Matrix vs Tensor Denoising Methods under Block Sparse Perturbations

Animashree Anandkumar, Prateek Jain, Yang Shi, U.N.Niranjan

EECS, UC Irvine; Microsoft Research, India; EECS, UC Irvine; ICS, UC Irvine

Highlights

- Robust tensor CP decomposition: separate a tensor into low rank and sparse components.
- Proposed method: alternating projections between power method and hard thresholding.
- Global convergence guarantees: under incoherence and bounded sparsity assumption.
- Improvements: can handle more gross corruptions than matrix-based methods.

Introduction

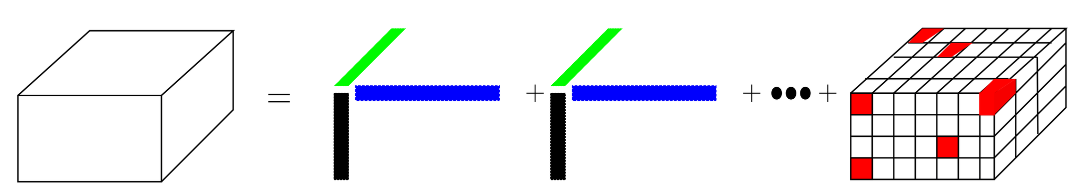

Figure 1: Robust tensor decomposition problem: decomposing a given tensor into low-rank and sparse components.

- **Problem:** Let $T, L, S \in \mathbb{R}^{n\times n\times n}$. Wlog, $\sigma_i^* > 0$. Given T, find L, S such that:

$$T = L + S, \quad L = \sum_{i=1}^{r} \sigma_i^* u_i^{\otimes 3}, \|S\|_0 \le d$$

such that L is a CP-rank r orthogonal tensor, ie, $\langle u_i, u_j\rangle = \delta_{i,j}$, where $\delta_{i,j} = 1$ iff $i = j$ and 0, else.

- **(L)** L^* is μ-incoherent, ie, $\|u_i\|_\infty \le \frac{\mu}{\sqrt{n}}$.
- **(S)** Random block sparsity pattern with B blocks of size d and overlap fraction η, ie,

$$\Psi = \sum_{i=1}^{B} \psi_i \otimes \psi_i \otimes \psi_i, \|\psi_i\|_0 \le d, \psi_i(j) = 0 \text{ or } 1$$

$$d = O(n/r\mu^3)^{2/3},\ B = O(\min(n^{2/3}r^{1/3},\ \eta^{-1.5}))$$

Under this model, the support tensor Ψ which encodes sparsity pattern, has rank B.

Tensor vs Matrix Method

- Under random block sparsity,

$$\frac{d_{ten}}{d_{\text{matrix}}} = \begin{cases} \Omega\left(n^{1/6}r^{4/3}\right), r < n^{0.25} \\ \Omega\left(n^{5/12}r^{1/3}\right), o.w. \end{cases}$$

- Thus, we can handle more gross corruptions than matrix methods.

Alternating Projection Algorithm

1: **for** Stage $l = 1$ to r **do**
2: **repeat**
3: $L^{(t+1)} = P_l(T - S^{(t)})$ (use power method and gradient ascent to compute the eigenvectors).
4: $S^{(t+1)} = \mathcal{H}_\zeta(T - L^{(t+1)})$ (hard thresholding of grossly corrupted entries).
5: **until** Convergence
6: **end for**

Gradient Ascent Algorithm

1: **Power method** [1] to land in spectral ball of sparse tensor: $v_i^{(t+1)} \leftarrow T_j(I, v_i^{(t)}, v_i^{(t)})/\|T_j(I, v_i^{(t)}, v_i^{(t)})\|_2$.
2: **Gradient ascent** iterations to compute tensor eigenvectors: $v_i^{(t+1)} \leftarrow v_i^{(t)} + \frac{1}{4\lambda(1+\lambda/\sqrt{n})} \cdot \left(\widetilde{L}(I, v_i^{(t)}, v_i^{(t)}) - \lambda\|v_i^{(t)}\|^2 v_i^{(t)}\right)$.
3: **Deflation** to obtain all leading components: $T_j \leftarrow T_j - \lambda_j\widehat{u}_j \otimes \widehat{u}_j \otimes \widehat{u}_j$.

