
Math + X Symposium on Data Science and Inverse Problems in Geophysics

Rice University · January 24–26, 2018

funded by

SIMONS FOUNDATION

Advancing Research in Basic Science and Mathematics

Math + X Symposium on Data Science and Inverse Problems in Geophysics

Rice University · January 24–26, 2018
BioScience Research Collaborative (BRC) Building, Room 280

Wednesday, January 24th

| | | |
|---------|---------------------------|---|
| 8:15am | Maarten de Hoop | <i>Welcome</i> |
| | Reginald DesRoches | <i>Opening Remarks</i> |
| 8:30am | Nathan Kutz | <i>Data-driven discovery of governing physical laws and their parametric dependencies in engineering, physics and biology</i> |
| 9:20am | Michel Campillo | <i>From "ambient noise" to seismic imaging and monitoring</i> |
| 10:10am | | ————— BREAK ————— |
| 10:35am | Andras Vasy | <i>Global analysis via microlocal tools: Fredholm problems in non-elliptic settings</i> |
| 11:25am | Andrew Stuart | <i>Large graph limits of learning algorithms</i> |
| 12:15pm | | ————— LUNCH ————— |
| 1:15pm | Matti Lassas | <i>Manifold learning and an inverse problem for a wave equation</i> |
| 2:05pm | Frederik Simons | <i>On the inversion of noisy, incomplete, scattered, and vector-valued satellite-data for planetary magnetic-field models</i> |
| 2:55pm | | ————— BREAK ————— |
| 3:20pm | Elchanan Mossel | <i>Hierarchical generative models and deep learning</i> |
| 4:10pm | Andrea Bertozzi | <i>Geometric graph-based methods for high dimensional data</i> |
| 6:00pm | | ————— RECEPTION, RICE FACULTY CLUB ————— |
| | | <i>Cohen House, 6100 Main St.</i> |
| | | <i>Cocktails and Appetizers, Event Parking</i> |

** Registrants are provided a daily boxed lunch
** For venue locations and parking information please see:
<https://earthscience.rice.edu/mathx2018/>

Thursday, January 25th

| | | |
|---------|---------------------------------|--|
| 8:30am | Gregory Beroza | <i>FAST: a data-mining approach for earthquake detection</i> |
| 9:20am | Ankur Moitra | <i>Robustness meets algorithms</i> |
| 10:10am | | ————— BREAK ————— |
| 10:35am | Yiran Wang | <i>Inverse problems for nonlinear acoustic and elastic wave equations</i> |
| 11:25am | Carola-Bibiane Schönlieb | <i>Model-based learning in imaging</i> |
| 12:15pm | | ————— LUNCH ————— |
| 1:15pm | Joan Bruna | <i>On the optimization landscape of neural networks</i> |
| 2:05pm | Lauri Oksanen | <i>Correlation based passive imaging with a white noise source</i> |
| 2:55pm | | ————— BREAK ————— |
| 3:20pm | Victor Pankratius | <i>Computer-aided discovery: can a machine win a Nobel Prize?</i> |
| 4:10pm | Robert Nowak | <i>Outranked: exploiting nonlinear algebraic structure in matrix recovery problems</i> |

Friday, January 26th

| | | |
|---------|---------------------------|--|
| 8:30am | Stéphane Mallat | <i>Unsupervised learning of stochastic models with deep scattering networks</i> |
| 9:20am | Peter Hintz | <i>Reconstruction of Lorentzian manifolds from boundary light observation sets</i> |
| 10:10am | | ————— BREAK ————— |
| 10:35am | Paul Johnson | <i>Probing fault frictional state with machine learning</i> |
| 11:25am | Ivan Dokmanić | <i>Regularization by multiscale statistics</i> |
| 12:15pm | | ————— LUNCH ————— |
| 1:15pm | Elizabeth Rampe | <i>Interpreting data collected by the NASA Mars Science Laboratory Rover</i> |
| 1:40pm | Michael McCann | <i>Convolutional neural networks for inverse problems in imaging</i> |
| 2:30pm | Joonas Ilmavirta | <i>Communication between theory and practice via deep learning and "deep teaching"</i> |
| 2:55pm | | ————— BREAK ————— |
| 3:20pm | Alexandros Dimakis | <i>Generative models and compressed sensing</i> |
| 4:10pm | Rich Baraniuk | <i>(Geo) Physics 101 for data scientists</i> |

