TRIPODS+X:RES: Investigations at the Interface of Data Science and Neuroscience

Project Summary

The research in this proposal lies at the interface between neuroscience and data science. The underlying theme is to develop a two-way channel between data science and cellular and cognitive neuroscience. In one direction, we will investigate how computational principles of data science can be used to understand recent empirical findings in neuroscience, associated with measurements at the cellular level in fruit flies, and brain imaging studies in humans. In the reverse direction, the project will view the processes and mechanisms of vision and cognition underlying these findings as a source for new mathematical frameworks for data analysis. The research will focus on four interrelated objectives:

Objective 1: Distributed processing. Recent work in machine learning has studied the effect of communication constraints and parallelization in distributed estimation. There is a close parallel in vision, where any given input is sensed by multiple parts of the retina and an accurate percept needs to be constructed. The project will consider different models for distributed processing, motivated by learning and perception in both lower-level organisms (visual processing in fruit flies) and higher-level cognition (visual cognition in humans).

Objective 2: Data representation. The brain stores the same information in several different ways, each emphasizing different dimensions of the input. Inspired by current understanding of representation of information in different regions of the cortex, the project will investigate how parallel lossy representations of the same inputs along different dimensions can be used for statistically and computationally efficient learning algorithms. In the other direction, we will investigate how embedding algorithms from machine learning might be used as mechanisms for processing massive cellular and brain imaging data.

Objective 3: Attentional filtering. The project will develop attention-based models in statistical learning, based on the use of lower-dimensional traces or curves through a high-dimensional input space. Attention curves have analogues in human cognition, where input dimensions are processed based on their inherent salience and relevance to a person's goals. The project will explore mathematical, computational and empirical models of attention. Experiments will focus on a public dataset of subjects watching episodes of Sherlock while being scanned with fMRI, using a frame-by-frame annotation of several dimensions of the movie as the basis for attention-based models.

Objective 4: Memory capacity. Evidence from studies of human behavior suggests that people store information about objects and events in long-term memory with incredible detail. How is this possible? We will consider cognitive studies and a current understanding of possible memory architectures in natural systems in order to inform approaches for reducing and sharing memory in artificial learning algorithms. A framework will be developed for establishing lower bounds on the risk of machine learning algorithms under memory constraints. Insights from this mathematical theory will be considered in the context of memory of complex organisms.

The ***intellectual merit*** of the proposed research includes the transfer of ideas between data science and neuroscience, with the goal of advancing knowledge in both domains. The ***broader impact*** of the research includes development of software that implements the advanced data science and machine learning algorithms, the development of labs for an undergraduate course in data science with examples drawn from neuroscience, and a series of workshops hosted at Yale and Brown Universities that expand the scope of the original TRIPODS effort at Brown.

TRIPODS+X:RES: Investigations at the Interface of Data Science and Neuroscience

Project Description

1. Introduction

2. Distributed Processing (Objective 1)

Recent work in machine learning has studied the effect of communication constraints and parallelization in distributed estimation. There is a close parallel in vision, where any given input is sensed by multiple parts of the retina and an accurate percept needs to be constructed. We will consider different models for distributed processing, motivated by learning and perception in both lower-level organisms (visual processing in fruit flies) and higher-level cognition (visual cognition in humans).

Neuroscience Background. Animals use visual cues to guide many behaviors, from navigation to foraging and courtship. The perception of these visual cues is an inference problem (Knill and Richards, 1996). In this problem, the animal obtains light intensity information from an array of photoreceptors focused on different points in space. The animal must combine these light intensity signals in a way that allows it to infer and respond the true state of the world, across many different parallel dimensions of inference. For instance, one dimension of inference might be the global motion of the visual scene, while another might be the existence or non-existence of a predator in the scene. The neuronal circuits in the visual system perform this inference task, at both low levels (is there an edge at this location and angle?) and at high levels (is that object a predator?) (Simoncelli and Olshausen, 2001). The operational processing of many visual neurons and circuits have been studied in depth, but it is frequently unknown how these operational descriptions relate to the inferences that guide behavior. In particular, these inferences require integrating distributed retinal information over space and over time, but we do not know how this integration relates to the statistics of the natural world, to channels of information flow within the circuits, or to noise or incomplete information about the world.

The fruit fly *Drosophila* has several advantages for studying the distributed visual processing that guides perception and behavior. First, there is a powerful genetic toolbox in fruit flies that allows researchers to genetically define, manipulate, and monitor specific classes of neurons (Luo et al., 2008). Those manipulations also allow specific neurons to be causally connected to behaviors. Second, the fly has a wealth of robust visual behaviors, including regulation of turning and speed, escape responses, and courtship behaviors (Card and Dickinson, 2008; Silies et al., 2014; Spieth, 1974). Third, the field has identified neuron types that appear to be making exactly the inferences described above: neurons sensitive to local motion direction and speed (Maisak et al., 2013); neurons sensitive to wide-field motion that corresponds to rotational self-motion of the fly about various axes (Joesch et al., 2008); neurons sensitive to looming (approaching) dark dots (de Vries and Clandinin, 2012; Klapoetke et al., 2017); and neurons sensitive to moving small dots (Keleş and Frye, 2017). In each of these cases, we can silence the neurons and observe behavioral deficits. We can also record neural activity in these individual neurons and measure their response properties with well-controlled visual stimuli (Salazar-Gatzimas et al., 2016). Thus, these neuron classes act as handles for understanding how visual inferences are made, and how neurons extract specific visual features from a spatiotemporally distributed set of inputs.

