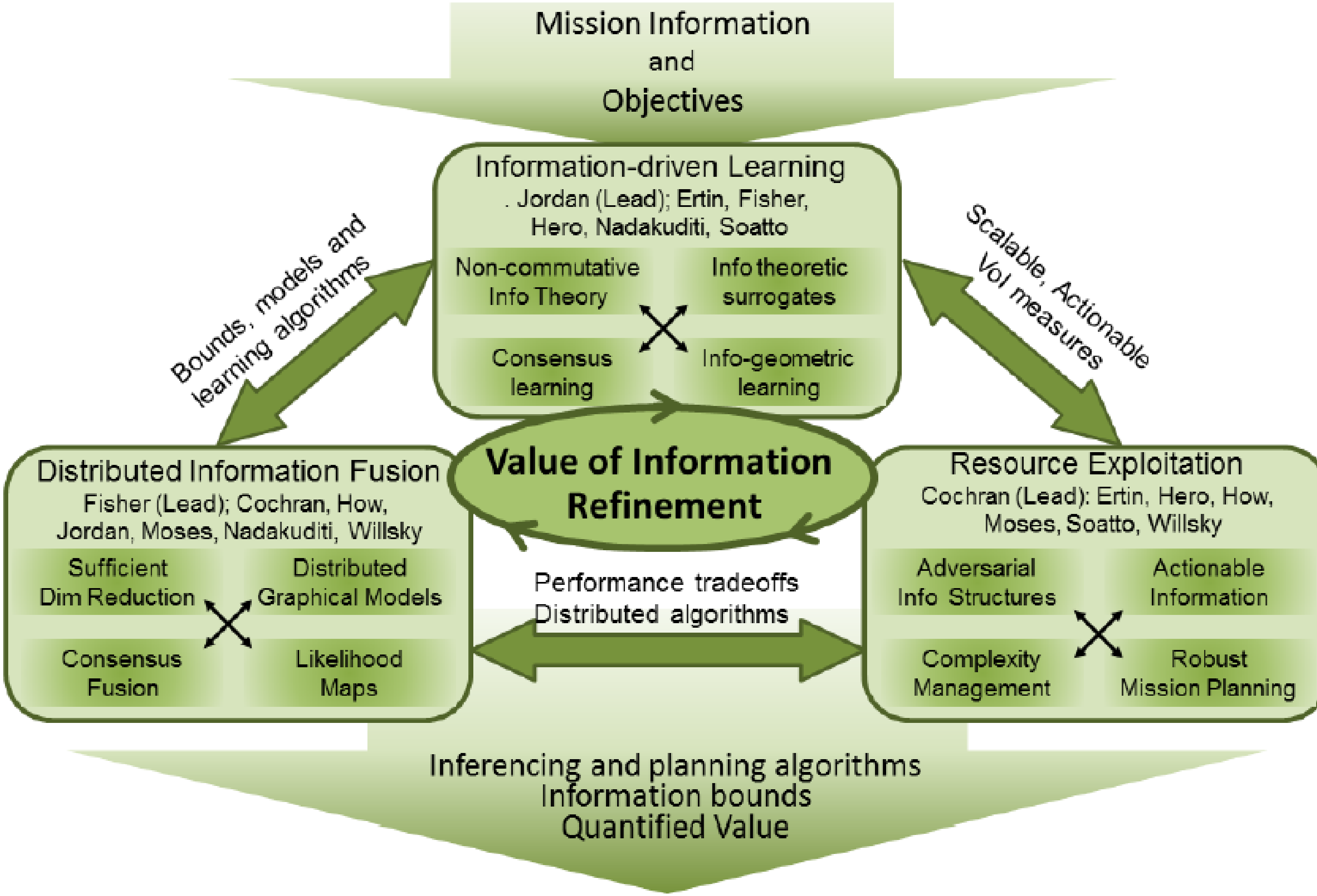
Algorithms, performance bounds and VoI metrics from non-commutative information theory