Abstracts – Wednesday, January 24th

Nathan Kutz (University of Washington) *Data-driven discovery of governing physical laws and their parametric dependencies in engineering, physics and biology*

We consider a data-driven method whereby low-rank reduced order models can be directly constructed from data alone using sparse regression algorithms. The methods can be enacted in an online and non-intrusive fashion using randomized (sparse) sampling. The method is ideal for parametric systems, requiring rapid model updates without recourse to the full high-dimensional system. Moreover, the method discovers the high-dimensional, white-box PDE model responsible for generating observed spatio-temporal data. Parametric changes only require adjusting the parameters of ROM or building a new ROM from new subsampled data, thus there is never recourse to the original high-dimensional system. We demonstrate our algorithms on a variety of data sets and models.

Michel Campillo (Université Grenoble-Alpes) *From "ambient noise" to seismic imaging and monitoring*

We present the basic arguments leading to the retrieval of the elastic Green function from 2-point cross-correlation of noise-like continuous seismic records. We explain the link between noise correlations and time reversal experiments. We discuss the practical limitations of the approach in seismology and specifically the precision of travel time measurements for the ballistic waves used for imaging. We present shortly applications for imaging the deep structures of the Earth. Monitoring slight temporal changes of elastic properties relies on the stable measurements made with (scattered) coda waves extracted from correlations. We present arguments showing that the scattered part of the Green function is retrieved from noise with an accuracy sufficient for monitoring the changes associated with actual deformation in the solid Earth at short time scales. We present examples of applications of passive monitoring.

András Vasy (Stanford) *Global analysis via microlocal tools: Fredholm problems in non-elliptic settings*

In this talk I will discuss recent developments in understanding non-elliptic problems, such as wave propagation and the (tensorial) X-ray transform, from a global perspective, but using tools of microlocal (phase space) analysis. As an illustration: for the wave equation, this might mean solving the equation globally on spacetime, rather than piecing together local solutions; for the X-ray transform it may mean working in a subregion of the domain we are interested in, but inverting the X-ray transform globally on this subregion — indeed in a suitable sense the boundary of the subregion is analytically pushed out to infinity. The talk will emphasize the overview, rather than the technical details.

Part of this talk is based on joint work with Peter Hintz, Plamen Stefanov and Gunther Uhlmann.

Andrew Stuart (Caltech) *Large Graph Limits of Learning Algorithms*

Many problems in machine learning require the classification of high dimensional data. One methodology to approach such problems is to construct a graph whose vertices are identified with data points, with edges weighted according to some measure of affinity between the data points. Algorithms such as spectral clustering, probit classification and the Bayesian level set method can all be applied in this setting. The goal of the talk is to describe these algorithms for classification, and analyze them in the limit of large data sets. Doing so leads to interesting problems in the calculus of variations, in stochastic partial differential equations and in Monte Carlo Markov Chain, all of which will be highlighted in the talk. These limiting problems give insight into the structure of the classification problem, and algorithms for it.

Collaboration with Andrea Bertozzi (UCLA), Michael Luo (UCLA), Kostas Zygalakis (Edinburgh); and Matt Dunlop (Caltech), Dejan Slepcev (CMU), Matt Thorpe (Cambridge).

Matti Lassas (University of Helsinki) *Manifold learning and an inverse problem for a wave equation*

We consider the following two problems on manifold learning and wave imaging.