Computation and Inference Background. Classic statistical theory studies the difficulty of estimation under various models, and attempts to find the optimal estimation procedures. Such studies usually assume that all of the collected data are available to construct the estimators. Recent research has begun to study the problem of statistical estimation with data residing at multiple machines. Estimation in distributed settings is becoming common in modern data analysis tasks, as the data can be collected or stored at different locations. In order to obtain an estimate of some statistical functional, information needs to be gathered and aggregated from the multiple locations to form the final estimate. However, the communication between machines may be limited. In such a setting, it is important to understand how the statistical risk of estimation degrades as the communication budget becomes more limited.

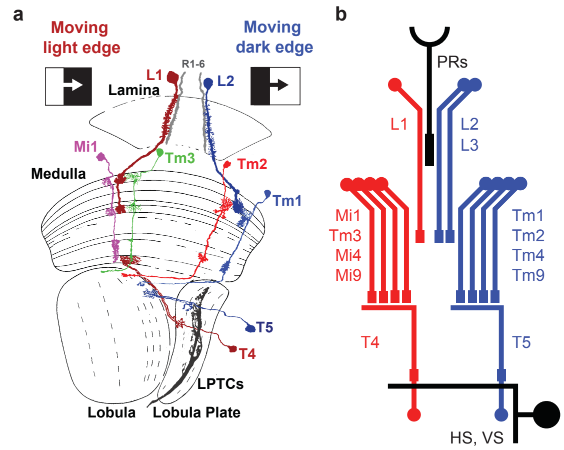

Figure 1: Motion circuits in the fly. (a) Light is detected at the retina (at the top of this diagram), and information is processed moving down through different neuropils. Each of the highlighted neurons is required for motion detection. Motion detection is split into one circuit that detects light edges moving over dark backgrounds and another that detects dark edges over light backgrounds. (b) Cartoon of neurons known to be involved in motion detection in the light edge (red) and dark edge (blue) pathways. There are four T4s and four T5s, selective for each of the four cardinal directions across the retina. This circuit repeats at all points in space.

The so-called CEO problem, first studied in the electrical engineering community in the context of rate-distortion-theory theory, treats a similar distributed estimation problem (Berger et al., 1996; Viswanathan and Berger, 1997). More recently, several studies have focused on more specific statistical tasks and models, including mean estimation, regression, principal eigenspace estimation, and discrete density estimation (Zhang et al., 2013; Shamir, 2014; Battey et al., 2015; Braverman et al., 2016; Diakonikolas et al., 2017; Fan et al., 2017; Lee et al., 2017; Shang and Cheng, 2017). Most of this existing research focuses on parametric and discrete models, where the parameter of interest has finite dimension. In a nonparametric setting, the effective dimension of the problem typically grows with the sample size, and these results no longer apply. Other results have been obtained on these problems in the normal means model of nonparametric estimation. The normal means model arises naturally when representing an estimator in terms of an orthogonal basis (Johnstone, 2002; Tsybakov, 2008). One result gives a sharp constrained minimax analysis of nonparametric regression under quantization constraints (Zhu and Lafferty, 2018b); another characterizes lower bounds and achievability for distributed nonparametric regression (Zhu and Lafferty, 2018a).

Proposed Research A: Parallel channels for local motion detection. The fly's eye is arranged in a hexagonal lattice of repeated circuit motifs, with each column of circuitry representing one retinotopic point in visual space. Each eye consists of an array of roughly 800 of these pixels, which together cover approximately one half of visual space. Two classes of local motion detection cells exist in every column: there are T4 cells, which detect light edges moving across dark backgrounds, T5 cells, which detect dark edges moving across light backgrounds. There are 4 of each class, one for each cardinal direction. Thus, there are 8 parallel channels at each point in space representing motion in two dimensions. Why is the system organized in this way? How are naturalistic motion signals distributed across the 8 channels, and what encoding or decoding advantages does this serve? Under what conditions are the channels redundant? How would an optimal observer partition signals among these parallel channels? Could a data driven approach predict or give insight into this encoding scheme? One approach to begin studying these questions will be to adapt the classical framework of sparse coding (Olshausen and Field, 1996) in a way that represents the neurobiology of the fly's visual system.

Proposed Research B: Detecting motion flow fields. When an animal moves or rotates in the world, its self-motion generates flow fields across its retina. These flow fields are often used as feedback to control orientation or speed. In *Drosophila*, some neurons downstream of the local motion detectors have large receptive fields that integrate motion signals across the retina. They appear to be selective for specific flow fields, which correspond to the flow fields created by the rotation of the fly about different axes. These neural signals have been proposed to be linear filters, matching their weighting for local motion to specific optical flow fields ("matched filters"). However, it is not clear whether a linear weighting of local motion estimates represents a best estimate of each flow field, or whether more complex dendritic computations could improve encoding. In particular, it is not clear how these neurons might optimally integrate motion signals in the presence of occlusions or differential velocity fields that would be caused by fly translation through space. We will investigate these issues using methods based on hierarchical sparse coding (Yu et al., 2011) and related computational methods for low rank decomposition.

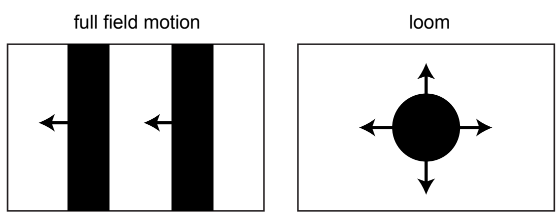

Figure 2: Global motion and loom may generate similar local motion signals, but local motion signals must be integrated over space to distinguish the two stimulus types.

