ELECTRICAL ENGINEERING AND COMPUTER SCIENCE
UNIVERSITY of MICHIGAN

Improved Robust PCA Using Low-Rank Denoising with Optimal Singular Value Shrinkage

Brian E. Moore, Raj Rao Nadakuditi, and Jeffrey A. Fessler

Signal-plus-noise-plus-outliers model

$$\widetilde{X} = L + S + X$$

- $\widetilde{X} = n \times m$ signal-plus-noise-plus-outliers matrix
- L = low-rank signal matrix

$$L = \sum_{i=1}^{r} \theta_i u_i v_i^H$$

- S is a sparse matrix of *outliers* modeled as

$$S_{ij} = \begin{cases} q_{ij} & \text{with probability } p \\ 0 & \text{with probability } 1-p, \end{cases}$$

- q_{ij} drawn from *unknown* distribution $q(s)$ with

$$\mathbb{E}[q] = 0 \qquad \mathbb{E}[q]^2 = \sigma_q^2 \qquad \mathbb{E}[q]^4 < \infty,$$

- X = noise-only matrix having i.i.d. elements with

$$\mathbb{E}[X_{ij}] = 0 \qquad \mathbb{E}[X_{ij}^2] = \sigma^2/m \qquad \mathbb{E}[X_{ij}^4] < \infty$$

Problem Statement

- Objective: Given $\widetilde{X}$, estimate L as accurately as possible
- Applications:
  - Foreground/background separation
  - Dynamic MR imaging

Need for Robust PCA

- Assume L is *incoherent* with Euclidean basis. Let

$$\widetilde{X} = \sum_i \widetilde{\sigma}_i \widetilde{u}_i \widetilde{v}_i^H,$$

Theorem: As $n/m_n \to c \in (0, \infty)$,

$$|\langle u_i, \widetilde{u}_i \rangle|^2 \xrightarrow{a.s.} \begin{cases} 1 - \dfrac{c\,(1+\bar{\theta}_i^2)}{\bar{\theta}_i^2\,(\bar{\theta}_i^2 + c)} & \text{if } \bar{\theta}_i > c^{1/4} \\ 0 & \text{otherwise,} \end{cases}$$

and

$$|\langle v_i, \widetilde{v}_i \rangle|^2 \xrightarrow{a.s.} \begin{cases} 1 - \dfrac{(c+\bar{\theta}_i^2)}{\bar{\theta}_i^2\,(\bar{\theta}_i^2 + 1)} & \text{if } \bar{\theta}_i > c^{1/4} \\ 0 & \text{otherwise,} \end{cases}$$

where

$$\bar{\theta}_i = \lim_{m \to \infty} \frac{\theta_i}{\sqrt{p\,m\,\sigma_q^2 + \sigma^2}}$$

- Insight:
  - If $\sigma_q^2 = O(1)$ and $pm \to \infty$, then $\bar{\theta}_i \to 0$
  - $\Rightarrow p \geq O(\log m/m)$ will "break" PCA

Nuclear-norm based Robust PCA

- Solves the optimization problem

$$\{\widehat{L}_{\text{svt}}, \widehat{S}_{\text{svt}}\} = \arg\min_{L,S} \frac{1}{2}\|\widetilde{X} - (L+S)\|_F^2 + \lambda_L \|L\|_\star + \lambda_S \|S\|_1,$$

via alternating minimization

$$\begin{aligned} L_{k,\text{svt}} &= \textbf{SVT}_{\tau_k \lambda_L}(M_{k-1,\text{svt}} - S_{k-1,\text{svt}}) \\ S_{k,\text{svt}} &= \textbf{soft}_{\tau_k \lambda_S}(M_{k-1,\text{svt}} - L_{k-1,\text{svt}}) \\ M_{k,\text{svt}} &= L_{k,\text{svt}} + S_{k,\text{svt}} - \tau_k(L_{k,\text{svt}} + S_{k,\text{svt}} - \widetilde{X}), \end{aligned}$$

where **SVT**(.) is the *singular value thresholding* operator

$$\textbf{SVT}_\lambda(Y) = \sum_{i=1}^{n} (\sigma_i - \lambda)_+ u_i v_i^H$$

- Disadvantages:
  - How to select parameters λ_L, λ_S, and τ_k?
  - Provably suboptimal low-rank matrix recovery

