] G. FEIGENBLAT, E. PORAT, and A. SHIFTAN. An Improved Query Time for Succinct Dynamic Dictionary Matching. In *Combinatorial Pattern Matching*, pp. 120–129, Springer, 2014.
[53] O. FERCOQ and P. RICHTÁRIK. Accelerated, parallel and proximal coordinate descent. *arXiv preprint arXiv: 1312.5799,* 2013. To appear in SIAM Journal on Optimization.
[54] O. FERCOQ and P. RICHTÁRIK. Smooth minimization of nonsmooth functions with parallel coordinate descent methods. *arXiv preprint arXiv:1309.5885,* 2013.
[55] M. P. FRIEDLANDER, I. MACEDO, and T. K. PONG. Gauge optimization, duality, and applications. *SIAM Journal on Optimization,* 24, 4, pp. 1999-2022, 2014.
[56] A. GALANIS, D. STEFANKOVIC, and E. VIGODA. Inapproximability for antiferromagnetic spin systems in the tree non-uniqueness region. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pp. 823–831, ACM, 2014.
[57] A. GALANIS, D. STEFANKOVIC, E. VIGODA, and L. YANG. Ferromagnetic Potts Model: Refined #BIS-hardness and Related Results. *APPROX-RANDOM,* pp. 677–691, 2014.
[58] A. C. GILBERT, Y. LI, E. PORAT, and M. J. STRAUSS. For-all sparse recovery in near-optimal time. In *Automata, Languages, and Programming*, 538–550. Springer, 2014.
[59] D. F. GLEICH and M. W. MAHONEY. Algorithmic Anti-differentiation: A case study with min-cuts, spectral, and flow. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pp. 1018–1025, 2014.
[60] D. F. GLEICH and M. W. MAHONEY. Using local spectral methods to robustify graph-based learning algorithms. Preprint. 2015.
[61] F. HAN and H. LIU. Optimal Rates of Convergence for Latent Generalized Correlation Matrix Estimation in Transelliptical Distribution. *Bernoulli,* 2015.
[62] F. HAN and H. LIU. Scale-Invariant Sparse PCA on High Dimensional Meta-elliptical Data. *Journal of American Statistical Association (Theory and Methodology),* 109, 505, pp. 275–287, 2014.
[63] M. HARDT. Robust Subspace Iteration and Privacy-Preserving Spectral Analysis. In *Allerton Conference*, 2013.
[64] M. HARDT. Understanding alternating minimization for matrix completion. In IEEE, editors, *55th Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 651–660, 2014.
[65] M. HARDT, R. MEKA, P. RAGHAVENDRA, and B. WEITZ. Computational Limits for Matrix Completion. In *Proceedings of the Conference on Learning Theory (COLT)*, pp. 703–725, 2014.
[66] M. HARDT and E. PRICE. The Noisy Power Method: A Meta Algorithm with Applications. In *Advances in Neural Information Processing Systems*, pp. 2861–2869, 2014.
[67] M. HARDT and E. PRICE. Tight Bounds for Learning a Mixture of Two Gaussians. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing (STOC)*, 2015.
[68] M. HARDT and J. ULLMAN. Preventing false discovery in interactive data analysis is hard. In *Proceedings of the 55th IEEE Symposium on Foundations of Computer Science (FOCS)*, 2014.
[69] M. HARDT and M. WOOTTERS. Fast Matrix Completion Without the Condition Number. In *Proceedings of The 27th Conference on Learning Theory*, pp. 638–678, 2014.
[70] A. W. HARROW, A. KOLLA, and L. J. SCHULMAN. Dimension-free L2 maximal inequality for spherical means in the hypercube. *Theory of Computing, Special Issue on Boolean Functions,* 10, 3, pp. 55–75, 2014.
[71] B. HE, H. LIU, Z. WANG, and X. YUAN. A Strictly Contractive Peaceman-Rachford Splitting Method for Convex Programming. *SIAM Journal on Optimization,* 24, 3, pp. 1011–1040, 2014.
