Novel Multiscale Representations of Data Sets for Interactive Learning

Mauro Maggioni, Eric Monson, R. Brady
Dept. of Mathematics and Computer Science
Duke University

FODAVA Annual Review 2010, Georgia Tech
12/9/2010

Joint work with: G. Chen.

Partial support: NSF/DHS, ONR

Ongoing efforts in several directions

- Using diffusion processes on graphs for (inter)active learning.
- Perform multiscale analysis on graphs: construction of graph-adaptive multiscale analysis, for graph visualization and exploration, and (inter)active learning.
- Sparse learning w.r.t. multiscale dictionaries on graphs.
- Construct data-adaptive dictionaries for data-modeling and exploration.

Ongoing efforts in several directions

- Using diffusion processes on graphs for (inter)active learning.
- Perform multiscale analysis on graphs: construction of graph-adaptive multiscale analysis, for graph visualization and exploration, and (inter)active learning.
- Sparse learning w.r.t. multiscale dictionaries on graphs.
- Construct data-adaptive dictionaries for data-modeling and exploration.

Random walks on data & graphs

- One may connect data points to form a graph, with edges weighted by the similarity of data points.
- One can then construct a random on the data points, which may be used for a variety of tasks:
  - construct local and global embeddings of the data in low dimensions,
  - perform learning tasks such as clustering, classification, regression, etc..
  - diffuse information (e.g. labels) on data
  - study geometric properties of data

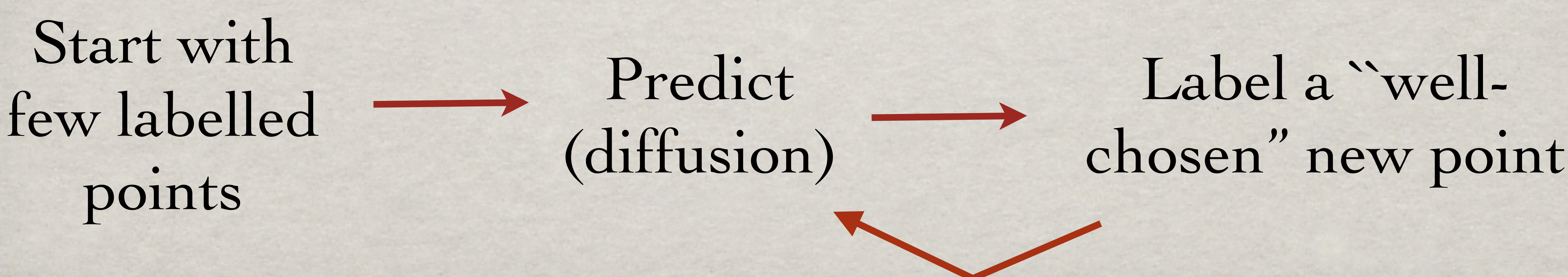
Active Learning

With E. Monson and R. Brady [C.S.]

Given: full data set (e.g. a body of text documents).