Main Result

Theorem (Convergence to Global Optimum)

Let L^*, S^* satisfy (L) and (S). The outputs $\bar{L}$ (and its parameters $\hat{u}_i$ and $\hat{\lambda}_i$) and $\bar{S}$ of the alternating projection algorithm satisfy w.h.p.:

$$\|\hat{u}_i - u_i\|_\infty \le \delta/\mu^2 r n^{1/2}\sigma^*_{\min}, |\hat{\lambda}_i - \sigma_i^*| \le \delta, \quad \forall i \in [n],$$
$$\|\bar{L} - L^*\|_F \le \delta, \quad \|\bar{S} - S^*\|_\infty \le \delta/n^{3/2}, \text{ and } \quad \operatorname{supp} \bar{S} \subseteq \operatorname{supp} S^*.$$

Error Plots

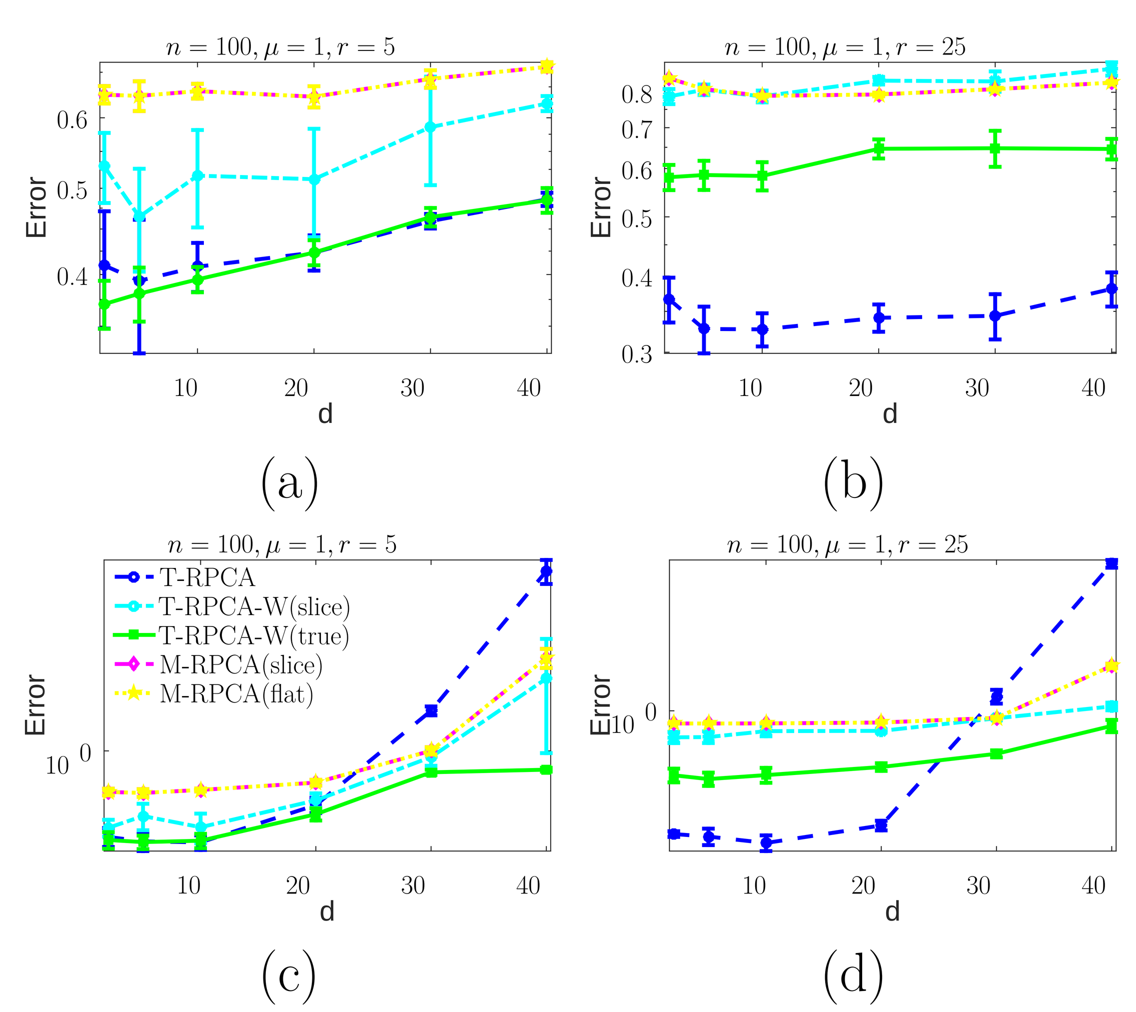

Figure 2: (a),(b) Error with deterministic sparsity. (c),(d) Error with block sparsity.