1. We study the geometric Whitney problem on how a Riemannian manifold (M, g) can be constructed to approximate a metric space (X, d_X) in the Gromov-Hausdorff sense. This problem is closely related to manifold interpolation (or manifold learning) where a smooth n -dimensional surface $S \subset \mathbb{R}^m$, $m > n$ needs to be constructed to approximate a point cloud in \mathbb{R}^m . These questions are encountered in differential geometry, machine learning, and in many inverse problems encountered in applications. The determination of a Riemannian manifold includes the construction of its topology, differentiable structure, and metric.
2. We study an inverse problem for a wave equation $(\partial_t^2 - \Delta_g)u(x, t) = F(x, t)$ on a compact manifold (M, g) . We assume that we are given an open subset $V \subset M$ and the source-to-solution map $L : F \rightarrow u|_{V \times \mathbb{R}_+}$ defined in for all sources F supported in $V \times \mathbb{R}_+$. This map corresponds to data obtained from measurements made on the set V . We use this data to construct in a stable way a discrete metric space X that approximates the manifold M in the Gromov-Hausdorff sense.

By combining these results we obtain that the source-to-solution map in an open set V , determines in a stable way the smooth manifold (M, g) .

The results on the first problem are done in collaboration with C. Fefferman, S. Ivanov, Y. Kurylev, and H. Narayanan, and the results on the second problem with R. Bosi and Y. Kurylev.

Frederik Simons (Princeton) *On the inversion of noisy, incomplete, scattered, and vector-valued satellite-data for planetary magnetic-field models*

When modeling satellite data to recover a global planetary magnetic or gravitational potential field, the method of choice remains their analysis in terms of spherical harmonics. When only regional data are available, or when data quality varies strongly with geographic location, the inversion problem becomes severely ill-posed. In those cases, adopting explicitly local methods is to be preferred over adapting global ones (e.g., by regularization). Here, we develop the theory behind a procedure to invert for planetary potential fields from vector observations collected within a spatially bounded region at varying satellite altitude. Our method relies on

the construction of spatio-spectrally localized bases of functions that mitigate the noise amplification caused by downward continuation (from the satellite altitude to the source) while balancing the conflicting demands for spatial concentration and spectral limitation. The 'altitude-cognizant' gradient vector Slepian functions enjoy a noise tolerance under downward continuation that is much improved relative to the 'classical' gradient vector Slepian functions, which do not factor satellite altitude into their construction. We have extended the theory to being able to handle both internal and external potential-field estimation. Solving simultaneously for internal and external fields under the limitation of regional data availability reduces internal-field artifacts introduced by downward-continuing unmodeled external fields.

We explain our solution strategies on the basis of analytic expressions for the behavior of the estimation bias and variance of models for which signal and noise are uncorrelated, (essentially) space- and bandlimited, and spectrally (almost) white. The 'altitude-cognizant' gradient vector Slepian functions are optimal linear combinations of vector spherical harmonics. Their construction is not altogether very computationally demanding when the concentration domains (the regions of spatial concentration) have circular symmetry, e.g., on spherical caps or rings — even when the spherical-harmonic bandwidth is large. Data inversion proceeds by solving for the expansion coefficients of truncated function sequences, by least-squares analysis in a reduced-dimensional space. Hence, our method brings high-resolution regional potential-field modeling from incomplete and noisy vector-valued satellite data within reach of contemporary desktop machines. Our examples are drawn from a study of the Martian magnetic field, where we present high-resolution local models for the crustal magnetic field of the Martian South Polar region derived from three-component measurements made by Mars Global Surveyor. Robust features of both models are magnetic stripes of alternating polarity in southern Terra Sirenum and Terra Cimmeria, ending abruptly at the rim of Prometheus Planum, an impact crater with a very weak or undetectable magnetic field.

Joint work with Alain Plattner (California State University, Fresno), Volker Michel (University of Siegen).

Elchanan Mossel (MIT) *Hierarchical Generative Models and deep learning*

We introduce Hierarchical Generative Models (HGMs), a family of models which generate data in a multi-layered fashion. These models include some classical models in molecular biology. We hypothesize that HGMs are good models for data such as natural images and speech. We discuss some initial mathematical results showing that the inference of such models requires "deep" algorithms.