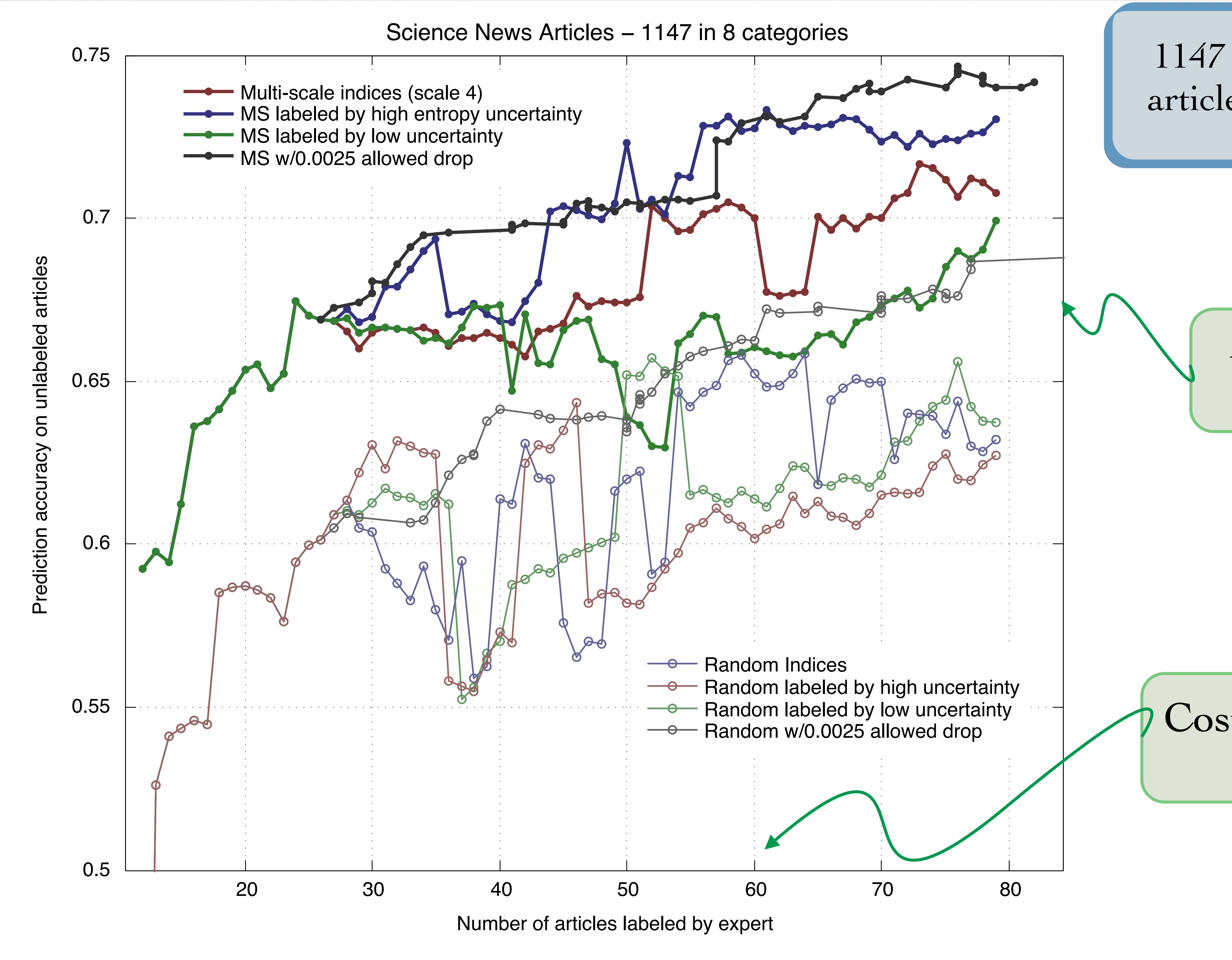
Goal: learn a categorization of the data (e.g. topics of the text documents). Cost for every label we obtain from an expert. Large scale here means: “labels are very expensive compared to very large amount of data available”.

Find points whose labels maximize the gain in prediction accuracy. Natural candidates: points with highly uncertain predictions + well-distributed on the data (standard idea) (our contribution) Points actually proceed in a multiscale fashion.



Active Learning

With E. Monson and R. Brady [C.S.]



1147 Science News articles, 8 categories

Accuracy of predictions

Cost: # of labeled points

Ongoing efforts in several directions

- Using diffusion processes on graphs for (inter)active learning.
- Perform multiscale analysis on graphs: construction of graph-adaptive multiscale analysis, for graph visualization and exploration, and (inter)active learning.
- Sparse learning w.r.t. multiscale dictionaries on graphs.
- Construct data-adaptive dictionaries for data-modeling and exploration.