[72] D. S. HOCHBAUM and P. BAUMANN. Sparse computation for large-scale data mining. In *2014 IEEE International Conference on Big Data*, pp. 354-363, Oct 2014.
[73] K. IWAMA and A. NAGAO. Read-Once Branching Programs for Tree Evaluation Problems. In *LIPIcs-Leibniz International Proceedings in Informatics*, volume 25, Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2014.
[74] K. IWAMA and Y. YOSHIDA. Parameterized Testability. In *Proceedings of the 5th Conference on Innovations in Theoretical Computer Science*, ITCS '14, pp. 507–516, ACM, 2014.
[75] M. JAGGI, V. SMITH, M. TAKÁC, J. TERHORST, S. KRISHNAN, T. HOFMANN, and M. I. JORDAN. Communication-efficient distributed dual coordinate ascent. In *Advances in Neural Information Processing Systems (NIPS)*, pp. 3068–3076, 2014.
[76] K. JAMIESON, M. MALLOY, R. NOWAK, and S. BUBECK. lil'UCB: An Optimal Exploration Algorithm for Multi-Armed Bandits. *Proceedings of the Conference on Learning Theory (COLT),* 2014.
[77] J. JIANG, M. MITZENMACHER, and J. THALER. Parallel peeling algorithms. In *SPAA*, pp. 319–330, 2014. Best Paper Award.

[78] V. KANADE and J. THALER. Distribution-Independent Reliable Learning. In *Proceedings of the Conference on Learning Theory (COLT)*, pp. 3–24, 2014.
[79] R. KANNAN, S. S. VEMPALA, and D. P. WOODRUFF. Principal Component Analysis and Higher Correlations for Distributed Data. In *Proceedings of The 27th Conference on Learning Theory*, pp. 1040–1057, 2014.
[80] A. K. H. KIM and R. J. SAMWORTH. Global rates of convergence in log-concave density estimation. *arXiv preprint arXiv:1404.2298v1,* 2014.
[81] V. KING. Byzantine agreement in Polynomial Expected Time. Submitted to JACM. 2015.
[82] V. KING, S. KUTTEN, and M. THORUP. Construction and impromptu repair of an MST in a distributed network with o(m) communication. *arXiv preprint arXiv:1502.03320v1,* 2015. Submitted to PODC 2015.
[83] M. KOLAR, H. LIU, and E. XING. Graph Estimation From Multi-attribute Data. *Journal of Machine Learning Research,* 15, 1, pp. 1713–1750, 2014.
[84] J. KONEČNÝ and P. RICHTÁRIK. Semi-Stochastic Gradient Descent Methods. *arXiv preprint arXiv: 1312.1666v1,* 2013.
[85] T. KOPELOWITZ, R. KRAUTHGAMER, E. PORAT, and S. SOLOMON. Orienting fully dynamic graphs with worst-case time bounds. In *Automata, Languages, and Programming*, 532–543. Springer, 2014.
[86] S. LACOSTE-JULIEN and M. JAGGI. An Affine Invariant Linear Convergence Analysis for Frank-Wolfe Algorithms. *arXiv preprint arXiv:1312.7864,* 2013.
[87] J. LI, Y. RABANI, L. J. SCHULMAN, and C. SWAMY. Learning Arbitrary Statistical Mixtures of Discrete Distributions. In *Proceedings of 47th Symposium on Theory of Computing (STOC)*, 2015. To appear.
[88] Y. LI, H. L. NGUYÊN, and D. P. WOODRUFF. Turnstile Streaming Algorithms Might as Well Be Linear Sketches. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing (STOC)*, STOC '14, pp. 174–183, ACM, 2014.
[89] E. LIBERTY, M. MITZENMACHER, J. THALER, and J. ULLMAN. Space Lower Bounds for Itemset Frequency Sketches. *arXiv preprint arXiv:1407.3740,* 2014.
[90] H. LIU, L. WANG, and T. ZHAO. Multivariate Regression with Calibration. *Journal of Machine Learning Research,* 2015.
[91] C. MA, V. SMITH, M. JAGGI, M. I. JORDAN, P. RICHTARIK, and M. TAKAC. Adding vs. averaging in distributed primal-dual optimization. In *Proceedings of the 32th International Conference on Machine Learning*, 2015.