Proposed Research C: Detecting looming stimuli. Looming stimuli are created by objects as they approach the observer: the object becomes larger, and if it is on a collision course, opposite edges of the object will move in opposite directions on the retina. Thus, detecting loom requires integrating information over space, as any local motion detector cannot know if local motion is due to a looming object or is just wide-field motion. Two neuron types in *Drosophila* have recently been described as loom detectors, responding selectively to objects that grow larger in their receptive fields. The receptive fields of these cells to local motion have been characterized, but the relationship between these receptive fields, the nonlinear computations of these neurons, and the statistics of natural loom stimuli remain unclear. This is in large part due to the absence of good statistics on natural loom stimuli. It is also possible that these loom detectors use features beyond motion to detect approaching objects; for instance, a detector might integrate motion signals with light intensity information, since light intensity is correlated with distance. Here, one might also ask whether the motion detecting neurons upstream of the loom-sensitive neurons could convey information about stimulus features beyond just motion direction and speed. We will abstract loom detection as a statistical testing problem, building work on such as (Arias-Castro et al., 2005; Arias-Castro et al., 2006; Donoho and Jin, 2004). Specifically, for a given object geometry such as a disk, we will study minimax rates for loom detection, in terms of the noise level and sparity of the number of boundary neuron measurements required. Fast hierarchical algorithms to achieve the minimax rates will be studied.

3. Data Representation (Objective 2)

The brain stores the same information in several different ways, each emphasizing different dimensions of the input. In visual cortex, two inputs with similar visual features (e.g., views of the same face, prairie landscapes, baseball mitts, etc.) will be stored together. In temporal and frontal cortex, two inputs with the same conceptual or semantic meaning/function will be stored together (e.g., sports balls, cooking utensils, animals, etc.). In the hippocampus, these kinds of sensory or semantic overlap are discounted by orthogonalizing similar inputs encountered at different times (pattern separation); instead, co-occurring inputs that are part of the same event get stored together. There are many other dominant organizing dimensions in different brain regions (e.g., reward, emotion, modality, task, action, etc.). What can we learn about how to represent data for efficient computation from this idea of parallel lossy representations along different dimensions?

Our broad goal in this component of the project will be to develop algorithms for machine representation of data that are informed by such understanding of higher-level cognition. Conversely, we will leverage developments in embedding algorithms in machine learning for advanced processing of fMRI and cellular recordings.

Neuroscience Background.

Computation and Inference Background. Exponential family embeddings are a new way to generalize classical methods of finding distributed representations in language (Rudolph et al., 2016). Consider a corpus of language $\boldsymbol{x} = \{x_1, \ldots, x_n\}$, where each x_i is a word from a vocabulary of terms. An exponential family embedding has three components: (a) a notion of *context* for each data point, e.g., a window of observed words around each word (b) a form of the *conditional distribution*, e.g., for text a categorical distribution over V items is appropriate and (c) an *embedding structure* that describes how parameters are shared across data, e.g., for text we typically posit that each term (such as "walnut" or "bicycle") shares the same representations wherever it appears in the collection. An exponential family embedding posits two d-dimensional latent representations for each term v, one is the *embedding vector* ρ_v and the other is the *context vector* α_v, where d is a hyperparameter. The model asserts that each observation is drawn from a conditional distribution given its context. Exponential family embeddings generalize many existing methods for learning distributed representations, including continuous bag of words, negative sampling, and the many other variants of word2vec (Bengio et al., 2003; Mikolov et al., 2013).

Fitting such models is difficult, and requires robust methods for computation and evaluation. We will lean on and adapt a line of research on black box variational inference methods (Ranganath et al., 2014), particularly for probabilistic programming (Kucukelbir et al., 2017; Tran et al., 2017), to be able to quickly develop and test our models. Variational inference approximates the posterior by fitting a family of distributions over the latent variables to be close in KL divergence (Jordan et al., 1999; Blei et al., 2017). For simplicity, consider the problem of fitting embeddings. The variational distribution is $q(\alpha_{1:V}, \rho_{1:V}; \nu)$ and we fit the variational parameters ν to be close in KL divergence to $p(\alpha_{1:V}, \rho_{1:V} \mid \boldsymbol{x})$. Recent innovations in probabilistic programming also us to do this *generically* and *scalably*, easily fitting many types of models to large data sets. This enables the exploration of many variants of the models, e.g., different types of contexts, different values of d, different underlying conditional distributions. We support an empirical approach to making these choices, using cross-validation with the held-out predictive log likelihood (Wallach et al., 2009). The intuition is that a model that provides a good density of the data is more likely to be one that is useful and interpretable.

Proposed Research A: Representations beyond co-occurrence statistics. Distributed embedding representations in machine learning are almost exclusively based on co-occurrence statistics. For instance, when constructing embeddings for words in text, names of colors ("red," "blue"...) will be embedded in nearby locations simply because they tend to be used together. How can a richer knowledge of representation in the brain be used to inspire algorithms for processing text, images, and audio? What are other brain-inspired features over which co-occurrence can be evaluated?

Proposed Research B: Embeddings for fMRI data. Exponential family embeddings open the door to developing more complex embeddings for problems in cognitive neuroscience: embeddings for fMRI and other types of data, embeddings with a complex notion of context, and embeddings that are themselves a part of a larger probabilistic model, such as where representations are shared or tied across tasks. For examples of some of these innovations see Rudolph et al. (2016). The approach will be based on the incorporation of a latent variable model, for which variational methods can be used for scalable inference, as described above.

Proposed Research C: Time-dependent representations. To capture dependence on time, we will build on one of the PIs earlier work on dynamic topic models (Blei and Lafferty, 2006). Dynamic topic models captured how the latent themes in a collection can grow and shrink and change over time. For example, the theme of "technology" in a scientific corpus might start with words about electricity and wires and end with words about computers and semiconductors. Dynamic topic models were developed specifically for language. We will generalize this idea to capture the evolution over time of distributed representations. Unlike dynamic topic models, the fitted model will capture multimodal data and the evolution of its latent characteristics. In the exponential family embedding framework, this amounts to placing a linear dynamic prior on the embedding vectors or the context vectors (or both).

