REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
27-08-2013	Final Report	1-Oct-2012 - 30-Jun-2013

4. TITLE AND SUBTITLE
High Dimensional Learning

5a. CONTRACT NUMBER
W911NF-12-1-0404

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

6. AUTHORS
Animashree Anandkumar

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
University of California - Irvine
5171 California Avenue, Suite 150

Irvine, CA 92697 -7600

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
62594-NS-II.8

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
The problem of high dimensional learning is considered. Efficient methods are developed for learning latent variable models and graphical models in high dimensions. Theoretical guarantees are established for the developed methods. The methods are applied to various domains including social networks and computational biology.

15. SUBJECT TERMS
high dimensional learning, graphical models, latent variable models, graph estimation, spectral methods

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
UU	UU	UU	UU		Animashree Anandkumar
					19b. TELEPHONE NUMBER
					949-824-9072

Standard Form 298 (Rev 8/98)
Prescribed by ANSI Std. Z39.18

Report Title

High Dimensional Learning

ABSTRACT

The problem of high dimensional learning is considered. Efficient methods are developed for learning latent variable models and graphical models in high dimensions. Theoretical guarantees are established for the developed methods. The methods are applied to various domains including social networks and computational biology.

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Received Paper

TOTAL:

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

Received Paper

TOTAL:

Number of Papers published in non peer-reviewed journals:

(c) Presentations

Number of Presentations: 0.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received	Paper

TOTAL:

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received		Paper
07/08/2013	4.00	Animashree Anandkumar, Daniel Hsu, Adel Javanmard, Sham M. Kakade. Learning Linear Bayesian Networks with Latent Variables, International Conference on Machine Learning. , . : ,
07/08/2013	5.00	A. Anandkumar, D. Hsu, F. Huang, S.M. Kakade.. Learning High-Dimensional Mixtures of Graphical Models , Neural Information Processing. , . : ,
07/08/2013	6.00	A. Anandkumar, R. Valluvan. Learning Loopy Graphical Models with Latent Variables: Efficient Methods and Guarantees, Neural INformation Processing. , . : ,
07/08/2013	7.00	A. Anandkumar, D. P. Foster, D. Hsu, S.M. Kakade, Y.K. Liu.. A Spectral Algorithm for Latent Dirichlet Allocation, Neural Information Processing. , . : ,
07/08/2013	1.00	Animashree Anandkumar, Rong Ge, Daniel Hsu, Sham M. Kakade. A Tensor Spectral Approach to Learning Mixed Membership Community Models, Conference on Learning Theory. , . : ,

TOTAL: **5**

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

Received	Paper

TOTAL:

Number of Manuscripts:

Books

Received   Paper

TOTAL:

Patents Submitted

Patents Awarded

Awards

Microsoft Faculty Fellowship 2013
ARO Young Investigator Award (YIP) 2013
NSF CAREER Award 2013

Graduate Students

NAME	PERCENT_SUPPORTED	Discipline
Majid Janzamin	1.00	
FTE Equivalent:	**1.00**	
Total Number:	**1**	

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Names of Faculty Supported

NAME	PERCENT_SUPPORTED	National Academy Member
Animashree Anandkumar	0.50	
FTE Equivalent:	**0.50**	
Total Number:	**1**	

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period:	0.00
The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:......	0.00
The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:......	0.00
Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):......	0.00
Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:......	0.00
The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense	0.00
The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields:......	0.00

Names of Personnel receiving masters degrees

NAME

Total Number:

Names of personnel receiving PHDs

NAME

Total Number:

Names of other research staff

NAME	PERCENT_SUPPORTED

FTE Equivalent:

Total Number:

Sub Contractors (DD882)

Inventions (DD882)

Scientific Progress

See Attachment

Technology Transfer

Report for W911NF1210404: High Dimensional Learning

Motivation

Today we are facing a "data deluge" in almost every domain. Online social networks have seen an explosion in activity and have fundamentally transformed the nature of human interaction. In the biological realm, modern genome sequencers can output data at a rate 400 times faster than the ones a decade ago, and so on. However, although having a transformative potential, the data deluge has not yet been exploited to the fullest extent. Ironically, the data deluge has also resulted in a "data desert". The collected data in many domains are noisy, subsampled, with typically a large number of variables or "unknowns" compared to the number of observations or the "knowns". Such high-dimensionality entails practical principled approaches for learning from ill-posed and ill-behaved data.