OptShrink: Low-rank recovery by optimal shrinkage

- Consider the oracle low-rank denoising problem

$$w^{\text{opt}} = \underset{[w_1,\ldots,w_r]^T \in \mathbb{R}^r}{\text{argmin}} \|L - \textstyle\sum_{i=1}^{r} w_i \widetilde{u}_i \widetilde{v}_i^T\|_F^2$$

- Under Theorem hypotheses and when $S = 0$, we have

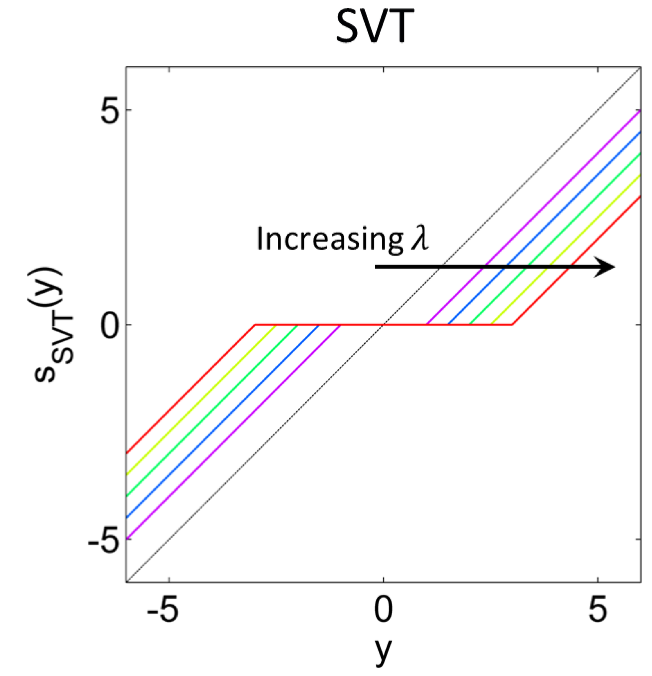
$$w_i^{\text{opt}} = \Re_+ \left\{ \sum_{j=1}^{r} \theta_j \left(\widetilde{u}_i^H u_j\right) \cdot \left(v_j^H \widetilde{v}_i\right) \right\} \xrightarrow{a.s.} -2\,\frac{D_{\mu_{\widetilde{X}}}(\widetilde{\sigma}_i)}{D'_{\mu_{\widetilde{X}}}(\widetilde{\sigma}_i)},$$

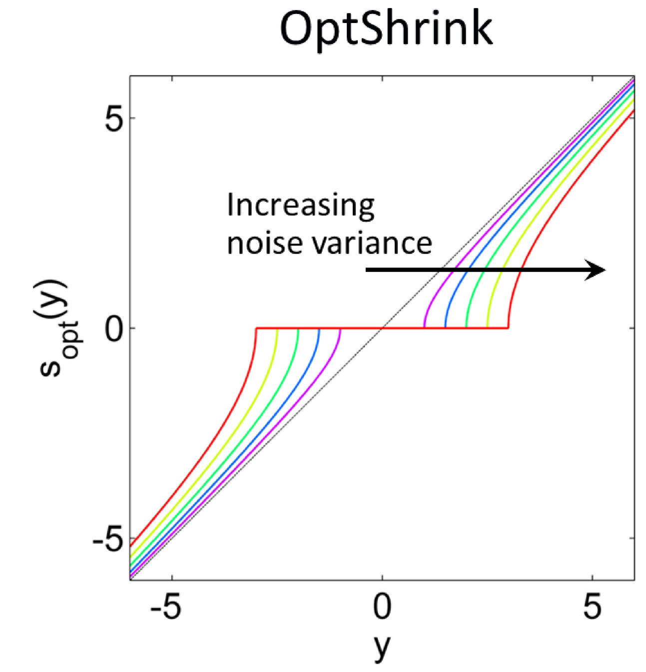
where $q = \min(m,n)$, and

$$D_{\mu_{\widetilde{X}}}(z) = \left[\frac{1}{q-r}\sum_{i=r+1}^{q} \frac{z}{z^2 - \widetilde{\sigma}_i^2}\right] \times \left[\frac{1-c}{z} + \frac{c}{q-r}\sum_{i=r+1}^{q} \frac{z}{z^2 - \widetilde{\sigma}_i^2}\right]$$

