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

Co-PI Raj Rao
University of Michigan
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

• **Signal-plus-Noise Model:**

\[ \tilde{X} = \sum_{i=1}^{r} \theta_i u_i v_i^H + X \]

• **Objective:** Optimally denoise \( S = \sum_{i=1}^{r} \theta_i u_i v_i^H \) (assume \( r \) known)

• **(Minimal) Assumptions:**
  – No structure on \( S \) (besides low-rank)
  – Noise-only matrix \( X \) has isotropically random singular vectors
The “optimality” of the truncated SVD

- Optimization problem:

\[
\hat{S}_{\text{eym}} = \arg \min_{\text{rank}(S)=r} \| \tilde{X} - S \|_F,
\]

- Eckart-Young-Mirsky Theorem:

\[
\hat{S}_{\text{eym}} = \sum_{i=1}^{r} \hat{\sigma}_i \hat{u}_i \hat{u}_i^H
\]

- Problem solved? No!
Two optimization problems

• **An observable** optimization problem:

\[
\hat{S}_{\text{eym}} = \arg \min_{\text{rank}(S)=r} \|\hat{X} - S\|_F,
\]

• **An unobservable** optimization problem:

\[
w^{\text{opt}} := \arg \min \| \sum_{i=1}^{r} \theta_i u_i v_i^H - \sum_i w_i \hat{u}_i \hat{v}_i^H \|_F.
\]

• **Key Insight**: SVD says how to best *represent* measured matrix

• No reason to expect optimal denoising!
An *unobservable* optimization problem:

\[
\begin{align*}
 w^\text{opt} &:= \arg \min \| \sum_{i=1}^r \theta_i u_i v_i^H - \sum_i w_i \hat{u}_i \hat{v}_i^H \|_F \\
 \|w\|_{\ell_0} &= r
\end{align*}
\]

**Oracle solution:**

\[
 w^\text{opt}_i = \left( \mathbb{R} \left\{ \sum_{j=1}^r \theta_j (\hat{w}_j^H u_j) (\nu_j^H \hat{v}_i) \right\} \right) + \frac{a.s.}{-2} \frac{D_{\mu_X}(\rho_i)}{D'_{\mu_X}(\rho_i)} \quad \text{if } \theta_i^2 > 1/D_{\mu_X}(b^+)
\]

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

- **Oracle solution:**

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

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

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

- \(c = \frac{\# \text{ Sensors}}{\# \text{ Measurements}} = \frac{\# \text{ Features}}{\# \text{ Measurements}}\)
- D-transform is analog of Fourier transform in NCIT!
Computing D-transform from Data

\[ D_{\mu_X}(z) := \left[ \int \frac{z}{z^2 - t^2} d\mu_X(t) \right] \times \left[ c \int \frac{z}{z^2 - t^2} d\mu_X(t) + \frac{1 - c}{z} \right] \]
OptShrink: Data-driven singular value shrinkage

- Optimal weights:

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

\[
\hat{S}_{\text{opt}} = \sum_{i=1}^{\hat{r}} \hat{w}_{i,\hat{r}}^{\text{opt}} \hat{u}_i \hat{u}_i^H
\]

\[
\hat{D}(z; X) := \frac{1}{n} \text{Tr} \left( z \left( z^2 I - XX^H \right)^{-1} \right) \cdot \frac{1}{m} \text{Tr} \left( z \left( z^2 I - X^HX \right)^{-1} \right)
\]

\[
\hat{D}'(z; X) := \frac{1}{n} \text{Tr} \left[ z \left( z^2 I - XX^H \right)^{-1} \right] \cdot \frac{1}{m} \text{Tr} \left[ -2z^2 \left( z^2 I - X^HX \right)^{-2} + \left( z^2 I - X^HX \right)^{-1} \right] \\
+ \frac{1}{m} \text{Tr} \left[ z \left( z^2 I - X^HX \right)^{-1} \right] \cdot \frac{1}{n} \text{Tr} \left[ -2z^2 \left( z^2 I - XX^H \right)^{-2} + \left( z^2 I - XX^H \right)^{-1} \right].
\]

A new VOI metric

- **Optimal weights:**

\[
\hat{w}_{i,\hat{r}}^{\text{opt}} = -2 \frac{\hat{D}(\hat{\sigma}_i; \hat{\Sigma}_r)}{\hat{D}'(\hat{\sigma}_i; \hat{\Sigma}_r)}
\]

\[
\hat{S}_{\text{opt}} = \sum_{i=1}^{\hat{r}} \hat{w}_{i,\hat{r}}^{\text{opt}} \hat{u}_i \hat{v}_i^H
\]

\[
\text{relMSE}_{\hat{r}} = 1 - \frac{\sum_{i=1}^{\hat{r}} (\hat{w}_{i,\hat{r}}^{\text{opt}})^2}{\sum_{i=1}^{\hat{r}} \frac{1}{\hat{D}(\hat{\sigma}_i; \hat{\Sigma}_r)}}
\]

- Data-driven estimate of relative denoising MSE!
Optimal weights
Denoising performance
Gains relative to Nuclear Norm regularization
Accuracy of new VoI metric

Normalized approximation error vs. \( \theta \)

Graph showing:
- **Alg. 1 Norm. Err.: Empirical**
- **Alg. 1 Norm. Err. Estimate from Data**

Legend:
- Triangle for empirical error
- Inverted triangle for predicted error
- Diamond for estimate error
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, "When are the most informative components for inference also the principal components," Under review

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