Ongoing efforts in several directions

- Using diffusion processes on graphs for (inter)active learning.
- Perform multiscale analysis on graphs: construction of graph-adaptive multiscale analysis, for graph visualization and exploration, and (inter)active learning.
- Sparse learning w.r.t. multiscale dictionaries on graphs.
- Construct data-adaptive dictionaries for data-modeling and exploration.

Example: text documents

With S. Mukherjee
and J. Guinney

X is $N \times D$, N documents in $\mathbb{R}^D$, compute multiscale dictionary Φ $(D \times M)$ on the D words. If f maps documents to their topic, write $f = X\Phi\beta + \eta$ and find β by

$$\text{argmin}_\beta ||f - X\Phi\beta||_2^2 + \lambda||\{2^{-j\gamma}\beta_{j,k}\}||_1 \,,$$

which is a form of sparse regression. (λ, γ) are determined by cross-validation.

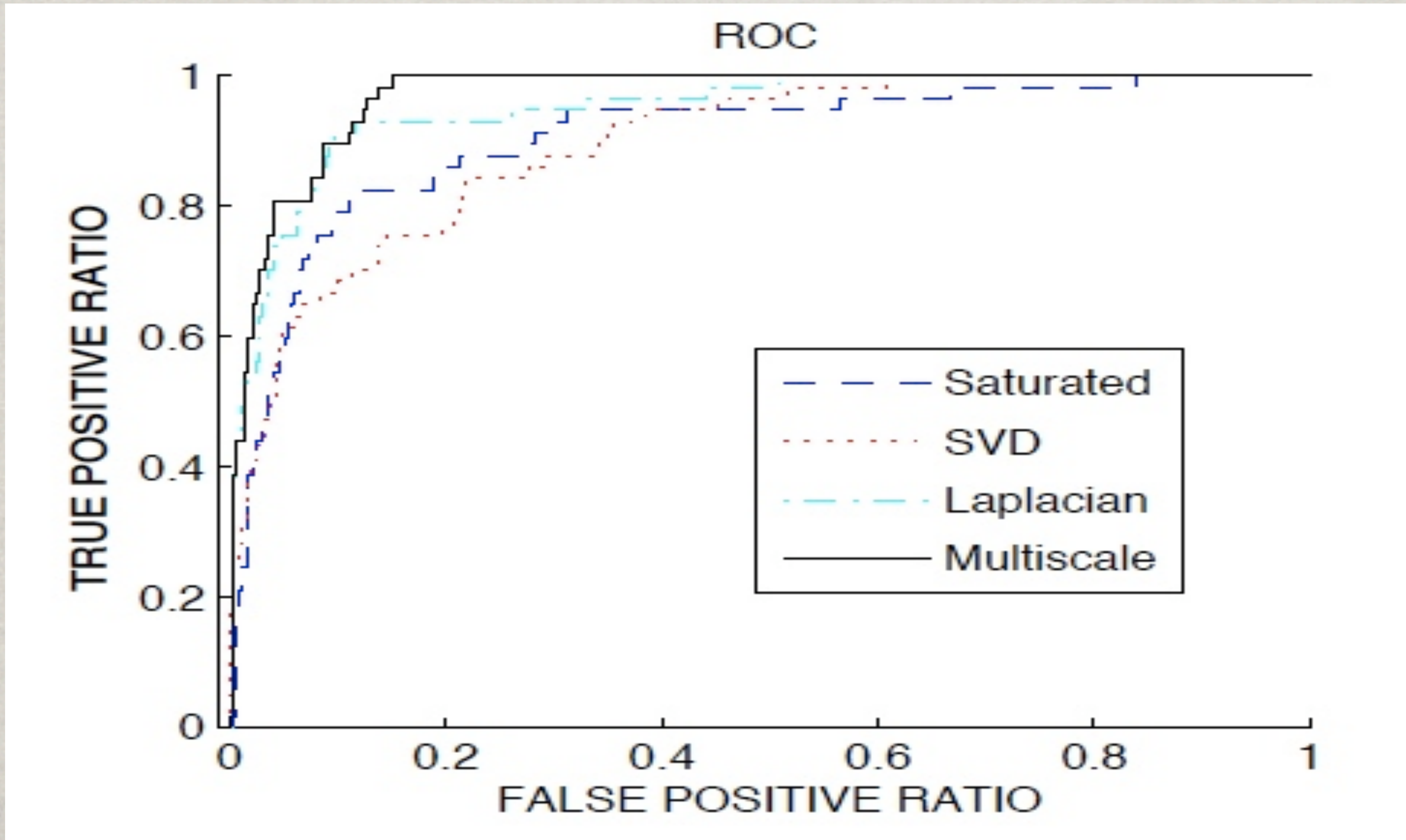
Example: text documents

With S. Mukherjee
and J. Guinney

X is $N \times D$, N documents in $\mathbb{R}^D$, compute multiscale dictionary Φ $(D \times M)$ on the D words. If f maps documents to their topic, write $f = X\Phi\beta + \eta$ and find β by

$$\text{argmin}_\beta ||f - X\Phi\beta||_2^2 + \lambda ||\{2^{-j\gamma}\beta_{j,k}\}||_1 \,,$$

which is a form of sparse regression. (λ, γ) are determined by cross-validation.



Example: gene microarray data

With S. Mukherjee
and J. Guinney

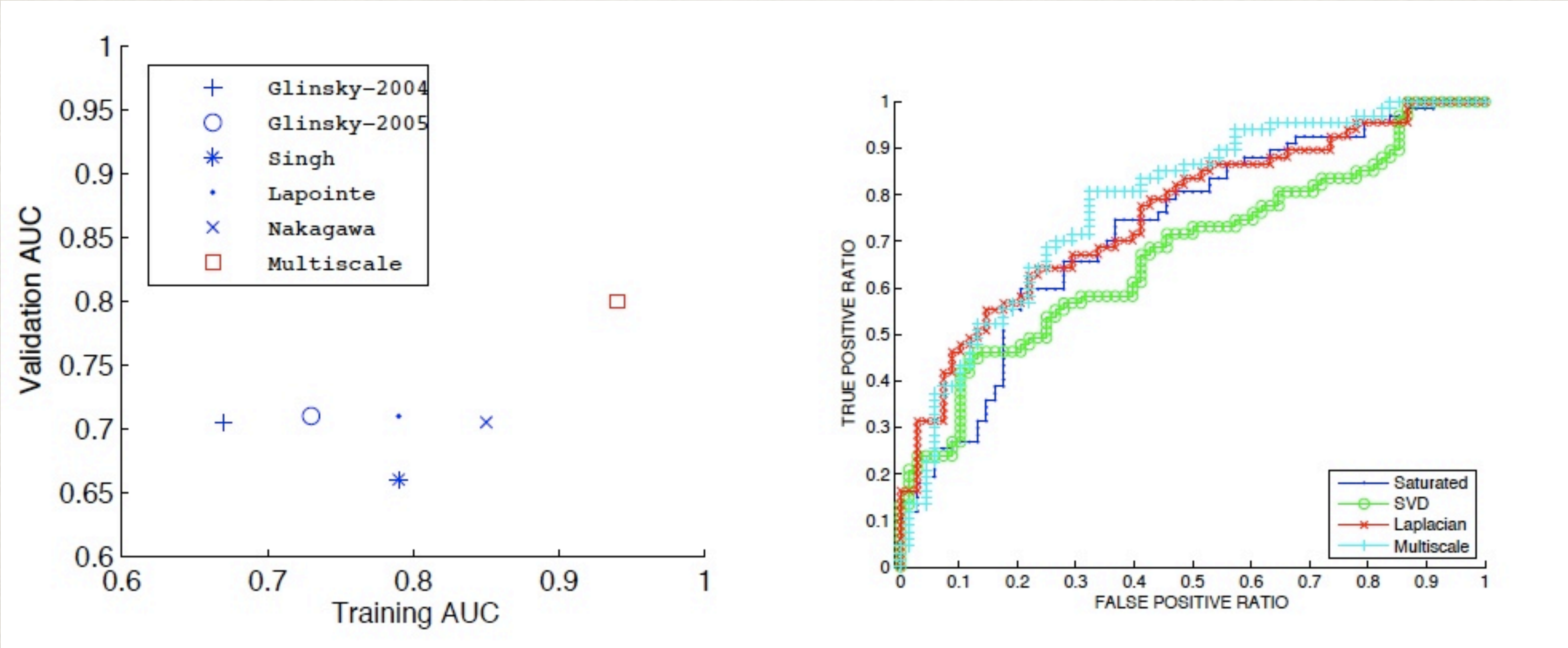
X is $N \times D$, N patients with D genes (here $N \sim 400$ and $D \sim 1000$).