4. Results from Prior NSF Support

John Lafferty was previously supported as co-PI under NSF grant IIS-1116730, "III: Small: Nonparametric Structure Learning for Complex Scientific Datasets," from August 1, 2011 to July 31, 2014. The PI of the grant was Han Liu (Princeton University) and the co-PIs were Lafferty and Larry Wasserman (Carnegie Mellon University). The total award amount was $499,344; the amount awarded to the University of Chicago was $118,750.

Intellectual Merit. The project focused on developing scalable methods for finding structure in complex scientific datasets, without making strong distributional assumptions. The project explored several aspects of nonparametric structure learning, including methods, theory, large-scale computing, and applications, with five concrete aims: (1) nonparametric structure learning in high dimensions, (2) nonparametric conditional structure learning, (3) regularization parameter selection, (4) parallel and online nonparametric learning, and (5) minimax theory for nonparametric structure learning problems. The outcomes included practical models and algorithms; application areas included genomics, cognitive neuroscience, climate science, astrophysics, and language processing. Publications resulting from this grant include (Balakrishnan et al., 2012; Chen et al., 2014; Chen and Lafferty, 2013; Fan et al., 2014, 2015; Fan and Liu, 2013; Fang et al., 2015; Foygel et al., 2012; Gu and Lafferty, 2012; Gu et al., 2014; Han and Liu, 2012a,b, 2013a,b,c, 2014, 2015a,b; Han et al., 2015, 2013; Han and Liu, 2013d; He et al., 2014; Jiang et al., 2014; Kalaitzis et al., 2013; Kolar and Liu, 2013, 2015; Kolar et al., 2013, 2014; Lafferty et al., 2012; Li et al., 2014; Liu and Han, 2012; Liu et al., 2012a,b,c; Liu and Wang, 2015; Liu et al., 2014a,b, 2015; Liu and Lafferty, 2014; Mishra et al., 2015a; Neale et al., 2012; Ning and Liu, 2013; Pang et al., 2014; Qiu et al., 2015a,b; Rosenblum et al., 2014; Shender and Lafferty, 2013; Shou et al., 2015; Song et al., 2015; Wang and Liu, 2014; Wang et al., 2014; Xu et al., 2014; Xu and Lafferty, 2012; Zhao and Liu, 2013, 2014; Zhao et al., 2012a,b, 2014a,b; Zhu and Lafferty, 2014a, 2018b).

Broader Impact. The broader impact of the project included interdisciplinary training for graduate students from biostatistics, computer science, statistics, and medical schools, strengthening the collaboration and interdisciplinary infrastructure between Carnegie Mellon, Johns Hopkins, the University of Chicago and

Princeton, and broadly disseminating the results from this research in journals from all relevant fields. The research had impact outside of machine learning and statistics. In a genomic study, PI Liu applied structured nonparametric methods to analyze high dimensional genomic data, identifying several gene mutations associated with autism. These results were published in Nature (Neale et al., 2012), and reported in the New York Times. In another neuroscience study, the PI developed an effective algorithm for predicting Attention Deficit Hyperactive Disorder (ADHD) disease. The research led to several statistical software packages in R, including (Zhao et al., 2012a; Pang et al., 2014; Li et al., 2014), all of which are freely available on CRAN.

Lafferty is currently supported as PI under NSF grant DMS-1513594, "Constrained Statistical Estimation and Inference: Theory, Algorithms and Applications," from June 29, 2015 to July 1, 2018. The total award amount was $320,000. After two years of the project, the remainder of the funds were transferred from the University of Chicago to Yale University, where the PI moved in July 2017.

Intellectual Merit. The project is studying constraints that are present in complex scientific data analysis problems, but that have not been thoroughly studied in traditional approaches. Different aspects of theory, algorithms, and applications of statistical procedures, with constraints imposed on the storage, runtime, shape, energy or physics of the estimators and applications. The overall goal of the research is to develop theory and tools that can help scientists to conduct more effective data analysis. Publications under this grant have included (Chatterjee and Lafferty, 2018; Mishra et al., 2018; Zhu and Lafferty, 2018c; Mishra et al., 2017; Yang et al., 2016; Chatterjee et al., 2016; Zheng and Lafferty, 2016; Mishra et al., 2015b; Zheng and Lafferty, 2015; Zhu and Lafferty, 2014b)

Broader Impact. The broader impact of the project is aimed in three directions. First is the development of flexible and principled large scale data analysis tools that can benefit many scientific domains. Second, is the development of software that is widely distributed, allowing others to build on the work. The third is to education, to allow the research to impact the training of students at both the graduate and undergraduate levels.

Lafferty was previously supported as PI under NSF grant DMS-1547396, "RTG: Computational and Applied Mathematics in Statistical Science" from July 1, 2016 to July 1, 2017. This grant did not transfer to Yale University; the current PI is Jonathan Weare at the University of Chicago. The total award amount is $1,697,320.

Intellectual Merit. This Research Training Group (RTG) project supports creation of a dynamic, interactive, and vertically integrated community of students and researchers working together in computational and applied mathematics and statistics. The work is motivated by the growing need to train the next generation of statisticians and computational and applied mathematicians in new ways, to confront data-centric problems in the natural and social sciences.

Broader Impact. The broader impact includes vertical integration of education and training from undergraduate to postdoctoral researchers, including activities at Toyota Technological Institute at Chicago and Argonne National Laboratory. Participants in the RTG will receive an educational experience that provides them with strong preparation for positions in industry, government, and academics, with an ability to adopt approaches to problem solving that are drawn from across the computational, mathematical, and statistical sciences.