Some of the fundamental questions in high-dimensional learning are: Can we design **scalable** models for efficiently representing and learning high-dimensional data? Here, scalability refers to low **computational** requirements and reduced **sampling** of high-dimensional data. Not all phenomena can be learnt in a scalable manner. Can we characterize the **fundamental limits** on complexity of learning complex phenomena? As part of this project, the PI has tackled the above challenges by exploiting "inherent data architecture". This can be in the form of structural relationships among the variables, represented as **graphs**, or as parametric forms, represented as **tensor decompositions**. The PI has developed novel approaches for handling such high-dimensional data.

1 Summary of Results: Tensor Approaches for Learning Latent Variable Models

Mixture Models: Classically, latent variables have been incorporated via mixture models. A mixture model can be thought of as selecting the distribution of the observed variables, based on a so-called latent choice variable. **Gaussian mixtures** are the most well studied class of mixture models. Recently the so-called class of **exchangeable topic models** such as **latent Dirichlet allocation** have been popular for modeling large word corpora [1]. These models incorporate documents with multiple hidden topics. We propose efficient methods for learning these popular mixture models.

Challenges: Learning general latent variable models through maximum likelihood is NP-hard. Previous methods with theoretical consistency guarantees have high computational and sample complexity which typically scale exponentially with the latent space dimensionality. The current practice for estimating latent variable models is mostly through local search heuristics (e.g., the EM algorithm) which are prone to failure in high dimensions.

Spectral Approach to Inverse Moment Methods: The method of moments presents a powerful alternative to EM and other heuristics. The basic paradigm of method of moments [2] is to: (i) compute certain statistics of the data — often empirical moments such as means and correlations — and (ii) find model parameters that give rise to (nearly) these moments. The second step of equation solving to obtain the parameters can typically reduced to operations on the "spectrum" of matrices and tensors obtained from the moments. Finally, these problems have efficient iterative methods to find the solutions, even though they are non-convex.

Single Topic Exchangeable Model: Consider a simple bag-of-words model for documents in which the words in the document are assumed to be *exchangeable*. Recall that a collection of random variables $x_1, x_2, \ldots, x_\ell$ are exchangeable if their joint probability distribution is invariant to permutation of the indices. The well-known De Finetti's theorem [3] implies that such exchangeable models can be viewed as mixture models in which there is a latent variable h such that $x_1, x_2, \ldots, x_\ell$ are conditionally i.i.d. given h (see Figure 1 for the corresponding graphical model) and the conditional distributions are identical at all the nodes.

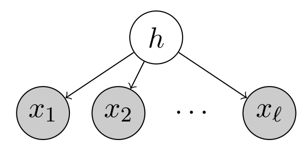


Figure 1: Exchangeable Topic Models.

In our simplified topic model for documents, the latent variable h is interpreted as the (sole) topic of a given document, and it is assumed to take only a finite number of distinct values. Let k be the number of distinct topics in the corpus, d be the number of distinct words in the vocabulary, and $\ell \geq 3$ be the number of words in each document. The generative process for a document is as follows: the document's topic is drawn according to the discrete distribution specified by the probability vector $w := (w_1, w_2, \ldots, w_k) \in \Delta^{k-1}$. This is modeled as a discrete random variable h such that $\Pr[h = j] = w_j$, for $j \in [k]$. Given the topic h, the document's ℓ words are drawn independently according to the discrete distribution specified by the probability vector $\mu_h \in \Delta^{d-1}$. It will be convenient to represent the ℓ words in the document by d-dimensional random *vectors* $x_1, x_2, \ldots, x_\ell \in \mathbb{R}^d$. Specifically, we set

$$x_t = e_i \quad \text{if and only if} \quad \text{the } t\text{-th word in the document is } i, \quad t \in [\ell],$$

where $e_1, e_2, \ldots e_d$ is the standard coordinate basis for $\mathbb{R}^d$. Because the words are conditionally independent given the topic, we can use this same property with conditional cross moments, say, of x_1 and x_2:

$$\mathbb{E}[x_1 \otimes x_2 | h = j] \;=\; \mathbb{E}[x_1 | h = j] \otimes \mathbb{E}[x_2 | h = j] \;=\; \mu_j \otimes \mu_j, \quad j \in [k].$$

