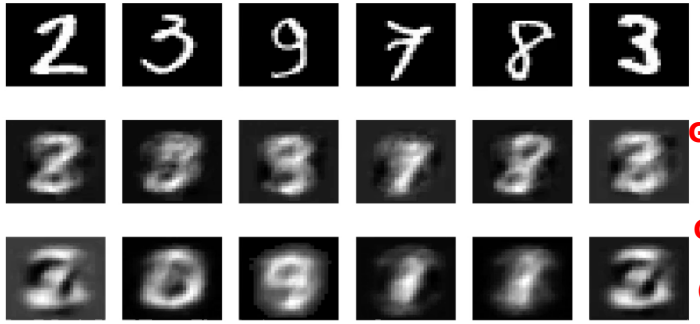
Gradient Information for Representation and Modeling

Goal: In this work we present a new set of information quantities which we refer to as gradient information. These measures serve as surrogates for classical information measures such as those based on logarithmic loss, Kullback-Leibler divergence, directed Shannon information, etc. in many data-processing scenarios of interest, and often provide significant computational advantage, improved stability, and robustness. We have applied them to faster training of RBM, tree approximation of high dimensional data, deep compression, etc.

Training of RBMs using Fisher-based versus Classical Approaches



Original MNIST Images

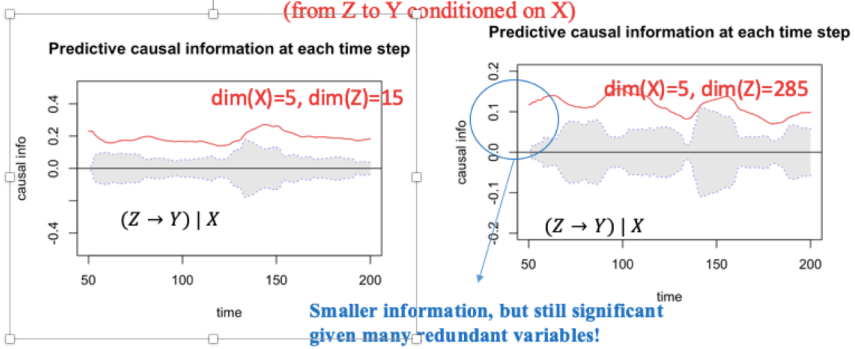
Generated Images using our method

Generated Images using Classical KL Divergence (approximated by Contrastive Divergence)

Causal Discovery using gradient information

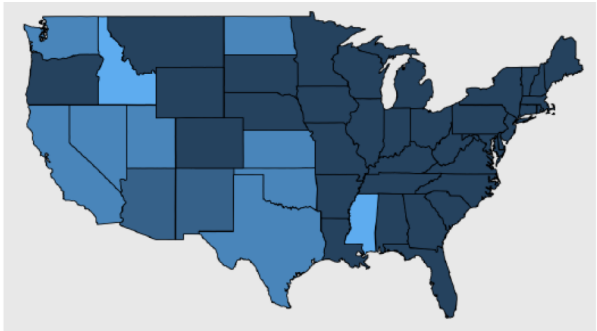
Example: $y = X\beta_1 + Z\beta_2 + e$,
where $[\beta_1, \beta_2] = [10, 5, 5, 2.5, 2.5, 1.25, 1.25, 0.675, 0.675, 0.3125, 0.3125, 0 \ldots, 0]$

Discovered significant causal information in both cases (from Z to Y conditioned on X)

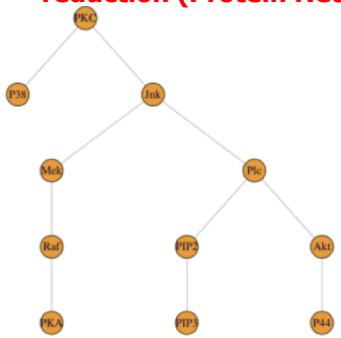


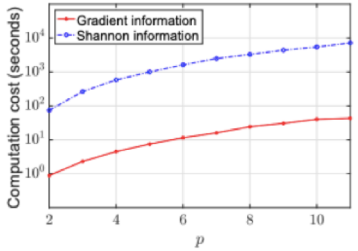
Smaller information, but still significant given many redundant variables!

Accurate Community Detection from quarterly growth rates of payroll employment



Faster and More Accurate Dimensionality reduction (Protein Network Example)





Gradient Information for Representation and Modeling

- 1- Video demonstrates the training of Restricted Boltzmann Machine on MNIST using our scheme versus the classical scheme (approximated contrastive divergence) with the same computational complexity at each time step. As it can even been seen visually, we have significant improvement in performance with the same computational complexity.
- 2- Causality calculation using new scheme for the model $y = X\beta_1 + Z\beta_2 + e$. The 95% confidence interval for no causal relation is in gray and we can not reject the causal relation. Stability to changes in dimension is also demonstrated.
- 3- A new Chow-Liu type result is developed using the new score and used for dimensionality reduction in "protein signaling flow cytometry" dataset. The dataset encodes the presence of p = 11 proteins in n = 7466 cells. The tree can be used as a graphical visualization tool to highlight the most highly correlated genes in the correlation network.
- 4- A new community detection scheme based on the new score applied to the data consists of quarterly growth rates of payroll employment for the U.S. states (excluding Alaska and Hawaii) from the second quarter of 1956 to the fourth of 2007.

Some of these results are in a paper accepted to NeurIPS 2019 and supported by DARPA grant number HR00111890040.

J. Ding, A. R. Calderbank, and V. Tarokh, "Gradient information for representation and modeling," *Conference on Neural Information Processing Systems (NeurIPS), 2019.*