References Cited

ARIAS-CASTRO, E., DONOHO, D. L. and HUO, X. (2005). Near-optimal detection of geometric objects by fast multiscale methods. *IEEE Trans. Information Theory* **51** 2402–2425.

ARIAS-CASTRO, E., DONOHO, D. L. and HUO, X. (2006). Adaptive multiscale detection of filamentary structures in a background of uniform random points. *Ann. Statist.* **34** 326–349.

BALAKRISHNAN, S., PUNIYANI, K. and LAFFERTY, J. D. (2012). Sparse additive functional and kernel CCA. In *Proceedings of the 29th International Conference on Machine Learning, ICML 2012*.

BATTEY, H., FAN, J., LIU, H., LU, J. and ZHU, Z. (2015). Distributed estimation and inference with statistical guarantees. *arXiv preprint arXiv:1509.05457* .

BENGIO, Y., DUCHARME, R., VINCENT, P. and JANVIN, C. (2003). A neural probabilistic language model. *Journal of Machine Learning Research* **3** 1137–1155.

BERGER, T., ZHANG, Z. and VISWANATHAN, H. (1996). The CEO problem. *IEEE Trans. Inform. Theory* **42** 887–902.

BLEI, D., KUCUKELBIR, A. and MCAULIFFE, J. (2017). Variational inference: A review for statisticians. *Journal of American Statistical Association* **112** 859–877.

BLEI, D. and LAFFERTY, J. (2006). Dynamic topic models. In *International Conference on Machine Learning*. ACM, New York, NY, USA.

BRAVERMAN, M., GARG, A., MA, T., NGUYEN, H. L. and WOODRUFF, D. P. (2016). Communication lower bounds for statistical estimation problems via a distributed data processing inequality. In *Proceedings of the Forty-eighth Annual ACM Symposium on Theory of Computing*. STOC '16.

CARD, G. and DICKINSON, M. H. (2008). Visually mediated motor planning in the escape response of *Drosophila*. *Current Biology* **18** 1300–1307.

CHATTERJEE, S., DUCHI, J. C., LAFFERTY, J. D. and ZHU, Y. (2016). Local minimax complexity of stochastic convex optimization. In Lee et al. (2016), 3423–3431.

CHATTERJEE, S. and LAFFERTY, J. D. (2018). Denoising flows on trees. *IEEE Trans. Information Theory* **64** 1767–1783.

CHEN, C., ZHAO, T., METAXAS, D. and LIU, H. (2014). Mode estimation for high dimensional discrete tree graphical models. In *Neural Information Processing Systems (NIPS)*.

CHEN, M. and LAFFERTY, J. D. (2013). Mismatched estimation and relative entropy in vector Gaussian channels. In *Proceedings of the 2013 IEEE International Symposium on Information Theory, Istanbul, Turkey, July 7-12, 2013*.

DE VRIES, S. E. and CLANDININ, T. R. (2012). Loom-sensitive neurons link computation to action in the *Drosophila* visual system. *Current Biology* **22** 353–362.

DIAKONIKOLAS, I., GRIGORESCU, E., LI, J., NATARAJAN, A., ONAK, K. and SCHMIDT, L. (2017). Communication-efficient distributed learning of discrete distributions. In *Advances in Neural Information Processing Systems*.

Donoho, D. and Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *The Annals of Statistics* **32** 962–994.

Fan, J., Han, F. and Liu, H. (2014). Challenges of big data analysis. *National Science Review* **2** 1–24.

Fan, J., Ke, T., Liu, H. and Xia, L. (2015). Quadro: A supervised dimension reduction method via rayleigh quotient. *The Annals of Statistics* Accepted for publication.

Fan, J. and Liu, H. (2013). Statistical analysis of big data on pharmacogenomics. *Advanced Drug Delivery Reviews* **65** 987–1000.

Fan, J., Wang, D., Wang, K. and Zhu, Z. (2017). Distributed estimation of principal eigenspaces. *arXiv preprint arXiv:1702.06488* .

Fang, E. X., He, B., Liu, H. and Yuan, X. (2015). Generalized alternating direction method of multipliers: New theoretical insight and application. *Mathematical Programming Computation* Accepted for publication.

Foygel, R., Horrell, M., Drton, M. and Lafferty, J. (2012). Nonparametric reduced rank regression. In *Advances in Neural Information Processing Systems 25* (P. Bartlett, F. Pereira, C. Burges, L. Bottou and K. Weinberger, eds.). NIPS Foundation, 1637–1645.

Gu, H. and Lafferty, J. D. (2012). Sequential nonparametric regression. In *Proceedings of the 29th International Conference on Machine Learning, ICML 2012*.

Gu, Q., Wang, Z. and Liu, H. (2014). Oracle sparse PCA and its inference. In *Neural Information Processing Systems (NIPS)*.

Han, F. and Liu, H. (2012a). High dimensional semiparametric scale-invariant principal component analysis. In *Neural Information Processing Systems (NIPS)*.

Han, F. and Liu, H. (2012b). Transelliptical principal component analysis for non-Gaussian data. In *Neural Information Processing Systems (NIPS)*.

Han, F. and Liu, H. (2013a). PCA on non-Gaussian dependent data. In *International Conference on Machine Learning (ICML)*.

Han, F. and Liu, H. (2013b). Robust sparse principal component regression. In *Neural Information Processing Systems (NIPS)*.

Han, F. and Liu, H. (2013c). Transition matrix estimation in high dimensional vector autoregressive models. In *International Conference on Machine Learning (ICML)*.