This and similar calculations lead one to the following results: If $M_2 := \mathbb{E}[x_1 \otimes x_2]$ and $M_3 := \mathbb{E}[x_1 \otimes x_2 \otimes x_3]$, then $M_2 = \sum_{i=1}^k w_i \; \mu_i \otimes \mu_i, \quad M_3 = \sum_{i=1}^k w_i \; \mu_i \otimes \mu_i \otimes \mu_i.$. In [4], the PI establishes that for many classes of latent variable models, including spherical **Gaussian mixtures, latent Dirichlet allocation** and **hidden Markov models**, using low-order moments (typically first, second- and third-order), we can obtain a *symmetric tensor form* as above. So the problem of parameter estimation reduces to finding the components of the tensor.

Reduction to Orthogonal Symmetric Tensors: While general tensor decomposition is NP-hard, the PI establishes that the symmetric tensor decomposition can be reduced to an orthogonal symmetric decomposition given the moments as above, when the number of topics $k \leq d$, where

d is the dimension of observed space (i.e., vocabulary size for topic models). Additionally we require non-degeneracy: the vectors $\mu_1, \mu_2, \ldots, \mu_k \in \mathbb{R}^d$ are linearly independent, and the scalars $w_1, w_2, \ldots, w_k > 0$ are strictly positive.

Now, let $W \in \mathbb{R}^{d \times k}$ be a linear transformation such that $M_2(W, W) = W^\top M_2 W = I$, *i.e.*, W whitens M_2. Let $\tilde{\mu}_i := \sqrt{w_i}\ W^\top \mu_i$. Observe that $M_2(W, W) = \sum_{i=1}^k \tilde{\mu}_i \tilde{\mu}_i^\top = I$, so the $\tilde{\mu}_i \in \mathbb{R}^k$ are orthonormal vectors. Now define $\tilde{M}_3 := M_3(W, W, W) \in \mathbb{R}^{k \times k \times k}$, so that

$$\tilde{M}_3 \;=\; \sum_{i=1}^k w_i\ (W^\top \mu_i)^{\otimes 3} \;=\; \sum_{i=1}^k \frac{1}{\sqrt{w_i}}\ \tilde{\mu}_i^{\otimes 3}$$

is an orthogonal symmetric tensor.

Tensor Power Iterations: Efficient Methods for Tensor Decomposition: The orthogonal tensor decomposition encountered in these models can be efficiently solved through a simple power iteration method. For a tensor T, consider the vector-valued map

$$u \mapsto T(I, u, u). \tag{1}$$

This can be explicitly written as $T(I, u, u) = \sum_{1 \le j, l \le d} T_{i,j,l} (e_j^\top u)(e_l^\top u) e_i$. Observe that (1) is *not* a linear map, which is a key difference compared to the matrix case.

An eigenvector u for a matrix M satisfies $M(I, u) = \lambda u$, for some scalar λ. We say a unit vector $u \in \mathbb{R}^n$ is an *eigenvector* of T, with corresponding *eigenvalue* $\lambda \in \mathbb{R}$, if $T(I, u, u) = \lambda u$. For orthogonally decomposable tensors $T = \sum_{i=1}^k \lambda_i v_i^{\otimes 3}$,

$$T(I, u, u) = \sum_{i=1}^k \lambda_i (u^\top v_i)^2 v_i\ .$$

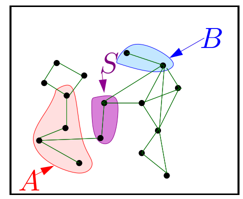
By the orthogonality of the v_i, it is clear that $T(I, v_i, v_i) = \lambda_i v_i$ for all $i \in [k]$. Therefore each (v_i, λ_i) is an eigenvector/eigenvalue pair. Thus, we can find *robust eigenvectors* through a simple power iteration: $\bar{\theta} \mapsto \frac{T(I, \bar{\theta}, \bar{\theta})}{\|T(I, \theta, \theta)\|}$ and it turns out that all the basis vectors turn out to be robust.

Thus, the PI presents guaranteed algorithms for learning latent variable models with low sample and computational complexities (as a low order polynomial in the latent space dimensionality). Additionally a subtle perturbation analysis controls the perturbation in multiple deflation stages of the power method. This can be seen as analogue of Weyl's and Wedin's theorems for singular value perturbation for matrices. Moreover, the proposed tensor power iteration algorithm is efficient for large-scale implementation and can be implemented using extremely simple linear algebra operations such as singular value decomposition and tensor power iterations.

