Robust M-Estimation Based Matrix Completion

Michael Muma, Wen-Jun Zeng, Abdelhak M. Zoubir

Signal Processing Group - Technische Universität Darmstadt - Email: {muma,zoubir}@spg.tu-darmstadt.de, wenjzeng@gmail.com

TECHNISCHE UNIVERSITÄT DARMSTADT

Signal Model and Motivation

Signal model: Observed matrix $\boldsymbol{X} \in \mathbb{R}^{n_1 \times n_2}$ modeled as

$$\boldsymbol{X} = \boldsymbol{M} + \boldsymbol{S} + \boldsymbol{N} \tag{1}$$

- $\boldsymbol{M}$: low-rank matrix of rank r,
- $\boldsymbol{S}$: column or entry-wise sparse outlier matrix
- $\boldsymbol{N}$: (impulsive) background noise

Goal: Recover the low-rank component $\boldsymbol{M}$ from partially observed entries of $\boldsymbol{X}$ corrupted by noise and outliers.

Applications: recommender systems, computer vision, image inpainting, biomedicine, information retrieval

Existing Robust Matrix Completion Approaches

Robust ℓ_p-loss based methods [1]:

⊕ robust and computationally efficient

⊖ statistically inefficient with respect to additional background noise

⊖ easily get stuck at an inferior solution (nonsmooth objective function)

Nuclear norm regularization of Huber's loss function approach [2]:

⊖ requires SVD at each iteration and has a high complexity

Proposed Robust M-Estimation Based Approach

Outlier-robust "norm" of $\boldsymbol{X}$ is defined as

$$\|\boldsymbol{X}\|_{\sigma,c} = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} \rho\left(\frac{x_{ij}}{\sigma}\right) \tag{2}$$

- $\sigma > 0$: scale parameter
- x_{ij}: (i,j)th entry of $\boldsymbol{X}$
- $\rho(\cdot)$: differentiable loss function, e.g.

Huber's

$$\rho_{\text{hub}}(x) = \begin{cases} \frac{1}{2}x^2, & |x| \le c \\ c|x| - \frac{1}{2}c^2, & |x| > c \end{cases}$$

Tukey's

$$\rho_{\text{tuk}}(x) = \begin{cases} \frac{1}{2}x^2 - \frac{x^4}{2c^2} + \frac{x^6}{6c^4}, & |x| \le c \\ \frac{c^2}{6}, & |x| > c \end{cases}$$

- c: tuning parameter trades off the efficiency and robustness.

Proposed robust *M*-estimation based matrix completion:

$$\min_{\boldsymbol{U},\boldsymbol{V}} \|(\boldsymbol{U}\boldsymbol{V})_\Omega - \boldsymbol{X}_\Omega\|_{\sigma,c} \tag{3}$$

- Computationally efficient direct matrix factorization $\widehat{\boldsymbol{M}} = \boldsymbol{U}\boldsymbol{V}$, where $\boldsymbol{U} \in \mathbb{R}^{n_1 \times r}$ and $\boldsymbol{V} \in \mathbb{R}^{r \times n_2}$ to make the estimate $\widehat{\boldsymbol{M}}$ low-rank
- $(\boldsymbol{X}_\Omega)_{ij} = 0$ if $(i,j) \notin \Omega$ and $(\boldsymbol{X}_\Omega)_{ij} = x_{ij}$ if $(i,j) \in \Omega$.
- σ: unknown and is estimated jointly with $(\boldsymbol{U},\boldsymbol{V})$
- c: constant that is set in advance

Algorithms

Algorithm 1: Huber's *M*-estimator

Input: $\boldsymbol{X}_\Omega$, Ω, and rank r

Initialize: Randomly initialize $\boldsymbol{U}^0 \in \mathbb{R}^{n_1 \times r}$

Determine $\{\mathcal{I}_j\}_{j=1}^{n_2}$ and $\{\mathcal{J}_i\}_{i=1}^{n_1}$ according to Ω.