Han, F. and Liu, H. (2014). Scale-invariant sparse PCA on high dimensional meta-elliptical data. *Journal of American Statistical Association* **109** 275–287.

Han, F. and Liu, H. (2015a). High dimensional semiparametric scale-invariant principal component analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Accepted for publication.

Han, F. and Liu, H. (2015b). Optimal rates of convergence for latent generalized correlation matrix estimation in transelliptical distributions. *Bernoulli* Accepted for publication.

HAN, F., LU, H. and LIU, H. (2015). A direct estimation of high dimensional stationary vector autogressions. *Journal of Machine Learning Research* Accepted for publication.

HAN, F., ZHAO, T. and LIU, H. (2013). CODA: High dimensional copula discriminant analysis. *Journal of Machine Learning Research* **14** 629–671.

HAN, Z. W. F. and LIU, H. (2013d). Sparse principal component analysis for high dimensional multivariate time series. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*.

HE, B., LIU, H., WANG, Z. and YUAN, X. (2014). A strictly contractive Peaceman-Rachford splitting method for convex program. *SIAM Journal on Optimization* **24** 1011–1040.

JIANG, X., YAO, Y., LIU, H. and GUIBAS, L. (2014). Compressive network analysis. *IEEE Transactions on Automatic Control* **59** 2946–2961.

JOESCH, M., PLETT, J., BORST, A. and REIFF, D. (2008). Response properties of motion-sensitive visual interneurons in the lobula plate of *Drosophila* melanogaster. *Current Biology* **18** 368–374.

JOHNSTONE, I. M. (2002). Function estimation and Gaussian sequence models. Unpublished manuscript.

JORDAN, M., GHAHRAMANI, Z., JAAKKOLA, T. and SAUL, L. (1999). Introduction to variational methods for graphical models. *Machine Learning* **37** 183–233.

KALAITZIS, A., LAFFERTY, J., LAWRENCE, N. and ZHOU, S. (2013). The bigraphical lasso. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)* (S. Dasgupta and D. Mcallester, eds.), vol. 28. JMLR Workshop and Conference Proceedings.

KELEŞ, M. F. and FRYE, M. A. (2017). Object-detecting neurons in *Drosophila*. *Current Biology* **27** 680–687.

KLAPOETKE, N. C., NERN, A., PEEK, M. Y., ROGERS, E. M., BREADS, P., RUBIN, G. M., REISER, M. B. and CARD, G. M. (2017). Ultra-selective looming detection from radial motion opponency. *Nature* **551** 237.

KNILL, D. C. and RICHARDS, W. (1996). *Perception as Bayesian inference*. Cambridge University Press.

KOLAR, M. and LIU, H. (2013). Feature selection in high-dimensional classification. In *International Conference on Machine Learning (ICML)*.

KOLAR, M. and LIU, H. (2015). Optimal feature selection in high-dimensional discriminant analysis. *IEEE Transactions on Information Theory* **61** 1063–1083.

KOLAR, M., LIU, H. and XING, E. (2013). Graph estimation from multi-attribute data. In *International Conference on Machine Learning (ICML)*.

KOLAR, M., LIU, H. and XING, E. (2014). Graph estimation from multi-attribute data. *Journal of Machine Learning Research* **15** 1713–1750.

KUCUKELBIR, A., TRAN, D., RANGANATH, R., GELMAN, A. and BLEI, D. M. (2017). Automatic differentiation variational inference. *Journal of Machine Learning Research* **18** 1–45.

LAFFERTY, J., LIU, H. and WASSERMAN, L. (2012). Sparse nonparametric graphical models. *Statist. Sci.* **27** 519–537.

LEE, D. D., SUGIYAMA, M., VON LUXBURG, U., GUYON, I. and GARNETT, R. (eds.) (2016). *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*.

LEE, J. D., LIU, Q., SUN, Y. and TAYLOR, J. E. (2017). Communication-efficient sparse regression. *Journal of Machine Learning Research* **18** 1–30.

LI, X., ZHAO, T., YUAN, X. and LIU, H. (2014). An R package flare for high dimensional linear regression and precision matrix estimation. *Journal of Machine Learning Research* Accepted for publication.

LIU, H. and HAN, F. (2012). High dimensional transelliptical graphical models. In *Neural Information Processing Systems (NIPS)*.

LIU, H., HAN, F., YUAN, M., LAFFERTY, J. and WASSERMAN, L. (2012a). High-dimensional semiparametric Gaussian copula graphical models. *Ann. Statist.* **40** 2293–2326.

LIU, H., HAN, F., YUAN, M., LAFFERTY, J. and WASSERMAN, L. (2012b). The nonparanormal SKEPTIC. In *Proceedings of the 29th International Conference on Machine Learning (ICML-12)* (J. Langford and J. Pineau, eds.). ICML '12, New York, NY, USA.

LIU, H., LAFFERTY, J. and WASSERMAN, L. (2012c). Exponential concentration inequality for mutual information estimation, with application to forest graphical models. In *Advances in Neural Information Processing Systems (NIPS)*, vol. Volume 25.

LIU, H. and WANG, L. (2015). A tuning-insensitive approach for optimally estimating Gaussian graphical models. *Electronic Journal of Statistics* Accepted for publication.

LIU, H., WANG, L. and ZHAO, T. (2014a). Multivariate regression with calibration. In *Neural Information Processing Systems (NIPS)*.

LIU, H., WANG, L. and ZHAO, T. (2014b). Sparse covariance matrix estimation with eigenvalue constraints. *Journal of Computational and Graphical Statistics* **23** 439–459.

LIU, H., WANG, L. and ZHAO, T. (2015). Multivariate regression with calibration. *Journal of Machine Learning Research* Accepted for publication.