Method of Moments for learning Community Models in Social Networks: The PI has also employed the method of moments for learning another class of latent variable models, viz., community models in social networks and has conducted some preliminary on-going work in [5]. A community generally refers to a group of individuals with shared interests (e.g. music, sports), or relationships (e.g. friends, co-workers). In [5], The PI considers a **mixed membership model** which incorporates overlapping communities, i.e., an agent can be part of multiple communities, which is realistic. The PI proposes a novel algorithm for learning these models, based on simple edge counts and "3-star" counts (i.e., a star with three leaves). This is the first work to present a guaranteed method for learning mixed membership community models. Moreover, the results are tight and match the best known bounds (e.g. for spectral clustering) in the special case of the **stochastic block model**, a well-studied model where individuals are present in only one community. The PI's research group has implemented these methods on graphics processing units (GPU), which makes it tractable to learn communities in social networks in an extremely fast manner [6].

2 Summary of Results: Learning Graph-Based Models

Probabilistic Graphical Models: One graphical framework for representing high-dimensional data is that of *probabilistic graphical models*, also known as *Markov random fields* or *Markov networks*. A Markov network represents complex relationships between data at different nodes in the form of a graph, known as the *dependency graph* [7–10]. Mathematically, any two sets of nodes A and B are conditionally independent, conditioned on the separator set S, as shown in the figure. Hence, the data at each node is influenced mainly by its neighbors in the dependency graph. A Markov representation is succinct with a much smaller number of parameters than the number of data dimensions (variables), and at the same time, it explicitly encodes the relationships between the variables.



$\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B | \mathbf{X}_S$

Formulation of Learning from Data: Given n i.i.d. data samples $\mathbf{x}^n := [\mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(n)]^T$ from a graphical model P with Markov graph G, the goal is the estimate the underlying graph. The PI proposes methods and provides consistency guarantees for graph estimation in the high dimensional regime.

Structure Learning with Hidden Variables: Developing tractable methods to discover hidden nodes and the overall graph structure(s) (and parameters) was an important goal of this project. The PI has developed efficient methods for learning latent variable models in a variety of settings. This includes the development of novel methods for learning hidden tree models [11, 12], which are especially relevant in *phylogenetics* [13]. Phylogenetics involves the estimation of the evolutionary tree process which resulted in the present-day species. The developed algorithms have low sample complexity and are much faster and more robust than the state of art. The algorithm, at a high level, maintains a tree model in each iteration and adds hidden variables by conducting local tests. This property is unique to our approach and makes it amenable for applying it to real data since we can tradeoff model complexity and data fitting in a principled and an efficient manner. The PI has extended these methods for learning latent loopy models with long cycles [14], and has demonstrated effectiveness in financial and topic modeling.

Bayesian Networks with Latent Variables: In addition to incorporating latent variables, it is important to model the complex dependencies among the variables. In [15], the PI provides novel methods for learning directed acyclic graphs (DAG) with hidden variables. The method is based on the intuition that learning is tractable when there is sufficient expansion in the DAG from hidden to observed variables (e.g. when it is latent tree or has small number of collidiers, i.e., nodes with multiple parents). This work combines sparse dictionary learning with method of moments in a novel manner and is the first work to provide guaranteed learning for latent Bayesian networks. This has implications in many practical settings, e.g. for learning **correlated topic models**.

Modeling Using Multiple Graphs: Modeling high-dimensional data involves a delicate trade-off between faithful representation and parsimony. Models which are sparse in some domain achieve a parsimonious representation but may poorly fit the given data. The PI has developed frameworks for relaxing the sparsity constraints without sacrificing on parsimony in high dimensions. One framework involves incorporating hidden factors which can change the structural (and parametric) relationships among the observed variables [16], thereby resulting in a mixture of probabilistic graphical models. The PI has developed methods with guaranteed recovery of mixture components which are also efficient for practical implementation. The PI has also considered another approach for modeling with multiple graphs. In [17], the observed data is fitted to a combination of a sparse graphical model and a sparse independence model, thereby incorporating different kinds of statistical relationships among the variables. The PI has developed novel decomposition methods based on convex relaxation with guaranteed recovery in both the domains.