Co-PI Raj Rao
University of Michigan


VOI
Mission Information
and
Objectives
Information-driven Learning
. Jordan (Lead); Ertin, Fisher,
Hero, Nadakuditi, Soatto
Non-commutative
Info Theory
Info theoretic
surrogates
Consensus
learning
Info-geometric
learning
Bounds, models and
learning algorithms
Scalable, Actionable
VoI measures
Value of Information
Refinement
Distributed Information Fusion
Fisher (Lead); Cochran, How,
Jordan, Moses, Nadakuditi, Willsky
Sufficient
Dim Reduction
Distributed
Graphical Models
Consensus
Fusion
Likelihood
Maps
Resource Exploitation
Cochran (Lead): Ertin, Hero, How,
Moses, Soatto, Willsky
Adversarial
Info Structures
Actionable
Information
Complexity
Management
Robust
Mission Planning
Performance tradeoffs
Distributed algorithms
Inferencing and planning algorithms
Information bounds
Quantified Value

Four major axes of progress

- **Progress 1**: Data-driven low-rank matrix denoising:
  - Outperforms truncated SVD (!) and nuclear norm reg.
  - Optimal singular value shrinkage via NCIT
- **Progress 2**: New eigen-VoI metrics from NCIT:
  - Consistent, data-driven estimator of denoising MSE
  - Use: Fusion of multiple modes with varying SNRs
  - Collaborations w/ Cochran & Hero
- **Progress 3**: Universality of eigen-spectrum behavior
  - Square-root decay at edge is universal
- **Progress 4**: Kronecker structured CS
  - Eigen-spectra of sections of Kroneckered Haar unitaries
  - Same behavior as sections of Haar unitaries!

The low-rank matrix denoising problem

- <u>Signal-plus-Noise Model:</u>

$$\widetilde{X} = \sum_{i=1}^{r} \theta_i u_i v_i^H + X$$

- <u>Objective</u>: Optimally denoise $S = \sum_{i=1}^{r} \theta_i u_i v_i^H$ (assume *r* known)

- <u>(Minimal) Assumptions</u>*:*
  - No structure on *S* (besides low-rank)
  - Noise-only matrix *X* has isotropically random singular vectors

The “optimality” of the truncated SVD

- Optimization problem:

$$\widehat{S}_{\text{eym}} = \underset{\text{rank}(S)=r}{\arg\min} \, ||\widetilde{X} - S||_F,$$

- Eckart-Young-Mirsky Theorem:

$$\widehat{S}_{\text{eym}} = \sum_{i=1}^{r} \widehat{\sigma}_i \widehat{u}_i \widehat{v}_i^H$$

- Problem solved? **No!**

Two optimization problems

- An ***observable*** optimization problem:

$$\widehat{S}_{\text{eym}} = \underset{\text{rank}(S)=r}{\arg\min} ||\widetilde{X} - S||_F,$$

- An ***unobservable*** optimization problem:

$$w^{\text{opt}} := \underset{||w||_{\ell_0}=r}{\arg\min} ||\sum_{i=1}^{r} \theta_i u_i v_i^H - \sum_i w_i \widehat{u}_i \widehat{v}_i^H ||_F.$$

- Key Insight: SVD says how to best *represent* measured matrix
- No reason to expect optimal denoising!

Oracle Denoising Solution

- An ***unobservable*** optimization problem:

$$w^{\text{opt}} := \arg\min_{||w||_{\ell_0}=r} || \sum_{i=1}^{r} \theta_i u_i v_i^H - \sum_i w_i \widehat{u}_i \widehat{v}_i^H ||_F.$$

- Oracle solution:

$$w_i^{\text{opt}} = \left(\Re\{ \sum_{j=1}^{r} \theta_j (\widehat{u}_i^H u_j)(v_j^H \widehat{v}_i) \} \right)_+ \xrightarrow{a.s.} -2 \frac{D_{\mu_X}(\rho_i)}{D'_{\mu_X}(\rho_i)} \quad \text{if } \theta_i^2 > 1/D_{\mu_X}(b^+)$$

- Optimal weight = D-transform of noise-only spectrum
- D-transform is analog of Fourier transform in NCIT!