LIU, Z. and LAFFERTY, J. (2014). Blossom tree graphical models. In *Advances in Neural Information Processing Systems 27* (Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence and K. Weinberger, eds.). Curran Associates, Inc., 1458–1465.

LUO, L., CALLAWAY, E. and SVOBODA, K. (2008). Genetic dissection of neural circuits. *Neuron* **57** 634–660.

MAISAK, M. S., HAAG, J., AMMER, G., SERBE, E., MEIER, M., LEONHARDT, A., SCHILLING, T., BAHL, A., RUBIN, G. M., NERN, A., DICKSON, B. J., RIEFF, D. F., HOPP, E. and BORST, A. (2013). A directional tuning map of *Drosophila* elementary motion detectors. *Nature* **500** 212–216.

MIKOLOV, T., CHEN, K., CORRADO, G. and DEAN, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* .

MISHRA, N., IMES, C., LAFFERTY, J. D. and HOFFMANN, H. (2018). CALOREE: learning control for predictable latency and low energy. In *Proceedings of the Twenty-Third International Conference on Architectural Support for Programming Languages and Operating Systems, ASPLOS 2018, Williamsburg, VA, USA, March 24-28, 2018* (X. Shen, J. Tuck, R. Bianchini and V. Sarkar, eds.). ACM.

MISHRA, N., LAFFERTY, J. D. and HOFFMANN, H. (2017). ESP: A machine learning approach to predicting application interference. In *2017 IEEE International Conference on Autonomic Computing, ICAC 2017, Columbus, OH, USA, July 17-21, 2017* (X. Wang, C. Stewart and H. Lei, eds.). IEEE Computer Society.

MISHRA, N., ZHANG, H., LAFFERTY, J. D. and HOFFMANN, H. (2015a). A probabilistic graphical model-based approach for minimizing energy under performance constraints. In *Proceedings of the Twentieth International Conference on Architectural Support for Programming Languages and Operating Systems*. ASPLOS '15.

MISHRA, N., ZHANG, H., LAFFERTY, J. D. and HOFFMANN, H. (2015b). A probabilistic graphical model-based approach for minimizing energy under performance constraints. In *Proceedings of the Twentieth International Conference on Architectural Support for Programming Languages and Operating Systems, ASPLOS '15, Istanbul, Turkey, March 14-18, 2015* (Ö. Özturk, K. Ebcioglu and S. Dwarkadas, eds.). ACM.

NEALE, B. M., KOU, Y., LIU, L., MA'AYAN, A., SAMOCHA, K. E., SABO, A., LIN, C.-F., STEVENS, C., WANG, L.-S., MAKAROV, V., POLAK, P. P., YOON, S., MAGUIRE, J., CRAWFORD, E. L., CAMPBELL, N. G., GELLER, E. T., VALLADARES, O., SHAFER, C., LIU, H., ZHAO, T., CAI, G., LIHM, J., DANNENFELSER, R., JABADO, O., PERALTA, Z., NAGASWAMY, U., MUZNY, D., REID, J. G., NEWSHAM, I., WU, Y., LEWIS, L., HAN, Y., VOIGHT, B. F., LIM, E., ROSSIN, E. J., KIRBY, A., FLANNICK, J. A., FROMER, M., SHAKIR, K., FENNELL, T., GARIMELLA, K., BANKS, E., POPLIN, R., GABRIEL, S., DEPRISTO, M., WIMBISH, J. R., BOONE, B. E., LEVY, S. E., BETANCUR, C., SUNYAEV, S. R., BOERWINKLE, E., BUXBAUM, J. D., COOK, E. H., DEVLIN, B., GIBBS, R. A., ROEDER, K., SCHELLENBERG, G. D., SUTCLIFFE, J. S. and JOSEPH, M. (2012). Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* **485** 242–245.

NING, Y. and LIU, H. (2013). Semiparametric matrix-normal graphical models. *Biometrika* **100** 655–670.

OLSHAUSEN, B. A. and FIELD, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381** 607–609.

PANG, H., LIU, H. and VANDERBEI, R. (2014). The fastclime package for linear programming and large-scale precision matrix estimation. *Journal of Machine Learning Research* **15** 489–493.

QIU, H., HAN, F., LIU, H. and CAFFO, B. (2015a). Joint estimation of multiple graphical models from high dimensional time series. *Journal of the Royal Statistical Society, Series B* Accepted for publication.

QIU, H., HAN, F., LIU, H. and CAFFO, B. (2015b). Robust estimation of transition matrices in high dimensional heavy-tailed vector autoregressive processes. In *International Conference on Machine Learning (ICML)*.

RANGANATH, R., GERRISH, S. and BLEI, D. (2014). Black box variational inference. In *International Conference on Artifical Intelligence and Statistics*.

ROSENBLUM, M., LIU, H. and YEN, E.-H. (2014). Optimal tests of treatment effects for the overall population and two subpopulations in randomized trials, using sparse linear programming. *Journal of the American Statistical Association* **109** 1216–1228.

RUDOLPH, M., RUIZ, F., MANDT, S. and BLEI, D. (2016). Exponential family embeddings. In *Neural Information Processing Systems*.

SALAZAR-GATZIMAS, E., CHEN, J., CREAMER, M. S., MANO, O., MANDEL, H. B., MATULIS, C. A., POTTACKAL, J. and CLARK, D. A. (2016). Direct measurement of correlation responses in *Drosophila* elementary motion detectors reveals fast timescale tuning. *Neuron* **92** 227–239.

SHAMIR, O. (2014). Fundamental limits of online and distributed algorithms for statistical learning and estimation. In *Proceedings of the 27th International Conference on Neural Information Processing Systems*. NIPS'14.