Andrea Bertozzi (UCLA) *Geometric graph-based methods for high dimensional data*

We present methods for segmentation of large datasets with graph based structure. The method combines ideas from classical nonlinear PDE-based image segmentation with fast and accessible linear algebra methods for computing information about the spectrum of the graph Laplacian. The goal of the algorithms is to solve semi-supervised and unsupervised graph cut optimization problems. I will present results for image processing applications such as image labeling and hyperspectral video segmentation, and results from machine learning and community detection in social networks, including modularity optimization posed as a graph total variation minimization problem.

Abstracts – Thursday, January 25th

Gregory Beroza (Stanford) *FAST: A Data-Mining Approach for Earthquake Detection*

The Fingerprint and Similarity Thresholding (FAST) earthquake detection algorithm finds small earthquakes in continuous seismic data through uninformed similarity search (Yoon et al., 2015). FAST does not assume prior knowledge of templates nor does it use labeled examples of earthquake waveforms, rather it treats earthquake detection as a data-mining problem. FAST extracts a set of features, called fingerprints, which are compact binary representations of short-duration time windows that span long-duration data sets. We design fingerprints to be discriminative such that similar waveforms produce similar waveform fingerprints, and fingerprints corresponding to noise, which dominates most seismic data sets, have low similarity. FAST uses locality-sensitive hashing to index the data, and queries the index to identify similar fingerprints. FAST outperforms naïve, brute force similarity search by carrying out approximate similarity search that identifies similar waveforms with high probability. The improved scalability that results allows us to search up to a decade of continuous data.

Because we are detecting weak signals in very long duration data sets, we are susceptible to false detections due, for example, to sources of persistent noise. To address the false detection problem, we developed a method for extending single-station similarity-based detection over a network (Bergen and Beroza, 2017). We designed pair-wise pseudo-association to leverage the pair-wise structure of FAST output. Unlike the association typically carried out for earthquake detection, pseudo-association does not explicitly account for move-out. Instead, we exploit the fact that the relative arrival time of a pair of events at two different stations will be the same at all receivers. Pairwise pseudo-association and the supporting techniques, event-pair extraction and event resolution, complete a post-processing pipeline that combines single-station similarity measures from each station in a network into a list of candidate events. We have applied network-FAST to the Iquique, Chile foreshock sequence and found that it is sensitive and maintains a low false detection rate: we identify nearly five times as many events as are present in the local seismicity catalog (including 95% of the catalog events), and less than 1% of these candidate events are false detections.

Co-authored with Karianne J. Bergen (Institute for Computational and Mathematical Engineering, Stanford University), and Kexin Rong, Hashem Elezabi, Philip Levis, Peter Bailis (Department of Computer Science, Stanford University).

Ankur Moitra (MIT) *Robustness meets algorithms*

In every corner of machine learning and statistics, there is a need for estimators that work not just in an idealized model but even when their assumptions are violated. Unfortunately in high-dimensions, being provably robust and efficiently computable are often at odds with each other.

In this talk, we give the first efficient algorithm for estimating the parameters of a high-dimensional Gaussian which is able to tolerate a constant fraction of corruptions that is independent of the dimension. Prior to our work, all known estimators either needed time exponential in the dimension to compute, or could tolerate only an inverse polynomial fraction of corruptions. Not only does our algorithm bridge the gap between robustness and algorithms, it turns out to be highly practical in a variety of settings.

Yiran Wang (University of Washington) *Inverse problems for nonlinear acoustic and elastic wave equations*

We consider wave propagation in acoustic and elastic medium with nonlinear properties. In various applications, it is observed that the nonlinear interactions of waves could generate new responses and such interactions have been studied using plane waves and spherical waves in the literature. In this talk, we analyze the nonlinear interactions using distorted plane waves and microlocal methods. The new approach gives us a precise characterization of the nonlinear responses, which lead us to the determination of medium properties, such as the location of an interface and the nonlinear parameters.

This talk is based on joint works with M. de Hoop and G. Uhlmann.