Oracle Denoising Solution

- Oracle solution:

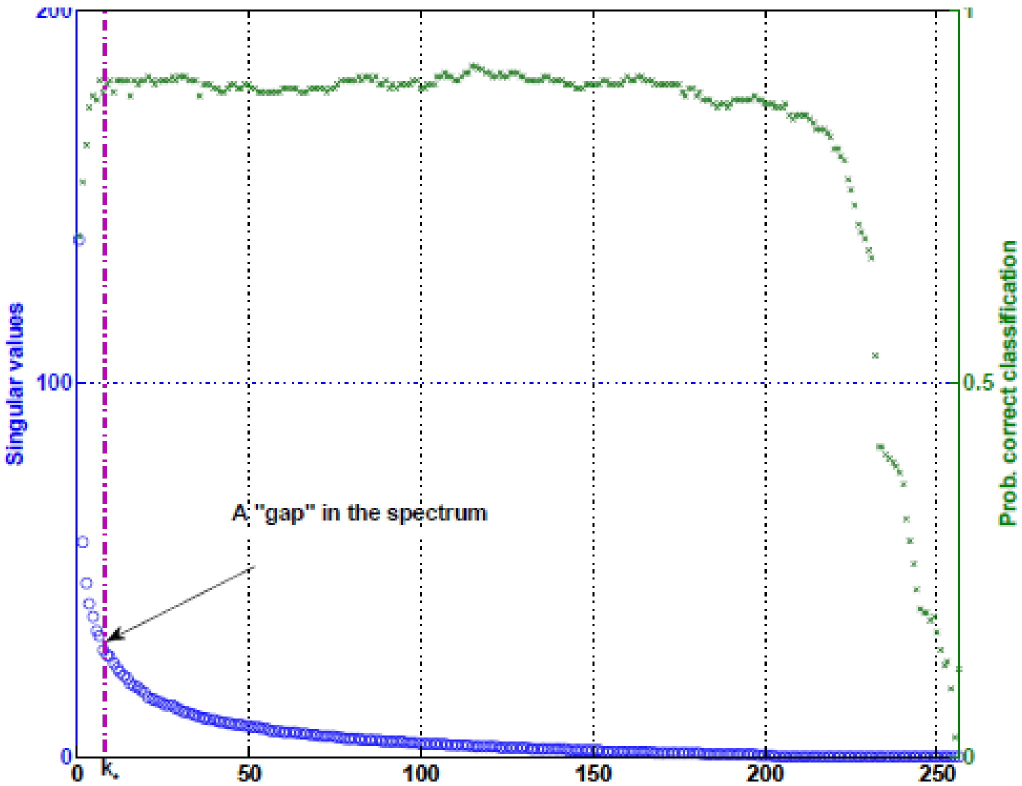
$$w_i^{\text{opt}} = \left(\Re\{ \sum_{j=1}^{r} \theta_j (\widehat{u}_i^H u_j)(v_j^H \widehat{v}_i) \} \right)_+ \xrightarrow{a.s.} -2 \frac{D_{\mu_X}(\rho_i)}{D'_{\mu_X}(\rho_i)} \qquad \text{if } \theta_i^2 > 1/D_{\mu_X}(b^+)$$

- Optimal weight = D-transform of noise-only spectrum
- D-transform is analog of Fourier transform in NCIT!

$$D_{\mu_X}(z) := \left[\int \frac{z}{z^2 - t^2} \mathrm{d}\mu_X(t) \right] \times \left[c \int \frac{z}{z^2 - t^2} \mathrm{d}\mu_X(t) + \frac{1-c}{z} \right]$$

- c = # Sensors / # Measurements = # Features / # Measurements
- D-transform is analog of Fourier transform in NCIT!