for $k = 0, 1, \cdots$ **do**

// Fix $\boldsymbol{U}^k$, optimize $\boldsymbol{V}$

$$\boldsymbol{v}_j^{k+1} = \arg\min_{\boldsymbol{v}_j,\sigma} \left\{ \sigma \sum_{i \in \mathcal{I}_j} \rho_{\text{hub}}\left(\frac{x_{ij} - (\boldsymbol{u}_i^\top)^k \boldsymbol{v}_j}{\sigma}\right) + |\mathcal{I}_j|(\alpha\sigma) \right\}$$

for all $j = 1, 2, \cdots, n_2$.

// Fix $\boldsymbol{V}^{k+1}$, optimize $\boldsymbol{U}$

$$(\boldsymbol{u}_i^\top)^{k+1} = \arg\min_{\boldsymbol{u}_i^\top,\sigma} \left\{ \sigma \sum_{j \in \mathcal{J}_i} \rho_{\text{hub}}\left(\frac{x_{ij} - \boldsymbol{u}_i^\top \boldsymbol{v}_j^{k+1}}{\sigma}\right) + |\mathcal{J}_i|(\alpha\sigma) \right\}$$

for all $i = 1, 2, \cdots, n_1$.

Stop if a termination condition is satisfied.

end for

Output: $\widehat{\boldsymbol{M}} = \boldsymbol{U}^{k+1}\boldsymbol{V}^{k+1}$

⊕ Per-iteration complexity of *M*-estimation based matrix completion using Huber's loss: $\mathcal{O}(|\Omega| r^2)$. → attractive tool for the "big data" setting.

⊕ guaranteed convergence to stationary point

Theorem *The sequence generated by Algorithm 1, i.e., $\{\boldsymbol{U}^k, \boldsymbol{V}^k\}$, converges to a stationary point of the nonconvex problem of* (3).

A proof is provided in the paper.

Algorithm 2: Tukey's *M*-estimator The estimate obtained from Algorithm 1 is used as starting point for Tukey's *M*-estimator, which solves

$$\min_{\boldsymbol{v}_j} L_{\text{tuk}}(\boldsymbol{v}_j, \sigma) \triangleq \sum_{i \in \mathcal{I}_j} \rho_{\text{tuk}}\left(\frac{x_{ij} - (\boldsymbol{u}_i^\top)^k \boldsymbol{v}_j}{\sigma_{\text{hub}}}\right) \tag{4}$$

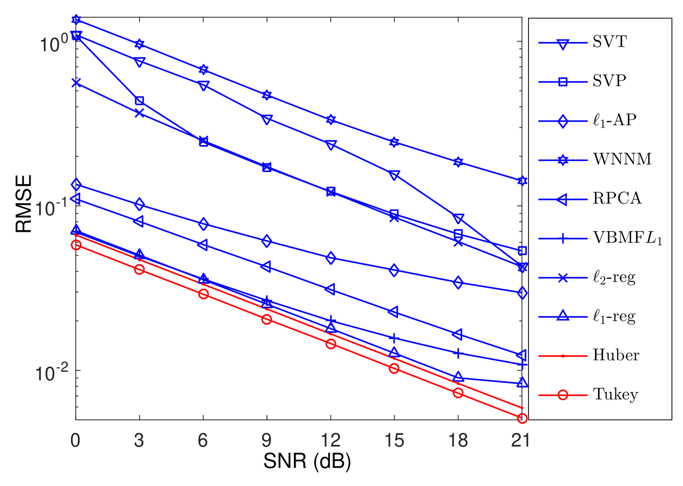
using an iteratively reweighted least-squares (IRWLS) algorithm.