- D-transform:
  - Matrix analogue of log-Fourier transform
- Insights:
  - Optimal shrinkage is *inversely related* to *eigen-SNR*

Why is OptShrink better than ℓ_1 type shrinkage?





Proposed Approach

- Replace SVT with OptShrink:

$$\begin{aligned} L_k &= \textbf{OptShrink}_r(M_{k-1} - S_{k-1}) \\ S_k &= \textbf{soft}_{\tau_k \lambda_S}(M_{k-1} - L_{k-1}) \\ M_k &= L_k + S_k - \tau_k(L_k + S_k - \widetilde{X}), \end{aligned}$$

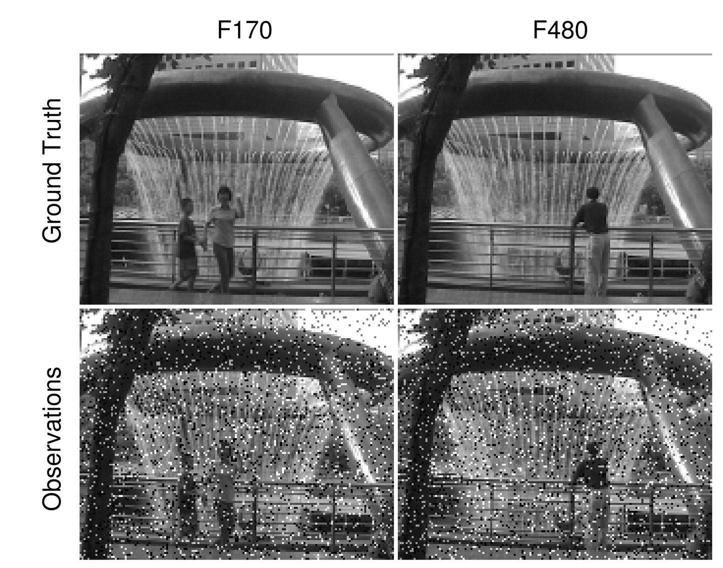
where

$$\textbf{OptShrink}_r(Y) = \sum_{i=1}^{r} \left(-2\,\frac{D_{\mu_Y}(\sigma_i)}{D'_{\mu_Y}(\sigma_i)}\right) u_i v_i^H$$

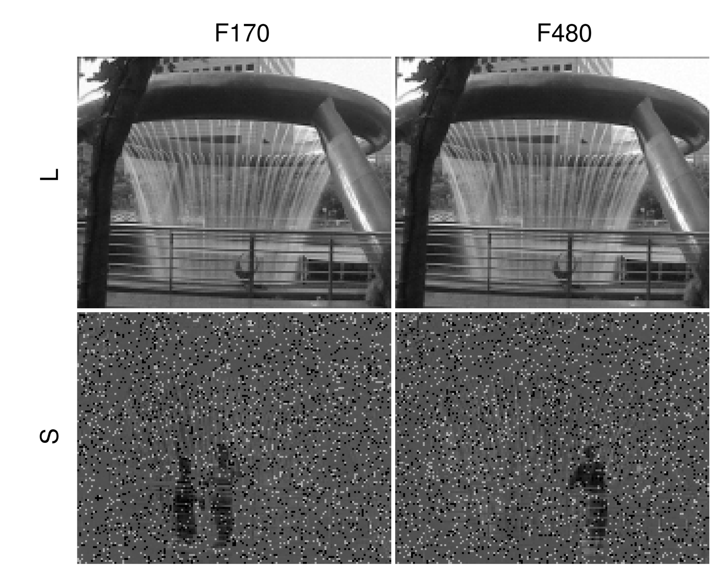
- Advantages:
  - *Explicit* rank regularization (cf. implicit SVT)
  - OptShrink is optimal low-rank denoiser
  - Small λ_L does not denoise
  - Large λ_L shrinks strong subspaces and leaves artifact