Source of data: Nakagawa T, Kollmeyer T, Morlan B, Anderson, S, Bergstralh E, et al, (2008)
A tissue biomarker panel predicting systemic progression after PSA recurrence post-definitive prostate cancer therapy, Plos One 3:e2318.

Example: gene microarray data

With S. Mukherjee
and J. Guinney

X is $N \times D$, N patients with D genes (here $N \sim 400$ and $D \sim 1000$).

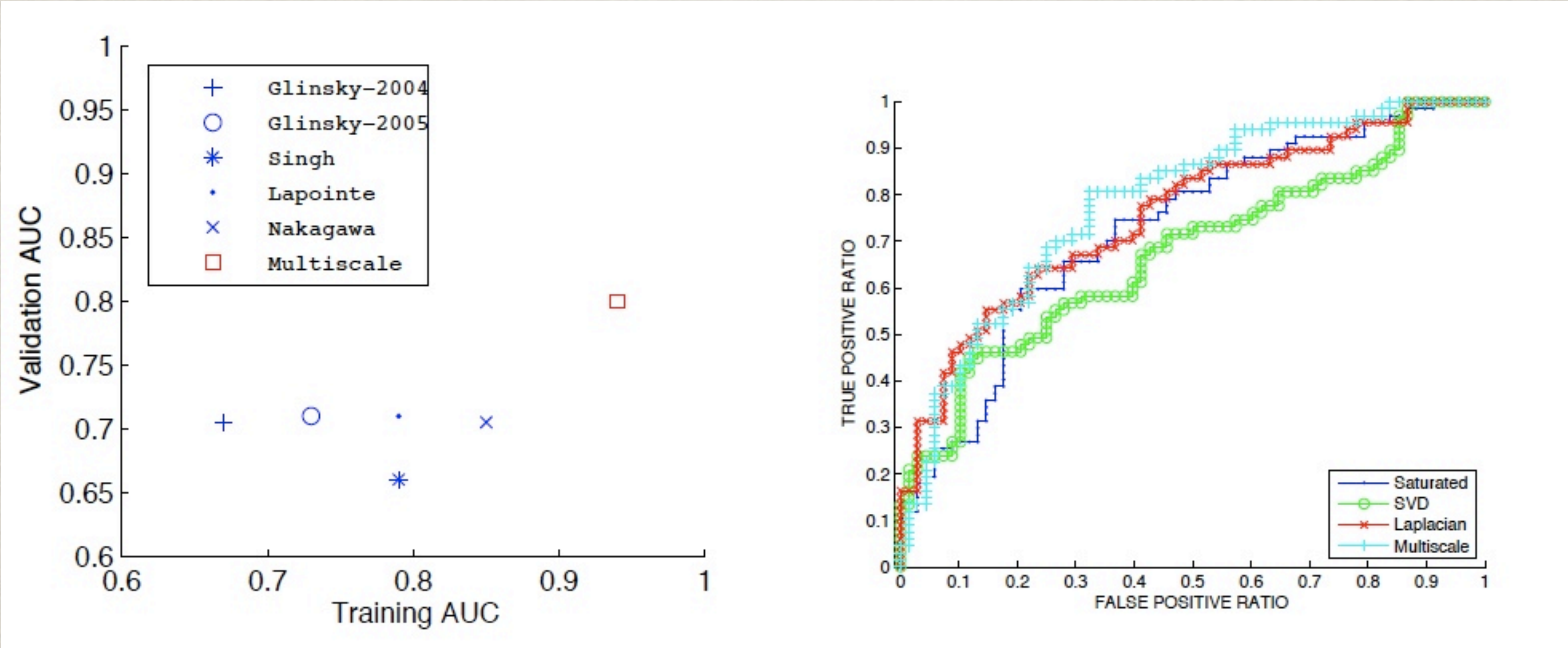


Source of data: Nakagawa T, Kollmeyer T, Morlan B, Anderson, S, Bergstralh E, et al, (2008)
A tissue biomarker panel predicting systemic progression after PSA recurrence post-definitive prostate cancer therapy, Plos One 3:e2318.

Example: gene microarray data

With S. Mukherjee
and J. Guinney

X is $N \times D$, N patients with D genes (here $N \sim 400$ and $D \sim 1000$).



Added advantage: the multiscale genes we construct are much interpretable than eigengenes, several of them match important pathways, and moreover both small scale and large scale genelets seem relevant.

Source of data: Nakagawa T, Kollmeyer T, Morlan B, Anderson, S, Bergstralh E, et al, (2008)
A tissue biomarker panel predicting systemic progression after PSA recurrence post-definitive prostate cancer therapy, Plos One 3:e2318.

Ongoing efforts in several directions

- Using diffusion processes on graphs for (inter)active learning.
- Perform multiscale analysis on graphs: construction of graph-adaptive multiscale analysis, for graph visualization and exploration, and (inter)active learning.
- Sparse learning w.r.t. multiscale dictionaries on graphs.
- Construct data-adaptive dictionaries for data-modeling and exploration.

Ongoing efforts in several directions

- Using diffusion processes on graphs for (inter)active learning.
- Perform multiscale analysis on graphs: construction of graph-adaptive multiscale analysis, for graph visualization and exploration, and (inter)active learning.
- Sparse learning w.r.t. multiscale dictionaries on graphs.
- Construct data-adaptive dictionaries for data-modeling and exploration.

Geometric Wavelets: Multiscale Data-Adaptive Dictionaries

- Many constructions of "general-purpose" dictionaries [Fourier, wavelets, curvelets, ...], especially for low-dimensional signals (sounds, images,...).

  Motivation: pretend we have rather good tractable models (e.g. function spaces), construct good dictionaries by hand.

  Goals: compression, signal processing tasks (e.g. denoising), etc...

- Recently, many constructions of data-adaptive dictionaries [K-SVD, K-planes, ...].