SHANG, Z. and CHENG, G. (2017). Computational limits of a distributed algorithm for smoothing spline. *The Journal of Machine Learning Research* **18** 3809–3845.

SHENDER, D. and LAFFERTY, J. (2013). Computation-risk tradeoffs for covariance-thresholded regression. In *Journal of Machine Learning Research*, vol. 28. Proceedings of International Conference on Machine Learning.

SHOU, H., SHINOHARA, R. T., LIU, H., REICH, D. S. and CRAINICEANU, C. M. (2015). Soft null hypotheses: A case study of image enhancement detection in brain lesions. *Journal of Computational and Graphical Statistics* Accepted for publication.

SILIES, M., GOHL, D. M. and CLANDININ, T. R. (2014). Motion-detecting circuits in flies: Coming into view. *Annual review of neuroscience* **37** 307–327.

SIMONCELLI, E. and OLSHAUSEN, B. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience* **24** 1193–1216.

SONG, L., LIU, H., PARIKH, A. and XING, E. (2015). Nonparametric latent tree graphical models: Inference, estimation, and learning. *Journal of Machine Learning Research* Accepted for publication.

SPIETH, H. T. (1974). Courtship behavior in *Drosophila*. *Annual review of entomology* **19** 385–405.

TRAN, D., HOFFMAN, M., SAURAUS, R., BREVDO, E., MURPHY, K. and BLEI, D. (2017). Deep probabilistic programming. In *International Conference on Learning Representations*.

TSYBAKOV, A. B. (2008). *Introduction to Nonparametric Estimation*. 1st ed. Springer Series in Statistics.

VISWANATHAN, H. and BERGER, T. (1997). The quadratic Gaussian ceo problem. *IEEE Transactions on Information Theory* **43** 1549–1559.

WALLACH, H., MURRAY, I., SALAKHUTDINOV, R. and MIMNO, D. (2009). Evaluation methods for topic models. In *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM New York, NY, USA.

WANG, Z. and LIU, H. (2014). Tightening after relax: Minimax-optimal sparse PCA in polynomial time. In *Neural Information Processing Systems (NIPS)*.

WANG, Z., LIU, H. and ZHANG, T. (2014). Optimal computational and statistical rates of convergence for sparse nonconvex learning problems. *The Annals of Statistics* **42** 2164–2201.

XU, M., CHEN, M. and LAFFERTY, J. (2014). Faithful variable screening for high-dimensional convex regression. arXiv:1411.1711; submitted to Annals of Statistics.

XU, M. and LAFFERTY, J. D. (2012). Conditional sparse coding and grouped multivariate regression. In *Proceedings of the 29th International Conference on Machine Learning, ICML 2012*.

YANG, F., BARBER, R. F., JAIN, P. and LAFFERTY, J. D. (2016). Selective inference for group-sparse linear models. In Lee et al. (2016), 2469–2477.

YU, K., LIN, Y. and LAFFERTY, J. D. (2011). Learning image representations from the pixel level via hierarchical sparse coding. In *CVPR 2011*.

ZHANG, Y., DUCHI, J., JORDAN, M. and WAINWRIGHT, M. J. (2013). Information-theoretic lower bounds for distributed statistical estimation with communication constraints. In *Advances in Neural Information Processing Systems*.

ZHAO, T. and LIU, H. (2013). Sparse inverse covariance estimation with calibration. In *Neural Information Processing Systems (NIPS)*.

ZHAO, T. and LIU, H. (2014). Calibrated precision matrix estimation for high dimensional elliptical distributions. *IEEE Transactions on Information Theory* **60** 7874–7887.

ZHAO, T., LIU, H., ROEDER, K., LAFFERTY, J. and WASSERMAN, L. (2012a). The `huge` package for high-dimensional undirected graph estimation in R. *The Journal of Machine Learning Research* **13** 1059–1062.

ZHAO, T., ROEDER, K. and LIU, H. (2012b). High-dimensional nonparanormal graph estimation via smooth-projected neighborhood pursuit. In *Neural Information Processing Systems (NIPS)*.

ZHAO, T., ROEDER, K. and LIU, H. (2014a). Positive semidefinite rank-based correlation matrix estimation with application to semiparametric graph estimation. *Journal of Computational and Graphical Statistics* **23** 895–922.

ZHAO, T., YU, M., WANG, Y., ARORA, R. and LIU, H. (2014b). Accelerated mini-batch randomized block coordinate descent method. In *Neural Information Processing Systems (NIPS)*.

ZHENG, Q. and LAFFERTY, J. D. (2015). A convergent gradient descent algorithm for rank minimization and semidefinite programming from random linear measurements. In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada* (C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama and R. Garnett, eds.).

ZHENG, Q. and LAFFERTY, J. D. (2016). Convergence analysis for rectangular matrix completion using burer-monteiro factorization and gradient descent. *arXiv* abs/1605.07051.

ZHU, Y. and LAFFERTY, J. (2014a). Quantized estimation of Gaussian sequence models in Euclidean balls. In *Advances in Neural Information Processing Systems 27 (NIPS)*. arXiv:1409.6833.

ZHU, Y. and LAFFERTY, J. (2018a). Distributed nonparametric regression under communication constraints. In *International Conference on Machine Learning*. arXiv:1803.01302.

ZHU, Y. and LAFFERTY, J. (2018b). Quantized nonparametric regression over Sobolev spaces. *Information and Inference* **7** 31–82.

ZHU, Y. and LAFFERTY, J. D. (2014b). Quantized estimation of gaussian sequence models in euclidean balls. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada* (Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence and K. Q. Weinberger, eds.).

ZHU, Y. and LAFFERTY, J. D. (2018c). Distributed nonparametric regression under communication constraints. *arXiv* abs/1803.01302.