NEURIPS 2026

Robust Noisy Inductive Matrix Completion with Local Linear Convergence

Xingcai Zhou[1,*], **Xin Dong**[1], **Linglong Kong**[2]

[1]School of Statistics and Data Science, Nanjing Audit University
[2]Department of Mathematical and Statistical Sciences, University of Alberta

Motivation & Challenges

Inductive Matrix Completion

Reconstruct high-dimensional matrices using auxiliary side information:

$$L^\star = X_L M^\star X_R^\top = X_L U^\star V^{\star\top} X_R^\top$$

- **Cold-Start Solution**: Predicts for unseen entities via features.
- **Low-Rank Core**: Reduces sample complexity via feature dimension.

Challenge: Heavy-Tailed Noise

Real-world observations are often corrupted by severe perturbations:

- **Adversarial Outliers**: Bot attacks, anomalies, and extreme behaviors.
- l_2**-Loss Sensitivity**: Standard least squares diverges under heavy tails.
- **Landscape Coupling**: Features distort the optimization space.

Core Question: *How to recover low-rank matrices from sparse observations under heavy-tailed noise with provable linear convergence?*

Problem Formulation

Optimization Objective

$$\min_{U,V} \; f_\Omega(U,V) := \frac{1}{2p} \sum_{(i,j)\in\Omega} \rho_\tau \left(x_{L,i}^\top U V^\top x_{R,j} - Y_{ij} \right) + \frac{\lambda}{8} \left\| U^\top U - V^\top V \right\|_F^2$$

Observation Model

Target matrix $L^\star = X_L M^\star X_R^\top$ is sampled via Bernoulli model:

$$Y_{ij} = L_{ij}^\star + \xi_{ij}, \quad (i,j) \in \Omega$$

- **Random Sampling**: Observed with probability $p \in (0,1]$.
- **Side Information**: Features X_L, X_R project data to low rank.

Heavy-Tailed Noise ξ_{ij}

Accommodating severe non-Gaussian perturbations:

- **Bounded Variance**: $\mathbb{E}[\xi_{ij}] = 0, \mathbb{E}[\xi_{ij}^2] \le \sigma^2$.
- **Outlier Resilience**: Handles Student's t or GMM contaminants.
- **Beyond Gaussian**: Overcomes l_2-loss failure.

Main Contributions

01 Robust Method

Adaptive Huber Loss

Accommodates heavy-tailed noise and eliminates extreme residual spikes.

Two-Stage RGDIMC

- Truncation initialization
- Huber gradient descent
- Balanced factorization

02 Convergence Theory

Local Linear Rate

Converges geometrically toward ground truth down to statistical floor.

Low Sample Complexity

- Linear in feature dimension n
- Up to $\log n$ factors
- Logarithmic in ambient d

03 Statistical & Empirical

Minimax Optimality

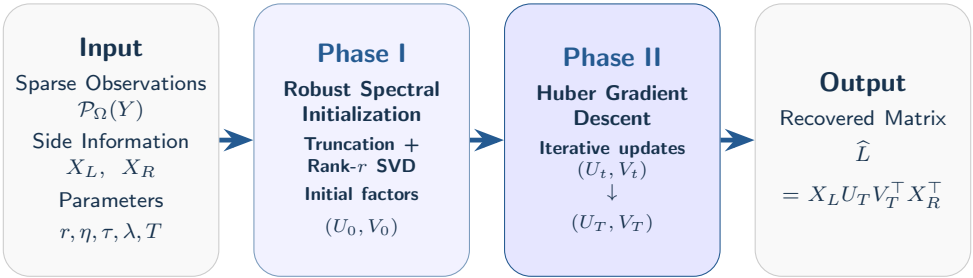
$$\mathcal{O}\left(\sigma\sqrt{\frac{rn}{p}}\right)$$

Empirical Validation

Superior robustness & $100\times$ speedup on benchmark tasks.

Key Takeaway: *RGDIMC achieves robustness, linear convergence, and minimax optimality in a unified inductive completion framework.*

RGDIMC Framework



Core Design Principle: Robust truncation controls extreme observations; Huber-based gradient descent refines the factors and achieves local linear convergence under suitable conditions.