[92] Q. MA, S. MUTHUKRISHNAN, and M. SANDLER. Frugal Streaming for Estimating Quantiles: One (or two) memory suffices. *arXiv preprint arXiv:1407.1121,* 2014.
[93] A. MCGREGOR. Graph stream algorithms: A survey. *ACM SIGMOD Record,* 43, 1, pp. 9–20, 2014.
[94] A. MCGREGOR and S. MUTHUKRISHNAN. *Algorithms for modern data: Sketching and streaming*. Book in preparation, 2015.
[95] A. MCGREGOR, E. PRICE, and S. VOROTNIKOVA. Trace Reconstruction Revisited. In *Algorithms-ESA 2014*, 689–700. Springer, 2014.
[96] S.-Y. OH, O. DALAL, K. KHARE, and B. RAJARATNAM. Optimization Methods for Sparse Pseudo-Likelihood Graphical Model Selection. *arXiv preprint arXiv:1409.3768,* 2014.
[97] X. PAN, J. E. GONZALEZ, S. JEGELKA, T. BRODERICK, and M. JORDAN. Optimistic Concurrency Control for Distributed Unsupervised Learning. In *Advances in Neural Information Processing Systems*, pp. 1403–1411, 2013.
[98] X. PAN, S. JEGELKA, J. E. GONZALEZ, J. K. BRADLEY, and M. I. JORDAN. Parallel double greedy submodular maximization. In *Advances in Neural Information Processing Systems*, pp. 118–126, 2014.
[99] S. PAUL, M. MAGDON-ISMAIL, and P. DRINEAS. Feature Selection for Linear SVM with Provable Guarantees. In *Proceedings of the 16th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2015. To appear.
[100] J. J. PFEIFFER III, S. MORENO, T. LA FOND, J. NEVILLE, and B. GALLAGHER. Attributed graph models: Modeling network structure with correlated attributes. In *Proceedings of the 23rd international conference on World wide web*, pp. 831–842, International World Wide Web Conferences Steering Committee, 2014.
[101] J. J. PFEIFFER III, J. NEVILLE, and P. BENNETT. Composite Likelihood Data Augmentation for Within-Network Statistical Relational Learning. In *Proceedings of the 14th IEEE International Conference on Data Mining (ICDM)*, 2014.
[102] M. PILANCI and M. J. WAINWRIGHT. Randomized Sketches of Convex Programs with Sharp Guarantees.

arXiv preprint arXiv:1404.7203v1, 2014. Presented in part at ISIT 2014 Conference.
[103] H. QIU, F. HAN, H. LIU, and B. CAFFO. Joint Estimation of Multiple Graphical Models from High Dimensional Time Series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* 2015.
[104] Z. QU, P. RICHTÁRIK, M. TAKÁC, and O. FERCOQ. SDNA: Stochastic Dual Newton Ascent for Empirical Risk Minimization. *arXiv preprint arXiv:1502.02268,* 2015.
[105] P. RICHTÁRIK and M. TAKÁC. Distributed coordinate descent method for learning with big data. *arXiv preprint arXiv:1310.2059,* 2013.
[106] P. RICHTÁRIK and M. TAKÁC. On optimal probabilities in stochastic coordinate descent methods. *arXiv preprint arXiv:1310.3438,* 2013.
[107] M. ROSENBLUM, H. LIU, and E.-H. YEN. Optimal Tests of Treatment Effects for the Overall Population and Two Subpopulations in Randomized Trials, using Sparse Linear Programming. *Journal of American Statistical Association (Theory and Methodology),* 109, 507, pp. 1216–1228, 2014.
[108] R. J. SAMWORTH. Big Data: a new era for Statistics. *The Eagle,* pp. 43–46, 2014.
[109] P. SARKAR and P. J. BICKEL. Role of normalization in spectral clustering for stochastic blockmodels. *The Annals of Statistics,* 43, 3, pp. 962–990, 2015.