Carola-Bibiane Schönlieb (Cambridge) *Model-based learning in imaging*

One of the most successful approaches to solve inverse problems in imaging is to cast the problem as a variational model. The key to the success of the variational approach is to define the variational energy such that its minimiser reflects the structural properties of the imaging problem in terms of regularisation and data consistency.

Variational models constitute mathematically rigorous inversion models with stability and approximation guarantees as well as a control on qualitative and physical properties of the solution. On the negative side, these methods are rigid in a sense that they can be adapted to data only to a certain extent.

Hence researchers started to apply machine learning techniques to “learn” more expressive variational models. In this talk we discuss two approaches: bilevel optimisation (which we investigated over the last couple of years and which aims to find an optimal model by learning from a set of desirable training examples) and quotient minimisation (which we only recently proposed as a way to incorporate negative examples in regularisation learning). Depending on time, we will review the analysis of these approaches, their numerical treatment, and show applications to learning sparse transforms, regularisation learning, learning of noise models and of sampling patterns in MRI. To finish I will also give a sneak-preview onto our recent efforts integrating deep learning in regularised image reconstruction.

This talk will potentially include joint work with S. Arridge, M. Benning, L. Calatroni, C. Chung, J. C. De Los Reyes, M. Ehrhardt, G. Gilboa, J. Grah, A. Hauptmann, S. Lunz, G. Maierhofer, O. Oektem, F. Sherry, and T. Valkonen.

Joan Bruna (NYU, Courant Institute) *On the optimization landscape of neural networks*

The loss surface of deep neural networks has recently attracted interest in the optimization and machine learning communities as a prime example of high-dimensional non-convex problem. Some insights were recently gained using spin glass models and mean-field approximations, but at the expense of simplifying the nonlinear nature of the model, as well as tensor decompositions.

In this talk, we study conditions on the data distribution and model architecture that prevent the existence of bad local minima. We first take a topological approach and characterize absence of bad local minima by studying the connectedness of the loss surface level sets. Our work quantifies and formalizes two important facts: (i) the landscape of deep linear networks has a radically different topology from that of deep half-rectified ones, and (ii) that the energy landscape in the non-linear case is fundamentally controlled by the interplay between the smoothness of the data distribution and model over-parametrization. Our main theoretical

contribution is to prove that half-rectified single layer networks are asymptotically connected, and we provide explicit bounds that reveal the aforementioned interplay.

The conditioning of gradient descent is the next challenge we address. We study this question through the geometry of the level sets, and we introduce an algorithm to efficiently estimate the regularity of such sets on large-scale networks. Our empirical results show that these level sets remain connected throughout all the learning phase, suggesting a near convex behavior, but they become exponentially more curvy as the energy level decays, in accordance to what is observed in practice with very low curvature attractors.

Joint work with Daniel Freeman (UC Berkeley), Luca Venturi and Afonso Bandeira (Courant, NYU).

Lauri Oksanen (UCL) *Correlation based passive imaging with a white noise source*

Since the first empirical seismological demonstrations were achieved by Campillo and Paul (2003) and Shapiro and Campillo (2004), ambient noise tomography has received a lot of attention. In the purely mathematical context, models of this technique have been studied by Bardos, Garnier and Papanicolaou (2008) and Colin de Verdière (2009), who showed that the leading order structure of Green's function can be recovered from correlations of ambient noise signals. We consider a mathematical model of ambient noise tomography and show that, in the context of this model, a spatially varying speed of sound can be fully recovered from the correlations of ambient noise signals.

The talk is based on a joint work with T. Helin, M. Lassas and T. Saksala.

Victor Pankratius (MIT Haystack Observatory) *Computer-aided discovery: can a machine win a Nobel Prize?*

How far can we push automated scientific discovery? Can a machine eventually win a Nobel Prize by generating theories and insight beyond human cognitive limits? Just like the Moon landing, we need this challenging goal as an inspiration and measure of success, but it will be impossible to achieve without concerted interdisciplinary research that goes far beyond the state-of-the-art in AI and Deep Learning.

