



# Discussion on Levina, Johnstone, Lafferty

Bin Yu

Statistics Department  
University of California, Berkeley

# The Data Deluge

## We are awash in Data

Science, engineering, research, education data

Medical and government records

Finance and commerce

Digital music, video, entertainment, recreation

...

## Data Characteristics

High-dimensionality

( large  $p$  and/or  $n$ , e.g.  $n = 36$  million,  $p = 860k$ )

High data rate

(streaming data) (e.g. digital sky survey)

# Major goals of statistical investigation

1. Prediction
2. Interpretation

## Common requirements

1. Data reduction
2. Fast computation

**PCA is often used for data reduction/denoising.  
RMT results warn PCA's potential dangers  
in large  $p$  case...**

**Questions:**

- 1. Are the subspaces of non-leading eigen vectors fine to use?**
- 2. Results for Bernoulli variables? (e.g. IME)**

# Regularization through sparsity

To avoid the instability of the largest eigenvalue/eigenvector, regularization is used.

Examples of regularization via sparsity (for prediction and interpretation)

Spiked model (sparsity) (Johnstone)

Band covariance matrices (sparsity) (Levina)

Regularization is used elsewhere as well:

L1-Penalized loss function (sparsity) (Levina)

Local gradient kernel estimation (sparsity) (Lafferty)

...

# Regularization vs computation

**There are computation constraints such that not all regularization methods are implementable for a particular computing device (e.g. streaming data, huge  $n$  and  $p$  for off-line)**

**Hope: bring computation into our analysis. For example,**

- 1. Analyze algorithmic versions of methods**
- 2. Show that computation considerations are forms of regularization**

# Hints of close connections between computation and regularization

**Ridge regression (Tikhonov regularization) was invented to solve a numerical instability problem.**

**Recent loss function based machine learning methods use convex optimization:**

- **Boosting was invented to solve a learning problem. Its regularization is carried out by stopping the iteration early – using test set or CV. L2 Boost was later shown to be adaptively minimax optimal for Sobolev classes.**
- **Support Vector Machine (SVM) employs the “kernel trick” to minimize a penalized hinge loss (convex optimization). It is also later found to be adaptively minimax optimal for Sobolev classes.**

# Hints (cont): from Donoho's talk

**Similar phase transition curves appear from both**

**computational side (solving  $l_0$  penalty problem using  $l_1$  penalty)**

**and**

**minimax estimation framework (regularization).**

**Question:**

**Coincidence or ?**

**An important development in science is occurring at the intersection of computer science and the sciences that has the potential to have a profound impact on science. It is a leap from the application of computing to support scientists to “do” science (i.e. “computational science”) to the integration of computer science concepts, tools, and theorems into the very fabric of science.**

**-- 2020 Science Report, 2005**