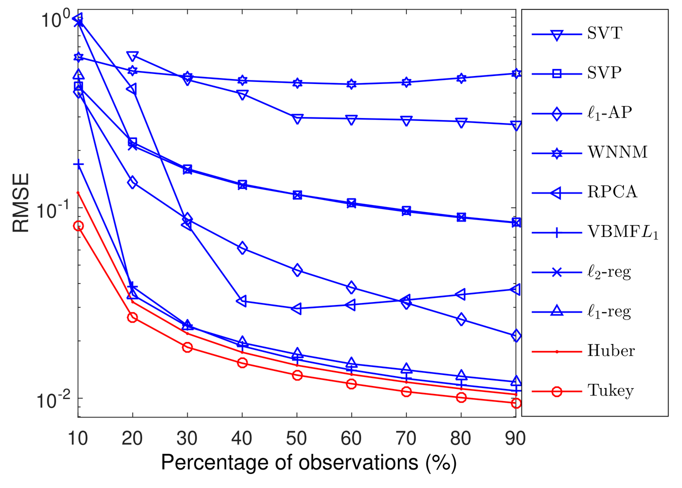
Download Matlab Robust Signal Processing Toolbox [3]:

https://github.com/RobustSP/

Results

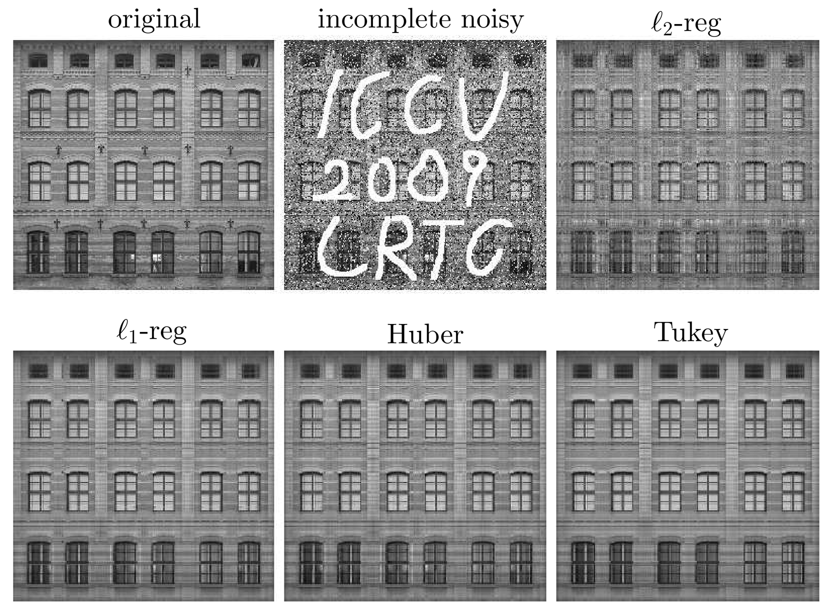
Results for synthetic random data:





- $n_1 = 150$, $n_2 = 300$, and $r = 10$.
- $\boldsymbol{M} = \boldsymbol{X}_1\boldsymbol{X}_2$ where $\boldsymbol{X}_1 \in \mathbb{R}^{n_1 \times r}$ and $\boldsymbol{X}_2 \in \mathbb{R}^{r \times n_2}$ are Gaussian random matrices.
- $\boldsymbol{N}$: impulsive Gaussian mixture model (GMM) noise

Image inpainting in salt-and-pepper noise:



Peak Signal-to-Noise Ratio (PSNR) in dB at SNR = 6 dB	
baseline	10.83
ℓ_2-regression	19.18
ℓ_1-regression	21.61
proposed Huber's *M*	23.29
proposed Tukey's *M*	23.71

References

[1] W.-J. Zeng and H. C. So, "Outlier-robust matrix completion via ℓ_p-minimization," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1125–1140, Mar. 2018.

[2] A. Elsener and S. van de Geer, "Robust low-rank matrix estimation," *Ann. Stat.*, vol. 46, no. 6, pp. 3481–3509, 2018.

[3] A. M. Zoubir, V. Koivunen, E. Ollila, and M. Muma, *Robust Statistics for Signal Processing.* Cambridge University Press, Cambridge, UK, 2018.