The above developed algorithms have been applied by the PI to a number of practical problems including financial and document modeling [11], object recognition in computer vision [18], to track the evolution of dynamic social networks [19] and to model gene associations [20]. The PI's approaches have shown a huge improvement over previous ones in all these instances.

3 Significance and Impact of Conducted Research

Impact on the theory of high-dimensional learning: The PI's recent contributions lie at the forefront of innovation in big data and high-dimensional machine learning. She has provided a new theoretical understanding of tractable models and regimes for high-dimensional learning, developed novel approaches for handling massive scale data and also analyzed fundamental limits on learning. Her work has direct implications to the areas of machine learning, statistics and algorithms, as well as to a number of applications such as social network analysis, document categorization, computer vision, recommendation systems, and computational biology.

The approaches employed by the PI involve a cross-pollination of tools and techniques from machine learning, statistics, signal processing, information theory, optimization, random graph models, and social sciences. In particular, her work brings together techniques from machine learning and statistics (e.g. probabilistic graphical models, mixture models), information theory (fundamental information limits), signal processing (e.g independent component analysis), optimization (e.g convex relaxation techniques and tensor algebra), statistical physics (e.g. phase transitions) and social sciences (e.g. community formation models). This cross-disciplinary fusion of methods allows the problem of data deluge to be tackled in ways far more effective than any individual approach.

Another significant contribution by the PI is to the area of **learning latent variable models**. It is widely recognized that incorporating latent or hidden variables is a crucial aspect of modeling. Latent variables can provide a succinct representation of the observed data through dimensionality reduction; the possibly many observed variables are summarized by fewer hidden effects. Further, they are central to predicting causal relationships and interpreting the hidden effects as unobservable concepts. For instance in sociology, human behavior is affected by abstract notions such as social attitudes, beliefs, goals and plans. As another example, medical knowledge is organized into casual hierarchies of invading organisms, physical disorders, pathological states and symptoms, and only the symptoms are observed. However, learning general latent variable models is challenging (in fact, it is NP-hard). Previous methods with theoretical consistency guarantees have high computational and sample complexity, which typically scale exponentially with the latent space dimensionality. The current practice for estimating latent variable models is mostly through local search heuristics (e.g., the EM algorithm) which are prone to failure in high dimensions.

The PI has been able to circumvent the above challenges and she has developed novel scalable approaches for learning a wide class of latent variable models, which are guaranteed to succeed, and require only polynomial sample and computational complexity [4,5,15,21,22]. The PI has been able to achieve these impressive results by invoking the underlying **tensor algebra** in many popular latent variable models such as *Gaussian mixture, latent Dirichlet allocation* and *hidden Markov models*. These models are relevant in a number of applications including document modeling, natural language processing, as well as detecting overlapping communities in social networks.

In particular, the PI's work on community detection [23] is the first guaranteed approach for **learning mixed membership community models**, which are highly relevant for modeling online social networks. Community detection is a classical problem studied in theoretical computer science, statistics and sociology (see [23] for a survey). Previous theoretical guarantees for community detection were mostly limited to the setting where each node belongs to a single community (popularly known as the *stochastic block model*). In contrast, The PI's innovative approach provides guaranteed recovery of hidden communities for a wide class of models where communities can overlap, and also provides tight guarantees for the special case of the stochastic block model. Thus, the PI's work significantly advances the state of art on community detection in social networks.

Impact on applications of high-dimensional learning: Research involving learning from high-dimensional data has widespread application. The PI has been actively involved in transforming her theoretical results to practical algorithms in several domains. For instance, her algorithms have been applied for **text modeling** [11, 14], to automatically categorize words into (local) hierarchies of topics. It has been applied for **object recognition in computer vision** [18], where robust detection is achieved by exploiting the contextual information in natural images using co-occurrence of objects. Another important application is to model the co-evolution of vertices and edges in **dynamic social networks** [19]. Recently, the PI is collaborating with domain experts to apply the developed algorithms for **modeling gene associations** and predicting relationships between regulators and genes [20]. The PI's research group has implemented tensor decomposition algorithms on graphics processing units (GPU), and can **detect overlapping communities in large graphs** efficiently [6]. In all these instances, The PI's approaches have shown a huge improvement in performance over previous ones. Thus, The PI has made great strides in pushing the boundaries of large-scale machine learning, on both theoretical and practical fronts.

References Cited

[1] D.M. Blei, A.Y. Ng, and M.I. Jordan. Latent dirichlet allocation. *J. of Machine Learning Research*, 3:993–1022, 2003.