Results: Background subtraction task

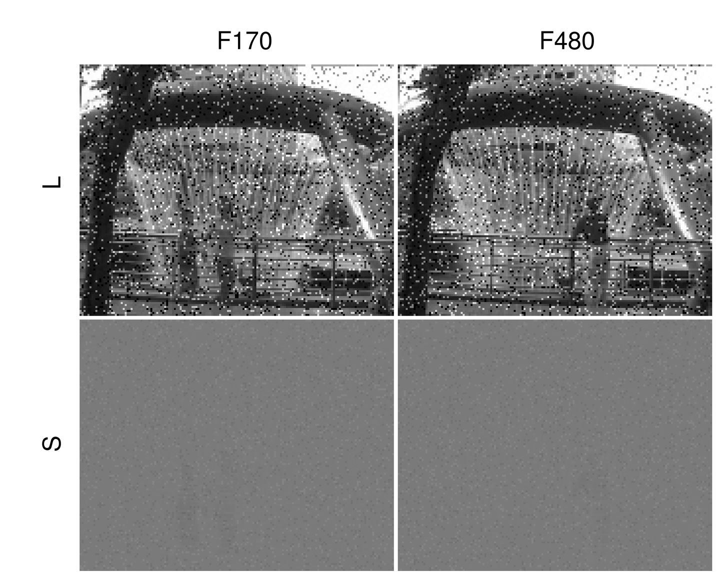
- Random saturations as outliers



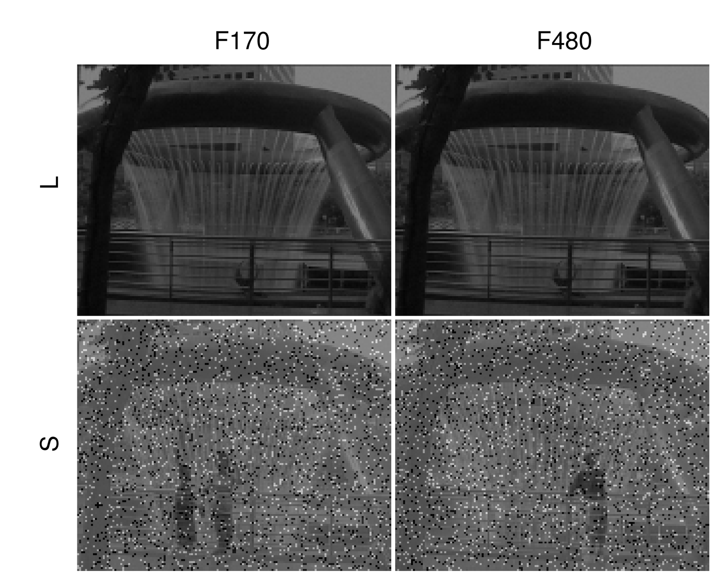
Top: ground truth. Bottom: $\widetilde{X}$



Proposed approach with $r = 1$



RPCA with $\lambda_L = 10$



RPCA with $\lambda_L = 300$

References

- R. R. Nadakuditi, "OptShrink: An algorithm for improved low-rank signal matrix denoising by optimal, data-driven singular value shrinkage," *IEEE Trans. of IT*, May 2014.
- B. Moore, R. R. Nadakuditi, and J. Fessler, "Dynamic MRI reconstruction using low-rank plus sparse model with optimal rank regularized eigen-shrinkage," *ISMRM*, May 2014.
- R. Otazo, E. J. Candès, and D. K. Sodickson, "Low-rank and sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components," *Magn. Reson. Med.*, 2014.

Acknowledgements: Work supported by ONR YIP Award N000141110660, AFOSR YI Award FA9550-12-1-0266, NSF award CCF-1116115, ARO MURI grant W911NF-11-1-0391, and NIH grant 1P01-CA87634.