[110] L. J. SCHULMAN and A. SINCLAIR. Analysis of a classical matrix preconditioning algorithm. In *Proceedings of 47th Symposium on Theory of Computing (STOC)*, 2015. To appear.
[111] H. SHOU, R. T. SHINOHARA, H. LIU, D. S. REICH, and C. M. CRAINICEANU. Soft Null Hypotheses: A Case Study of Image Enhancement Detection in Brain Lesions. *Journal of Computational and Graphical Statistics,* 2015.
[112] A. SINCLAIR, P. SRIVASTAVA, D. STEFANKOVIC, and Y. YIN. Spatial mixing and the connective constant: Optimal bounds. In *SODA*, pp. 1549–1563, 2015.
[113] L. SONG, H. LIU, A. PARIKH, and E. XING. Nonparametric Latent Tree Graphical Models: Inference, Estimation, and Structure Learning. *Journal of Machine Learning Research,* 2014.
[114] V. STRNADOVA, A. BULUÇ, J. GONZALEZ, S. JEGELKA, J. CHAPMAN, J. GILBERT, D. ROKHSAR, and L. OLIKER. Efficient and accurate clustering for large-scale genetic mapping. In *The IEEE International Conference on Bioinformatics and Biomedicine (BIBM'14)*, 2014.
[115] R. TAPPENDEN, P. RICHTÁRIK, and B. BUKE. Separable approximations and decomposition methods for the augmented Lagrangian. *Optimization Methods and Software,* 2013.
[116] J. THALER. Semi-Streaming Algorithms for Annotated Graph Streams. *arXiv preprint arXiv:1407.3462,* 2014.
[117] T. WANG, Q. BERTHET, and R. J. SAMWORTH. Statistical and computational trade-offs in estimation of sparse principal components. *arXiv preprint arXiv:1408.5369v1,* 2014.
[118] Z. WANG, H. LIU, and T. ZHANG. Optimal Computational and Statistical Rates of Convergence for Sparse Nonconvex Learning Problems. *The Annals of Statistics,* 42, 6, pp. 2164-2201, 2014.
[119] Y. YU, T. WANG, and R. J. SAMWORTH. A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika,* 2015. To appear.
[120] Y. ZHANG, X. CHEN, D. ZHOU, and M. I. JORDAN. Spectral methods meet EM: A provably optimal algorithm for crowdsourcing. In *Advances in Neural Information Processing Systems*, pp. 1260–1268, 2014.
[121] Y. ZHANG, J. DUCHI, M. JORDAN, and M. J. WAINWRIGHT. Information-theoretic lower bounds for distributed statistical estimation with communication constraints. In *Advances in Neural Information Processing Systems*, pp. 2328–2336, 2013.
[122] Y. ZHANG, M. J. WAINWRIGHT, and M. I. JORDAN. Lower bounds on the performance of polynomial-time algorithms for sparse linear regression. *Proceedings of the Conference on Learning Theory (COLT),* 2014.
[123] Y. ZHANG, M. J. WAINWRIGHT, and M. I. JORDAN. Distributed Estimation of Generalized Matrix Rank: Efficient Algorithms and Lower Bounds. *arXiv preprint arXiv:1502.01403v2,* 2015.
[124] Y. ZHANG, M. J. WAINWRIGHT, and M. I. JORDAN. Optimal prediction for sparse linear models? Lower bounds for coordinate-separable M-estimators. *arXiv preprint arXiv:1503.03188v1,* 2015.
[125] T. ZHAO, M. KOLAR, and H. LIU. A General Framework for Robust Testing and Confidence Regions in High-Dimensional Quantile Regression. *arXiv preprint arXiv:1412.8724v2,* 2015.
[126] Y. ZHOU, X. CHEN, and J. LI. Optimal PAC Multiple Arm Identification with Applications to Crowdsourcing. In *Proceedings of The 31st International Conference on Machine Learning*, volume 32, pp. 217–225, 2014.
[127] P. ZWIERNIK, C. UHLER, and D. RICHARDS. Maximum Likelihood Estimation for Linear Gaussian Covariance Models. *arXiv preprint arXiv:1408.5604v1,* 2014.