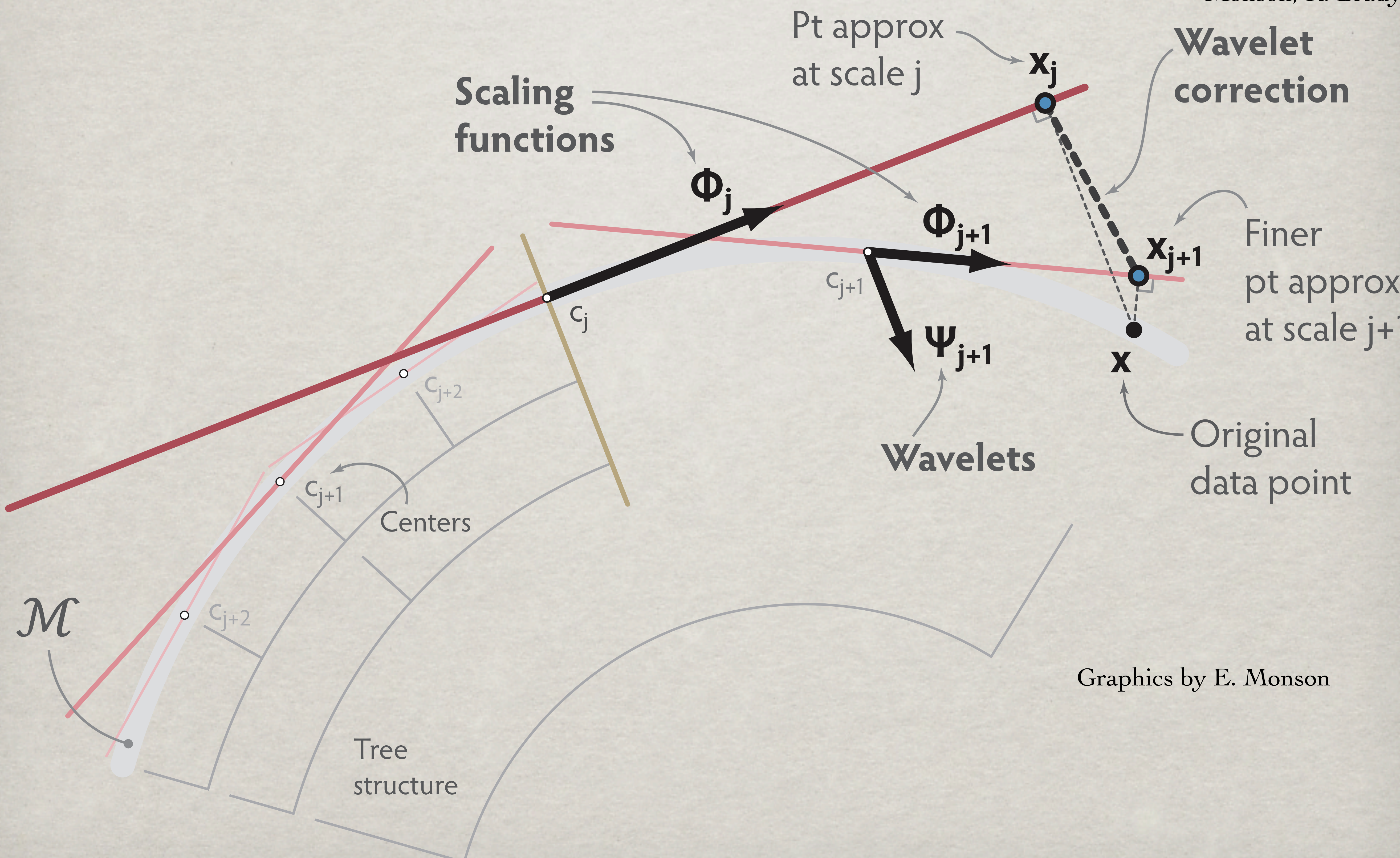
  Motivation: we do not have tractable good models, need to adapt to data.

  Goals: as before, albeit hopes for more general types of high-dimensional data.

- Important role of sparsity in statistics, learning, design of measurements, ...: seek dictionaries that yield sparse representations of the data.