Computing D-transform from Data



$$D_{\mu_X}(z) := \left[\int \frac{z}{z^2 - t^2} \mathrm{d}\mu_X(t) \right] \times \left[c \int \frac{z}{z^2 - t^2} \mathrm{d}\mu_X(t) + \frac{1-c}{z} \right]$$

OptShrink: Data-driven singular value shrinkage

- <u>Optimal weights:</u>

$$\widehat{w}_{i,\widehat{r}}^{\text{opt}} = -2\,\frac{\widehat{D}(\widehat{\sigma}_i;\widehat{\Sigma}_{\widehat{r}})}{\widehat{D}'(\widehat{\sigma}_i;\widehat{\Sigma}_{\widehat{r}})}$$

$$\widehat{S}_{\text{opt}} = \sum_{i=1}^{\widehat{r}} \widehat{w}_{i,\widehat{r}}^{\text{opt}}\,\widehat{u}_i\,\widehat{v}_i^H$$

$$\widehat{D}(z;X) := \frac{1}{n}\operatorname{Tr}\left(z\,(z^2\,I - X X^H)^{-1}\right)\cdot\frac{1}{m}\operatorname{Tr}\left(z\,(z^2\,I - X^H X)^{-1}\right)$$

$$\widehat{D}'(z;X) := \frac{1}{n}\operatorname{Tr}\left[z\,(z^2\,I - X X^H)^{-1}\right]\cdot\frac{1}{m}\operatorname{Tr}\left[-2z^2\,(z^2\,I - X^H X)^{-2} + (z^2\,I - X^H X)^{-1}\right]$$
$$+\frac{1}{m}\operatorname{Tr}\left[z\,(z^2\,I - X^H X)^{-1}\right]\cdot\frac{1}{n}\operatorname{Tr}\left[-2z^2\,(z^2\,I - X X^H)^{-2} + (z^2\,I - X X^H)^{-1}\right].$$

http://arxiv.org/abs/1306.6042 http://web.eecs.umich.edu/~rajnrao/optshrink/

A new VoI metric

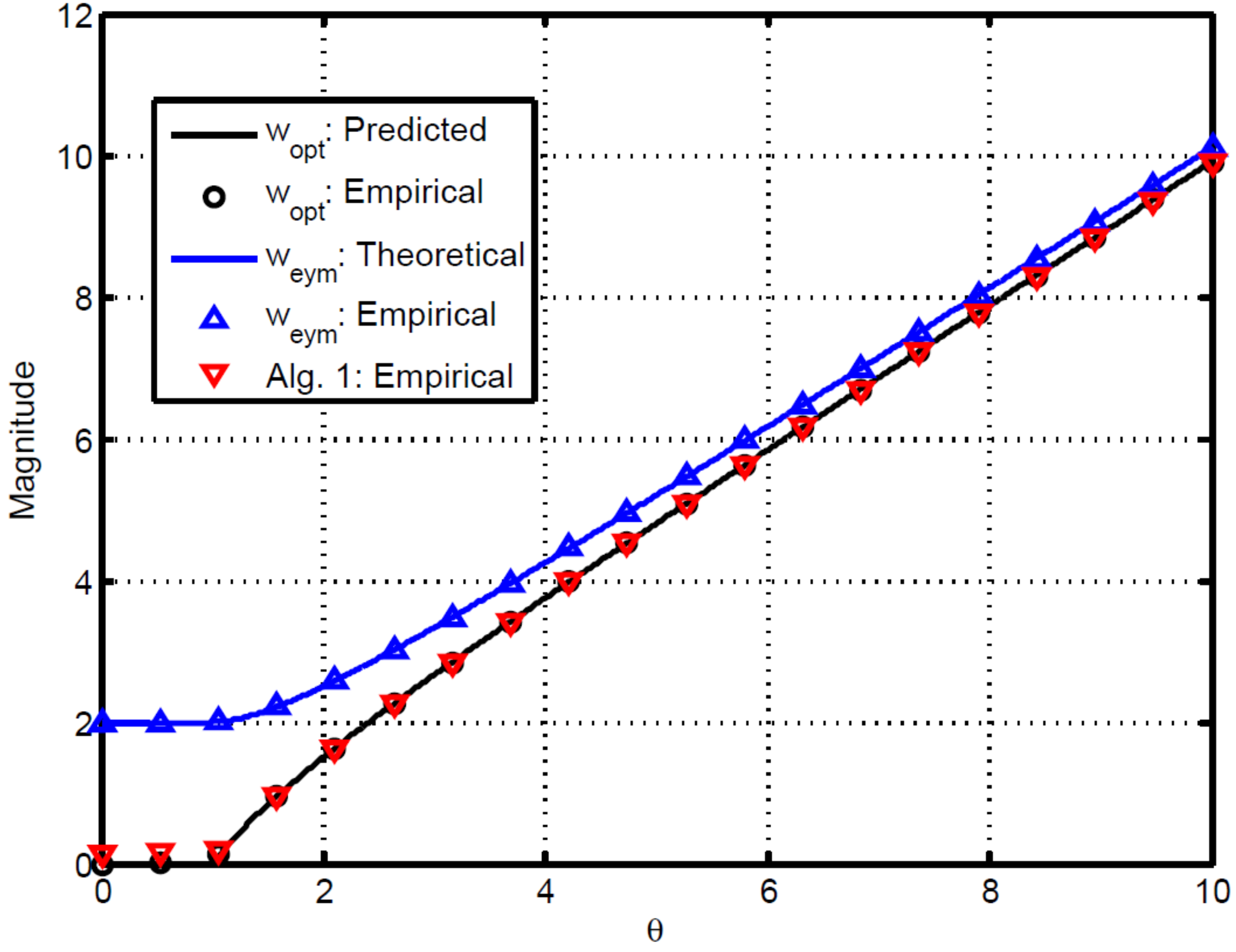
- Optimal weights:

$$\widehat{w}_{i,\widehat{r}}^{\text{opt}} = -2\,\frac{\widehat{D}(\widehat{\sigma}_i;\widehat{\Sigma}_{\widehat{r}})}{\widehat{D}'(\widehat{\sigma}_i;\widehat{\Sigma}_{\widehat{r}})} \quad \Leftrightarrow \quad \widehat{S}_{\text{opt}} = \sum_{i=1}^{\widehat{r}} \widehat{w}_{i,\widehat{r}}^{\text{opt}}\,\widehat{u}_i\,\widehat{v}_i^H$$

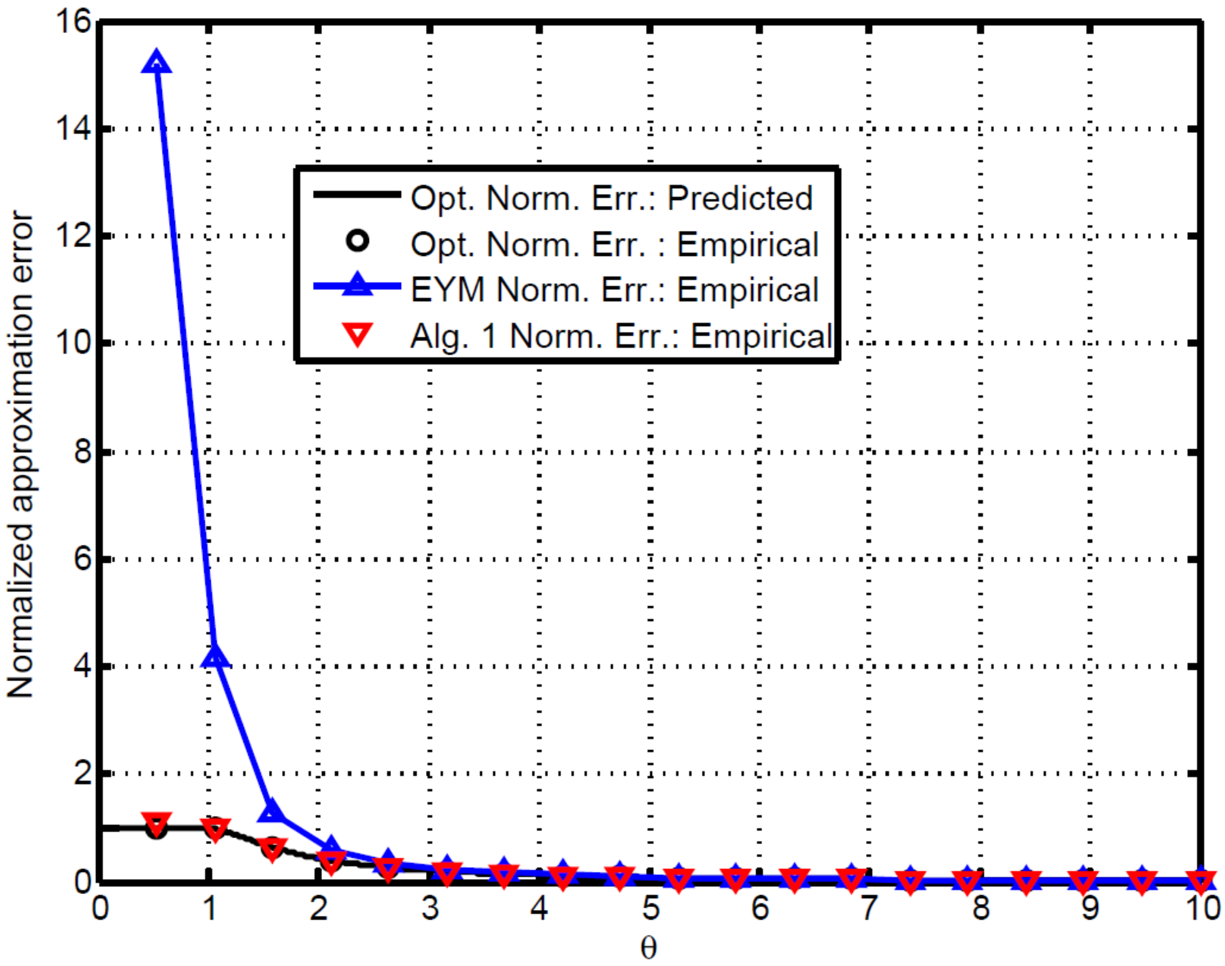
$$\text{rel}\widehat{\text{MSE}}_{\widehat{r}} = 1 - \frac{\sum_{i=1}^{\widehat{r}} (\widehat{w}_{i,\widehat{r}}^{\text{opt}})^2}{\sum_{i=1}^{\widehat{r}} \frac{1}{\widehat{D}(\widehat{\sigma}_i;\widehat{\Sigma}_{\widehat{r}})}}$$

- Data-driven estimate of relative denoising MSE!