[2] K. Pearson. Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society, London, A.*, page 71, 1894.

[3] T. Austin et al. On exchangeable random variables and the statistics of large graphs and hypergraphs. *Probab. Surv*, 5:80–145, 2008.

[4] A. Anandkumar, R. Ge, D. Hsu, S. M. Kakade, and M. Telgarsky. Tensor Methods for Learning Latent Variable Models. *Under Review. J. of Machine Learning. Available at arXiv:1210.7559*, Oct. 2012.

[5] A. Anandkumar, R. Ge, D. Hsu, and S. M. Kakade. A Tensor Spectral Approach to Learning Mixed Membership Community Models. *ArXiv 1302.2684*, Feb. 2013.

[6] F. Huang, U.N. Niranjan, M. Hakeem, P. Verma, and A. Anandkumar. Fast Detection of Overlapping Communities via Online Tensor Methods on GPUs. *Under submission*, Aug. 2013.

[7] S.L. Lauritzen. *Graphical models*. Clarendon Press, 1996.

[8] J. Pearl. *Probabilistic Reasoning in Intelligent Systems—Networks of Plausible Inference*. Morgan Kaufmann, 1988.

[9] M.J. Wainwright and M.I. Jordan. Graphical Models, Exponential Families, and Variational Inference. *Foundations and Trends in Machine Learning*, 1(1-2):1–305, 2008.

[10] M. I. Jordan. *Learning in Graphical Models*. The MIT Press, Cambridge, MA, 1998.

[11] M.J. Choi, V.Y.F. Tan, A. Anandkumar, and A. Willsky. Learning latent tree graphical models. *J. of Machine Learning Research*, 12:1771–1812, May 2011.

[12] A. Anandkumar, K. Chaudhuri, D. Hsu, S.M. Kakade, L. Song, and T. Zhang. Spectral Methods for Learning Multivariate Latent Tree Structure. *Preprint, ArXiv 1107.1283*, July 2011.

[13] R. Durbin, S. R. Eddy, A. Krogh, and G. Mitchison. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge Univ. Press, 1999.

[14] A. Anandkumar and R. Valluvan. Learning Loopy Graphical Models with Latent Variables: Efficient Methods and Guarantees. *Ann. Statist.*, 41(2):401–435, 2013.

[15] A. Anandkumar, D. Hsu, and A. Javanmard S. M. Kakade. Learning Topic Models and Latent Bayesian Networks Under Expansion Constraints. *Preprint. ArXiv:1209.5350*, Sept. 2012.

[16] A. Anandkumar, D. Hsu, F. Huang, and S.M. Kakade. Learning High-Dimensional Mixtures of Graphical Models. *Accepted to Annals of Statistics.*, Feb. 2013.

[17] M. Janzamin and A. Anandkumar. High-Dimensional Covariance Decomposition into Sparse Markov and Independence Domains. In *Proc. of International Conf. on Machine Learning*, June 2012.

[18] M.J. Choi, A. Torralba, and A.S. Willsky. Context models and out-of-context objects. *Pattern Recognition Letters*, 2012.

[19] R. Valluvan, Z. W. Almquist, C. T. Butts, and A. Anandkumar. Semi-parametric vertex set prediction for dynamic networks using latent tree models. In *Intl. Conf. for Social Network Analysis (Sunbelt 2012)*, Redondo Beach, CA, March 2012.

[20] A. Gitter, F. Huang, R. Valluvan, E. Fraenkal, and A. Anandkumar. Unsupervised Learning of Transcriptional Regulatory Networks via Latent Tree Graphical Models. *In Preparation*, 2013.

[21] A. Anandkumar, D. Hsu, and S.M. Kakade. A Method of Moments for Mixture Models and Hidden Markov Models. In *Proc. of Conf. on Learning Theory*, June 2012.

[22] A. Anandkumar, D. P. Foster, D. Hsu, S. M. Kakade, and Y. K. Liu. A Spectral Algorithm for Latent Dirichlet Allocation. In *Proc. of Neural Information Processing (NIPS)*, Dec. 2012.

[23] A. Anandkumar, R. Ge, D. Hsu, and S. M. Kakade. A Tensor Spectral Approach to Learning Mixed Membership Community Models. In *Conference on Learning Theory (COLT)*, June 2013.