In this talk, I will address key problems in scientific insight generation and current gaps in AI. I will discuss the steps we have taken at MIT with NASA and NSF support to close these gaps and develop a Computer-Aided Discovery system that is being tested on real Geoscience problems. Our work contributes foundations that facilitate the answering of scientific questions, including how empirical detections fit into hypothesized models and model variants. Successful algorithmic techniques will be discussed in the context of volcanology, groundwater studies, atmospheric phenomena, with an emphasis on how this approach has led to a new discovery of volcanic inflation events.

Robert Nowak (University of Wisconsin-Madison) *Outranked: exploiting nonlinear algebraic structure in matrix recovery problems*

This talk considers matrix completion in cases where columns are points on a nonlinear algebraic variety (in contrast to the commonplace linear subspace model). A special case arises when the columns come from a union of subspaces, a model that has numerous practical applications. We propose a new approach to this problem based on data tensorization (i.e., products of original data) in combination with standard low-rank matrix completion methods. The key

insight is that while the original data matrix may not exhibit low-rank structure, often the tensorized data matrix does. The challenge, however, is that the missing data patterns in the tensorized representation are highly structured and far from uniformly random. We show that, under mild assumptions, the observation patterns are generic enough to enable exact recovery.

Abstracts – Friday, January 26th

Stéphane Mallat (Collège de France) *Unsupervised learning of stochastic models with deep scattering networks*

A major challenge in physics and inverse problems is to model high-dimensional non-Gaussian data with long range interactions, and hence estimate their probability density. This density is often continuous to actions of deformations on the data, and usually locally or globally invariant to translations and sometimes rotations. We shall first introduce maximum entropy models based on multiscale scattering moments, which capture the multiscale regularity of stationary probability densities. Applications will be shown on fluid and gas turbulence as well as image and audio textures. We will then concentrate on non-stationary and highly structured processes such as faces or geometrical shapes. Generative-adversarial deep neural networks have obtained remarkable results to model such random processes. We show that one can obtain similar results by transforming a Gaussian white noise with a non-linear operator, computed by inverting a multiscale scattering transform.

Peter Hintz (Berkeley) *Reconstruction of Lorentzian manifolds from boundary light observation sets*

In joint work with Gunther Uhlmann, we consider the problem of reconstructing the topological, differentiable, and conformal structure of subsets S of Lorentzian manifolds M with boundary from the collection of boundary light observation sets: these are the intersections of light cones emanating from points in S with a fixed subset of the boundary of M ; here, light rays get reflected according to Snell's law upon hitting the boundary. This can be viewed as a simple model of wave propagation in the interior of the Earth. We solve this inverse problem under a convexity assumption on the boundary of M .

Paul Johnson (Los Alamos) *Probing fault frictional state with machine learning*

We are applying machine learning techniques to probe the frictional state of laboratory faults. We use continuous recordings of the sound emitted from the fault zone to determine the frictional state, and when an upcoming laboratory earthquake may occur. Indeed, the machine learning approach allows us to develop a constitutive relation between the statistical characteristics and the fault friction. This approach also allows us to predict how large the laboratory earthquake will be in some instances. It is remarkable that such information can be gleaned from the sound the fault makes. Our goal is to scale these approaches to faults in Earth.

This work is co-authored with Claudia Hulbert and Bertrand Rouet-LeDuc.

Ivan Dokmanić (UIUC) *Regularization by multiscale statistics*

I will talk about a new approach to learning-based regularization of ill-posed linear inverse problems. Unlike emerging data-driven approaches such as directly learning a stable inverse, unrolling iterative algorithms into neural nets, or learning projectors and gradients in iterative schemes, our approach is arguably more classical—we find the solution as a minimizer of a non-convex cost functional. We regularize the problem by requiring that the solution reproduce a vector of generalized moments estimated from the measured data. The moments are computed by the non-linear multiscale scattering transform—a complex convolutional network which discards the phase and thus exposes spectral correlations otherwise hidden beneath the phase fluctuations. This regularizer can be interpreted as promoting “correct” conditional statistics in the image space of the transform that maps signals to moments. Scale separation provided by the scattering transform ensures that the moment representation is stable to deformations; it also gives us a scale-dependent sparsity specification of the signal class of interest. For a given realization, the moment vector is linearly estimated from a reconstruction in a stable subspace and it encodes the unstable part of the signal. We demonstrate that our approach stably recovers the missing spectrum in super-resolution and tomography.