Dictionary Learning

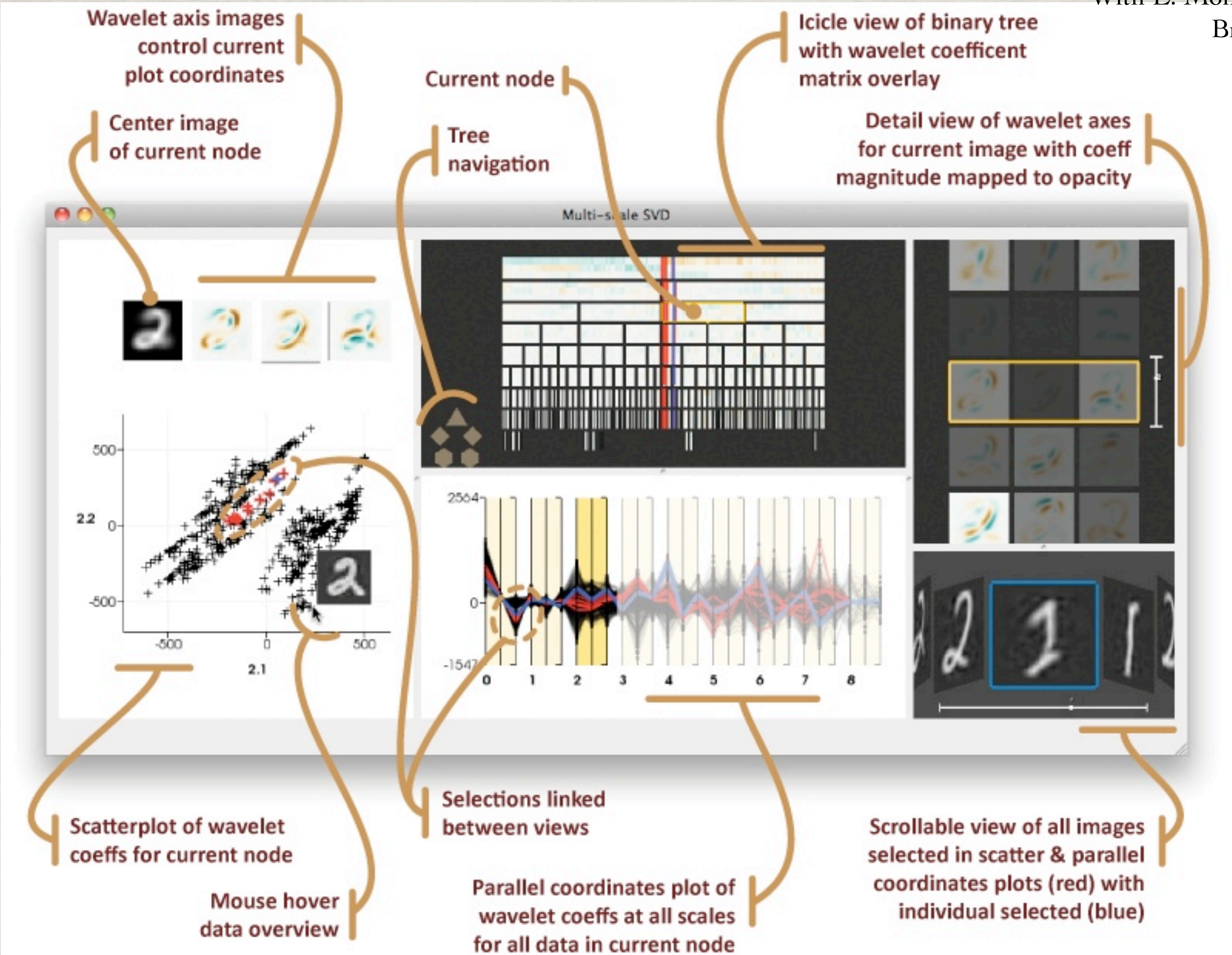
With G. Chen, E. Monson, R. Brady



Graphics by E. Monson

UI for Geometric Wavelets

With E. Monson and R. Brady [C.S.]



Open problems & future dir.'s

- Geometric wavelets meet interactive learning.
- Multiscale analysis on graphs meets interactive learning.
- Better visualization of multiscale analysis of graphs [E. Monson, R. Brady]
- Towards a toolbox of highly robust geometric analysis tools for data sets [A. Little, G. Chen].
- Dynamic graphs [J. Lee].
- Wrap up toolboxes; scale part of the code.

Collaborators: E. Monson, R. Brady (Duke C.S.); A. V. Little, K. Balachandrian (Math grad, Duke), J. Lee (Math undergrad, Duke); L. Rosasco (CS, MIT and Universita' di Genova).
Funding: NSF, ONR, Sloan Foundation, Duke.

www.math.duke.edu/~mauro