Time Plots

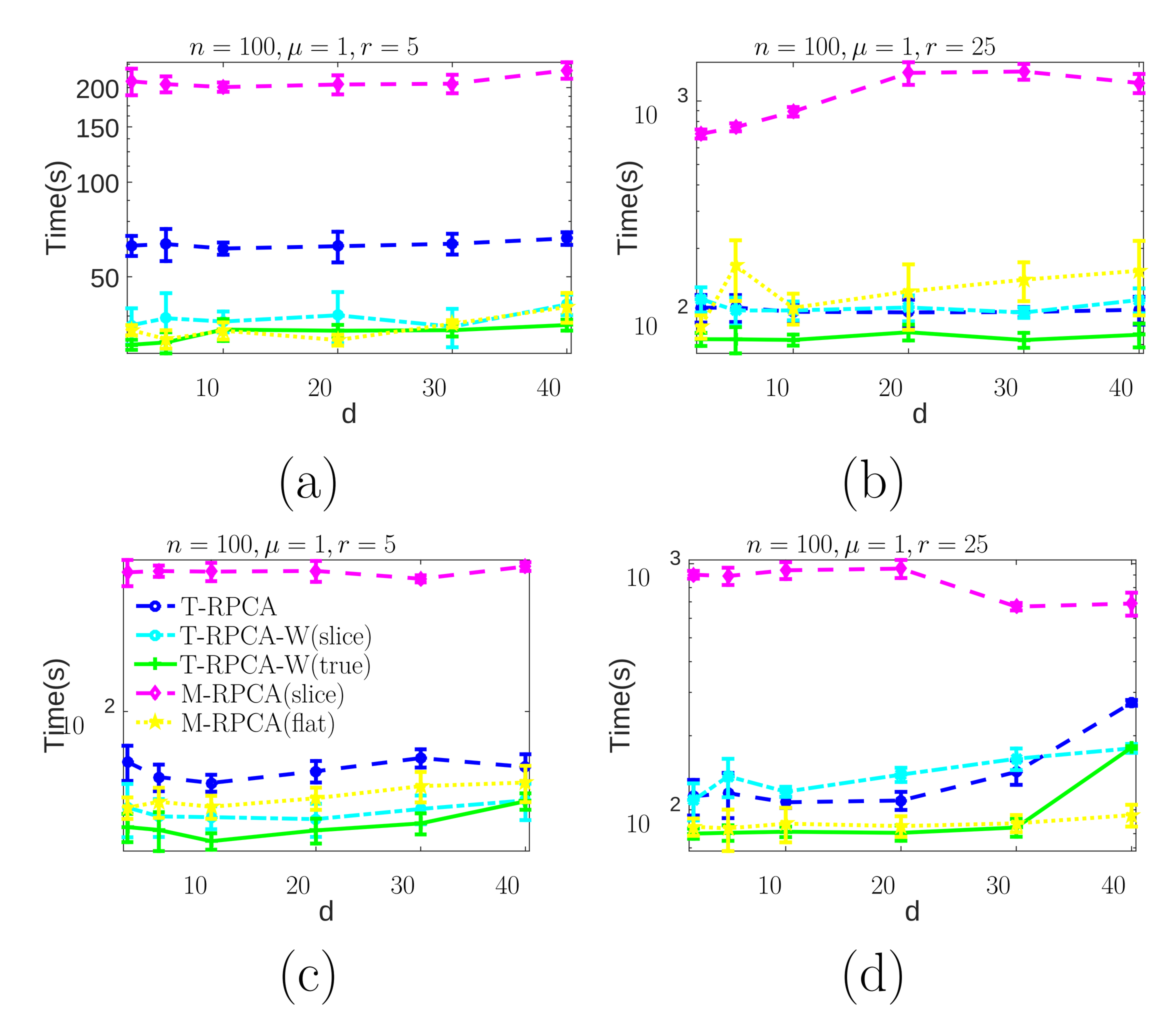

Figure 3: (a),(b) Running time with deterministic sparsity. (c),(d) Running time with block sparsity.

Foreground-background separation

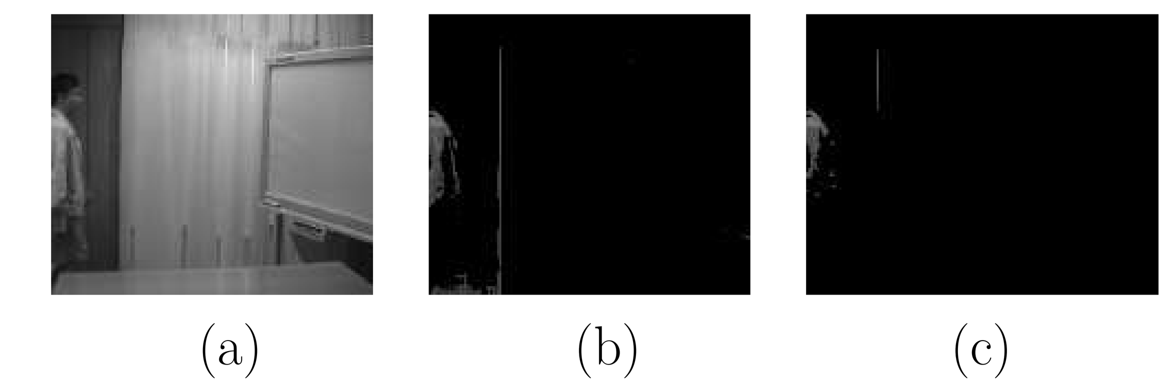

Figure 4: Foreground filtering in the *Curtain* video dataset. (a): Original image frame. (b): Foreground filtered using tensor method; time taken is $5.1s$. (c): Foreground filtered using matrix method; time taken is $5.7s$.

Proof Outline

- The inductive proof for alternating projections is along the similar lines of [2].
- Assuming $T = L^* + S^*$, $L^* = u \otimes u \otimes u$
- Update $L_{t+1} = T - S_t = L^* + S^* - S_t = L^* + E_t$
- Consider tensor fixed point equation $L_{t+1}(u_{t+1}, u_{t+1}, I) = \lambda_{t+1}u_{t+1}$ and apply perturbation arguments from [1].
- Goal is to prove $\|L_{t+1} - L^*\|_\infty \le \epsilon''\|L_t - L^*\|_\infty$:
- Prove $\|L_{t+1} - L^*\|_\infty \le \epsilon\|E_t\|_\infty$ using inductive assumption $\|E_t\|_\infty \le \epsilon\|L_t - L^*\|_\infty$ and then prove $\|E_{t+1}\|_\infty \le \epsilon'\|L_{t+1} - L^*\|_\infty$. Piecing these together we obtain the full proof.
- The challenge in the tensor case is to prove the validity of the tensor fixed point equation.

References

[1] Animashree Anandkumar, Rong Ge, Daniel Hsu, Sham M Kakade, and Matus Telgarsky. Tensor decompositions for learning latent variable models. *The Journal of Machine Learning Research*, 15(1):2773–2832, 2014.

[2] Praneeth Netrapalli, UN Niranjan, Sujay Sanghavi, Animashree Anandkumar, and Prateek Jain. Non-convex robust pca. In *Advances in Neural Information Processing Systems*, pages 1107–1115, 2014.