Michael McCann (EPFL) *Convolutional neural networks for inverse problems in imaging*

When the first convolutional neural network (CNN)-based method entered the ImageNet Large-Scale Visual Recognition Challenge in 2012, its error rate was 15.3%, as compared to an error rate of 26.2% for the next closest method and 25.8% for the 2011 winner. In subsequent competitions (2013–2017), the majority of the entries (and all of the winners) were CNN-based and continued to improve substantially, with the 2017 winner achieving an error rate of just 2.25%. A plethora of CNN-based approaches are now being applied to inverse problems in imaging. Should we expect the same dramatic improvements here? For example, can we expect denoising results to improve by an order of magnitude (20 dB) in the next few years?

In this talk, I will survey some of the recent CNN-based approaches to inverse problems in imaging, focusing on key design questions such as (1) From where do the training data come? (2) What is the architecture of the CNN? and (3) How is the learning problem formulated and solved? I will also present our recent work on using a CNN for X-ray computed tomography reconstruction.

Joonas Ilmavirta (University of Jyväskylä) *Communication between theory and practice via deep learning and “deep teaching”*

Before one computes anything from data, it would be good to be sure that under ideal circumstances the measurements actually determine the desired quantity uniquely. Proving results of this kind is what mathematical inverse problems is all about. However, these proofs do not always produce implementable algorithms. Deep learning can provide a way to communicate between theorists and experimentalists and steer both fields in new directions. This requires that insight gained through deep learning be taught back to humans in what I would call deep teaching. I will discuss this with concrete examples arising from a practical problem: How to determine the interior structure of an object from the spectrum of its free oscillations?

Alexandros Dimakis (University of Texas at Austin) *Generative models and compressed sensing*

The goal of compressed sensing is to estimate a vector from an under-determined system of noisy linear measurements, by making use of prior knowledge in the relevant domain. For most results in the literature, the structure is represented by sparsity in a well-chosen basis.

We show how to achieve guarantees similar to standard compressed sensing but without employing sparsity at all. Instead, we assume that the unknown vectors lie near the range of a generative model, e.g. a GAN or a VAE. We show how the problems of image inpainting and super-resolution are special cases of our general framework.

We show how to generalize the RIP condition for generative models and that random gaussian measurement matrices have this property with high probability. A Lipschitz condition for the generative neural network is the key technical issue for our results.

Time permitting we will discuss follow-up work on how GANs can model causal structure in high-dimensional probability distributions.

This talk is based on collaborations with Ashish Bora, Ajil Jalal, Murat Kocaoglu, Christopher Snyder and Eric Price.

Rich Baraniuk (Rice University) *(Geo) Physics 101 for data scientists*

A grand challenge in machine learning is the development of computational algorithms that match or outperform humans in perceptual inference tasks that are complicated by nuisance variation. For instance, visual object recognition involves the unknown object position, orientation, and scale in object recognition while speech recognition involves the unknown voice pronunciation, pitch, and speed. Recently, a new breed of deep learning algorithms have emerged for high-nuisance inference tasks that routinely yield pattern recognition systems with near- or super-human capabilities. But a fundamental question remains: Why do they work? Intuitions abound, but a coherent framework for understanding, analyzing, and synthesizing deep learning architectures has remained elusive. We answer this question by developing a new probabilistic framework for deep learning based on the Deep Rendering Model: a generative probabilistic model that explicitly captures latent nuisance variation. By relaxing the generative model to a discriminative one, we can recover two of the current leading deep learning systems, deep convolutional neural networks and random decision forests, providing insights into their successes and shortcomings, a principled route to their improvement, and new avenues for exploration.