Theoretical Guarantees

Theorem 3.4: Robust Spectral Initialization

Under sample complexity $p \geq \mathcal{O}\left(\frac{\tau^2 \mu_1^2 n_1 n_2 \log d}{(n_1+n_2)\sigma_r^{\star 2} d_1 d_2}\right)$, Phase 1 initialization satisfies with high probability:

$$\|U_0 V_0^\top - M^\star\|_2 \leq \gamma \frac{\sigma_r^\star}{\sqrt{r}} + C_1 \sqrt{\frac{(n_1+n_2)\sigma^2 \log d}{p}} + C_2 \tau \sqrt{\frac{(n_1+n_2)\log d}{p}}$$

Theorem 3.6: Local Linear Rate

Iterates decay geometrically to the statistical error floor:

$$\mathcal{D}_{t+1}^2 \leq \rho \mathcal{D}_t^2 + \mathcal{C}_{\text{stat}}$$

- **Contraction Factor**: $\rho = 1 - \mathcal{O}(\eta \sigma_r^{\star 2}) \in (0,1)$.
- **Statistical Floor $\mathcal{C}_{\text{stat}}$:**

$$\mathcal{C}_{\text{stat}} = \frac{C \sigma_1^* \sigma^2 r(n_1+n_2)\log d \cdot \eta^2}{p}$$

Corollary 3.8: Minimax Optimality

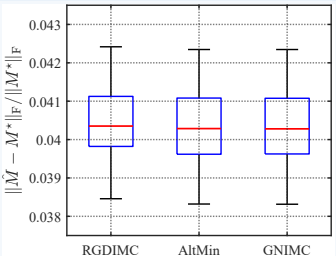
Under step size $\eta = c_\eta / \sigma_1^\star$, estimation error satisfies:

$$\|\hat{L} - L^\star\|_F \leq \tilde{\mathcal{O}}\left(\sigma \sqrt{\frac{r(n_1+n_2)}{p}}\right)$$

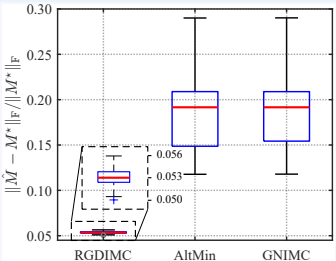
- **Optimal Bound**: Matches standard MC statistical rate under sub-Gaussian noise.
- **Relaxed Condition**: Depends only on bounded 2nd moment.

Simulation Experiments

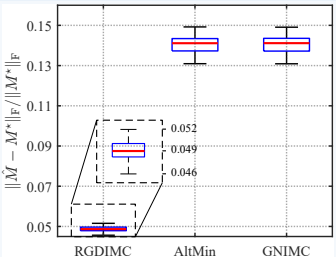
Gaussian Noise



Student's $t(3)$ Noise



GMM Contamination



Additional Experimental Insights & Robustness Analysis

Minimax Scaling & Baseline Superiority:

- **Geometric Decay**: RGDIMC converges linearly down to $\mathcal{C}_{\text{stat}}$ across all noise regimes.
- **Performance Gain**: Significantly outperforms AltMin & GNIMC under heavy tails; avoids extreme outlier divergence.

Parameter Stability & Efficiency:

- **Huber Threshold (τ)**: Maintains a robust U-shaped error basin over two orders of magnitude ($\tau \approx 3$).
- **Speedup**: RGDIMC achieves up to $\sim 100\times$ faster CPU runtime compared to convex relaxation.

Real-World Experiments

01 MovieLens-1M Recommendation ($d_1 = 6040, d_2 = 3952$)

Low-Rank & Heavy-Tails Verification:

- Scree plot confirms underlying low-rank preference structure.
- QQ-plot of residuals clearly exhibits heavy-tailed noise characteristics.

Cold-Start & Robustness Superiority:

- Accurately maps user/item metadata into latent space.
- Outperforms least-squares baseline (IMC-LS) across all ranks r.

02 Image Inpainting under Mixture Noise (Peppers & House)

		Peppers			House		
Noise	**Metric**	**Noisy Obs.**	**AltMin**	**RGDIMC (Ours)**	**Noisy Obs.**	**AltMin**	**RGDIMC (Ours)**
GMM	PSNR (↑)	16.25	12.33	**20.82**	16.41	10.81	**18.05**
	SSIM (↑)	0.6693	0.4360	**0.8477**	0.4898	0.1629	**0.5733**

Key Takeaways: Under complex mixture contamination, RGDIMC effectively suppresses extreme impulsive spikes and preserves structural details, significantly outperforming AltMin.

Conclusion & Future Work

Conclusion

RGDIMC Framework

- We develop RGDIMC, a provably efficient two-stage nonconvex algorithm under heavy-tailed noise, combining robust spectral initialization with Huber-based gradient descent.

Theoretical & Empirical Success

- Establishes local linear convergence and minimax-optimal estimation error under the stated assumptions.
- Experiments demonstrate robust recovery and substantial computational speedups on recommendation and image inpainting tasks.

Future Directions

Adversarial & Infinite-Variance Noise

- Extend the theoretical framework to ϵ-fraction adversarial corruption models.
- Establish convergence guarantees under infinite-variance noise using fractional moments or median-of-means estimators.

Algorithmic Enhancements

- Explore scalable Newton-type methods exploiting generalized Jacobian sparsity for faster convergence.

Thank You!

Questions & Discussion

Xingcai Zhou
School of Statistics and Data Science, Nanjing Audit University