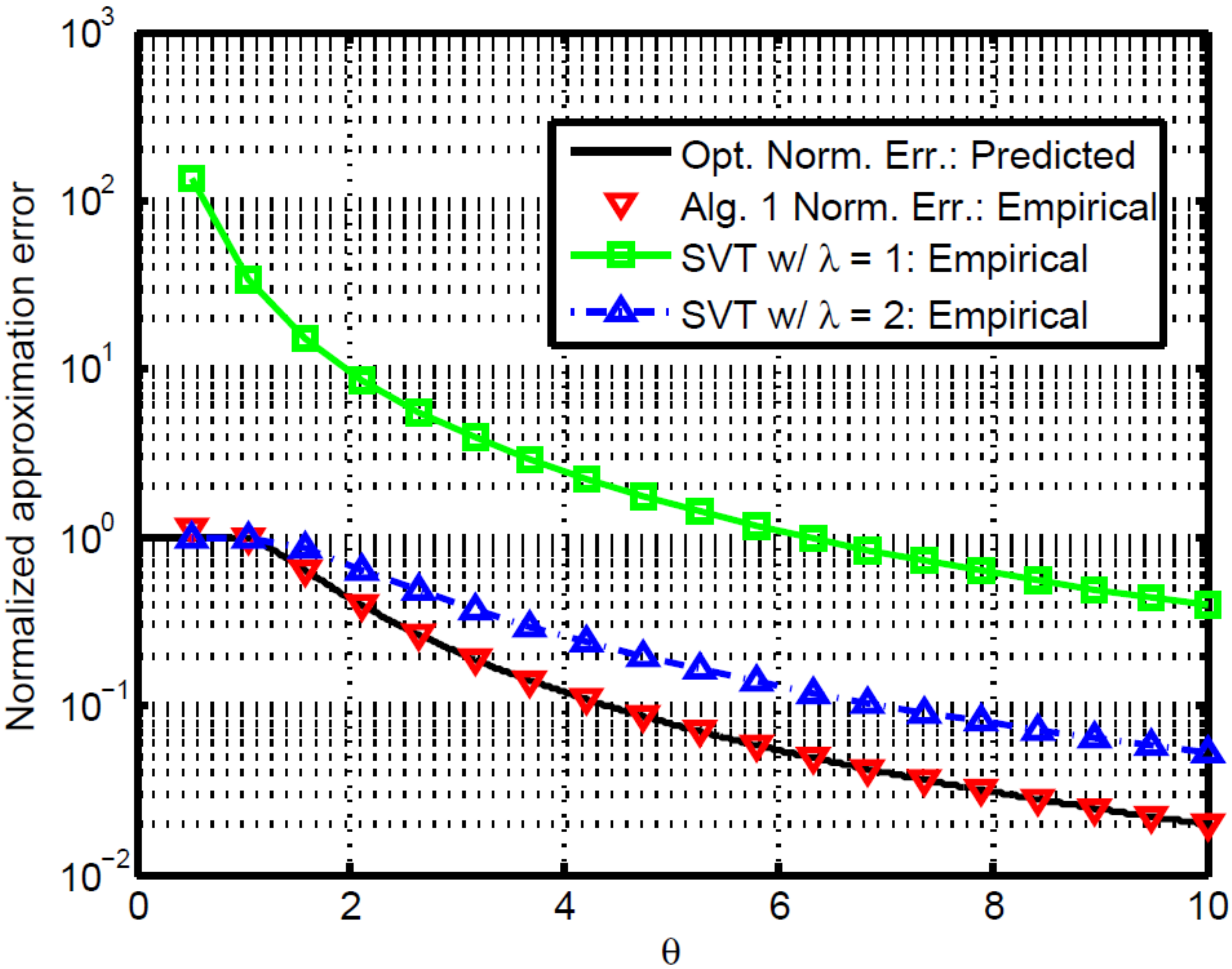
Optimal weights



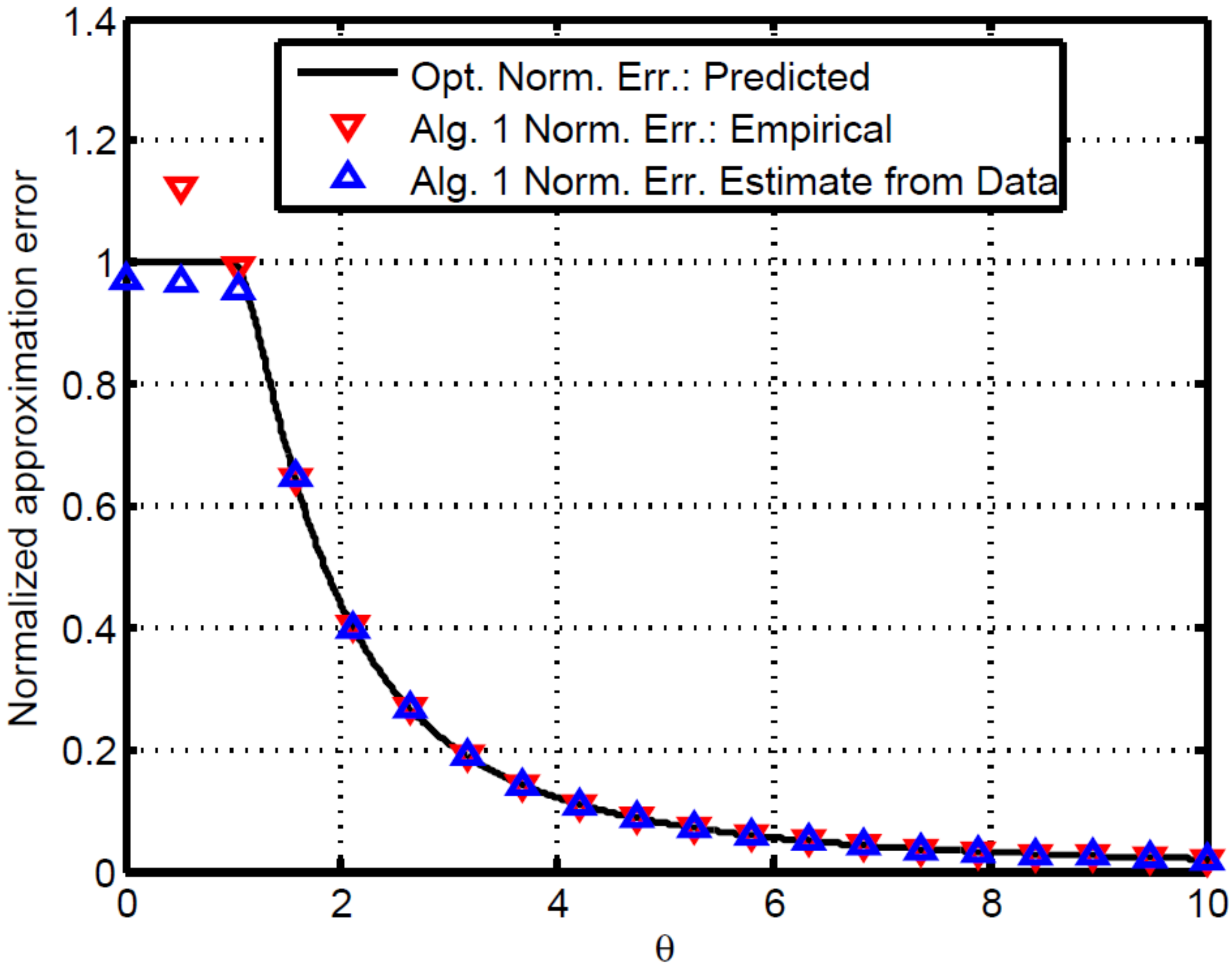
Denoising performance



Gains relative to Nuclear Norm regularization



Accuracy of new VoI metric



Summary of year 2 activities

- This year's research directly impacts
  - Information-driven learning
    - Improved denoising of low-rank signals
    - "Blackbox" type algorithm for post-processing trunc. SVD
  - Information exploitation
    - Using new VoI metric to rank quality of denoised estimate
    - Using new VoI metric to fuse different estimates

Future and ongoing focus areas and collaborations

- Tradeoffs between analyze-and-fuse versus fuse-and-analyze (UM/ASU)
- Multimodality fusion with different SNRs (UM/ASU/OSU)
- Optimal denoising in sparse plus low-rank plus noise type matrices (UM/MIT)

Publications in Y2

- B. Farrell and R. Nadakuditi,"Local spectrum of truncations of Kronecker products of Haar distributed unitary matrices," Under review
- R. Nadakuditi, "OptShrink: An algorithm for improved low-rank signal matrix denoising by optimal, data-driven singular value shrinkage,“ Under review
- R. Nadakuditi, "When are the most informative components for inference also the principal components,“ Under review
- R. Nadakuditi and M. Newman, "Spectra of random graphs with arbitrary expected degrees," Phys. Rev. E 87, 012803 (2013)

- Poster presentations today by
  - Nick Asendorf – “Informative versus useful components”
  - Raj Tejas Suryaprakash – “DOA performance of MUSIC with missing data”